

**UNIVERSIDADE DE SOROCABA  
PRÓ-REITORIA DE PÓS-GRADUAÇÃO, PESQUISA, EXTENSÃO E INOVAÇÃO  
PÓS-GRADUAÇÃO EM PROCESSOS TECNOLÓGICOS E AMBIENTAIS**

**Edson Martin Feitosa**

**MÉTODOS DE MINERAÇÃO DE DADOS NA IDENTIFICAÇÃO DOS MOTIVOS DA  
EVASÃO NA UNIVERSIDADE DE SOROCABA, NO PERÍODO DE 2012 A 2018**

**Sorocaba/SP  
2020**

**Edson Martin Feitosa**

**MÉTODOS DE MINERAÇÃO DE DADOS NA IDENTIFICAÇÃO DOS MOTIVOS DA  
EVASÃO NA UNIVERSIDADE DE SOROCABA, NO PERÍODO DE 2012 A 2018**

Dissertação apresentada à Banca Examinadora de Qualificação do Programa de Pós-Graduação em Processos Tecnológicos e Ambientais da Universidade de Sorocaba, como exigência para obtenção do título de Mestre em Processos Tecnológicos e Ambientais.

Orientador: Dr. Waldemar Bonventi Junior

**Sorocaba/SP  
2020**

### Ficha Catalográfica

Feitosa, Edson Martin  
F336m Métodos de mineração de dados na identificação dos motivos da evasão na Universidade de Sorocaba, no período de 2012 a 2018 / Edson Martin Feitosa. – 2020.  
252 f. : il.

Orientador: Prof. Dr. Waldemar Bonventi Junior  
Dissertação (Mestrado em Processos Tecnológicos e Ambientais)  
– Universidade de Sorocaba, Sorocaba, SP, 2020.

1. Mineração de dados (Computação). 2. Inteligência competitiva (Administração). 3. Evasão universitária – Universidade de Sorocaba - Indicadores. 4. Ensino superior – Sorocaba (SP). I. Bonventi Junior, Waldemar, orient. II. Universidade de Sorocaba. III. Título.

**Edson Martin Feitosa**

**MÉTODOS DE MINERAÇÃO DE DADOS NA IDENTIFICAÇÃO DOS MOTIVOS DA  
EVASÃO NA UNIVERSIDADE DE SOROCABA, NO PERÍODO DE 2012 A 2018**

Dissertação aprovada como requisito para  
obtenção do grau de Mestre no Programa  
de Pós-Graduação em Processos  
Tecnológicos e Ambientais da  
Universidade de Sorocaba.

**BANCA EXAMINADORA:**

Prof. Dr. Waldemar Bonventi Junior  
Universidade de Sorocaba

Prof. Dr. Fernando de Sá Del Fiol  
Universidade de Sorocaba

Prof. Dr. José Roberto Garcia  
Universidade de Sorocaba

Dedico esse trabalho aos meus filhos  
Victor e Vinícius, a minha esposa Renata  
e minha mãe Geralda.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus que me guiou em todas as decisões que me levaram a ser o que sou hoje.

Aos meus filhos Victor e Vinícius, pois são os principais motivos do meu esforço em ser cada vez melhor a fim de tornar-me um exemplo para suas vidas.

A minha esposa Renata a quem amo muito.

A minha mãe Geralda que sempre me mostrou o valor do estudo e me proporcionou todo o amor que um filho possa querer.

Ao meu falecido pai Manuel, que sei que um dia vou reencontrá-lo.

Ao meu orientador, pela sua paciência e suporte na criação deste trabalho.

Ao Prof.<sup>o</sup> Dr. Rafael Moreno, pela sua valiosa amizade e ajuda em todos os momentos que precisei.

Ao Prof.<sup>o</sup> Dr. Rogério Profeta, por todo incentivo e auxílio na elaboração deste trabalho.

Ao meu colega de trabalho Leandro Gamero, pela ajuda e motivação no desenvolvimento do trabalho.

A Universidade de Sorocaba, seu corpo docente e direção, lugar que considero como meu segundo lar.

“E conhecereis a verdade, e a verdade  
vos libertará. ”  
(João 8:32)

## RESUMO

A evasão é a saída precoce do discente do curso escolhido. Para os gestores das instituições de ensino superior (IES) tem-se caracterizado um problema devido à dificuldade na previsão dos fatos geradores. Logo, essa pesquisa procura apresentar um estudo sobre a evasão nos cursos na UNISO entre 2012 e 2018, identificando padrões e regras encontradas no histórico dos discentes evadidos. Para tal utilizou técnicas para medir a evolução da evasão com o auxílio da equação de Evasão Anual e aplicação de métodos de Business Intelligence (BI), dentre eles árvores de decisão e análise de agrupamentos (*clustering*) para identificar padrões ocultos no grande volume de dados gerados. Foi analisado 29.303 discentes, que cursaram 31.799 cursos, totalizando 37.621 registros comportamentais analisados, onde no estudo mostra-se a evolução da evasão no contexto da universidade, das grandes áreas e especificamente dos cursos, descrição do perfil do discente na faixa de tempo estudada, padrões relacionados as tendências a evadir no aspecto da faixa etária, períodos cursados, média de notas, quantidade de faltas, grande área dos cursos, diferenças dos evadidos para os cursos da manhã e da noite, aspectos relacionados a quanto foram concluído dos componentes curriculares do curso, dentre outras variáveis, trazendo como resultado os perfis dos discentes evadidos, onde um dos perfis de maior relevância é o discente do sexo masculino entre 21 e 25 anos, da grande área de Engenharias que evade no 2º ou 5º semestre, com padrão de comportamentos apontados nas variáveis relacionadas a cancelamentos, trancamentos e média de notas.

**Palavras-chave:** Evasão. Business Intelligence. Mineração de dados. Árvores de decisão. Análise de agrupamentos.

## ABSTRACT

Evasion is the student's early departure from the chosen course. For managers of higher education institutions (HEIs), a problem has been characterized due to the difficulty in predicting the generating facts. Therefore, this research seeks to present a study on dropout in courses at UNISO between 2012 and 2018, identifying patterns and rules found in the history of dropout students. To this end, he used techniques to measure evasion evolution with the aid of the Annual Evasion equation and the application of Business Intelligence (BI) methods, including decision trees and clustering analysis to identify hidden patterns in the large volume of data generated. 29,303 students were analyzed, who attended 31,799 courses, totaling 37,621 behavioral records analyzed, where the study shows the evolution of dropout in the context of the university, the large areas and specifically the courses, description of the student profile in the studied time range, patterns related to the trends to evade in the age group aspect, periods taken, average grades, number of absences, large area of courses, differences between those who dropped out for morning and evening courses, aspects related to how much of the curricular components of the course were completed course, among other variables, bringing as a result the profiles of dropout students, where one of the most relevant profiles is the male student between 21 and 25 years old, from the large area of Engineering that drop out in the 2nd or 5th semester, with a pattern of behaviors pointed out in variables related to cancellations, lockouts and average grades.

**Keywords:** Evasion. Business Intelligence. Datamining. Decision trees. Clustering.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Visão geral dos passos que compõem o processo de KDD .....	27
Figura 2 - Exemplo de Modelo Estrela .....	30
Figura 3 - Classificação simplificada dos métodos de e agrupamentos .....	34
Figura 4 - Classificação simplificada dos métodos de agrupamentos .....	35
Figura 5 - Exemplo de Clustering com auxílio de grafos .....	37
Figura 6 - Formato geral de uma árvore de decisão .....	39
Figura 7 - Modelo Dimensional do Data Mart da evasão da UNISO .....	53
Figura 8 - Seleção dos atributos para mineração de dados .....	60
Figura 9 - Exemplo de sugestão da ferramenta Analisys Service .....	61
Figura 10 - Definição do tipo de dado na ferramenta Analisys Service .....	62
Figura 11 - Configuração do conjunto de treinamento e teste no Analisys Service... 63	63
Figura 12 – Configuração do Excel na criação de matriz de correlação .....	65
Figura 13 - Distribuição de alunos/curso por turno .....	74
Figura 14 – Visualização do modelo de mineração do algoritmo de <i>Clustering</i> para todos os dados .....	104
Figura 15 - Porcentagem de casos por grande área .....	105
Figura 16 - Clusters com os casos de alunos evadidos .....	106
Figura 17 – Intensidade de ligação dos clusters com casos de evasão .....	106
Figura 18 - Características do cluster 8.....	108
Figura 19 - Características do cluster 10.....	109
Figura 20 - Características do cluster 6.....	110
Figura 21 - Comparação entre o cluter 8 e o 1 .....	111
Figura 22 - Árvore de decisão para previsão de bolsa de estudo (turno da noite, grande área de Engenharias).....	112
Figura 23 - Árvore de decisão para previsão de bolsa de estudo (turno da noite, grande área de ciência da saúde) .....	112
Figura 24 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, com valor médio acima de R\$1.212,80) .....	113
Figura 25 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, curso médio abaixo de R\$1.212,80) .....	113
Figura 26 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, grande área de Linguística, Letras e Artes).....	114
Figura 27 - Árvore de decisão para previsão de bolsa de estudo (turno da manhã para o Curso Superior de Tecnologia em Estética e Cosmética) .....	114
Figura 28 - Árvore de decisão da população do turno da manhã (previsão de evasão) .....	116
Figura 29 - Árvore de decisão do turno da noite (previsão de evasão) .....	117
Figura 30 - Visualização do modelo de mineração do algoritmo de Clustering para todos os fatos categorizados como evadido.....	119
Figura 31 - Visualização do modelo de mineração do algoritmo de Clustering para todos os fatos categorizados como evadido e diminuindo a força de ligação .....	120
Figura 32 - Comparação entre o cluster 1 e 2 da população de evadidos .....	130
Figura 33 - Comparação entre o cluster 1 e 3 da população de evadidos .....	131
Figura 34 - Comparação entre o cluster 1 e 4 da população de evadidos .....	132
Figura 35 - Árvore de decisão da análise do estado civil dos casos evadidos .....	134
Figura 36 - Árvore de decisão para análise de evadidos segundo a faixa de idade.....	137

Figura 37 - Árvore de decisão para os casos evadidos predizendo cursos da área de Ciências Sociais Aplicadas.....	137
Figura 38 - Árvore de decisão para os casos evadidos predizendo cursos da área de Engenharias .....	139
Figura 39 - Árvore de decisão para os casos de evasão predizendo a variável de instituição de segundo grau.....	140
Figura 40 - Árvore de decisão dos casos evadidos configurada para previsão de outro curso .....	141
Figura 41 - Árvore de decisão de evadidos para previsão de quantidade de períodos .....	141
Figura 42 - Probabilidade de evasão por quantidade de períodos letivos no turno da noite .....	142
Figura 43 - Faixa de força da correlação.....	143
Figura 44 - Árvore de decisão para previsão do número total de matrículas (curso de logística).....	233
Figura 45 - Árvore de decisão para previsão do número total de matrículas (curso de farmácia) .....	234
Figura 46 - Árvore de decisão para previsão do número total de matrículas (curso de gestão de produção industrial) .....	235
Figura 47 - Árvore de decisão para previsão do número total de matrículas (curso de Fisioterapia).....	236
Figura 48 - Árvore de decisão para previsão do número total de matrículas (curso de Engenharia Elétrica).....	237
Figura 49 - Árvore de decisão para previsão do número total de matrículas (curso de Ciência da Computação).....	238
Figura 50 - Árvore de decisão para previsão do número total de matrículas (curso de Gestão Ambiental).....	239
Figura 51 - Árvore de decisão gerada para os casos de evasão ao prever a quantidade de aprovações .....	241
Figura 52 - Árvore de decisão dos casos evadidos na previsão da quantidade de aprovações.....	243
Figura 53 - Dendograma .....	246
Figura 54 - Árvore de Decisão para exemplo (Jogar Tênis).....	252

## LISTA DE TABELAS

Tabela 1 - Cursos de graduação da UNISO.....	24
Tabela 2 - Quantidade de alunos dos cursos da área de Ciências Agrárias dos anos de 2012 a 2018 .....	68
Tabela 3 - Quantidade de alunos dos cursos de Ciências Biológicas dos anos de 2012 a 2018 .....	68
Tabela 4 - Quantidade de alunos dos cursos de Ciências da Saúde dos anos de 2012 a 2018 .....	68
Tabela 5 - Quantidade de alunos dos cursos de Ciências Exatas e da terra dos anos de 2012 a 2018 .....	68
Tabela 6 - Quantidade de alunos dos cursos de Ciências Humanas dos anos de 2012 a 2018 .....	69
Tabela 7 - Quantidade de alunos nos cursos de Ciências Sociais Aplicadas dos anos de 2012 a 2018 .....	69
Tabela 8 - Quantidade de alunos dos cursos de Engenharias dos anos de 2012 a 2018 .....	70
Tabela 9 - Quantidade de alunos dos cursos de Linguística, Letras e Artes.....	70
Tabela 10 - Estado Civil dos alunos entre 2012 e 2018 .....	72
Tabela 11 – Turno / Grande área / Curso da população total .....	74
Tabela 12 - Total da população analisada por Grande Área/sexo .....	79
Tabela 13 - Distribuição dos evadidos nas grandes áreas (%) .....	79
Tabela 14 - Percentual de Evasão distribuído por grande área .....	80
Tabela 15 - Visualização dos dados computados no relatório de evasão geral da UNISO nos períodos de 2012 a 2018 .....	82
Tabela 16 - Distribuição dos casos de evadidos não solteiros .....	135
Tabela 17 - Distribuição dos casos dos alunos evadidos solteiros abaixo de 4 períodos letivos .....	135
Tabela 18 - Distribuição de casos de alunos solteiros entre 4 e 9 períodos letivos	136
Tabela 19 - Distribuição dos casos de alunos solteiros acima de 9 períodos letivos .....	136
Tabela 20 - Status dos alunos da área de Ciências Agrárias dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	160
Tabela 21 - Status dos alunos da área de Ciências Biológicas dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	162
Tabela 22 - Status dos alunos da área de Ciências da Saúde dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	164
Tabela 23 - Status dos alunos da área de Ciências Exatas e da Terra dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	167
Tabela 24 - Status dos alunos da área de Ciências Humanas dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	169
Tabela 25 - Status dos alunos da área de Ciências Sociais Aplicadas dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	171
Tabela 26 - Status dos alunos da área de Engenharias dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	174
Tabela 27 - Status dos alunos da área de Linguística, Letras e Artes dos semestres de 2012 1º Semestre a 2018 1º Semestre .....	177
Tabela 28 - Índice de evasão por período letivo para os cursos da grande área de Engenharias .....	180

Tabela 29 - Índice de evasão por período letivo para os cursos da grande área de Ciências Sociais Aplicadas.....	189
Tabela 30 - Índice de evasão por período letivo para os cursos da grande área de Ciências Agrárias .....	205
Tabela 31 - Índice de evasão por período letivo para os cursos da grande área de Ciências da Saúde .....	209
Tabela 32 – Índice de evasão por período letivo para os cursos da grande área de Ciências Exatas e da Terra .....	216
Tabela 33 - Índice de evasão por período letivo para os cursos da grande área de Ciências Humanas .....	221
Tabela 34 – Índice de evasão por período letivo para os cursos da grande área de Linguística, Letras e Artes .....	225
Tabela 35 - Índice de evasão por período letivo para os cursos da grande área de Ciências Biológicas .....	230
Tabela 36 - Dados de exemplo para explicação do método de árvore de decisão .	249

## LISTA DE QUADROS

Quadro 1 - Resultado do algoritmo de Clustering aplicado aos casos evadidos entre o 1º Sem. de 2012 e o 1º Sem. de 2018 .....	122
Quadro 2 - Cursos analisados por grande área .....	157
Quadro 3 - Status dos discentes nos cursos.....	159
Quadro 4 - Descrição dos atributos usados na mineração de dados .....	245

## LISTA DE GRÁFICOS

Gráfico 1 - Quantidade de alunos por grande área entre 2012 e 2018 .....	71
Gráfico 2 - Quantidade de alunos por sexo .....	71
Gráfico 3 - Faixa de idade dos alunos entre 2012 e 2018 .....	72
Gráfico 4 – Tipo de instituição de segundo realizado pelos alunos analisados .....	73
Gráfico 5 - Cor/raça declarada entre os alunos analisados .....	73
Gráfico 6 - Índice de evasão geral para o 1º semestre de cada ano .....	81
Gráfico 7 - Índice de evasão geral para o 2º semestre de cada ano .....	81
Gráfico 8 - Quantidade de veteranos e evadidos no 1º semestre de cada ano .....	81
Gráfico 9 - Quantidade de veteranos e evadidos no 2º semestre de cada ano .....	82
Gráfico 10 - Índice de evasão por grande área para o 1º semestre entre 2012/01 e 2018/01 .....	84
Gráfico 11 - Índice de evasão por grande área para o 2º semestre entre 2012/02 e 2017/02 .....	84
Gráfico 12 – Índices de evasão dos cursos de Engenharias entre 2012 e 2018 para o 1º semestre de cada ano .....	87
Gráfico 13 - Índices de evasão de Engenharias entre 2012 e 2018 para o 2º semestre de cada ano .....	87
Gráfico 14 - Índices de evasão de Ciências Sociais Aplicadas entre 2012 e 2015 para o 1º semestre de cada ano .....	89
Gráfico 15 - Índices de evasão de Ciências Sociais Aplicadas entre 2016 e 2018 para o 1º semestre de cada ano .....	89
Gráfico 16 – Índice de evasão de Ciências Sociais Aplicadas entre 2012 e 2015 para o 2º semestre de cada ano .....	90
Gráfico 17 – Índice de evasão de Ciências Sociais Aplicadas entre 2016 e 2018 para o 2º semestre de cada ano .....	90
Gráfico 18 - Índice de evasão dos cursos da grande área de Ciências Agrárias entre 2012 e 2018 para o 1º semestre de cada ano .....	91
Gráfico 19 - Índice de evasão dos cursos da grande área de Ciências Agrárias entre 2012 e 2018 para o 2º semestre de cada ano .....	91
Gráfico 20 - Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2012 e 2015 para o 1º semestre de cada ano .....	92
Gráfico 21 - Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2016 e 2018 para o 1º semestre de cada ano .....	93
Gráfico 22 – Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2012 e 2015 para o 2º semestre de cada ano .....	93
Gráfico 23 - Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2016 e 2018 para o 2º semestre de cada ano .....	94
Gráfico 24 - Índice de evasão dos cursos da grande área de Ciências Exatas e da Terra entre 2012 e 2018 para o 1º semestre de cada ano .....	95
Gráfico 25 - Índice de evasão dos cursos da grande área de Ciências Exatas e da Terra entre 2012 e 2018 para o 2º semestre de cada ano .....	95
Gráfico 26 - Índice de evasão dos cursos da grande área de Ciências Humanas entre 2012 e 2018 para o 1º semestre de cada ano .....	96
Gráfico 27 - Índice de evasão dos cursos da grande área de Ciências Humanas entre 2012 e 2018 para o 2º semestre de cada ano .....	97
Gráfico 28 - Índice de evasão dos cursos da grande área de Linguística, Letras e Artes entre 2012 e 2018 para o 1º semestre de cada ano .....	98

Gráfico 29 - Índice de evasão dos cursos da grande área de Linguística, Letras e Artes entre 2012 e 2018 para o 2º semestre de cada ano .....	98
Gráfico 30 - Índice de evasão para o curso de Ciências Biológicas entre 2012 e 2018 para o 1º semestre de cada ano .....	99
Gráfico 31 - Índice de evasão para o curso de Ciências Biológicas entre 2012 e 2018 para o 2º semestre de cada ano .....	99
Gráfico 32 – Índice de evasão dos cursos tecnológicos de Design de Interiores a Gestão Ambiental (EAD) entre 2012 a 2018 para os 1º semestres de cada ano ....	100
Gráfico 33 - Índice de evasão dos cursos tecnológicos de Design de Interiores a Gestão Ambiental (EAD) entre 2012 a 2018 para os 2º semestres de cada ano ....	100
Gráfico 34 - Índice de evasão dos cursos tecnológicos de Gestão Comercial a Gestão Financeira entre 2012 a 2018 para os 1º semestres de cada ano .....	101
Gráfico 35 - Índice de evasão dos cursos tecnológicos de Gestão Comercial a Gestão Financeira entre 2012 a 2018 para os 2º semestres de cada ano .....	101
Gráfico 36 - Índice de evasão dos cursos tecnológicos de Hotelaria a Segurança no Trabalho entre 2012 a 2018 para os 1º semestres de cada ano .....	102
Gráfico 37 - Índice de evasão dos cursos tecnológicos de Hotelaria a Segurança no Trabalho entre 2012 a 2018 para os 2º semestres de cada ano .....	102

## LISTA DE SIGLAS E ABREVIATURAS

IES	Instituição de Ensino Superior
UNISO	Universidade de Sorocaba
BI	Business Intelligence
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
MEC	Ministério da Educação
SQL	<i>Structured Query Language</i>
OLAP	On-Line Analytical Processing
SGBD	Sistema Gerenciador de Banco de Dados
KDD	knowledge Discovery in databases
FAFI	Faculdade de Filosofia, Ciências e Letras de Sorocaba
FACCAS	Faculdade de Ciências Contábeis e Administrativas de Sorocaba
NDPH	Núcleo de documentação e pesquisa histórica
NEAS	Núcleo de pesquisas ambientais
NET	Núcleo de estudos Tropeiro
MESAU	Núcleo de educação em saúde
NUCAB	Núcleo de cultura afro-brasileira
FIDA	Faculdades Integradas Dom Aguirre

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>19</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>21</b>
<b>2.1</b>	<b>EVASÃO NAS INSTITUIÇÕES DE ENSINO SUPERIOR</b> .....	<b>21</b>
<b>2.2</b>	<b>CÁLCULOS PARA EVASÃO</b> .....	<b>22</b>
<b>2.3</b>	<b>UNIVERSIDADE DE SOROCABA (UNISO)</b> .....	<b>23</b>
<b>2.4</b>	<b>EXTRAÇÃO DE CONHECIMENTO</b> .....	<b>26</b>
2.4.1	SELEÇÃO.....	27
2.4.2	PRÉ-PROCESSAMENTO (PREPROCESSING).....	28
2.4.3	TRANSFORMAÇÃO ( <i>TRANSFORMATION</i> ) .....	28
2.4.4	<i>DATA WAREHOUSE</i> E <i>DATA MART</i> .....	28
2.4.5	MINERAÇÃO DE DADOS (DATA-MINING) .....	31
2.4.6	INTERPRETAÇÃO / AVALIAÇÃO .....	41
<b>2.5</b>	<b>MATRIZ DE CORRELAÇÃO DE PEARSON</b> .....	<b>41</b>
<b>2.6</b>	<b>CONCLUSÃO DO CAPÍTULO</b> .....	<b>42</b>
<b>3</b>	<b>METODOLOGIA</b> .....	<b>43</b>
<b>3.1</b>	<b>CONTEXTUALIZAÇÃO DA EVASÃO NA UNISO</b> .....	<b>43</b>
<b>3.2</b>	<b>ÍNDICES DE EVASÃO NA UNISO</b> .....	<b>43</b>
<b>3.3</b>	<b>KDD APLICADO PARA ANÁLISE DA EVASÃO</b> .....	<b>45</b>
3.3.1	Seleção.....	46
3.3.2	Pré-processamento .....	48
3.3.3	Transformação.....	49
3.3.4	<i>Data Warehouse</i> e <i>Data Mart</i> .....	52
3.3.5	Mineração de Dados ( <i>DATA – MINING</i> ).....	54
<b>3.4</b>	<b>MATRIZ DE CORRELAÇÃO DE PEARSON E DENDOGRAMA</b> .....	<b>63</b>
<b>3.5</b>	<b>CONCLUSÃO DO CAPÍTULO</b> .....	<b>66</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>67</b>
<b>4.1</b>	<b>Análise do perfil dos alunos/cursos</b> .....	<b>67</b>
4.1.1	Discriminação do perfil dos alunos/curso no turno .....	74
4.1.2	Discriminação do perfil dos alunos evadidos .....	79
<b>4.2</b>	<b>Análise dos índices de evasão</b> .....	<b>80</b>
4.2.1	Índices de evasão das grandes áreas de graduação da UNISO .....	83
4.2.2	Índices de evasão dos cursos de graduação .....	86
<b>4.3</b>	<b>Mineração dos dados</b> .....	<b>103</b>

4.3.1	Clustering aplicado a todos os dados .....	103
4.3.2	Árvore de decisão aplicado a todos os dados .....	111
4.3.3	Clustering aplicado aos dados dos evadidos .....	119
4.3.4	Árvore de decisão aplicada aos dados dos evadidos .....	134
4.4	<b>MATRIZ DE CORRELAÇÃO DE PEARSON E DENDOGRAMA .....</b>	<b>143</b>
4.5	<b>CONCLUSÃO DO CAPÍTULO .....</b>	<b>145</b>
5	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>146</b>
6	<b>REFERÊNCIAS .....</b>	<b>151</b>
	<b>APÊNDICE A – SCRIPT SQL DO DATA MART DA EVASÃO.....</b>	<b>153</b>
	<b>APÊNDICE B – CURSOS DE GRADUAÇÃO ANALISADOS.....</b>	<b>157</b>
	<b>APÊNDICE C – STATUS DOS ALUNOS ENTRE O 1º SEMESTRE DE 2012 AO 1º SEMESTRE DE 2018.....</b>	<b>160</b>
	<b>APÊNDICE D – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE ENGENHARIAS.....</b>	<b>180</b>
	<b>APÊNDICE E – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS SOCIAIS APLICADAS.....</b>	<b>189</b>
	<b>APÊNDICE F – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS AGRÁRIAS .....</b>	<b>205</b>
	<b>APÊNDICE G – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS DA SAÚDE .....</b>	<b>209</b>
	<b>APÊNDICE H – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS EXATAS E DA TERRA.....</b>	<b>216</b>
	<b>APÊNDICE I – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS HUMANAS.....</b>	<b>221</b>
	<b>APÊNDICE J – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE LÍNGUÍSTICA, LETRAS E ARTES.....</b>	<b>225</b>
	<b>APÊNDICE K – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS BIOLÓGICAS.....</b>	<b>230</b>
	<b>APÊNDICE I – ANÁLISE DO PARÂMETRO TOTAL DE MATRÍCULAS COM A TÉCNICA ÁRVORES DE DECISÃO EM TODOS OS DADOS .....</b>	<b>233</b>
	<b>APÊNDICE J – ANÁLISE DO PARÂMETRO QUANTIDADE DE APROVAÇÕES APLICADA COM TÉCNICAS DE ÁRVORE DE DECISÃO NOS DADOS RELACIONADOS AOS CASOS DE EVASÃO.....</b>	<b>241</b>
	<b>APÊNDICE K – DESCRIÇÃO DOS ATRIBUTOS USADOS PELA FERRAMENTA DE MINERAÇÃO DE DADOS (ANALISYS SERVICE) .....</b>	<b>245</b>
	<b>APÊNDICE L – DENDOGRAMA .....</b>	<b>246</b>
	<b>ANEXO A – LEIAUTE DE IMPORTAÇÃO DOS DADOS DO CENSO PARA O INEP (VERSÃO 2017).....</b>	<b>248</b>
	<b>ANEXO B – EXEMPLO DE FUNCIONAMENTO DO ALGORITMO DE ÁRVORE DE DECISÃO .....</b>	<b>249</b>

## 1 INTRODUÇÃO

A evasão nas instituições de ensino superior (IES) tem se tornado um dos problemas mais relevantes enfrentados pela gestão atualmente, no Brasil existe pouca orientação oficial sobre como sistematizar e administrar esse fenômeno, sendo a fonte mais confiável atualmente publicada pelo Ministério da Educação, por intermédio do Censo da Educação Superior. (HOFFMANN, 2016).

Entretanto, os dados disponibilizados não refletem de maneira aprofundada o entendimento dos possíveis motivos que levam um discente a evadir do curso de sua escolha.

As IES demandam grandes investimentos em propaganda para a divulgação de seus cursos, seus prêmios e sua infraestrutura preparada para atender seus alunos no desenvolvimento de suas competências, mas estão com facilidade perdendo esses mesmos alunos evadidos sem motivo claro dessa decisão.

Quando o aluno desiste do curso que escolheu, o prejuízo não é só dele que demandou tempo e dinheiro no sonho de se tornar um profissional em sua área de escolha, mas também do professor que não conseguiu atingir sua meta com aquele discente, da instituição de ensino que não cumpriu sua missão institucional, da sociedade que perde com o desenvolvimento social e econômico e até mesmo da família que não tem o sonho de ver seu filho formado (SILVA FILHO et al., 2007).

Muitas iniciativas particulares são realizadas pelas IES tentando se aprofundar no entendimento da evasão, como questionários minuciosamente preparados para pontuar os diversos motivos, mas mesmo assim não estão sendo efetivos em categorizá-los. Esses questionários em sua maioria são aplicados nos discente no ato da evasão, mas nem sempre são respondidos com a devida importância pelo discente que já não vê sentido em contribuir com a IES que decidiu deixar.

Pensando nesse problema, nesse trabalho procura-se desenvolver um estudo sobre a evasão sem tabular questionários aplicado aos evadidos, e sim, utilizar o vasto banco de dados da Universidade de Sorocaba (Uniso), aplicado em um conjunto de métodos para medir a evolução da evasão e com o auxílio de técnicas de *Business Intelligence* (BI) conhecidas como mineração de dados, identificar padrões que possam ser utilizados para entender o comportamento do discente evadido nos cursos de graduação da Uniso.

**Objetivo:** desenvolver formas de acompanhamento e controle da evasão na Universidade de Sorocaba, uso de mineração dos dados existentes nas bases de alunos da UNISO, identificar padrões que evidenciem possíveis causas da evasão, para auxiliar no entendimento do comportamento de um discente evadido. Com esses dados, descobrir padrões que sirvam para auxiliar nas tomadas de decisão e na categorização dos evadidos.

Para tanto, foi necessário atingir os seguintes **objetivos específicos**:

- Identificar e quantificar a variação dos números de casos de evasão nos cursos de graduação de 2012 a 2018, categorizando de maneira geral, por grande área e curso.
- Estruturar os fatos de evasão na dimensão temporal do semestre letivo e nas dimensões espaciais com variáveis comportamentais de âmbito pessoal, acadêmico e financeiro do discente com o auxílio do processo de KDD (*Knowledge Discovery in Database*), conhecido em português como Descoberta de Conhecimento em Banco de Dados
- Identificar regras e padrões com o auxílio de técnicas de mineração de dados conhecidas como árvore de decisão e *clustering* aplicada na etapa do KDD chamada de mineração de dados e matriz de correlação.

O relato da pesquisa está dividido em cinco partes distribuídas ao longo deste texto da seguinte maneira:

I – Introdução: no capítulo 1 procura-se demonstrar a importância do trabalho a ser desenvolvido e situar o leitor nos diversos tópicos relacionados ao tema.

II – Evasão nas Instituições de Ensino Superior: no capítulo 2, item 1, procura-se descrever o problema da evasão e no capítulo 2, item 2, são apresentados os cálculos que são utilizados para quantificá-lo.

III – Universidade de Sorocaba (UNISO): no capítulo 2, item 3, é explicada de maneira histórica o surgimento da UNISO que será utilizada como estudo de caso.

IV – Extração de Conhecimento: no capítulo 2, item 4, são detalhados todos os passos para a utilização da técnica de mineração de dados com auxílio do processo de KDD, que será utilizada pela coleta de regras para a evasão na UNISO e no capítulo 2, item 5, passos para o desenvolvimento de uma matriz de correlação dos dados gerados no processo de KDD.

V – Desenvolvimento: aplica-se a técnica da maneira descrita no capítulo 3 e são analisados os resultados encontrados no capítulo seguinte.

## 2 REFERENCIAL TEÓRICO

Este capítulo apresenta os principais conceitos que fundamentam a elaboração e execução dessa pesquisa. São abordados os seguintes temas: evasão nas instituições de ensino superior; modalidades de evasão; fatores que levam à evasão; os principais cálculos para sua identificação; a Universidade de Sorocaba e sua história; cursos utilizados para a base de cálculo; extração de conhecimento e as técnicas árvores de decisão e *clustering* para mineração de dados.

### 2.1 EVASÃO NAS INSTITUIÇÕES DE ENSINO SUPERIOR

“A palavra *evasão* vem do latim *evasio* e foi utilizada pela primeira vez em meados do século XV significando fuga, saída, fora, abandono, fracasso, insucesso, mas vale ressaltar que as nomenclaturas referentes à evasão variam conforme o estudioso adotado e que todas estas terminologias levam a um único caminho que é a não finalização de um curso, treinamento, qualificação, especialização ou qualquer modalidade educacional que conduza o alunado a um conhecimento especializado [...]” (HOFFMANN, 2016, p. 22).

No ensino superior a evasão é caracterizada pela saída do aluno matriculado de um curso, seja por motivo justificado ou não (MONTEIRO, CARLOS A. ET AL., 2011, p. 47).

Segundo a Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas (1997, p.16), a evasão é caracterizada:

- Evasão de curso: quando o estudante se desliga do curso superior em situações diversas tais como: abandono (deixa de matricular-se), desistência, transferência ou reopção (mudança de curso), exclusão por norma institucional;
- Evasão da instituição: quando o estudante se desliga da instituição na qual está matriculado;
- Evasão do sistema: quando o estudante abandona de forma definitiva ou temporária o ensino superior.

Entendendo essas delimitações fica mais fácil compreender o foco da pesquisa que procura identificar os motivos da evasão de curso, não considerando a exclusão por norma institucional como motivo segundo Bueno (1993). (BUENO, 1993).

Diversos fatores além dos citados pela Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas (1997) caracterizam a evasão de curso e será

contemplado na pesquisa destacam-se também o cancelamento da matrícula, trancamento da matrícula ou a transferência de curso interna, não considerando a externa que ocorre quando se muda de uma IES para outra no mesmo curso, ou a transferência de instituição de ensino superior (BRASILEIRAS, 1997; MONTEIRO et al., 2011; SILVA et al., 2008).

No estudo de Paredes (1994) que procurou avaliar cursos de duas universidades brasileiras, sendo uma pública e outra privada ficou claro que o abandono definitivo do sistema de ensino superior corresponde apenas a 12,8% dos alunos dos grupos estudados, enquanto que 64% dos mesmos concluiu o 3º grau de ensino em outro curso ou em outra Universidade, o que indica que a evasão, da universidade e do sistema é menor do que a evasão de curso.

Com isso em mente, se faz necessário uma análise dos evadidos por curso, a fim de identificar se existem padrões que prevejam o momento que a evasão acontecerá e por quais motivos tendo apenas com base de estudo variáveis internas do discente.

## **2.2 CÁLCULOS PARA EVASÃO**

Para melhor entender o escopo da evasão na universidade de Sorocaba, faz-se necessário quantificar o grau de evasão do ponto de vista geral, ou seja, compreender qual é a média de evasão existente, se esses indicadores se mantem no decorrer do tempo e se estão de acordo com estudos já realizados em outras IES no Brasil e no mundo.

Após essa primeira análise é importante para um estudo mais conciso discriminar esses índices de maneira categórica nos seguintes tópicos:

- Período letivo
- Área de conhecimento
- Curso

O intuito desse aprofundamento é coletar métricas que exponham qual é o montante passado, atual e futuro previsto do problema da evasão, para com isso conseguir quantificar e qualificar a evasão de maneira que facilite o gerenciamento do mesmo.

Será utilizado o seguinte cálculo para atingir esse objetivo:

1. **Evasão anual média:** tem como objetivo medir qual a porcentagem de alunos matriculados que não tendo se formado também não se matriculou no ano seguinte, ou seja, cálculo da perda de alunos de um ano para o outro (LOBO, MOTEJUNAS, *et al.*, 2007).

Para que esse cálculo seja enquadrado na realidade da UNISO, o mesmo será adaptado para trabalhar de maneira semestral e não anual como na versão original de Lobo (2007).

O cálculo do percentual da evasão referente ao semestre  $n$  é dado por:

$$E_{(n)} = 1 - \frac{[M_{(n)} - I_{(n)}]}{[M_{(n-1)} - C_{(n-1)}]} \quad (1)$$

onde E é evasão, M é número de matriculados, C é o número de concluintes, I é o número de ingressantes,  $n$  é o ano em estudo e  $(n-1)$  é o ano anterior. (LOBO, MOTEJUNAS, *et al.*, 2007, p. 645).

Com a conclusão dessas análises será possível indicar a taxa de evasão geral nas diversas categorias mencionadas acima e complementá-la com a análise individual discente, que será realizada com o auxílio de técnicas de Business Intelligence (BI) detalhado no item 2.4.

### 2.3 UNIVERSIDADE DE SOROCABA (UNISO)

A UNISO foi implantada em 15 de setembro de 1994, tendo sido a primeira universidade da região (VANNUCCHI, 2012). É mantida pela Fundação Dom Aguirre atualmente situada na rodovia Raposo Tavares, Km 92,5, mesmo local da UNISO – Cidade Universitária Professor Aldo Vannucchi.

A Uniso possui como missão

Ser uma Universidade Comunitária que, por meio da integração do ensino, da pesquisa e da extensão, produza conhecimentos e forme profissionais, em Sorocaba e região, para serem agentes de mudanças sociais, à luz de princípios cristãos. (Relatório Social 2017, 2017)

Sua história está vinculada a Faculdade de Filosofia, Ciências e Letras de Sorocaba (FAFI) que começou a funcionar em 1954 com os cursos de Pedagogia e Letras Neolatinas e com a Faculdade de Ciências Econômicas (FACCAS) inaugurada

em 1966, inicialmente com os cursos de Administração de Empresas, Ciências Contábeis e Ciências Econômicas (Relatório Social 2017, 2017). Neste mesmo ano foram criados os primeiros núcleos de pesquisa e extensão, como o núcleo de documentação e pesquisa histórica – NDPH, o núcleo de pesquisas ambientais - NEAS, o núcleo de estudos Tropeiro – NET, o núcleo de educação em saúde – NESAU e o núcleo de cultura afro-brasileira – NUCAB, muito importantes nesse processo de institucionalização da pesquisa (Relatório Social 2017, 2017).

Em 1992 as duas faculdades passaram a ser uma única instituição chamada de Faculdades Integradas Dom Aguirre (FIDA) e, em 1994 pela portaria ministerial nº 1.364, de 13 de setembro de 1994, publicada no Diário Oficial em 15 de setembro de 1994 passou a se chamar UNISO.

Em 1996 foram inaugurados os campos Trujillo localizado na avenida General Osório número 35 e Seminário localizado na avenida Doutor Eugênio Salerno número 100.

Em 1999 foi inaugurada a Cidade Universitária localizada na rodovia Raposo Tavares, Km 92,5.

Em 2002 foi inaugurada o campus Tietê localizada na rua João Alves número 101, Jardim Santa Cruz em Tietê.

A partir de 2002 foi criado o primeiro Mestrado em educação.

Em 2006 foi reconhecido pela Capes o curso de Mestrado em Comunicação e Cultura.

Em 2010 foi reconhecido pela Capes o curso de Doutorado em Educação.

Em 2013 foi aprovado pela Capes o mestrado em Processos Tecnológicos e Ambientais

Com relação aos cursos de graduação, foco da pesquisa, segue uma relação separada por áreas e categorias, mostrada abaixo na tabela 1. Esta tabela servirá como base para a análise dos resultados de evasão por área em função da quantidade de cursos em cada uma delas.

**Tabela 1 - Cursos de graduação da UNISO**

Áreas	Bacharelados	Licenciaturas	Tecnológicos
Ciências Biológicas e da Saúde	Biomedicina		
	Ciências Biológicas		Gestão Ambiental(EAD)
	Educação Física	Educação Física	
	Enfermagem		
	Farmácia		Estética e Cosmética

Ciências Exatas, Agrárias e da Terra	Fisioterapia		
	Nutrição		
	Odontologia		
	Terapia Ocupacional		
	Ciência da Computação	Física	Jogos Digitais
	Química Industrial	Matemática	
	Sistemas de Informação	Química	Gestão de Equinocultura
	Medicina Veterinária		
Ciências Humanas	Psicologia	Filosofia	
		Geografia	
	Relações Internacionais	História	
		Pedagogia	
Ciências Sociais Aplicadas	Administração		Design Gráfico
	Arquitetura e Urbanismo		Gastronomia
	Ciências Contábeis		Hotelaria
	Ciências Econômicas		Processos Gerenciais
	Comércio Exterior		Gestão Comercial
	Jornalismo		Marketing
	Publicidade e Propaganda		Gestão de Qualidade
	Relações Públicas		Gestão de Recursos Humanos
	Design		Gestão Financeira
	Design de Moda		Logística
	Direito		
Linguística, Letras e Artes		Artes Visuais	
		Dança	
		Letras: Português e Inglês	
		Letras: Português e Espanhol	
		Música	
	Teatro		
Engenharias	Engenharia Agrônoma		
	Engenharia de Alimentos		
	Engenharia Ambiental		
	Engenharia de Bioprocessos e Biotecnologia		
	Engenharia Civil		
	Engenharia da Computação		Gestão de Produção Industrial
	Engenharia de Controle e Automação		
	Engenharia de Materiais		
	Engenharia de Produção		
	Engenharia Elétrica		
	Engenharia Química		
	Engenharia Mecânica		

Fonte: RELATÓRIO Social 2017. Uniso, 2017. Disponível em <[http://uniso.br/publicacoes/pdfs/Relatorio-Social/relatorio\\_social\\_2017.pdf](http://uniso.br/publicacoes/pdfs/Relatorio-Social/relatorio_social_2017.pdf)>. Acessado em 31 Agosto 2018.

Essa foi uma breve descrição histórica da UNISO, a fim de contextualizar o caso de estudo do trabalho.

## 2.4 EXTRAÇÃO DE CONHECIMENTO

O processo de extração de conhecimento (padrões) de dados é um método cujo objetivo é construir hipóteses por intermédio da análise de fatos, ou seja, identificar tendências com o auxílio de inferência lógica para extrair conhecimento do conjunto de dados analisados. Como consequência, auxilia em tarefas tais como a tomada de decisão em uma empresa. Esse conjunto de métodos é conhecido como *Knowledge Discovery in Database* (KDD), que pode ser traduzido como “descoberta de conhecimento em banco de dados” (CORADINE, LOPES e MACIEL, 2011).

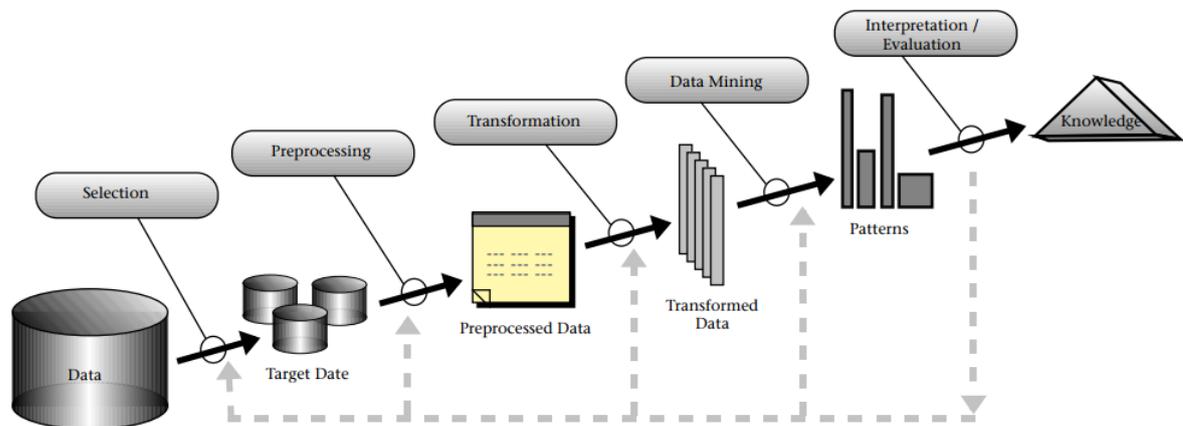
Diferentemente da estatística inferencial, o objetivo do KDD não é testar hipóteses previamente elaboradas e responder o quão verdadeiro ela é de um intervalo de confiança, e sim descobrir regras e relações entre os dados, ou seja, hipóteses não imaginadas de antemão.

A definição de KDD foi feita por um de seus criadores (FAYYAD, SHAPIRO e SMYTH, 1996), “processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”, ou seja, entende-se como não trivial pelo fato de exigir técnicas de busca e/ou inferência por parte do analista, é descrito como previamente desconhecidas e potencialmente úteis, pois os dados sempre estiveram com a empresa, porém ainda não trabalhadas para gerar conhecimento útil.

O KDD surgiu em resposta às necessidades de análise de grandes bancos de dados, devido a sua complexidade (ou não trivialidade), esse processo procura perceber e interpretar adequadamente inúmeros fatos e sanar a dificuldade em conjugar interpretações de forma a decidir que ações devem ser tomadas em cada caso. Ou seja, sua principal função é produzir conhecimento a partir de dados relacionados.

O processo é constituído de cinco etapas: seleção, pré-processamento e limpeza, transformação, mineração de dados e interpretação/avaliação, mostrados na Figura 1.

Figura 1 - Visão geral dos passos que compõem o processo de KDD



Fonte: FAYYAD, U.; SHARIPO, G. P.; SMYTH, P. From Data Mining to Knowledge Discovery in Database. *AI Magazine*, v.17, p54, Julho 1996

Apesar dos passos serem executados na ordem apresentada, esta pode ser alterada, pois o processo é *iterativo*, ou seja, o usuário pode intervir quando necessário no curso das atividades e *iterativo*, ou seja, possui uma sequência finita de ações onde cada passo depende de seus predecessores (FAYYAD, SHAPIRO e SMYTH, 1996).

Nos próximos tópicos será detalhado os passos do KDD.

### 2.4.1 SELEÇÃO

Nessa fase é definido o escopo do domínio dos dados a serem estudados, nela são levantados os casos que serão analisados e os atributos (características) que se deseja monitorar.

Essa parte do processo é de extrema importância, pois a mineração de dados para descoberta do conhecimento ocorre a partir dos dados disponíveis (MAIMON e ROKACH, 2010).

Após sintetizar essas informações, as mesmas são gerenciadas dentro de um *Data Warehouse* ou *Data Mart* explicada no tópico 2.4.4.

### 2.4.2 PRÉ-PROCESSAMENTO (PREPROCESSING)

Normalmente, os dados disponíveis para análise não estão em um formato adequado para a Extração de Conhecimento. Além disso, em razão de limitações de memória ou tempo de processamento, muitas vezes não é possível a aplicação direta dos algoritmos de extração de padrões aos dados. Dessa maneira, torna-se necessária a aplicação de métodos para tratamento, eliminação de dados redundantes, recuperação de dados incompletos, análise de dados discrepantes, limpeza e redução do volume de dados antes de iniciar a etapa de Extração de Padrões.

É importante salientar que a execução das transformações deve ser guiada pelos objetivos do processo de extração, a fim de que o conjunto de dados gerado apresente as características necessárias para que os objetivos sejam cumpridos. (DINIZ e LOUZADA, 2000).

### 2.4.3 TRANSFORMAÇÃO (*TRANSFORMATION*)

Essa etapa se preocupa com dados transformados por intermédio de combinações, também conhecidos como dados derivados, ex.: dados calculados como idade que é calculada pela data de nascimento, quantidades de determinado item que depende da contagem, ou somatória de valores, ou intervalos diversos.

### 2.4.4 *DATA WAREHOUSE E DATA MART*

*Data Warehouse* ou um *Data Mart* são maneiras estruturais de organizar um banco de dados gerenciado por um sistema gerencial de banco de dados (DBMS) que pode ser definido como

[...] uma coleção de dados inter-relacionados e um conjunto de programas para acessar esses dados. A coleção de dados, normalmente chamada de banco de dados, contém informações relevantes a empresa. O principal objetivo do DBMS é fornecer uma maneira de recuperar informações de banco de dados que seja tanto conveniente quanto eficiente (ABRAHAM SILVERSCHEATZ, 2010, p. 1).

*Data Warehouses* são armazéns de informações relativas às atividades de uma organização em banco de dados de forma consolidada. O *Data Warehouse* possibilita

a análise de grandes volumes de dados, também chamadas de séries históricas, que possibilitam uma melhor análise de eventos passados, oferecendo suporte às tomadas de decisões presentes e a previsão de eventos futuros.

Segundo Palestino (2001, p. 49), um *Data Warehouse* pode ser definido como um banco de dados destinado a sistemas de apoio a decisão. Seus dados são armazenados em estruturas lógicas dimensionais, possibilitando seu processamento por ferramentas especiais (OLAP e *Data Mining*), definidas nos tópicos a seguir.

Por definição, os dados em um *Data Warehouse* não são voláteis, ou seja, eles não mudam, são armazenados apenas para leitura e não podem ser alterados (INMON, 1997).

Após o processo de transformação dos dados analisados nos passos anteriores, os mesmos são armazenados em um *Data Warehouse* ou *Data Mart*.

A diferença de um *Data Warehouse* com relação a um *Data Mart* está apenas no fato de um *Data Warehouse* corresponder a um banco dimensional para a empresa como um todo, enquanto um *Data Mart* é uma parte específica desse banco, ou seja, um dos diversos módulos, uma das medidas das diversas analisadas pelo *Data Warehouse*.

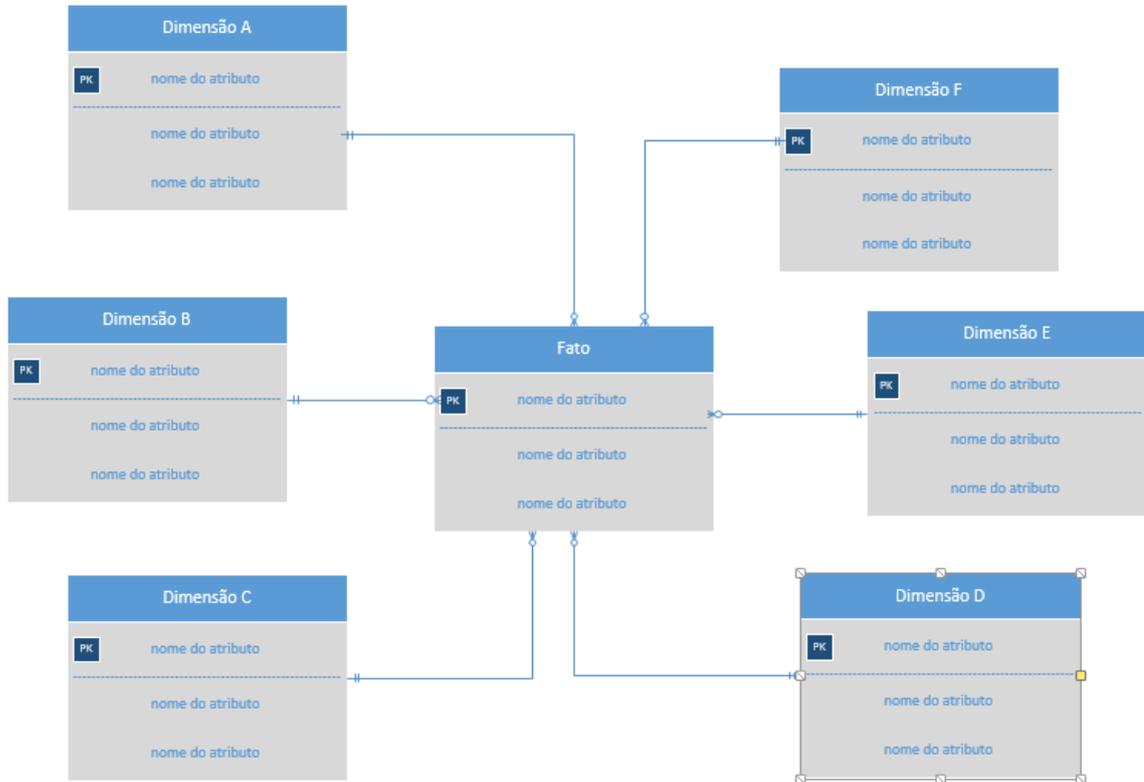
Outra característica importante a ser considerada para o desenvolvimento de um *Data Mart* diz respeito a forma como esse novo banco deve ser modelado. A maioria dos bancos de dados utilizados hoje em dia, e até mesmo o atual banco de dados utilizado na UNISO, possui arquitetura relacional, porém, para a criação de um projeto de *business intelligence* (BI) eficiente, sua modelagem de dados deve ser do tipo dimensional. Para esse tipo de projeto, esta forma de modelagem tem o objetivo de modificar a ordem de distribuição dos dados para trabalhar de maneira mais eficiente com os algoritmos de mineração de dados que será utilizada (PALESTINO, 2001).

A modelagem dimensional é composta basicamente de tabelas-fatos e tabelas-dimensões, além das medidas que também são conhecidas como métricas.

Cada registro em uma tabela-fato pode ser representado como um evento do negócio modelado, enquanto as dimensões podem ser caracterizadas como elementos que compõem o fato registrado, ou seja, uma dimensão é um ponto de vista do fato registrado ou mesmo uma parte do assunto na tabela-fato (NERY, 2013).

Na modelagem dimensional existem dois modelos predominantes para estruturar o banco de dados: o modelo *estrela*, também conhecido como *star schema* mostrado na Figura 2, e o modelo *floco de neve* também conhecido como *snowflake*.

**Figura 2 - Exemplo de Modelo Estrela**



Fonte: Elaborado pelo autor

Dentre as principais características do modelo estrela estão:

- grande redundância de informação armazenada nas tabelas dimensões, pois todas as características dos registros armazenados são replicadas para outros registros da mesma tabela, independentemente de serem iguais ou não, ou seja, comparando com os bancos de dados relacionais que são utilizados no ambiente operacional, esse detalhe de modelagem dimensional é chamado de *desnormalização*;
- ganho expressivo no desempenho para consulta, pois para a coleta das informações são necessárias muito menos ligações de tabelas;
- gera facilidade de entendimento por usuários leigos sobre modelagem.

O modelo floco de neve pode ser considerado uma variação do modelo estrela onde difere nos seguintes pontos:

- As tabelas dimensões estão normalizadas<sup>1</sup>, seguindo os preceitos de normalização de dados, logo não gera redundância de informações como no caso do modelo estrela.
- Perde desempenho nas consultas, pois para coletar as informações necessárias para o processamento dos algoritmos de análise são necessárias muitas ligações entre tabelas.
- Gera complexidade no entendimento do modelo se comparado com o modelo estrela por causa da quantidade de tabelas criadas para a realização da normalização das tabelas dimensões.
- Exige maior processamento na tarefa de mineração de dados.

O modelo que será adotado no desenvolvimento dessa pesquisa foi o modelo estrela.

#### 2.4.5 MINERAÇÃO DE DADOS (DATA-MINING)

*Data-Mining* é um conjunto de métodos de análise estatística e de Inteligência Artificial que são aplicados sobre os dados da empresa. Também denominado mineração de dados, pode ser entendido como um processo automático de descoberta de padrões ou regularidades em grandes conjuntos de dados que

[...] se baseia fortemente em técnicas conhecidas de aprendizagem de máquina, reconhecimento de padrões e estatísticas, para encontrar padrões de dados na etapa de mineração de dados do processo de KDD (FAYYAD, SHAPIRO e SMYTH, 1996; HAN, KAMBER e PEI, 2012; (DINIZ e LOUZADA, 2000, pag. 169).

Dentre os diversos exemplos de utilização podem ser citados: a descoberta de relações entre produtos, classificação de consumidores, previsão de vendas, localização de áreas geográficas potencialmente lucrativas e outros, como mencionado por Luís Alfredo Vidal de Carvalho (CARVALHO, 2001).

Nos tempos atuais, com a criação de um volume cada vez maior de informações gerado nas grandes empresas, ficou cada vez mais difícil adquirir todo o conhecimento necessário para os gestores na tomada de decisão e mesmo identificar conhecimentos que possivelmente estão ocultos nesse grande repositório em tempo

---

<sup>1</sup> Consiste em um conjunto de regras de modelagem de banco de dados relacionais que tem como objetivo evitar redundância nos dados armazenados e garantir integridade e desempenho no banco de dados.

hábil. Logo o conjunto de técnicas de Data Mining veio para esse fim, que é descobrir informações relevantes, como padrões, associações, mudanças, anomalias e estruturas, em grandes quantidades de dados armazenados em banco de dados, depósitos de dados ou outros depósitos de informação.

O Mecanismo de Transformação de dados de baixo nível em informação de alto nível ajuda no processo de tomada de decisões organizacionais, pelo uso de técnicas automáticas de exploração de grandes quantidades de dados, de forma a descobrir novos padrões e relações, que devido a esse volume não seriam descobertos utilizando uma técnica tradicional.

Nesse contexto a etapa de extração de padrões é direcionada ao cumprimento dos objetivos definidos na identificação do problema. Nessa etapa é realizada a escolha, a configuração e execução de um ou mais algoritmos para extração do conhecimento. Por esta etapa ser um processo iterativo, pode ser necessário que seja executada diversas vezes para ajustar o conjunto de parâmetros, visando à obtenção de resultados mais adequados aos objetivos preestabelecidos.

Existem vários algoritmos de mineração de dados, e os utilizados neste trabalho estão caracterizados em dois paradigmas de aprendizado em máquinas (*machine learning*), os quais, conforme descritos por Carvalho (2002) são:

- Supervisionados ou diretos – quando se têm em mãos as consequências possíveis das ações em estudo. Nesta sub-classe o objetivo é utilizar os dados para construir um modelo que explique uma variável de interesse, em função dos dados restantes disponíveis. Na classe de algoritmos analíticos supervisionados encontram-se as árvores de classificação e as árvores de regressão - que são chamados de algoritmos de árvores de decisão – e o algoritmo de seleção de colunas – *Column importance*, e alguns algoritmos de redes neurais.
- Não-supervisionados ou indiretos – quando nada se sabe a respeito do comportamento dos dados e deseja-se fazer um estudo exploratório. Nesta sub-classe o objetivo é encontrar alguma relação entre todas as variáveis. Com algoritmos analíticos não-supervisionados podem ser incluídos os de associação, os de agrupamento – clustering, e alguns algoritmos de redes neurais [...](CARVALHO, 2002, pag. 9)

Para a escolha adequada é necessário ter bem definidos o problema e as funções que especificam cada algoritmo. Dentre as funções mais comumente utilizadas podem ser citadas:

- Classificação: onde sua função é reconhecer fatos com as mesmas características e organizá-los em classes, por exemplo, se um registro

de dados contém o campo Região, então Norte, Sul, Leste e Oeste pode definir uma classe (DEVEDZIC, 2001). Esse processo é realizado em duas etapas, na primeira é definido as classes de dados e essa etapa é vista como um aprendizado supervisionado. Na segunda etapa após o modelo classificador ser estimado, é feita a classificação dos dados que poderá ser usado de forma preditiva para classificar novos registros nas mesmas classes pré-definidas (Han e Kamber, 2006; Diniz e Louzada-Neto, 2000).

- Regressão: conjunto de métodos que permite a interpretação da relação entre variáveis considerando a existência de uma relação entre as variáveis em questão, muito utilizada para explorar todos os tipos de relação de dependência.
- Análise de associação: método que procura determinar uma representação explícita entre objetos, ou seja, conexões presentes nos conjuntos de dados. Explicando com outras palavras, a presença de um item implica necessariamente na presença de outro item no mesmo fato (transação), em geral é representado formalmente através do tipo, se X então Y, considerados corpo e cabeça da regra (DINIZ e LOUZADA, 2000).
- Análise de Sequência: variação da análise associativa que objetiva extrair e registrar desvios e tendências no tempo a fim de reconhecer sequências relevantes que possam ser utilizadas para prever comportamentos. Por exemplo, se todo consumidor que volta a uma loja compra cerveja, todos os futuros consumidores que retornarem à loja do exemplo provavelmente comprará cerveja (CORADINE, LOPES e MACIEL, 2011).
- Sumarização: o objetivo desse método é proporcionar a limpeza dos dados, muito utilizada na etapa de pré-processamento para identificar valores inválidos com o auxílio de medidas estatísticas para variáveis quantitativas e distribuição de frequência de valores para variáveis categóricas (HÄRDLE e SIMAR, 2003).
- Visualização: conjunto de ferramentas visuais para visualização dos dados analisados. Pode ser classificado de maneira resumida em

histogramas, gráficos relacionando atributos ou resumos destes ou representações icônicas onde um dado é representado por uma figura.

Os algoritmos utilizados para análise levando em consideração suas principais funções estão descritas nos tópicos abaixo.

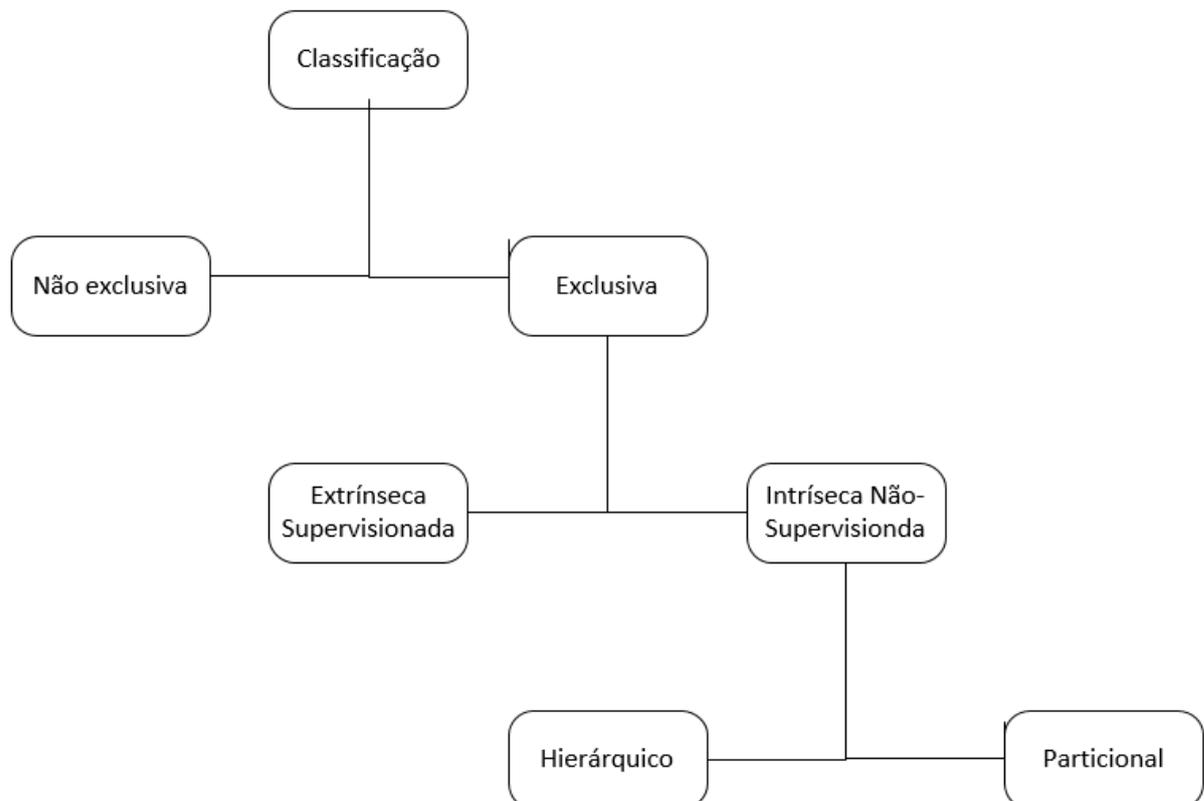
#### 2.4.5.1 Clustering

É um dos algoritmos implementados no trabalho e representa um conjunto de métodos que tem como objetivo agrupar objetos com características similares de uma população heterogênea.

O agrupamento é realizado sem classes pré-definidas. Sua abordagem é difundida em diversas áreas de conhecimento, como por exemplo ciências biológicas (no caso de taxonomia biológica para a classificação de vários grupos de animais), ciências humanas no mapeamento de perfis psiquiátricos dentre outros.

Os métodos de agrupamentos são classificados segundo a figura abaixo:

**Figura 3 - Classificação simplificada dos métodos de e agrupamentos**



Fonte: JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. New Jersey: Prentice Hall, 1988

A diferença da classificação exclusiva para a não exclusiva se dá no fato dos objetos pertencerem a apenas um ou mais grupos (*cluster*).

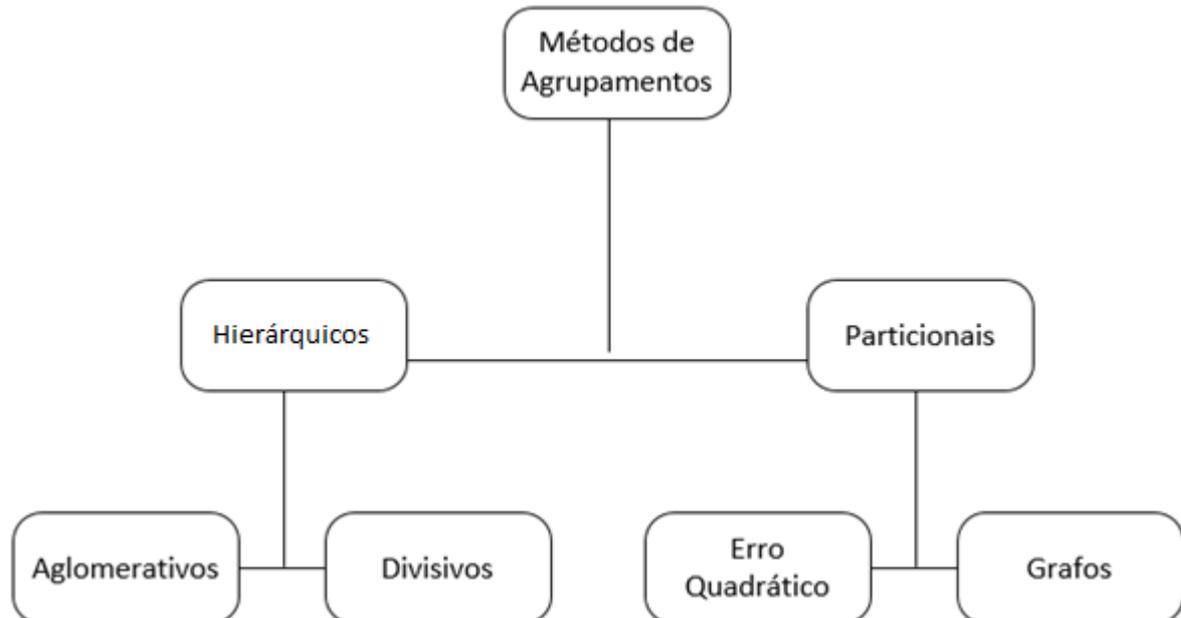
Na classificação exclusiva todos os objetos pertencem a apenas um grupo como por exemplo no agrupamento de pessoas por sexo ou idade.

Na classificação não exclusiva, um objeto pode pertencer a mais de um grupo, como por exemplo no caso de agrupamento por categoria de doença, pois uma pessoa pode ter várias doenças simultaneamente.

No que diz respeito a classificação intrínseca (com relação a classificação extrínseca) a diferença está no fato da classificação extrínseca depender de rótulos para categorizar os objetos, ou seja, depende de um “professor” para parametrizar o processo de classificação como comentado por Jain e Dubes (1998), enquanto que a classificação intrínseca não depende de rótulos. Todo o processo é feito por aprendizado não supervisionado, diferenciando os grupos por intermédio da matriz de proximidade para realizar a classificação.

Como mostrado na Figura 3 a classificação intrínseca se subdivide em dois métodos de agrupamento hierárquico e particional que são detalhados na Figura 4 e explicado com detalhes abaixo.

**Figura 4 - Classificação simplificada dos métodos de agrupamentos**



Fonte: JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. New Jersey: Prentice Hall, 1988

O agrupamento hierárquico trata os objetos como uma estrutura de partições hierarquicamente organizados segundo suas similaridades (JAIN e DUBES, 1988).

Todos os métodos necessitam de uma medida de similaridade entre os objetos a serem agrupados, que é expresso por uma função de distância ou métrica.

Dentre as métricas mais utilizadas estão:

- Distância euclidiana:  $d_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- Quadrado da Euclidiana:  $d_{xy} = \sum_{i=1}^n (x_i - y_i)^2$
- Manhattan:  $d_{xy} = \sum_{i=1}^n |x_i - y_i|$
- Chebychev:  $d_{xy} = \max(|x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|)$
- Correlação:  $r = \frac{\sum xy - \frac{\sum x \cdot \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$

A fim de evitar ruídos e erros nos atributos todos são normalizados no intervalo [0, 1]. Existem várias outras métricas descritas no trabalho de Jain e Dubes (1988), muitas das quais não necessitam normalização e muitas se aplicam somente após uma transformação nos dados.

Depois que os atributos forem mensurados com uma das medidas de similaridade é utilizado o método aglomerativo, que pode utilizar uma das seguintes técnicas:

- Ligação Individual ou Simples ou abordagem do vizinho mais próximo: onde a distância entre dois clusters é a distância entre os pontos mais próximos.
- Ligação completa ou abordagem do vizinho mais distante: onde a distância entre dois clusters é a distância entre os pontos mais distantes.
- Ligação média: Distância entre os clusters é a distância entre os centroides que nada mais é que o centro do agrupamento.
- Método de Ward: Distância entre os clusters é a soma dos quadrados entre os dois clusters.
- Método Centroide: Distância geralmente euclidiana quadrada ou euclidiana simples entre os centroides dos clusters em questão.

No método de agrupamento hierárquico divisivo faz o inverso do método aglomerativo, começando com um grupo contendo todos os objetos e, em passos sucessivos, vão separando e criando agrupamentos menores (CORADINE, LOPES e MACIEL, 2011).

O método de agrupamento Particional faz a divisão de um conjunto de objetos em um número pré-definido de *clusters*. Dentre os métodos que existem, o mais popular é o *k-means*.

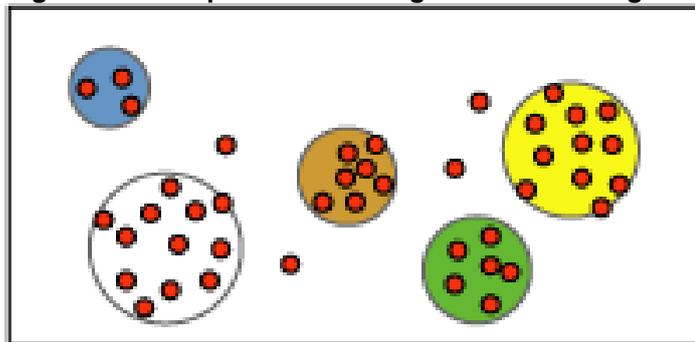
O método *K-means* funciona da seguinte maneira:

1. Escolhe um número K de agrupamentos a particionar o conjunto de dados e posiciona aleatoriamente K centroides;
2. Calcula o centroide para cada um dos K agrupamentos;
3. Atribui cada objeto ao centroide mais próximo;
4. Recalcula o centroide para cada agrupamento;
5. Verifica se atingiu a estabilidade, caso não volta para o passo 3.

O valor do parâmetro K é definido pelo usuário, que deverá escolher o melhor valor para a análise, os dados devem ser numéricos e devem ser comparados através da distância euclidiana já comentada acima para normalização. No entanto, este método foi aperfeiçoado para se obter o melhor número de grupos K a dividir o conjunto, de forma a minimizar a soma das distâncias dos objetos aos respectivos centroides.

No método de particionamento utilizando grafos, todos os vértices são os próprios objetos e as arestas interligam objetos que são considerados similares. As arestas podem ser reduzidas modificando o grau de similaridade, essa configuração de modificação do grau de similaridade é conhecida como corte de peso mínimo, a Figura 5 demonstra um exemplo de clustering com o auxílio de grafos.

**Figura 5 - Exemplo de Clustering com auxílio de grafos**



Fonte: SQL Server 2017. **Microsoft**, 2017. Disponível em < <https://docs.microsoft.com/pt-br/sql/analysis-services/data-mining/microsoft-clustering-algorithm?view=sql-server-2017>>. Acessado em 25 Janeiro 2019.

O trabalho emprega o método de classificação exclusiva e intrínseca que segundo Jain e Dubes (1998) é considerado a essência da análise de agrupamentos.

A análise de agrupamentos ou *clustering* é um conjunto de algoritmos baseados no paradigma de aprendizado não-supervisionado, descrito por Carvalho (2002) e citados na página 32 por extrair algum tipo de organização na estrutura dos dados.

Uma forma usada no trabalho de representar graficamente relações de distância ou correlação é através de dendograma em que cada bifurcação se encontra no nível de distância ou correlação que separa dois objetos (no primeiro nível) e nos demais níveis as bifurcações exprimem as distâncias entre os centros dos grupos. O dendograma é construído hierarquicamente, por meio de aglutinações sucessivas de objetos, com a medida da distância ou correlação a cada aglutinação realizada (JAIN e DUBES, 1988). Assim, um dendograma demonstra a correlação pela distância das bifurcações, onde no primeiro nível separa dois objetos e nos próximos níveis exprimem as distâncias entre os centros dos grupos.

#### 2.4.5.2 Árvores de Decisão

Um dos algoritmos de extração de padrões que também será utilizada no projeto é árvores de decisão. Esse algoritmo fornece uma representação gráfica para um processo de decisão, permitindo mostrar as possíveis consequências e os resultados esperados a partir de uma decisão tomada.

Como definido por Norvig e Russell (2013, p. 811) uma árvore de decisão

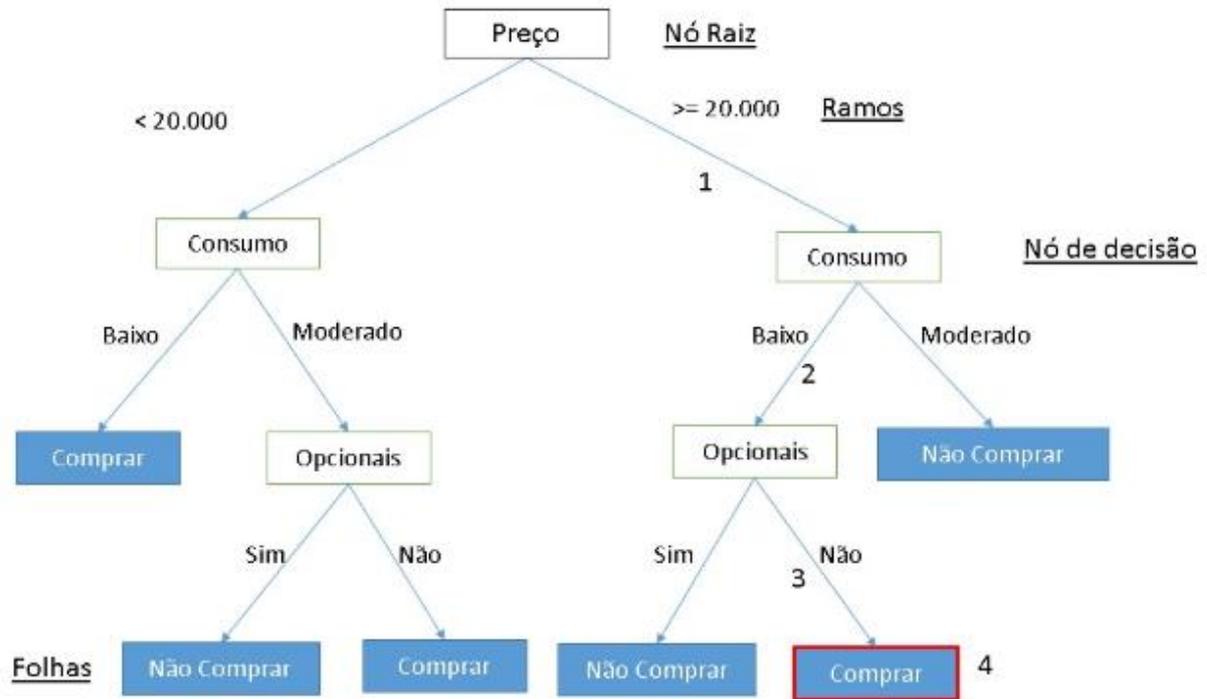
[...] alcança sua decisão executando uma sequência de testes. Cada nó interno na árvore corresponde a um teste do valor de um dos atributos de entrada,  $A_i$ , e as ramificações dos nós são classificadas com os valores possíveis do atributo,  $A_i = V_{ik}$ . Cada nó de folha na árvore especifica o valor a ser retornado pela função.

Colocando com outras palavras segundo Silva et al. (2008, p. 81)

[...] a Árvore de Decisão é um método que permite indicar, de forma gráfica e cronológica, um caminho a ser seguido em um processo de decisão, explicitando etapas a serem cumpridas para o alcance do objetivo pretendido.

Uma árvore de decisão segue o seguinte formato mostrado abaixo:

Figura 6 - Formato geral de uma árvore de decisão



Fonte: GRANATYR, J. et al. Mineração de dados com árvores de decisão. **DevMedia**. Disponível em: <https://www.devmedia.com.br/mineracao-de-dados-com-arvores-de-decisao/31397>. Acesso em: 05 Fevereiro 2019

A Figura 6 mostra um exemplo de árvore de decisão para um processo de escolha de um automóvel por um comprador, onde segundo Granatyr et al. (2019) para a montagem de uma árvore de decisão é importante conhecer algumas definições como ramos, nós, raiz, folha e decisão.

O nó raiz define a primeira condição usada na escolha do automóvel, nesse caso o preço, os nós decisão são divididos conforme as condições apresentadas nos ramos, gerada por intermédio do algoritmo de entropia (mostrado abaixo). O último nó apresentado em azul na Figura 6 demonstra o resultado caso o comprador tome as decisões que o levam por esse caminho.

Dentre os algoritmos para construção da árvore de decisão encontram-se o ID3 – *Induction of Decision trees* apresentado por Quinlan (1979) e seu sucessor C4.5 – *Classifier* apresentado por Quinlan (1993).

Outra forma de representação é gerar as saídas da árvore de decisão como um conjunto de regras em formato SE-ENTÃO (IF-THEN), que procura facilitar o entendimento das regras geradas.

No caso da árvore mostrada na Figura 6, o conjunto de regras em formato SE-ENTÃO ficaria da seguinte maneira:

Se Preço < 20.000 e (Consumo = baixo  
ou Consumo = Moderado e Opcionais = não)  
ou Preço >= 20.000 e Consumo = Baixo e Opcionais = não  
então Comprar.  
senão Não Comprar.

Em sua construção, o algoritmo faz a escolha do melhor atributo para ser usado como raiz da árvore, utilizando para isso o cálculo da entropia da informação (eq.2):

$$Entropia(S) \equiv p_+ \log_2 p_+ - p_- \log_2 p_- \quad (2)$$

Onde:

- S é a coleção de valores do atributo.
- $p_+$  é a soma de exemplos positivos de S.
- $p_-$  é a soma de exemplos negativos de S.

Isso quando os resultados dos atributos tiverem valores booleanos, ou seja, os exemplos variarem entre 0 e 1.

Para os casos onde o atributo aceita diferentes valores a fórmula da entropia a ser seguida é definida como:

$$Entropia(S) \equiv \sum_{i=1}^c -p_i \log_2 p_i \quad (3)$$

Onde  $p_i$  é a proporção de S pertencendo a classe i.

Após o cálculo da entropia deve ser calculado o ganho de informação com a fórmula:

$$Gain(S, A) \equiv Entropia(S) - \sum_{V \in \text{Valores}(A)} \frac{S_v}{S} Entropia(S_v) \quad (4)$$

Árvores de decisão são baseados em *aprendizado supervisionado*, pois obtêm regras lógicas que explicam a dependência das variáveis de interesse (no caso, a evasão) em função dos dados disponíveis.

Para otimizar o tamanho da árvore e evitar o ajuste “viciado” para aquele conjunto de dados (*overfitting*), os dados são subdivididos aleatoriamente em dois conjuntos, um conjunto de treinamento e um conjunto de testes, com base na porcentagem de dados para teste e no número máximo de casos no conjunto de dados

de teste fornecido. O conjunto de treinamento é usado para criar o modelo de mineração, enquanto o conjunto de testes é usado para verificar a precisão do modelo.

#### 2.4.6 INTERPRETAÇÃO / AVALIAÇÃO

Nesta etapa as árvores de decisão e regras de produção são analisadas, para se ter certeza de que foi feita a escolha certa, caso contrário volta-se para uma das etapas anteriores.

### 2.5 MATRIZ DE CORRELAÇÃO DE PEARSON

Matriz de correlação é uma técnica de estatística descritiva com o intuito de identificar o grau de correlação entre todas as variáveis estudadas.

Segundo Larson e Farver (2013, p. 395) uma correlação

[...] é uma relação entre duas variáveis. Os dados podem ser representados por pares ordenados  $(x,y)$ , onde  $x$  é a variável independente (ou exploratória) e  $y$  é a variável dependente (ou resposta).

A aplicação é feita no cruzamento entre todas as variáveis estudadas, onde o cálculo disponibiliza um número chamado coeficiente de correlação, que segundo Larson e Farber (2013, p. 398)

[...] é uma medida da força e direção de uma relação linear entre duas variáveis. O símbolo  $r$  representa o coeficiente de correlação amostral. Uma fórmula para  $r$  é:

$$r = \frac{n \cdot \sum_{i=1}^n (x_i \cdot y_i) - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left[ n \cdot \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] \cdot \left[ n \cdot \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2 \right]}} \quad (5)$$

Onde  $n$  é o número de pares de dados.

O coeficiente de correlação populacional é representado por  $\rho$  (a letra minúscula grega rô, pronuncia-se "row").

O número gerado pelo coeficiente de correlação possui uma amplitude que vai de -1 a 1. Quanto mais próximo o número gerado estiver do -1 ou do 1, maior será a

correlação entre as duas variáveis, quanto mais próximo do zero, mais fraca a correlação linear, sendo 0 a inexistência de correlação e -1 ou 1 a correlação perfeita.

Quando a coeficiente gerar um número negativo, significa que a correlação é inversamente proporcional, ou seja, enquanto uma variável tem aumento de valor, a outra variável diminui.

Quanto a correlação gerar um número positivo, significa que o aumento de valor em uma variável estudada gera aumento de valor na outra variável.

Outro detalhe salientado por Larson e Farber (2013) está no fato de a baixa ou inexistência de correlação entre as variáveis estudadas pela proximidade do valor 0 no coeficiente de correlação nem sempre significar que não existe correlação, pode significar que não há correlação linear.

## **2.6 CONCLUSÃO DO CAPÍTULO**

No capítulo 2 foi mostrado todo o referencial teórico necessário para o entendimento da evasão, desde um aprofundamento sobre sua definição, uma discussão sobre os diversos tipos de evasão de uma instituição de ensino superior e as melhores formas de quantificá-la.

Foi descrita de maneira sucinta a história da universidade de Sorocaba – UNISO, focando em alguns fatos históricos relacionados aos diversos cursos existentes.

Foi descrita, de forma mais detalhada, toda a teoria aplicada aos passos relacionados ao processo de extração de conhecimento, delimitando na etapa de mineração de dados nos algoritmos de *Clustering* e *Árvores de Decisão*.

No capítulo 3 será mostrada a aplicação do embasamento teórico aplicado na análise da evasão dos cursos de graduação da UNISO.

### **3 METODOLOGIA**

O trabalho trata-se de uma pesquisa aplicada de abordagem quantitativa com utilização de métodos e procedimentos estatísticos, cujo objetivo geral é exploratório e analítico, com o auxílio de análise dos dados dos alunos dos cursos de graduação da UNISO nos anos de 2012 a 2018.

No capítulo 3 serão descritos todos os detalhes dos métodos utilizados para a análise dos casos de evasão onde foi subdividido em: contextualização da evasão na UNISO, índices de evasão na UNISO e KDD aplicado para análise da evasão.

#### **3.1 CONTEXTUALIZAÇÃO DA EVASÃO NA UNISO**

A pesquisa foi desenvolvida em duas etapas.

A primeira procurou identificar fórmulas utilizadas para controlar a evolução da evasão para com isso monitorar a evasão no tempo tanto em nível geral, ou seja, evasão dos cursos de graduação como um todo, como em nível de grande área e curso, categorizações que fizeram necessária.

Essa segmentação em índice de evasão geral, grande área e curso foi realizada para identificar exceções, pois mesmo um índice geral não significa que se aplica para todas as grandes áreas ou curso específico e um índice mostrado em uma grande área pode não ser regra para todos os cursos.

A segunda etapa foi aplicar os passos de extração de conhecimento comentados no item 2.4 para criar uma base de dados propícia a geração de conhecimento oculto sobre o tema da evasão na UNISO.

Maiores detalhes sobre a primeira etapa estão descritos no item 3.2 e os passos do KDD para extração de conhecimento sobre a evasão estão descritos no item 3.3.

#### **3.2 ÍNDICES DE EVASÃO NA UNISO**

Para o desenvolvimento dos índices de evasão da UNISO foi utilizada a equação (1) comentada com detalhes no item 2.2.

Para o seu desenvolvimento foi utilizada a ferramenta de geração de relatórios *Reporting Service*, complemento da ferramenta *Visual Studio Enterprise 2015* na versão 14.0.25431.01 *Update 3*.

No relatório gerado para a evasão geral, mostrado no Gráfico 6 e Gráfico 7, pag. 81, foi parametrizado para requisitar dois parâmetros:

- Período Letivo Inicial.
- Período Letivo Final.

O relatório mostra três informações:

- O percentual de evasão anual em formato de gráfico de colunas, onde no eixo x é visualizado os anos letivos, conforme parâmetros selecionados e no eixo y o percentual propriamente dito, os gráficos podem ser visualizados no tópico 4.2 - Análise dos índices de evasão, página 80.
- Um gráfico de comparação entre discentes veteranos e evadidos em formato de gráfico de barras, onde no eixo x é visualizado os anos letivos e no eixo y a quantidade de discentes, como mostrado no Gráfico 8 e Gráfico 9, no tópico 4.2 - Análise dos índices de evasão, página 80.
- Uma tabela com os valores: número de matrícula atual, número de ingressantes atuais, número de matrículas do período letivo anterior, número de concluintes e a taxa de evasão segundo Lobo (2007, p.645).

A visualização do índice de evasão por grande área foi desenvolvida requisitando os seguintes parâmetros:

- Período Letivo Inicial;
- Período Letivo Final;
- Área

O resultado do relatório por grande área, visualizado no tópico 4.2.1- Índices de evasão das grandes áreas de graduação da UNISO, pag. 83, mostra o índice de evasão no período definido, especificamente na grande área selecionada, aplicado em um gráfico de coluna, seguindo o padrão do Gráfico 6, pag. 81, e uma tabela com o ano letivo, número de matrículas, número de ingressantes daquele ano letivo, número de matrículas para o ano letivo anterior, número de concluintes e a taxa de evasão utilizada no gráfico em coluna.

No relatório de evasão por curso foi criado com os seguintes parâmetros:

- Período Letivo Inicial;
- Período Letivo Final;
- Área;
- Curso

A única diferença entre o relatório dos índices de evasão por grande área e curso está no acréscimo do parâmetro curso e seu resultado pode ser visto no tópico 4.2.2, pag. 86.

Onde os gráficos em questão são retirados dos seguintes números gerados que também são mostrados pelo relatório visto na Tabela 15 - Visualização dos dados computados no relatório de evasão geral da UNISO nos períodos de 2012 a 2018, presente na página 82.

Dentre as principais funções dos relatórios mostrados estão:

- base inicial para estudo de discrepâncias como altos e baixos índices de evasão que servirão para direcionar na parametrização das ferramentas utilizadas para extração de conhecimento,
- controle detalhado dos índices da evasão,
- primeiras previsões sobre a evasão.

Todos os relatórios tomam como base o ano letivo, apesar de todo o estudo ser concentrado na análise dos fatos de evasão ocorridos no nível de detalhamento do período letivo.

No relatório da análise da evasão por curso o parâmetro área só é utilizado como filtro para o parâmetro curso.

Maiores detalhes sobre os índices gerados são mostrados no capítulo 4 e no apêndice D ao K, pag. 180.

### **3.3 KDD APLICADO PARA ANÁLISE DA EVASÃO**

Nesse tópico serão detalhados todos os passos necessários para a mineração de dados dos alunos evadidos dos cursos de graduação da UNISO de 2012 a 2018, pontuando todos os passos do KDD – *Knowledge Discovery in Database* desde sua definição até a parametrização dos algoritmos de análise deixando a interpretação dos dados para o tópico 4 onde será mostrado os resultados obtidos.

### 3.3.1 Seleção

No processo de seleção, instruído por Palestino (2001, p. 83) foi necessário definir qual seria o nível de granularidade<sup>2</sup> desejada para os alunos evadidos, ou seja, que atributos seriam importantes registrar no fato da evasão além da definição de qual nível temporal esse fato aconteceu.

Na pesquisa foi descartado o nível temporal “evasão por dia ou mês”, porque a maneira como o banco de dados foi modelado não facilitava esse estudo. Todo plano de aula, que nada mais é que o registro do aluno no dia, naquela disciplina, daquele turno, naquele curso só existe no período letivo corrente, sendo descartado após seu término.

A única informação que fica guardada no período letivo passado são as faltas do aluno, juntamente com o total de carga horária da disciplina cursada. Por esse motivo o nível de granularidade temporal escolhido foi o período letivo que possui o registro de todas as alterações do aluno independente de ser o período corrente ou passado.

Com relação aos atributos da contextualização do aluno no momento da evasão, sua escolha levou em consideração as necessidades na instituição, juntamente com a escolha dos atributos que demonstraram menor índice de problemas de cadastro e redundâncias de informação, chegando nos seguintes atributos:

- Período Letivo: Período letivo do fato registrado, no caso, ano/período, por exemplo, 2018 1º Semestre. No banco de dados foi registrado da seguinte maneira, <ANO><PERÍODO>, ou seja, no exemplo fica 20181.
- Registro do aluno: código que caracteriza o aluno no sistema da UNISO.
- Status no período letivo: Os possíveis status assumidos pelo aluno no período letivo em questão, os possíveis status são:
  - Matriculado;
  - Cancelado
  - Trancado

---

<sup>2</sup> Nível de detalhamento do dado estudado, ou seja, quando menor a granularidade, maior o detalhamento do fato, em contrapartida, quando maior a granularidade, maior a sumarização do fato estudado. Por exemplo, posso estudar quando o discente evade de um curso, tentando identificar o dia, o mês, o semestre ou o ano, quando mais abrangente, maior a granularidade.

- Transferência interna
- Evadido
- Transferência de curso
- Status no curso: a diferença desse campo com o status no período letivo está no fato do aluno ter um curso e um conjunto de períodos letivos. Todo status atribuído ao curso replica em todos os períodos letivos, que replica em todas as disciplinas, mas o status atribuído ao período letivo não replica no curso, ou seja, funciona internamente como um conjunto de processos para manter íntegro as regras acadêmicas atribuídas. O status atribuídos nesse campo são:
  - Cancelado
  - Concluinte
  - Evadido
  - Ingressante
  - Matriculado
  - Trancado
  - Transferência de Curso

Maiores detalhes sobre o número de alunos atribuído a cada status pode ser analisado no apêndice C.

- Nome do curso: Nome do curso do aluno. A listagem de cursos analisados pode ser visualizada no apêndice B.
- Turno: O turno que o aluno estava cursando no momento do período letivo fato. No banco contemplam os seguintes turnos:
  - Integral;
  - Manhã;
  - Noite;
  - Tarde;
- Grade: A matriz curricular do curso. Informação importante, pois caracteriza o conjunto de disciplinas, a carga horária de cada disciplina e o conhecimento/habilidade cobrado.
- Período letivo ingressante: Período letivo que o aluno ingressou.
- Último período letivo: Último período letivo computado no registro do aluno no curso.

- Nome do aluno;
- Tipo de instituição de ensino no segundo grau: Caracteriza se o aluno terminou o segundo grau em uma instituição de ensino pública ou privada.
- Estado civil: Estado civil informado pelo aluno no momento do cadastro, os possíveis estados são:
  - Casado
  - Desquitado
  - Divorciado
  - Não informado
  - Outros
  - Separado
  - Solteiro
  - Viúvo
- Sexo: Masculino ou feminino informado pelo ingressante no momento da inscrição;
- Cor/raça: cor/raça declarada pelo aluno ingressão, os possíveis estados são:
  - Amarela
  - Branca
  - Indígena
  - Não declarado
  - Parda
  - Preta/Negra
- Data de nascimento;
- Grande área: Grande área do curso;

### **3.3.2 Pré-processamento**

Em meados do ano 2000 a Universidade de Sorocaba fez a aquisição de um novo sistema acadêmico e administrativo do segmento educacional e, por esse motivo, teve que realizar todo um processo de migração dos dados que anteriormente eram gerenciados pelo sistema anterior. Essa mudança trouxe diversos benefícios para a instituição, porém enquanto não estabilizada, acarretou em um período com

diversos problemas no trabalho dos dados, como, registros incompletos, falhas geradas no transporte das informações de um sistema para o outro, falta de treinamento dos funcionários que estavam passando por essas mudanças dentre outros motivos. Por esses fatos citados é que a pesquisa se inicia no ano de 2012, término do processo de implantação do sistema.

O ano limite da pesquisa, no caso 2018 - 1º semestre, foi definido levando em consideração que os fatos são registrados por período letivo e a técnica de mineração de dados não trabalha com dados correntes, ou seja, suscetíveis a alterações como mencionado por Inmon (1997).

Para uma perfeita aplicabilidade das técnicas é necessário trabalhar com históricos e, como a mineração foi realizada no ano de 2018 no 2º semestre, a limitação do trabalho ficou para o semestre anterior.

As informações relacionadas ao registro do ingressante antes de se tornar aluno, ou seja, no processo de vestibular, foram descartados para o trabalho, por se tratarem de registros menos completos e não validados com entrega de documentações no ato da matrícula.

Como o trabalho está relacionado a entender qual o estado do aluno academicamente no ato da evasão poucas informações de relevância foram levantadas da questão financeira e as poucas que foram consideradas tiveram que ser transformadas como detalhado no próximo tópico.

### **3.3.3 Transformação**

Nesta etapa, os dados são transformados por intermédio de combinações, também conhecidos como dados derivados, ex.: dados calculados como: idade que é calculada pela data de nascimento, quantidades de determinado item que depende da contagem, ou somatória de valores.

Muitos dos dados definidos para o registro do fato da evasão do aluno foram derivados de transformações para facilitar o entendimento da ocorrência. Abaixo segue uma listagem dos atributos transformados:

- Idade: calculado por intermédio da data de nascimento registrada.
- Carga Horária cursada: somatória da carga horária de todas as disciplinas realizadas pelo aluno.

- Percentual do curso realizado: cálculo que indica a porcentagem do curso realizado pelo aluno, levando em consideração a fórmula, onde  $CHCURSADA$  é a somatória da carga horária de todas as disciplinas realizadas pelo aluno até o semestre letivo analisado e  $CHTOTAL$  é a carga horária total do curso, considerando a somatória da carga horária de todas as disciplinas da matriz curricular do aluno.
- Percentual de faltas: cálculo que indica a porcentagem de faltas do aluno até o registro fato analisado, para o cálculo foi realizado a fórmula  $\frac{QTDFALTAS*100}{CHCURSADA}$ , onde  $QTDFALTAS$  é a somatória das faltas em todas as disciplinas cursadas pelo aluno até aquele momento,  $CHCURSADA$  é a somatória da carga horária de todas as disciplinas realizada pelo aluno até o semestre letivo analisado.
- Média de notas: Valor médio da nota do aluno até aquele fato analisado, para geração desses dados foi considerado a seguinte fórmula  $\frac{SNota}{CNota}$ , onde  $SNota$  é a somatória de todas as disciplinas cursadas pelo aluno até o período letivo analisado e  $CNota$  é a contagem das disciplinas concluídas até o período letivo analisado.
- Quantidade de faltas: quantidade de faltas que o aluno teve no acumulado de todas as disciplinas cursadas até o registro fato analisado.
- Quantidade de aprovadas: somatória de disciplinas aprovadas no aluno até o registro fato analisado.
- Quantidade de reprovadas: somatória de disciplinas reprovadas do aluno até o registro fato analisado.
- Quantidade de reprovações por falta: somatória de disciplinas reprovadas, cujo motivo está relacionado a ultrapassar o limite de faltas estipulados pela universidade. Todo aluno que ultrapassa o limite de 75% de faltas automaticamente é reprovado por faltas.
- Quantidade de dispensas: somatória de disciplinas dispensadas do aluno, isso acontece por aproveitamento de disciplinas quando o aluno é transferido de curso ou instituição de ensino ou realocamento de matriz curricular.

- Quantidade de cancelamentos: somatória de disciplinas canceladas pelo aluno até o fato analisado.
- Quantidade de trancamentos: somatória das disciplinas trancadas pelo aluno até o fato analisado.
- Quantidade de inconcluídos: somatória das disciplinas inconcluídas pelo aluno até o fato analisado. Essa situação acontece com disciplinas como estágio e atividades complementares, essas disciplinas só são concluídas no histórico após o término do curso.
- Total de matrículas: somatória do total de disciplinas matriculadas pelo aluno até o fato analisado.
- Valor médio do curso: Média do valor do curso do aluno no período de 2012 até o período atual.
- Quantidade de períodos: quantidade de períodos cursados pelo aluno até o fato analisado. Todos os cursos de graduação analisados funcionam em períodos letivos e cada período equivale a um semestre.
- Tem bolsa: atributo para identificar se o aluno no período letivo analisado tinha bolsa de estudo.
- Quantidade de acordos: Somatória da quantidade de acordos financeiros realizados pelo aluno até o registro fato analisado.
- Quantidade de dias: Quantidade de dias computados desde o momento que o aluno ingressou no curso até o fato analisado.
- Quantidade de meses: Quantidade de meses computados desde o momento que o aluno ingressou no curso até o fato analisado.
- Outro curso: Status booleano que indica que o aluno já cursou um período letivo em um curso diferente do atual, onde a situação no período letivo estava como *matriculado*, *cancelado* ou *trancado*.
- Indicação de transferência: Indica que o aluno realizou ou não no período letivo analisado um dos tipos de transferências citadas abaixo:
  - Transferência interna: Mudança de matriz curricular.
  - Transferência de Turno: Mudança de turno.
  - Transferência de Curso: Mudança de curso.
  - Evadido: campo *booleano* que indica evasão do aluno, apesar de existir no sistema um estado evadido como mostrado no status de

curso e situação de período letivo no tópico 3.3.1, para o ministério da educação (MEC) é passado como evadido os alunos categorizados como evadido e cancelado e para atender a categoria de evasão de curso como comentado no tópico 2.1, também é considerado evadido no curso os alunos que realizaram transferência de Curso. Maiores detalhes sobre o leiaute de importação dos dados para o INEP podem ser visualizados no anexo A.

### **3.3.4 Data Warehouse e Data Mart**

Os dados selecionados e pré-processados são transformados e armazenados no *Data Mart* preparado especificamente para registrar fatos sobre a evasão dos cursos de graduação da UNISO.

Como mencionado no item 2.4.4 para a utilização de ferramentas de mineração é necessário criar um banco de dados estruturado de maneira dimensional e para a pesquisa foi definida a estrutura segundo a Figura 7 na pag. 53.

As tabelas e campos seguiram o padrão de apresentação em letras maiúsculas, a tabela fato, necessária para o registro de todos os casos de evasão foi chamada de FATOEVASAO e as tabelas com as dimensões iniciaram seu nome com a palavra DIMENSAO, onde foi definido as seguintes tabelas:

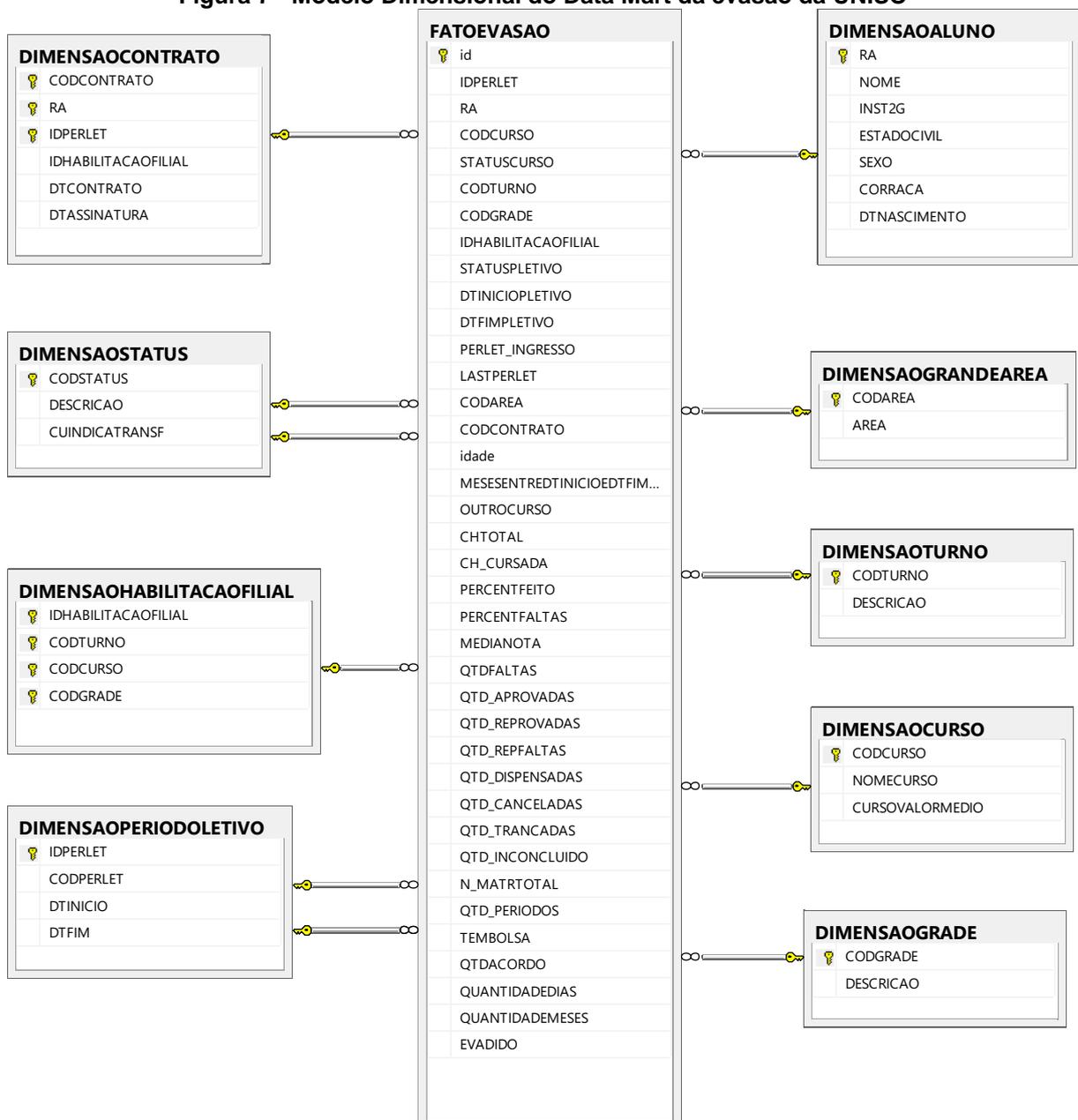
- Contrato: Com informações sobre todos os contratos relacionados as questões financeiras dos alunos de graduação.
- Status: Registro de todos os status utilizados na tabela fato, possui duas ligações na tabela fato devido ao seu relacionamento com o status no período letivo e status no curso.
- HabilidadeFilial: Detalhes sobre todas as matrizes curriculares dos cursos nos turnos.
- PeríodoLetivo: períodos letivos utilizados pela tabela fato. Possui ligação com o período letivo do aluno no fato, registrando duas informações: período letivo de ingresso e último período letivo cursado.
- Aluno: Detalhes sobre todos os alunos vinculados ao registro.
- GrandeArea: Listagem com as grandes áreas dos cursos mensurados na tabela fato.

- Turno: Informações sobre os turnos categorizados.
- Curso: Informações pertinentes aos cursos descritos.
- Grade (matriz curricular): Listagem com detalhes de todas as matrizes curriculares.

Na Figura 7 é mostrada o diagrama do modelo dimensional do *Data Mart* gerado, estruturado em formato de modelo estrela.

Os comandos em linguagem SQL necessários para a criação do *Data Mart* estão detalhados no apêndice A.

**Figura 7 - Modelo Dimensional do Data Mart da evasão da UNISO**



Fonte: Elaborado pelo autor

### 3.3.5 Mineração de Dados (*DATA – MINING*)

Para a realização da mineração de dados, foi necessário o auxílio de um conjunto de ferramentas e técnicas que no trabalho em questão foi definido o sistema gerenciador de banco de dados *SQL Server* e a ferramenta *Analisis Service*.

Nos próximos tópicos serão explicados o que são cada um, em qual momento foram utilizados e quais as configurações necessárias para atingir os resultados mostrados no capítulo 4.

#### 3.3.5.1 Sistema de gerenciamento de banco de dados SQL Server

A UNISO utiliza atualmente o sistema de gerenciamento de banco de dados SQL Server 2008 R2 e para acesso aos dados e desenvolvimento do trabalho foi utilizado o SQL Server 2014.

Essa ferramenta foi necessária para a realização de consultas dos dados para execução dos processos de seleção, pré-processamento e transformação dos dados com o auxílio da linguagem SQL<sup>3</sup>.

Também foi necessária a criação do *Data Mart*, criado especificamente para registrar os fatos de evasão importantes no processo de mineração.

Os registros dos fatos de evasão são processados uma vez por semestre e sempre do semestre anterior ao atual, pois além de ser um processo que demanda tempo do servidor, podendo acarretar travamentos na parte operacional.

Outro motivo de não trabalhar com registros do semestre corrente, está no fato da mineração de dados trabalhar com registros de histórico, ou seja, alterações constantes no histórico afetam diretamente no resultado da mineração.

Após a criação do *Data Mart* a mineração foi feita com o auxílio de:

---

<sup>3</sup> Structured Query Language, traduzida como linguagem de consulta estruturada, sua finalidade é fornecer uma linguagem de manipulação, definição e controle de dados para criação e administração dos diversos bancos de dados controlados pelo sistema gerenciador de banco de dados de arquitetura relacional.

- Desenvolvimento de relatório com a ferramenta *reporting service*, discutido no tópico 3.3.5.2,
- Consulta OLAP (*On-Line Analytical Processing*), discutido no tópico 3.3.5.3,
- Algoritmos de *Clustering* e árvore de decisão com o auxílio da ferramenta *Analisis Service*, discutido no tópico 3.3.5.4.

### 3.3.5.2 Reporting Service

Para desenvolvimento dos relatórios de análise de evasão geral, por grande área e por curso citados no tópico 3.2 foi utilizado a ferramenta Visual Studio na versão 2015, *update 3*, com instalação da extensão *SQL Server Data Tools in Visual Studio 2015*.

Essa extensão foi necessária, pois disponibilizou a IDE<sup>4</sup> a possibilidade de criação de projetos de *Reporting Service* e *Analisis Service*, que será discutida no próximo tópico.

O *Reporting Service* é uma ferramenta de desenvolvimento de relatórios, que utiliza a linguagem SQL, especificamente o comando *selects*, para trazer as informações necessárias e disponibilizar um conjunto de recursos para a visualização adequada, possibilitando demonstrações de gráficos e planilhas com filtros pré-definidos pelo desenvolvedor.

### 3.3.5.3 OLAP – On-Line Analytical Processing

Termo em inglês que se refere a *On-Line Analytical Processing*, e trata-se da capacidade de analisar grandes volumes de informações por várias perspectivas de um *Data Warehouse* e/ou *Data Mart*.

No projeto, o OLAP foi realizado pela ferramenta Microsoft Visual Studio, com o tipo de projeto de *Business Intelligence*.

---

<sup>4</sup> Integrated Development Environment ou ambiente de Desenvolvimento Integrado.

<sup>5</sup> Um dos comandos da linguagem SQL. Pertence ao conjunto de comando relacionados a manipulação de dados em banco de dados relacional e serve especificamente para realizar consultas.

Após a configuração para acesso ao *Data Mart* criado, para o recurso em questão foi selecionado o item *Cubes*, onde foi configurado para a criação de cubos e dimensionados para demonstrar os dados sobre diversas perspectivas (dimensões).

Com o auxílio da criação de cubos foi possível dispor de recursos de operadores dimensionais como:

- Conceitos de ponto e plano (*Slicing*): importante para identificar por exemplo os alunos por grande área específica ou estado civil específico.
- Cubo (*Dicing*): cruzamento de várias dimensões no montante de discentes analisados para realizar o filtro do resultado, dentre os vários momentos utilizados no desenvolvimento do trabalho, pode destacar a contagem de evadidos por curso, turno, faixa de idade, dentre outras dimensões utilizadas.
- Pivotamento ou rotação: Em muitos momentos o conceito de cubo aplicado, explicado no item acima, precisou ser rotacionado para variar as visões dos resultados.
- *Drill-Down*: conceito relacionado ao fato de sair de um nível mais alto do resultado mostrado para buscar informações mais detalhadas, pode ser exemplificado na pesquisa quando de uma visão do total de evadidos, detalhamos o resultado por grande área e de uma visão da grande área detalhamos descendo mais um nível na granularidade para atingirmos o mesmo resultado em uma visão específica de curso.
- *Drill-Up*: É o inverso do conceito de *drill-down*, onde de um resultado específico, vai subindo o nível de granularidade para um resultado geral, ou seja, do curso pode-se subir para uma visão de grande área e da grande área para a visão de todos os evadidos.

Outra forma utilizada foi com o auxílio da ferramenta Microsoft Excel versão 2013, com a realização de uma conexão direta no *Data Mart* e manuseio das informações com o auxílio do recurso do Excel chamado de tabela dinâmica.

A conexão direta no *Data Mart*, pode ser realizada no Excel pelo menu Dados, opção Conexões. Essa opção também pode conectar diretamente no cubo criado.

#### 3.3.5.4 *Analisis Service*

Como já comentado no tópico 3.3.5.2, o *Analisis Service* 2015, também é um dos recursos instalados na IDE Visual Studio 2015 com o auxílio da extensão *SQL Server Data Tools in Visual Studio 2015*.

Antes de sua utilização foi criado o *Data Mart* dos casos de evasão, pois a primeira configuração necessária na ferramenta de análise é a conexão com o banco de dados.

Esse tipo de projeto disponibiliza recursos de criação de cubos, recurso para visualização de consultas OLAP e métodos de mineração, dentre elas:

- Árvores de Decisão da Microsoft,
- *Microsoft Clustering*,
- MSC,
- MTS,
- Naive Bayes da Microsoft,
- Rede Neural da Microsoft,
- Regras de Associação da Microsoft,
- Regressão Linear da Microsoft,
- Regressão Logística da Microsoft.

Nessa pesquisa, os algoritmos utilizados foram: *Microsoft Clustering* e *Árvores de Decisão da Microsoft*.

A escolha se fez devido aos tipos de tarefas que tradicionalmente cada algoritmo realiza, tomando por base o objetivo da pesquisa.

Dentre as principais tarefas realizadas pelo algoritmo de árvore de decisão cruciais em sua escolha, pode-se destacar a previsão de atributos discretos e contínuos (Microsoft Docs, 2017).

A escolha do algoritmo de *clustering* foi feita principalmente pela necessidade de localização de grupos semelhantes (Microsoft Docs, 2017).

A ferramenta *Microsoft Analisis Service* fornece

dois métodos para criar clusters e atribuir pontos de dados aos clusters. O primeiro, o algoritmo *K-means*, é um método de cluster rígido. Isso significa que um ponto de dados pode pertencer somente a um cluster e que uma única probabilidade é calculada para a associação de cada ponto de dados nesse cluster. O segundo

método, de *Maximização de Expectativa*(EM), é um método de *cluster flexível*. Isso significa que um ponto de dados sempre pertence a vários clusters e que uma probabilidade é calculada para cada combinação de ponto de dados e cluster.

Você pode escolher o algoritmo que será usado definindo o parâmetro `CLUSTERING_METHOD`. O método padrão de cluster é o EM evolutivo (Microsoft Docs, 2017).

O método de *clustering* foi configurado para utilizar o método de maximização de expectativa (EM).

No cluster EM, o algoritmo refina de modo iterativo um modelo de *clustering* inicial para ajustar os dados e determina a probabilidade de um ponto de dados existir no cluster. O algoritmo termina o processo quando o modelo de probabilidade ajusta os dados. A função usada para determinar o ajuste é a probabilidade de log dos dados de acordo com o modelo (Microsoft Docs, 2017).

Outro detalhe destacado pela Microsoft é que

Os resultados do método de cluster EM são probabilidades. Isso significa que cada ponto de dados pertence a todos os clusters, mas cada atribuição de um ponto de dados a um cluster tem uma probabilidade diferente. Como o método permite a sobreposição dos clusters, a soma dos itens de todos os clusters pode ultrapassar o total de itens do conjunto de treinamento. Nos resultados de modelo de mineração, as pontuações que indicam suporte são ajustadas para cumprir esse requisito (Microsoft Docs, 2017).

A Microsoft utiliza o algoritmo EM como padrão no lugar do *K-means* pelos seguintes motivos:

- Requer no máximo uma verificação de banco de dados.
- Funciona apesar da memória limitada (RAM).
- Pode usar um cursor de somente avanço.
- Supera as abordagens de amostragem.

Com relação ao algoritmo de árvores de decisão a Microsoft dá as seguintes referências técnicas.

O algoritmo Árvores de Decisão da Microsoft é um híbrido que incorpora métodos diferentes para a criação de uma árvore e dá suporte a várias tarefas analíticas, incluindo regressão, classificação e associação. O algoritmo Árvores de Decisão da Microsoft dá suporte à modelagem de atributos discretos e contínuos (Microsoft Docs, 2017).

Com relação a pontuação do ganho de informação do método de árvore de decisão e *clustering* da Microsoft trabalha com três fórmulas para colunas com dados

discretos e diferenciados: entropia de Shannon, rede Bayesiana com K2 e rede Bayesiana com uma distribuição Dirichlet uniforme (Microsoft Docs, 2017).

A pontuação do ganho de informação, também conhecida pela ferramenta Microsoft *Analysis Service* como seleção de recursos é utilizado para impor um corte no número de atributos procurando identificar as melhores colunas, a fim de melhorar a qualidade dos modelos de várias maneiras como:

- Dados com ruído ou redundantes dificultam a descoberta de padrões significativos.
- Se o conjunto de dados for de grande dimensão, a maioria que algoritmos de mineração de dados exigirá um conjunto de dados de treinamento muito maior (Microsoft Docs, 2017).

A sugestão da ferramenta com relação a pontuação do ganho de informação para a pesquisa pode ser visualizada na Figura 9.

Para configuração do projeto, na ferramenta foi selecionado o tópico *Mining Structures* e configurado os atributos a serem utilizados para análise como mostrado na Figura 8.

Figura 8 - Seleção dos atributos para mineração de dados

Structure ↑	Geral	Geral arvore
	☉☉ Microsoft_Clustering	☐☐☐ Microsoft_Decision_Trees
AREA	➔ Input	☒ Predict
CODCURSO	➔ Input	➔ Input
CORRACA	➔ Input	☒ Predict
CURSO	➔ Input	☒ Predict
CURSOVALORMEDIO	➔ Input	☒ Predict
DESCSTATUSPLETIVO	➔ Input	➔ Input
DTFIMPLETIVO	➔ Input	☒ Ignore
ESTADOCIVIL	➔ Input	☒ Predict
EVADIDO	☒ PredictOnly	☒ Predict
FAIXA IDADE	➔ Input	☒ Predict
Id	☒ Key	☒ Key
INST2G	➔ Input	☒ Predict
LASTPERLET	➔ Input	➔ Input
MESESENTREDTINICIOEDT...	➔ Input	➔ Input
N MATRTOTAL	➔ Input	☒ Predict
QTD APROVADAS	➔ Input	➔ Input
QTD CANCELADAS	➔ Input	➔ Input
QTD DISPENSADAS	➔ Input	➔ Input
QTD PERIODOS	➔ Input	➔ Input
QTD REPFALTAS	➔ Input	➔ Input
QTD REPROVADAS	➔ Input	➔ Input
QTD TRACADAS	➔ Input	➔ Input
QTDACORDO	➔ Input	➔ Input
QUANTIDADEMESES	➔ Input	➔ Input
SEXO	➔ Input	☒ Predict
STATUSCURSO	➔ Input	☒ Predict
TEMBOLSA	➔ Input	☒ Predict
TURNO	➔ Input	☒ Predict

Fonte: Próprio autor

A descrição dos atributos mostrados na Figura 8 podem ser visualizadas no Quadro 4 - Descrição dos atributos usados na mineração de dados que se encontra na pág. 245

Na Figura 8, pode-se observar que a segunda e terceira colunas além de um nome definido pelo usuário do sistema, demonstra qual algoritmo será aplicado e em cada coluna, como o atributo será trabalhado. Dentre as possibilidades existentes foram utilizadas: *Input*, *PredictOnly*, *Key* e *Predict*.

A configuração *Input*, define no algoritmo que o atributo será utilizado como um dado de entrada a ser trabalhado, do caso do *Clustering* pode ser utilizado como

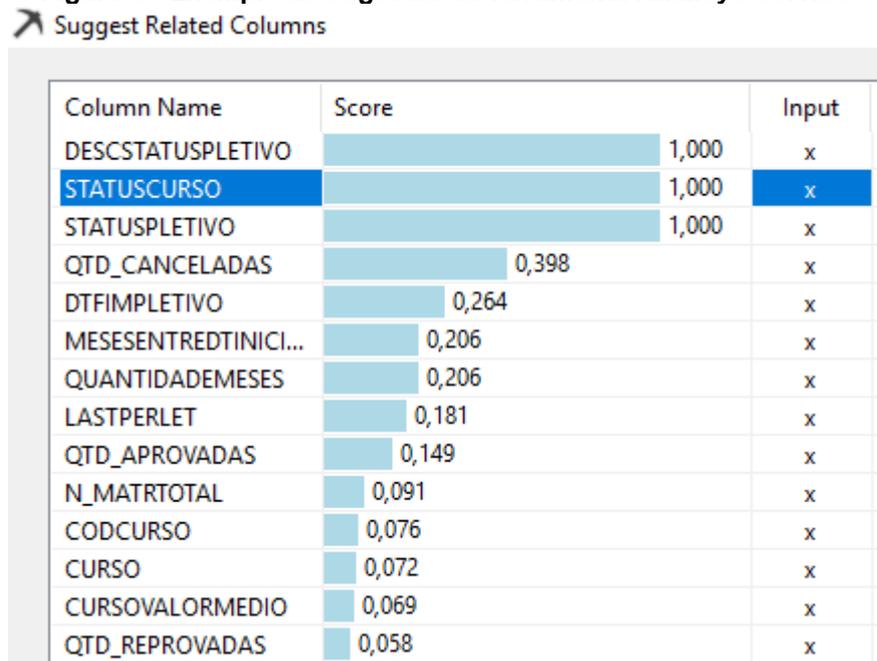
característica a ser considerado para o agrupamento e no algoritmo de árvore de decisão uma possibilidade de nó.

A configuração *Key*, define qual atributo é a chave do registro do *Data Mart*.

A configuração *PredictOnly*, define qual é o atributo a ser considerado para previsão, a diferença entre a configuração *Predict* está no fato do atributo configurado como *PredictOnly* não ser considerado um nó no algoritmo de árvore de decisão, apenas resultado para previsão.

A configuração *Ignore*, ignora o atributo selecionado no algoritmo utilizado.

A própria ferramenta sugere colunas para executar o algoritmo, após a seleção do atributo chave e do atributo a serem previsto, como mostrado na Figura 9.

**Figura 9 - Exemplo de sugestão da ferramenta Analisys Service**  


Column Name	Score	Input
DESCSTATUSPLETIVO	1,000	x
STATUSCURSO	1,000	x
STATUSPLETIVO	1,000	x
QTD_CANCELADAS	0,398	x
DTFIMPLETIVO	0,264	x
MESESENTREDTINICI...	0,206	x
QUANTIDADEMESES	0,206	x
LASTPERLET	0,181	x
QTD_APROVADAS	0,149	x
N_MATRTOTAL	0,091	x
CODCURSO	0,076	x
CURSO	0,072	x
CURSOVALORMEDIO	0,069	x
QTD_REPROVADAS	0,058	x

Fonte: Próprio autor

A descrição dos nomes de colunas mostrados na Figura 9 podem ser visualizados no Quadro 4 que se encontra na página 245.

Após a escolha dos atributos a serem analisados é necessário definir se os atributos são contínuos ou discretos, como mostrado na Figura 10. Por padrão, essa etapa já sugere os tipos, mas pode ser alterado pelo usuário.

**Figura 10 - Definição do tipo de dado na ferramenta Analysis Service**

Data Mining Wizard

**Specify Columns' Content and Data Type**  
Specify mining structure columns' content and data type.

Mining model structure:

Columns	Content Type	Data Type
AREA	Discrete	Text
CODCURSO	Discrete	Text
CORRACA	Discrete	Text
CURSO	Discrete	Text
CURSOVALORMEDIO	Continuous	Double
DESCSTATUSPLETIVO	Discrete	Text
DTFIMPLETIVO	Continuous	Date
ESTADOCIVIL	Discrete	Text
EVADIDO	Discrete	Text
FAIXA IDADE	Discrete	Text
Id	Key	Text
INST2G	Discrete	Text
LASTPERLET	Discrete	Text
MESESENTREDTINICIOEDTFIMPLETIVO	Continuous	Long
N MATRTOTAL	Continuous	Long
QTD APROVADAS	Continuous	Long
QTD CANCELADAS	Continuous	Long
QTD DISPENSADAS	Continuous	Long
QTD PERIODOS	Continuous	Long
QTD REPFALTAS	Continuous	Long
QTD REPROVADAS	Continuous	Long
QTD TRANCADAS	Continuous	Long
QTDACORDO	Continuous	Long
QUANTIDADEMESES	Continuous	Long
SEXO	Discrete	Text
STATUSCURSO	Discrete	Text
TEMBOLSA	Discrete	Text
TURNNO	Discrete	Text

Detect continuous or discrete for numeric columns:

Detect

< Back   Next >   Finish >>|   Cancel

Fonte: Próprio autor

A descrição das colunas mostradas na Figura 10 pode ser visualizada no Quadro 4 que se encontra na página 245.

A próxima etapa de configuração é definir a porcentagem do conjunto de treinamento e o limite do número total de casos no conjunto de testes (Figura 11).

Para a mineração foi deixado o valor padrão.

**Figura 11 - Configuração do conjunto de treinamento e teste no Analisis Service**

**Create Testing Set**  
Specify the number of cases to be reserved for model testing.

Percentage of data for testing:  %

Maximum number of cases in testing data set:

Description:

Input data will be randomly split into two sets, a training set and a testing set, based on the percentage of data for testing and maximum number of cases in testing data set you provide. The training set is used to create the mining model. The testing set is used to check model accuracy.

[Percentage of data for testing] specifies percentages of cases reserved for testing set.  
[Maximum number of cases in testing data set] limits total number of cases in the testing set.  
If both values are specified, both limits are enforced.

< Back   **Next >**   Finish >>|   Cancel

Fonte: Próprio autor

### 3.4 MATRIZ DE CORRELAÇÃO DE PEARSON E DENDOGRAMA

Para criação da matriz de correlação de Pearson foi utilizado a ferramenta Microsoft Excel na versão 2013.

Os dados foram extraídos diretamente do *Data Mart* para a faixa de tempo de estudo do 1º semestre de 2012 ao 1º semestre de 2018.

A fórmula de correlação trabalha apenas com campos numéricos, porém vários campos extraídos do banco de dados estão no formato nominal onde para aplicação da fórmula fez-se necessário realizar a conversão de nominal para binário.

A conversão é necessária para algoritmos que dependem de dados somente numéricos, como *clustering* e redes neurais. As árvores de decisão fazem contagem da ocorrência de cada valor nominal, não sendo necessária a conversão.

Para o tratamento desses casos foram adotadas duas medidas:

- Em casos onde os campos traziam dois valores possíveis foi codificado como 0 (falso) e 1 (verdadeiro), por exemplo: no campo evadido quando a resposta é afirmativa o valor apresentado é o número 1.
- Em casos onde os campos traziam mais de dois valores, para cada valor foi criado um novo campo, seguindo a mesma lógica do verdadeiro e falso para quando essa informação estava afirmativa para aquela categoria, por exemplo: no campo turno, se os possíveis valores são manhã, tarde e noite, quando a resposta for noite, será apresentado três campos com o nome dos turnos, porém apenas no campo com título noite será apresentado o número 1, enquanto nos campos manhã e tarde a resposta será o número 0.

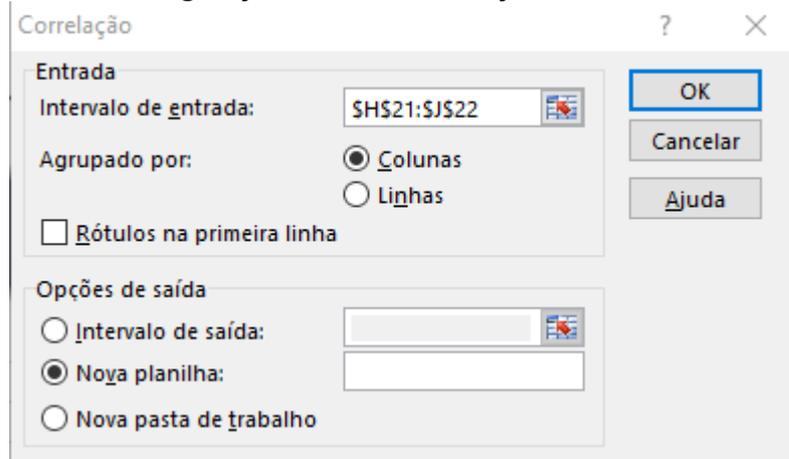
Na ferramenta Excel, após colocar os dados do *Data Mart* em formato de colunas em uma planilha do Excel, foi necessário seguir os seguintes passos para liberar o recurso utilizado:

1. Abrir o menu de opções.
2. Selecionar no menu lateral a opção “Suplementos”, o item “Ferramenta de Análise” e clicar no botão “Ir”.
3. Após aberto o menu de suplementos disponíveis, seleciona a opção “Ferramentas de Análise” e “Ferramentas de Análise – VBA”.

Terminado essa configuração, aparecerá no menu superior chamado “Dados” o item “Análise” e nesse tópico o botão “Análise de Dados”, ao selecioná-lo, dentre as várias opções relacionadas à área de estatística, utilize a opção “Correlação”.

Selecione a entrada de dados previamente acrescentado do *Data Mart*, mantém a seleção de agrupamento por colunas como mostrado na Figura 12 e seleciona o item “Rótulos na primeira linha”, caso tenha descrição no início de cada coluna exportada do *Data Mart*.

**Figura 12 – Configuração do Excel na criação de matriz de correlação**



Fonte: Próprio autor

Para a criação do dendrograma foi gerado um gráfico em escala de cores conhecido como *mapa de calor*, onde a diagonal principal representa a correlação perfeita devido ao cruzamento de uma variável com ela mesma.

No gráfico gerado a cor vermelha representa correlação positiva, a cor azul representa correlação negativa e a cor verde neutralidade ou não correlacionada.

Para a geração do gráfico os dados do *Data Mart* precisaram ser convertidos para um arquivo em formato CSV<sup>6</sup> e carregados em Python<sup>7</sup>, com o auxílio do código abaixo:

```
import pandas as pd # biblioteca para dataframes (tabelas)
import seaborn as sb # biblioteca para gráficos

dfDados = pd.read_csv('_sheetDadosDaMatriz.csv', index_col='id') # carga do
CSV

# apagando a coluna 0, que é um contador (números de linhas):
dfDados = dfDados.drop(dfDados.columns[[0]], axis=1)

colunas = dfDados.columns # para verificação
amostra = dfDados.iloc[0] # para verificação
idx = dfDados.index # para verificação
```

<sup>6</sup> Texto separado por um delimitador (vírgula), e quebra de linha para delimitação do registro.

<sup>7</sup> Linguagem de programação.

```
# agrupamento por semelhança:

# dendrogram + heatmap:
sb.set_context('paper', font_scale=1.0)
sb.clustermap(dfDados.corr(), metric='euclidean', method = 'ward',
              cmap='rainbow', figsize=(40, 40) )

# sb.clustermap(dfDados.corr(), metric='correlation', method = 'average',
#               cmap='rainbow', figsize=(30, 30) )
```

### 3.5 CONCLUSÃO DO CAPÍTULO

Neste capítulo foram detalhados todos os métodos necessários para a identificação dos índices de evasão discriminados por curso, grande área e geral na dimensão temporal semestral. Foram discutidos os motivos das escolhas desses parâmetros e mostrado as ferramentas utilizadas e desenvolvidas para o levantamento dos resultados. Foi detalhado como foi aplicado o processo de KDD para a extração de conhecimento desde a seleção dos dados até o desenvolvimento do *Data Mart*.

## 4 RESULTADOS E DISCUSSÃO

Neste capítulo são mostrados todos os resultados obtidos com a pesquisa, separados em *resultados previsíveis e regras ocultas*.

São considerados resultados previsíveis todo resultado gerado por intermédio de consultas *ad hoc* realizadas no Data Mart e pontuadas com o auxílio de OLAP (*On-Line Analytical Processing*), ou aplicação de fórmulas mostradas no tópico 2.2 com o intuito de mensurar a evasão da faixa de tempo analisada, ou seja, medir a evasão entre os anos de 2012 a 2018. São obtidos diretamente pela visualização de gráficos ou aplicando modelos de regressão, por exemplo.

As regras ocultas são as regras geradas pelo processo de mineração de dados por intermédio da indução de todo histórico das evasões ocorridas entre os anos de 2012 e 2018 com o auxílio das técnicas de *Clustering* e *Árvores de Decisão*.

### 4.1 Análise do perfil dos alunos/cursos

Foi contabilizado na análise um total de 29.303 alunos, presentes entre os períodos letivos do 1º semestre de 2012 ao 1º semestre de 2018. Como os dados foram analisados no final do ano de 2018, não foi considerado para a análise o período letivo do 2º semestre de 2018, pois para um correto estudo do período letivo em questão seria necessário realizar a análise a partir de fevereiro de 2019, prazo necessário para a definição dos status dos alunos atuais no 2º semestre de 2018.

Como o registro de alunos é único, no sistema da UNISO, ou seja, não muda mesmo que o aluno termine e comece novo curso, a quantidade de cursos realizados pelos 29.303 alunos é de 31.799 dentre o montante de cursos existentes mostrados no apêndice B, esse número não indica que o aluno terminou um curso para iniciar outro, pois uma transferência de curso também contabilizará o mesmo aluno em dois cursos, mas indica que existiu na faixa de tempo estudada o montante de 31.799 matrículas realizadas nos cursos demonstrando nas tabelas abaixo, dispostos por grande área.

**Tabela 2 - Quantidade de alunos dos cursos da área de Ciências Agrárias dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
AGRONOMIA (BACHARELADO)	98
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	403
ENGENHARIA AGRÔNOMICA	78
MEDICINA VETERINÁRIA	1069
<b>Total Geral</b>	<b>1648</b>

Fonte: Próprio autor

**Tabela 3 - Quantidade de alunos dos cursos de Ciências Biológicas dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
CIÊNCIAS BIOLÓGICAS (BACHARELADO)	232
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO AMBIENTAL (EAD)	39
GESTÃO AMBIENTAL	185
<b>Total Geral</b>	<b>456</b>

Fonte: Próprio autor

**Tabela 4 - Quantidade de alunos dos cursos de Ciências da Saúde dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
BIOMEDICINA	122
BIOTECNOLOGIA	67
CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	504
EDUCAÇÃO FÍSICA	233
EDUCAÇÃO FÍSICA (BACHARELADO)	157
ENFERMAGEM	419
FARMÁCIA	448
FISIOTERAPIA	629
NUTRIÇÃO	677
ODONTOLOGIA	116
TERAPIA OCUPACIONAL	316
<b>Total Geral</b>	<b>3688</b>

Fonte: Próprio autor

**Tabela 5 - Quantidade de alunos dos cursos de Ciências Exatas e da terra dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
CIÊNCIA DA COMPUTAÇÃO	390
FÍSICA	49
JOGOS DIGITAIS	220
MATEMÁTICA	181
QUÍMICA	161
QUÍMICA INDUSTRIAL	178
SISTEMAS DE INFORMAÇÃO	241
<b>Total Geral</b>	<b>1420</b>

Fonte: Próprio autor

**Tabela 6 - Quantidade de alunos dos cursos de Ciências Humanas dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
FILOSOFIA	283
GEOGRAFIA	48
HISTÓRIA	487
PEDAGOGIA	1022
PSICOLOGIA	601
RELAÇÕES INTERNACIONAIS	119
<b>Total Geral</b>	<b>2560</b>

Fonte: Próprio autor

**Tabela 7 - Quantidade de alunos nos cursos de Ciências Sociais Aplicadas dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
ADMINISTRAÇÃO	2278
ARQUITETURA E URBANISMO	2118
CIÊNCIAS CONTÁBEIS	917
CIÊNCIAS ECONÔMICAS	448
COMÉRCIO EXTERIOR	630
COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	459
COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	862
COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	114
CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	253
CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	287
Curso Superior de Tecnologia em Hotelaria	20
CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	126
DESIGN	464
DESIGN DE INTERIORES	85
DESIGN DE MODA	43
DIREITO	2392
EVENTOS	10
GESTÃO COMERCIAL	124
GESTÃO DA QUALIDADE	315
GESTÃO DE RECURSOS HUMANOS	345
GESTÃO FINANCEIRA	233
HOTELARIA	121
JORNALISMO	271
LOGÍSTICA	482
MARKETING	106
MODA	60
RELAÇÕES PÚBLICAS	117
<b>Total Geral</b>	<b>13680</b>

Fonte: Próprio Autor

**Tabela 8 - Quantidade de alunos dos cursos de Engenharias dos anos de 2012 a 2018**

Curso	Soma de quantidade de discentes
ENGENHARIA AMBIENTAL	781
ENGENHARIA CIVIL	1768
ENGENHARIA DA COMPUTAÇÃO	444
ENGENHARIA DE ALIMENTOS (BACHARELADO)	141
ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	127
ENGENHARIA DE CONTROLE E AUTOMAÇÃO	407
ENGENHARIA DE MATERIAIS (BACHARELADO)	63
ENGENHARIA DE PRODUÇÃO	1597
ENGENHARIA ELÉTRICA	376
ENGENHARIA MECÂNICA	104
ENGENHARIA QUÍMICA	888
GESTÃO DA PRODUÇÃO INDUSTRIAL	402
SEGURANÇA NO TRABALHO	30
<b>Total Geral</b>	<b>7128</b>

Fonte: Próprio autor

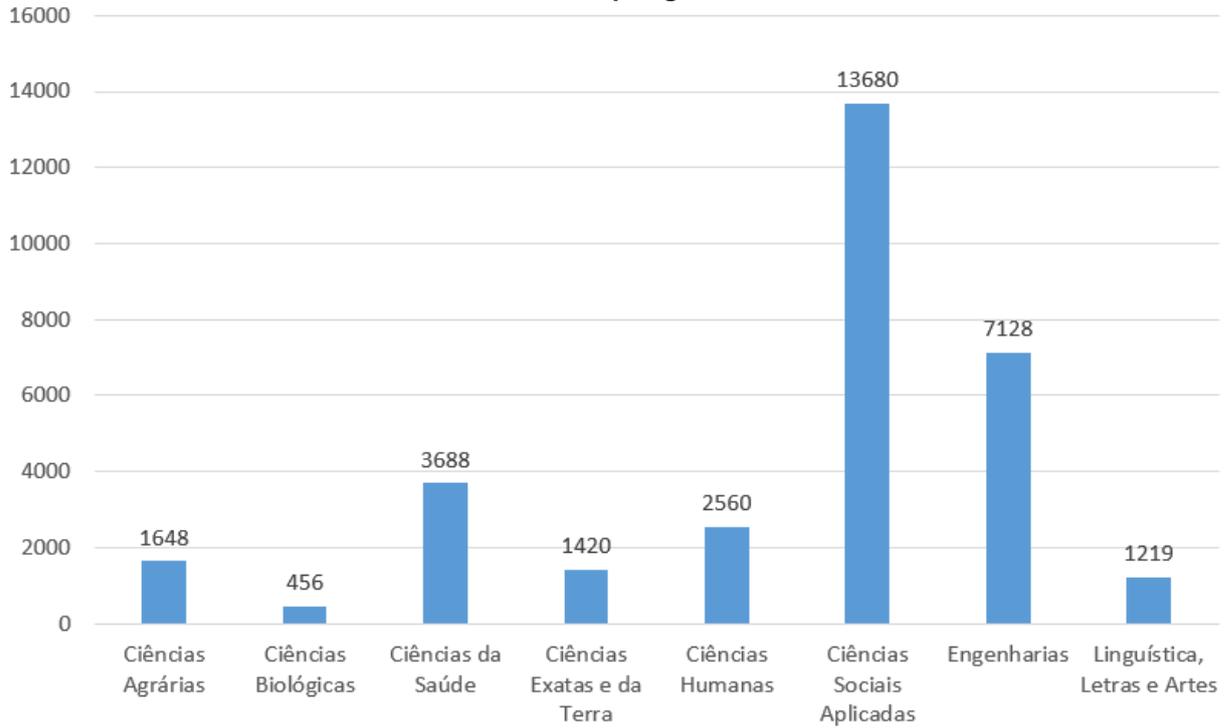
**Tabela 9 - Quantidade de alunos dos cursos de Linguística, Letras e Artes**

Curso	Soma de quantidade de discentes
ARTES VISUAIS	137
DANÇA	79
LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	35
LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	503
MÚSICA	273
TEATRO	192
<b>Total Geral</b>	<b>1219</b>

Fonte: Próprio autor

Pode-se confirmar com a Tabela 7 que a grande área de Ciências Sociais Aplicadas possui o maior montante de cursos realizados na graduação da UNISO e também contabiliza o maior número de alunos matriculados na faixa de tempo do 1º semestre de 2012 ao 1º semestre de 2018 como mostrado no Gráfico 1 - Quantidade de alunos por grande área entre 2012 e 2018.

**Gráfico 1 - Quantidade de alunos por grande área entre 2012 e 2018**

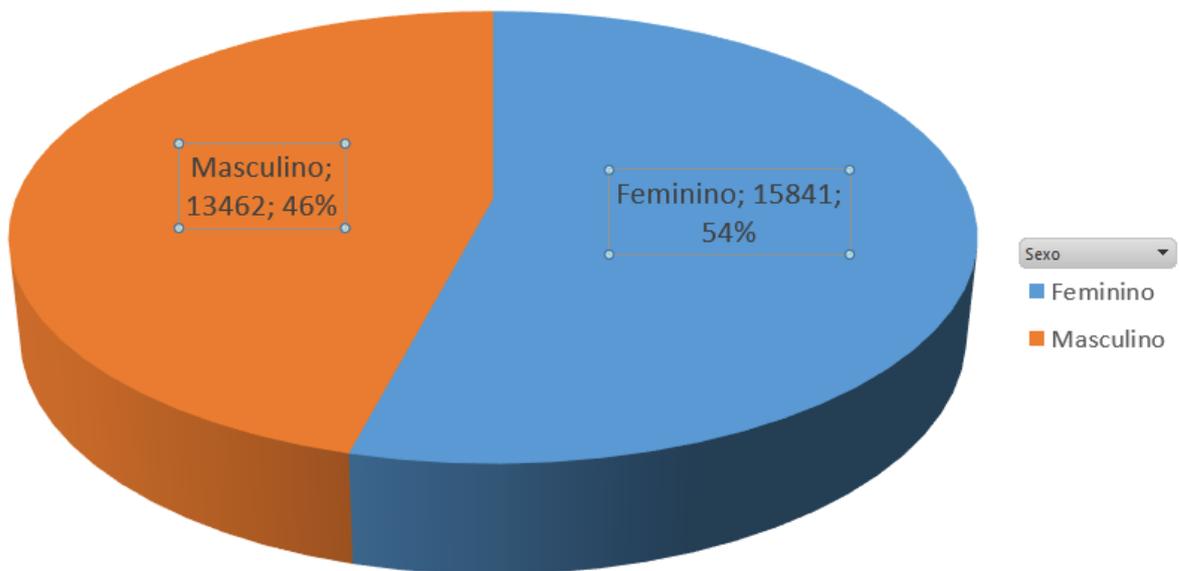


Fonte: Próprio autor

Com relação a quantidade de alunos analisados, categorizando-os por sexo, pode-se verificar que a quantidade de mulheres supera a de homens, como mostrado no Gráfico 2 abaixo.

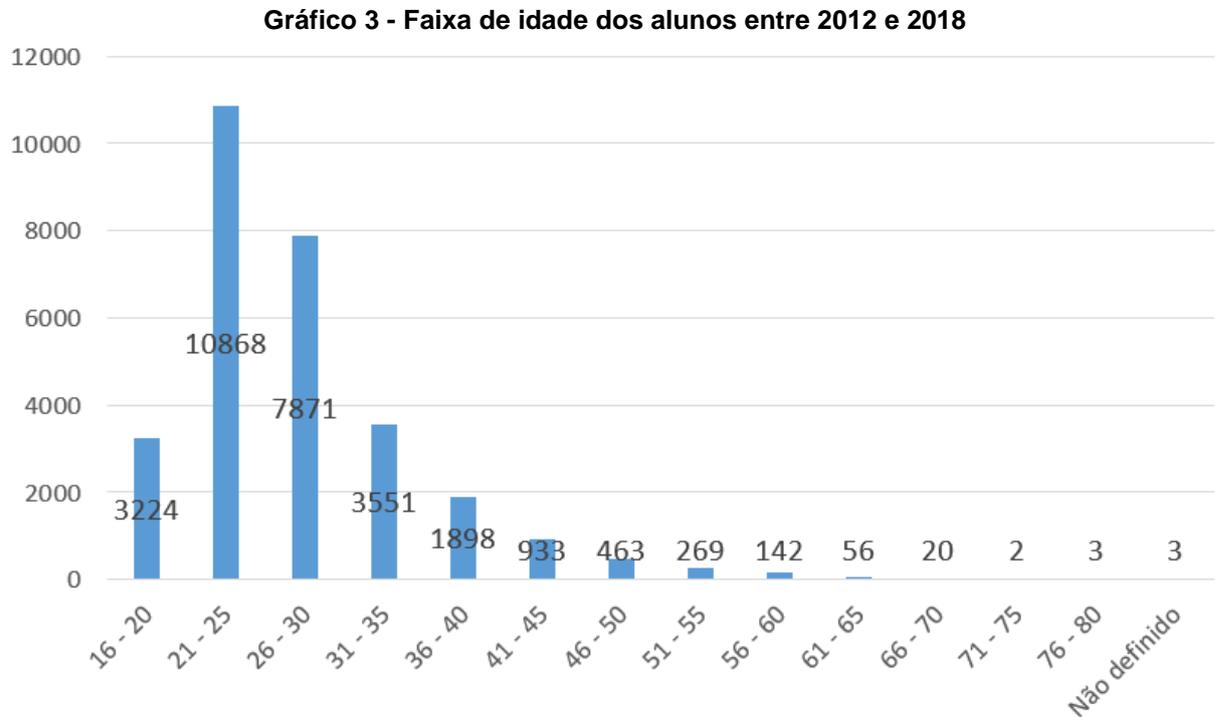
**Gráfico 2 - Quantidade de alunos por sexo**

Total de alunos: 29303



Fonte: próprio autor

E o público alvo se concentra na faixa de idade de 21 a 30 anos, como mostrado no Gráfico 3.



Fonte: Próprio autor

Sendo que a grande maioria apresenta o estado civil solteiro, mostrado na Tabela 10.

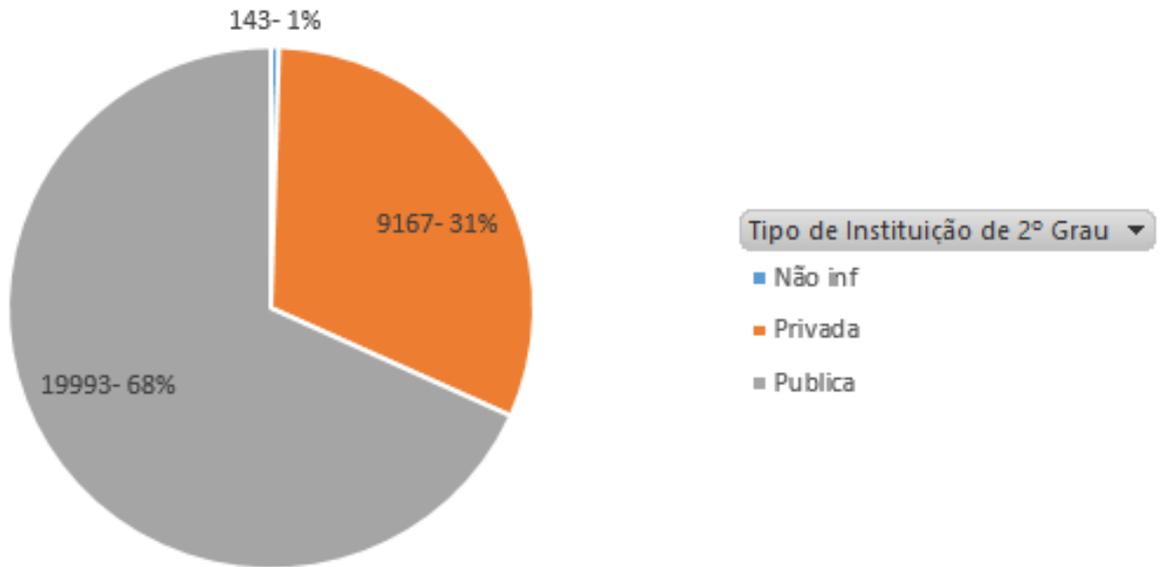
**Tabela 10 - Estado Civil dos alunos entre 2012 e 2018**

Estado Civil	Quantidade
Casado	3387
Desquitado	15
Divorciado	371
Não informado	233
Outros	422
Separado	3
Solteiro	24815
União Estável	29
Viúvo	28
<b>Total de alunos</b>	<b>29303</b>

Fonte: Próprio autor

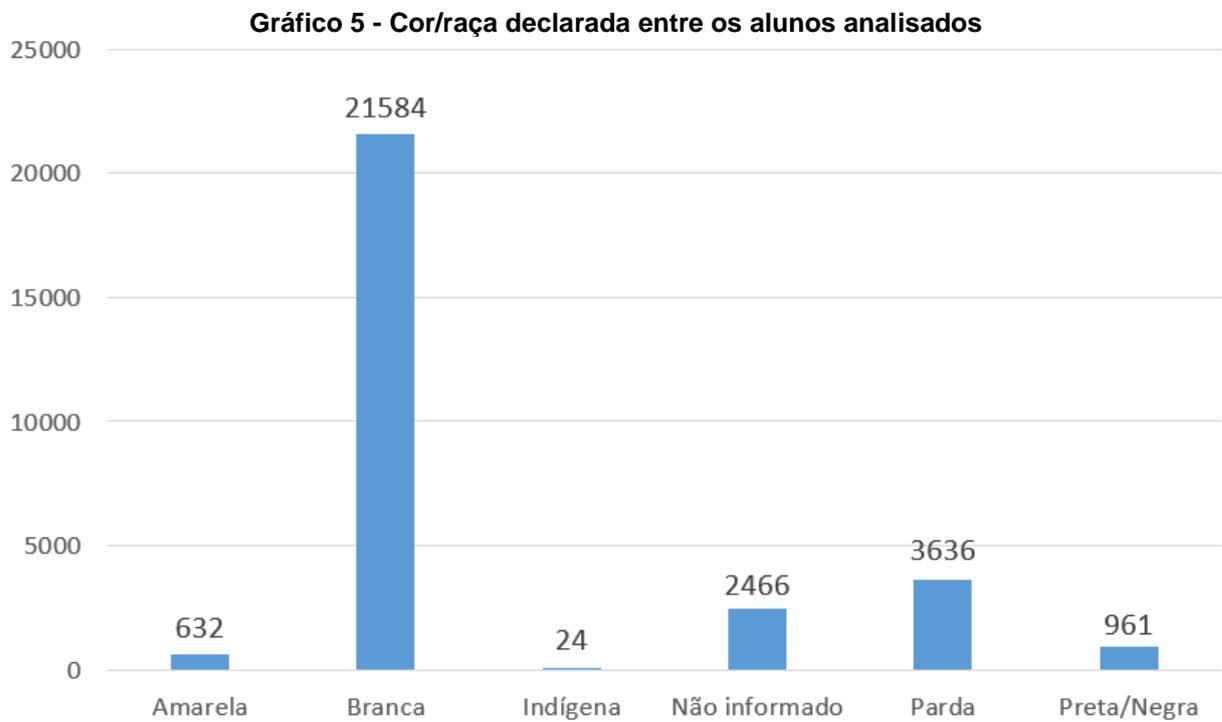
Quase 70% dos alunos concluiu o segundo grau em escola pública, mostrado no Gráfico 4.

**Gráfico 4 – Tipo de instituição de segundo realizado pelos alunos analisados**  
Total de alunos: 29303



Fonte: Próprio autor

E a grande maioria dos alunos declarou-se branca com relação a cor/raça, como mostrado no Gráfico 5.

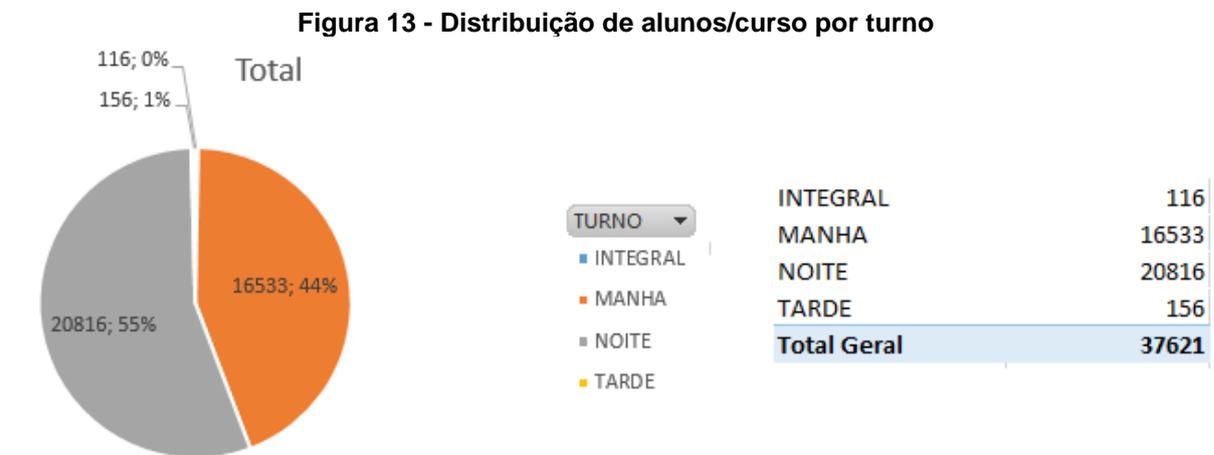


Fonte: Próprio autor

#### 4.1.1 Discriminação do perfil dos alunos/curso no turno

Da população de 29.303 alunos na faixa de tempo do 1º semestre de 2012 ao 1º semestre de 2018, ao qual cursaram um total de 31.799 cursos, ao acrescentar a variável turno esse valor sobe para 37.621, ou seja, uma diferença de 5.822 que representa as mudanças de turno na população estudada, uma representatividade de 15,5%.

A distribuição de alunos/curso por turno pode ser visualizada na Figura 13, onde comprova que o turno da noite possui mais da metade do público de discentes da UNISO para os cursos de graduação no período de tempo estudado.



Fonte: Próprio autor

A disposição por grande área e curso pode ser visualizada na Tabela 11.

**Tabela 11 – Turno / Grande área / Curso da população total**

Linhas	Contagem de RA
<b>INTEGRAL</b>	<b>116</b>
<b>Ciências da Saúde</b>	<b>116</b>
ODONTOLOGIA	116
<b>MANHA</b>	<b>16533</b>
<b>Ciências Agrárias</b>	<b>1211</b>
AGRONOMIA (BACHARELADO)	98
ENGENHARIA AGRONÔMICA	44
MEDICINA VETERINÁRIA	1069
<b>Ciências Biológicas</b>	<b>271</b>
CIÊNCIAS BIOLÓGICAS (BACHARELADO)	190
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO AMBIENTAL (EAD)	39

GESTÃO AMBIENTAL	42
<b>Ciências da Saúde</b>	<b>2628</b>
BIOMEDICINA	103
BIOTECNOLOGIA	67
CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	289
EDUCAÇÃO FÍSICA	95
EDUCAÇÃO FÍSICA (BACHARELADO)	121
ENFERMAGEM	330
FARMÁCIA	410
FISIOTERAPIA	499
NUTRIÇÃO	540
TERAPIA OCUPACIONAL	174
<b>Ciências Exatas e da Terra</b>	<b>452</b>
CIÊNCIA DA COMPUTAÇÃO	266
JOGOS DIGITAIS	157
SISTEMAS DE INFORMAÇÃO	29
<b>Ciências Humanas</b>	<b>587</b>
FILOSOFIA	280
PSICOLOGIA	307
<b>Ciências Sociais Aplicadas</b>	<b>6977</b>
ADMINISTRAÇÃO	982
ARQUITETURA E URBANISMO	1580
CIÊNCIAS CONTÁBEIS	342
CIÊNCIAS ECONÔMICAS	216
COMÉRCIO EXTERIOR	304
COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	269
COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	579
COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	49
CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	108
CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	138
DESIGN	242
DESIGN DE INTERIORES	7
DESIGN DE MODA	43
DIREITO	1471
GESTÃO COMERCIAL	2
GESTÃO DA QUALIDADE	99
GESTÃO DE RECURSOS HUMANOS	5
GESTÃO FINANCEIRA	31
JORNALISMO	223
LOGÍSTICA	125
MARKETING	1
MODA	60
RELAÇÕES PÚBLICAS	101
<b>Engenharias</b>	<b>3726</b>
ENGENHARIA AMBIENTAL	346

ENGENHARIA CIVIL	1047
ENGENHARIA DA COMPUTAÇÃO	282
ENGENHARIA DE ALIMENTOS (BACHARELADO)	103
ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	115
ENGENHARIA DE CONTROLE E AUTOMAÇÃO	218
ENGENHARIA DE MATERIAIS (BACHARELADO)	20
ENGENHARIA DE PRODUÇÃO	652
ENGENHARIA ELÉTRICA	173
ENGENHARIA MECÂNICA	63
ENGENHARIA QUÍMICA	577
GESTÃO DA PRODUÇÃO INDUSTRIAL	130
<b>Linguística, Letras e Artes</b>	<b>681</b>
ARTES VISUAIS	137
DANÇA	79
MÚSICA	273
TEATRO	192
<b>NOITE</b>	<b>20816</b>
<b>Ciências Agrárias</b>	<b>437</b>
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	403
ENGENHARIA AGRONÔMICA	34
<b>Ciências Biológicas</b>	<b>259</b>
CIÊNCIAS BIOLÓGICAS (BACHARELADO)	94
GESTÃO AMBIENTAL	165
<b>Ciências da Saúde</b>	<b>1670</b>
BIOMEDICINA	20
CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	263
EDUCAÇÃO FÍSICA	146
EDUCAÇÃO FÍSICA (BACHARELADO)	60
ENFERMAGEM	197
FARMÁCIA	127
FISIOTERAPIA	309
NUTRIÇÃO	340
TERAPIA OCUPACIONAL	208
<b>Ciências Exatas e da Terra</b>	<b>1159</b>
CIÊNCIA DA COMPUTAÇÃO	244
FÍSICA	49
JOGOS DIGITAIS	125
MATEMÁTICA	181
QUÍMICA	161
QUÍMICA INDUSTRIAL	178
SISTEMAS DE INFORMAÇÃO	221
<b>Ciências Humanas</b>	<b>2046</b>
FILOSOFIA	5
GEOGRAFIA	48
HISTÓRIA	487

PEDAGOGIA	1022
PSICOLOGIA	365
RELAÇÕES INTERNACIONAIS	119
<b>Ciências Sociais Aplicadas</b>	<b>9616</b>
ADMINISTRAÇÃO	1736
ARQUITETURA E URBANISMO	1278
CIÊNCIAS CONTÁBEIS	766
CIÊNCIAS ECONÔMICAS	322
COMÉRCIO EXTERIOR	489
COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	347
COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	544
COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	89
CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	162
Curso Superior de Tecnologia em Hotelaria	20
CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	126
DESIGN	280
DESIGN DE INTERIORES	78
DESIGN DE MODA	1
DIREITO	1551
EVENTOS	10
GESTÃO COMERCIAL	124
GESTÃO DA QUALIDADE	286
GESTÃO DE RECURSOS HUMANOS	343
GESTÃO FINANCEIRA	221
HOTELARIA	121
JORNALISMO	115
LOGÍSTICA	434
MARKETING	105
MODA	24
RELAÇÕES PÚBLICAS	44
<b>Engenharias</b>	<b>5091</b>
ENGENHARIA AMBIENTAL	595
ENGENHARIA CIVIL	1264
ENGENHARIA DA COMPUTAÇÃO	247
ENGENHARIA DE ALIMENTOS (BACHARELADO)	82
ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	41
ENGENHARIA DE CONTROLE E AUTOMAÇÃO	276
ENGENHARIA DE MATERIAIS (BACHARELADO)	51
ENGENHARIA DE PRODUÇÃO	1250
ENGENHARIA ELÉTRICA	276
ENGENHARIA MECÂNICA	48
ENGENHARIA QUÍMICA	601
GESTÃO DA PRODUÇÃO INDUSTRIAL	330
SEGURANÇA NO TRABALHO	30
<b>Linguística, Letras e Artes</b>	<b>538</b>

LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	35
LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	503
<b>TARDE</b>	<b>156</b>
<b>Ciências Sociais Aplicadas</b>	<b>156</b>
CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	156
<b>Total Geral</b>	<b>37621</b>

Fonte: próprio autor

Ao observar a Tabela 11, levando em consideração o período da manhã pode-se destacar:

- A área de Ciências Sociais Aplicadas no período da manhã é responsável por 18,5% da população total, onde considerando apenas o período da manhã esse percentual corresponde a 42,2% do total.
- Na área de Ciências Sociais Aplicadas dos 18 cursos mostrados na Tabela 11, a soma dos discentes dos cursos de Arquitetura e Urbanismo, Direito e Administração correspondem a aproximadamente 58% do total da grande área.

Considerando o período da noite, destaca-se:

- A área de Ciências Sociais Aplicadas é responsável por 25,6% da população total,
- o curso de Administração que apesar de ser um dos maiores do período da manhã, tem valor inferior a Arquitetura e Urbanismo e Direito, no período da noite que possui a maior quantidade de alunos dentre os cursos dessa grande área.
- Os cursos de Arquitetura e Urbanismo, Direito e Administração juntos são responsáveis por 47,5% do total da grande área, sendo a área de Ciências Sociais Aplicadas responsável por 33% do total no período noturno.

Relacionado a variável sexo da população analisada, o sexo feminino se destaca em todas as grandes áreas, perdendo apenas para Ciências Exatas e da Terra e Engenharias, como mostrado na tabela abaixo:

**Tabela 12 - Total da população analisada por Grande Área/sexo**

Grande Área	Feminino	Masculino	Total Geral
Ciências Agrárias	1047	601	1648
Ciências Biológicas	285	245	530
Ciências da Saúde	3556	858	4414
Ciências Exatas e da Terra	397	1214	1611
Ciências Humanas	1787	846	2633
Ciências Sociais Aplicadas	9662	7087	16749
Engenharias	2880	5937	8817
Linguística, Letras e Artes	747	472	1219
<b>Total Geral</b>	<b>20361</b>	<b>17260</b>	<b>37621</b>

Fonte: próprio autor

#### 4.1.2 Discriminação do perfil dos alunos evadidos

Quando utilizado apenas os discentes evadidos da população total de 37.624 o valor resultante é 9.628 casos, ou seja, cerca de 25,6% dos casos são evasão. Fazendo uma distribuição por sexo, do total de 20.361 discentes do sexo feminino 23,29% são evadidos e da população de 17.260 discentes do sexo masculino 28,30% são evadidos como mostrado na Tabela 13 - Distribuição dos evadidos nas grandes áreas (%).

**Tabela 13 - Distribuição dos evadidos nas grandes áreas (%)**

Grande Áreas	evasão Feminina na área	evasão Masculina na área
Ciências Agrárias	24,36	27,12
Ciências Biológicas	24,56	27,76
Ciências da Saúde	21,77	25,41
Ciências Exatas e da Terra	27,46	29,41
Ciências Humanas	21,04	24,11
Ciências Sociais Aplicadas	23,51	28,55
Engenharias	23,30	28,47
Linguística, Letras e Artes	28,92	34,32
<b>Total Geral</b>	<b>23,29</b>	<b>28,30</b>

Fonte: Próprio autor.

Onde os casos são divididos em 4.743 (49%) evadidos do sexo feminino e 4.885 (51%) do sexo masculino, tomando por base o Gráfico 2 - Quantidade de alunos por sexo, apresentado na página 71, que mostra que os discentes de sexo masculino representam 46% da população estudada, comprova-se o resultado apresentado no tópico 4.3.1, página 115, que por intermédio do estudo de *Cluster* afirmou que o sexo masculino se destaca no ato de evadir do curso.

Levando em consideração a Tabela 12 - Total da população analisada por Grande Área/sexo que mostra que discentes do sexo masculino possuem maior concentração nas áreas de Ciências Exatas e da Terra e Engenharias, não desconsiderando que apesar da área de Ciências Sociais Aplicadas possuírem mais discentes do sexo feminino, também possuem um número elevado de alunos homens podemos concluir o resultado que foi apresentado na tabela abaixo, onde pode-se observar em quais grandes áreas existem as maiores concentrações de evasão para ambos os sexos.

**Tabela 14 - Percentual de Evasão distribuído por grande área**

Grande Áreas	Evadido (Fem.)	Evadido (Masc.)	Perc. De Evasão (Fem.)	Perc.De Evasão (Masc.)
Ciências Agrárias	255	163	24,4	27,1
Ciências Biológicas	70	68	24,6	27,8
Ciências da Saúde	774	218	21,8	25,4
Ciências Exatas e da Terra	109	357	27,5	29,4
Ciências Humanas	376	204	21,0	24,1
Ciências Sociais Aplicadas	2272	2023	23,5	28,5
Engenharias	671	1690	23,3	28,5
Linguística, Letras e Artes	216	162	28,9	34,3
<b>Total Geral</b>	<b>4743</b>	<b>4885</b>	<b>23,3</b>	<b>28,3</b>

Fonte: próprio autor

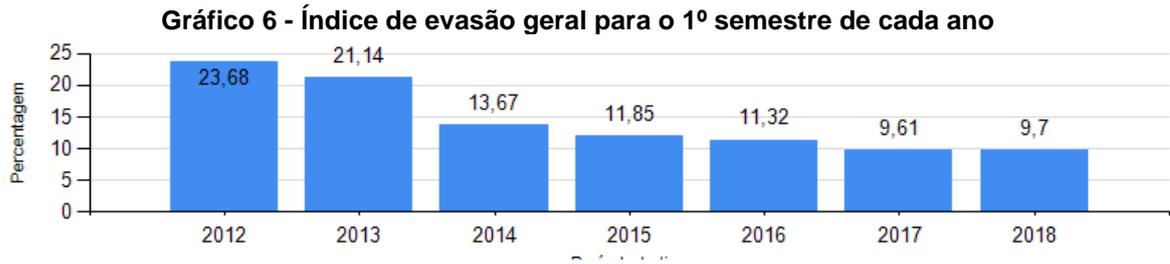
Comparando a Tabela 12 - Total da população analisada por Grande Área/sexo com a Tabela 14 - Percentual de Evasão distribuído por grande área, fica nítido que o sexo masculino possui um índice maior de evasão, independentemente da quantidade de discente do sexo feminino ser maior que a quantidade do sexo masculino.

#### **4.2 Análise dos índices de evasão**

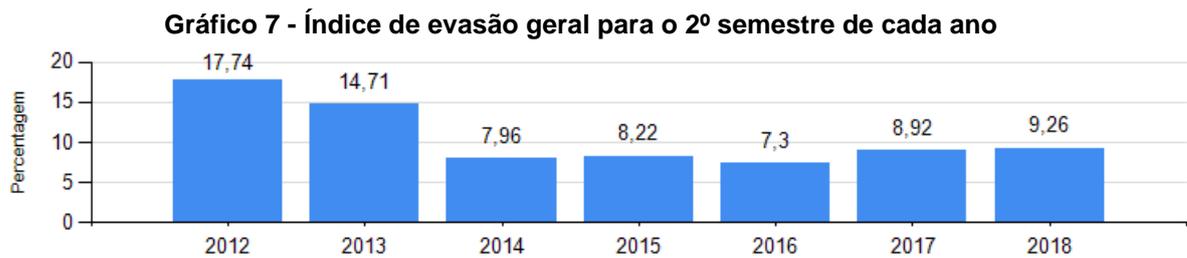
Foi computado no Data Mart um total de 9.628 fatos de evasão em cursos, que como comentado no tópico 3.3.3 sobre transformação de dados, encapsula as categorias de status no curso: evadido, cancelado e transferência de curso. Isso não significa que em todos os casos ocorreu um abandono do aluno na instituição, pois como discutido no tópico 2.1, explicado pela Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas (1997, p.16), o presente trabalho não está considerando abandono da instituição ou do sistema, apenas do curso. Maiores detalhes sobre os status dos alunos nos cursos podem ser visualizados no apêndice C. Os parâmetros de envio das informações de aluno para o INEP podem ser visualizados no anexo A, pag. 248.

A fim de pontuar a evolução da evasão na UNISO, foram criados, com o auxílio da equação de Lobo (2007), gráficos que demonstrem os índices de evasão por ano letivo, onde sua visualização foi categorizada de três maneiras: geral (evasão considerando todos os cursos), por grande área (evasão por grande área) e curso (índice de evasão do curso específico no período de tempo estudado).

Na evasão geral, chegou-se nos seguintes resultados.



Fonte: Próprio autor

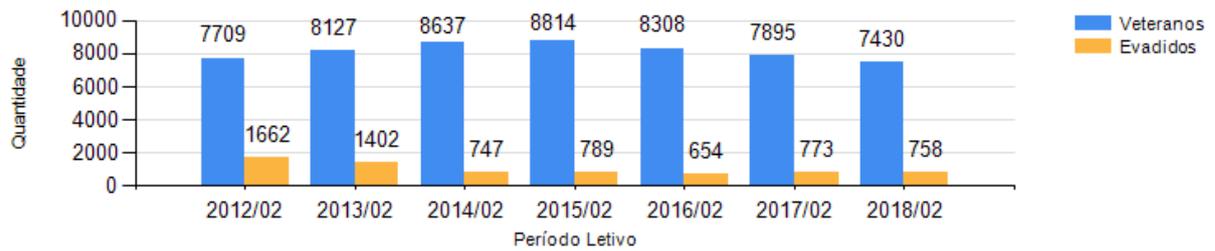


Onde pode-se observar no Gráfico 6 e no Gráfico 7 um alto índice no ano de 2012, diminuindo até 2014 e mantendo-se por volta dos 10% até o final do estudo em 2018.

Essas mesmas informações podem ser observadas comparando as quantidades de discentes veteranos e evadidos no Gráfico 8 e Gráfico 9.



Fonte: Próprio autor

**Gráfico 9 - Quantidade de veteranos e evadidos no 2º semestre de cada ano**

Com esses gráficos temos uma visualização da quantidade de alunos veteranos, ou seja, desconsiderando os alunos ingressantes, em comparação com a quantidade de evadidos e pode-se observar que a mesma estabilização mostrada no Gráfico 6 e Gráfico 7, pag. 81, com um detalhe a mais, o número de alunos veteranos que a Uniso trabalha nos cursos de graduação na faixa de tempo estudada.

Os dados utilizados para realização do cálculo do percentual de evasão mostrado nos Gráfico 6 e Gráfico 7 e da quantidade geral de alunos veteranos e evadidos mostrado no Gráfico 8 e Gráfico 9 utilizaram como base os dados apresentados na Tabela 15.

**Tabela 15 - Visualização dos dados computados no relatório de evasão geral da UNISO nos períodos de 2012 a 2018**

Período Letivo	Nº Matrículas Atuais	Nº Ingressantes No Período	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	9371	3065	8266	3	23,68
2012/02	8633	924	9371	0	17,74
2013/01	9761	3013	8633	76	21,14
2013/02	9190	1063	9761	232	14,71
2014/01	10191	2824	9190	656	13,67
2014/02	9535	898	10191	808	7,95
2015/01	10340	2683	9535	849	11,85
2015/02	9451	637	10340	738	8,21
2016/01	9572	2025	9451	941	11,32
2016/02	8894	586	9572	610	7,3
2017/01	9252	2130	8894	1015	9,61
2017/02	8521	626	9252	584	8,92
2018/01	8746	2014	8521	1066	9,7
2018/02	8149	719	8746	558	9,26

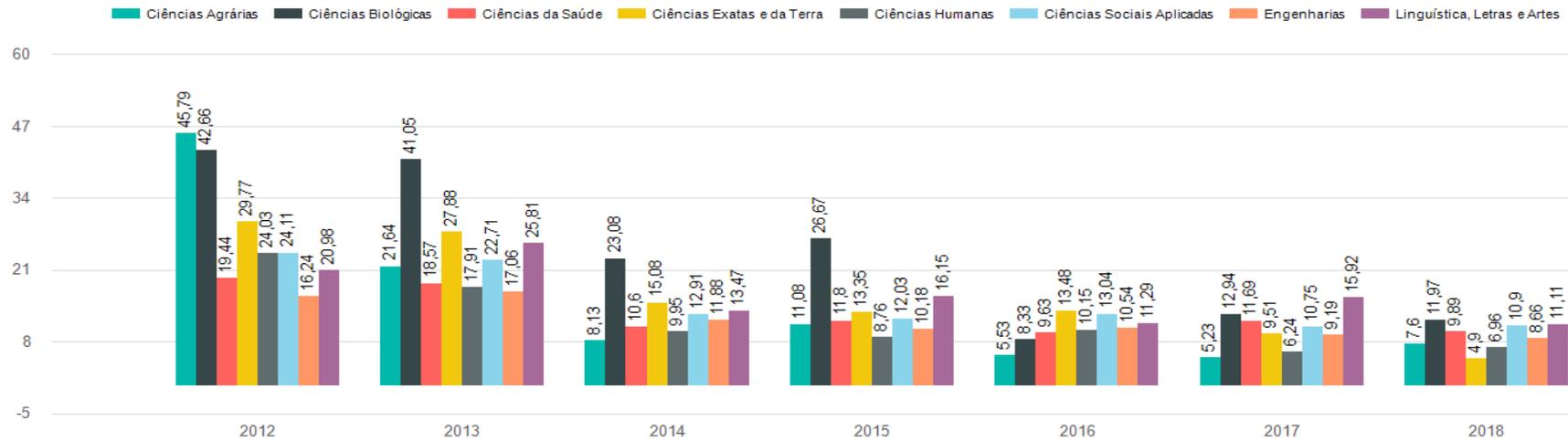
Fonte: Próprio autor

No tópico 4.2.1 são mostrados os índices de evasão de grande área em comparação com o índice de evasão geral apresentado.

#### **4.2.1 Índices de evasão das grandes áreas de graduação da UNISO**

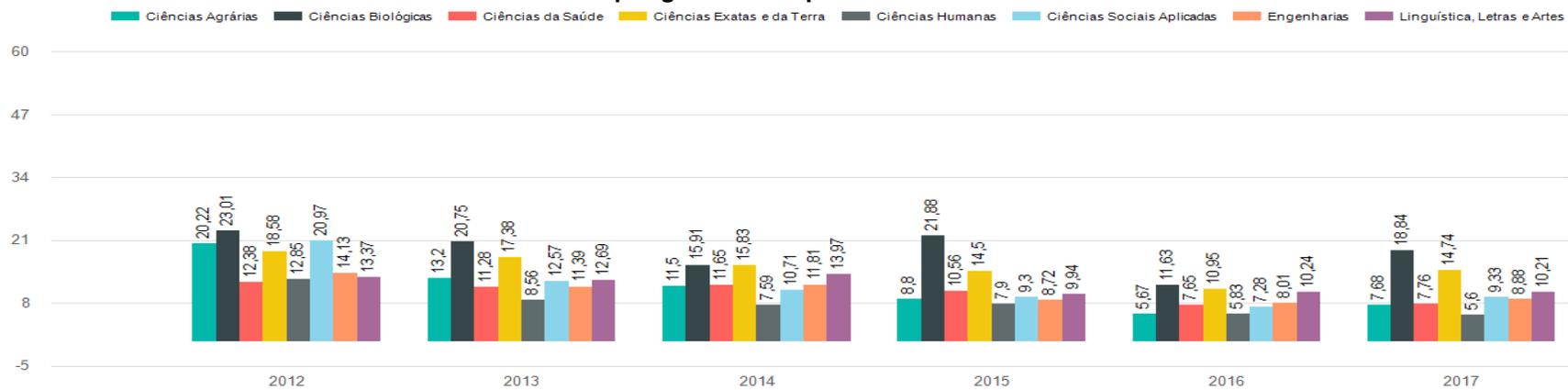
Trabalhando os mesmos dados do ponto de vista das grandes áreas, os índices de evasão mostrado para as áreas de Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias e Linguística, Letras e Artes podem ser visualizados no Gráfico 10 e Gráfico 11.

**Gráfico 10 - Índice de evasão por grande área para o 1º semestre entre 2012/01 e 2018/01**



Fonte: Próprio autor

**Gráfico 11 - Índice de evasão por grande área para o 2º semestre entre 2012/02 e 2017/02**



Fonte: Próprio autor

Os índices de evasão por grande área foram divididos entre dois gráficos, pois o comportamento das variáveis número de ingressantes e número de concluintes possui uma variação grande entre o 1º e o 2º semestre letivo de cada ano devido a quantidade de matrículas no início do ano ser maior que a mesma quantidade na metade do ano.

Os dados utilizados na elaboração dos gráficos podem ser visualizados separados por grande área entre os Apêndices D ao K.

Todas as grandes áreas demonstraram uma diminuição na taxa de evasão, estabilizando por volta dos anos de 2014 a 2015.

Em comparação com o Gráfico 8 e Gráfico 9 na pag.81, mantiveram o padrão de início alto, no ano de 2012 e final por volta dos 10% no ano de 2018.

A grande área de Engenharias foi a que teve o início mais baixo em comparação com as outras, porém, não foi a que teve o término mais baixo no ano de 2018 estabilizando por volta dos 9%.

A grande área de Ciências Biológicas iniciou o estudo com o maior índice de evasão, por volta dos 33% na média do 1º e 2º semestre do ano de 2012 e terminando em 2018 com um índice de 9,5%.

A grande área de Ciências Sociais Aplicadas iniciou em 2012 com uma taxa próxima dos 23% na média anual, mantendo-se próximo a média dos 10% a partir de 2014.

A grande área de Ciências da Saúde iniciou em 2012 com uma taxa de evasão de 16,5% na média do 1º e 2º semestre, caindo a partir de 2014 próximo dos 10% e mantendo-se nesse índice até o final do estudo em 2018.

A grande área de Linguística, Letras e Artes começou em 2012 apresentando uma taxa de evasão de por volta dos 17% na média anual e terminando em 2018 na taxa de 8,5%.

A grande área de Ciências Humanas em 2012 teve uma taxa de evasão anual de 19%, apresentando no decorrer do tempo uma queda, onde no ano de 2018 terminou o estudo com a taxa de 6,5%.

A grande área de Ciências Exatas e da Terra, não demonstrou comportamento diferentes das outras grandes áreas, iniciando em 2012 com o índice de 24% de evasão anual, diminuindo no decorrer do tempo até chegar em 2018 com o índice de 9,5%.

A grande área de Ciências Agrárias começou em 2012 com a taxa de 30% anual, terminando o estudo em 2018 por volta dos 10%.

#### **4.2.2 Índices de evasão dos cursos de graduação**

Abaixo são mostrados os índices de evasão dos cursos de graduação da Uniso do período de 2012 a 2018 segundo a fórmula de Lobos (2007) para um comparativo específico entre curso, grande área e evasão geral.

Os valores utilizados para o cálculo podem ser contemplados nos Apêndices D ao K, pag. 180, separados por período letivo e curso, número de matrículas, número de ingressantes atuais, número de matrículas do período anterior, número de concluintes e taxa de evasão.

Os índices demonstram quanto de evasão ocorreu no curso do período letivo atual para o anterior, por padrão esses valores são positivos, indicando perda de discentes veteranos, mas há casos onde esse índice gera valores negativos e isso acontece pelos seguintes motivos:

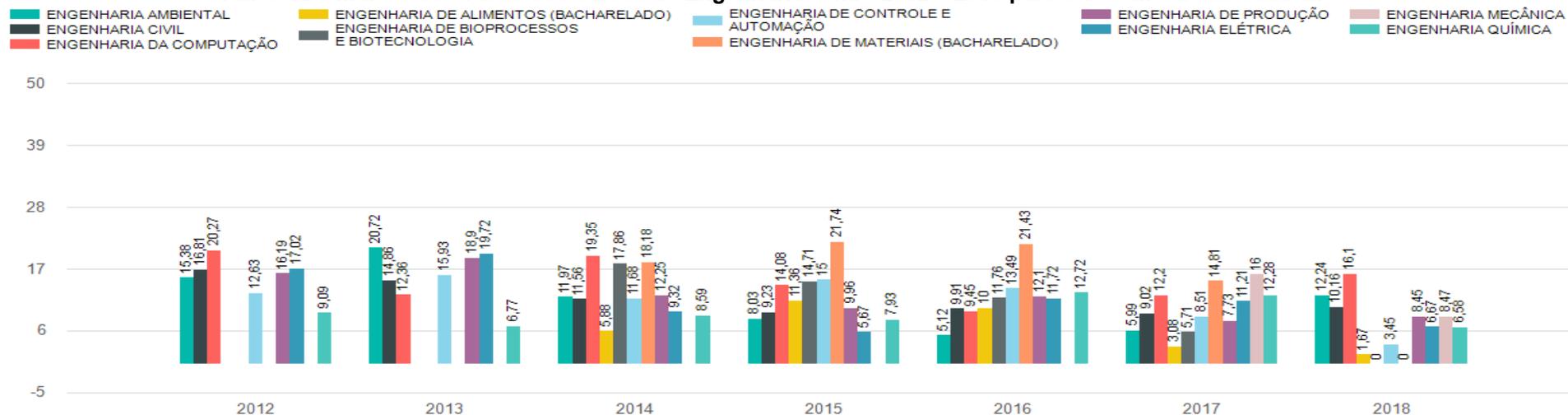
- Volta de alunos trancados, ou seja, não computados no período anterior, porém não calouros no período letivo corrente (volta de aluno);
- Alunos computados como concluintes no último período letivo cursado, porém nesse período não estavam registrados como alunos matriculados;
- Erro operacional dos responsáveis pela categorização do aluno/curso como concluído.

##### **4.2.2.1 Índices de evasão dos cursos de Engenharias**

Para a grande área de Engenharias, a tabela com os dados correspondentes encontra-se no apêndice D, pág. 180.

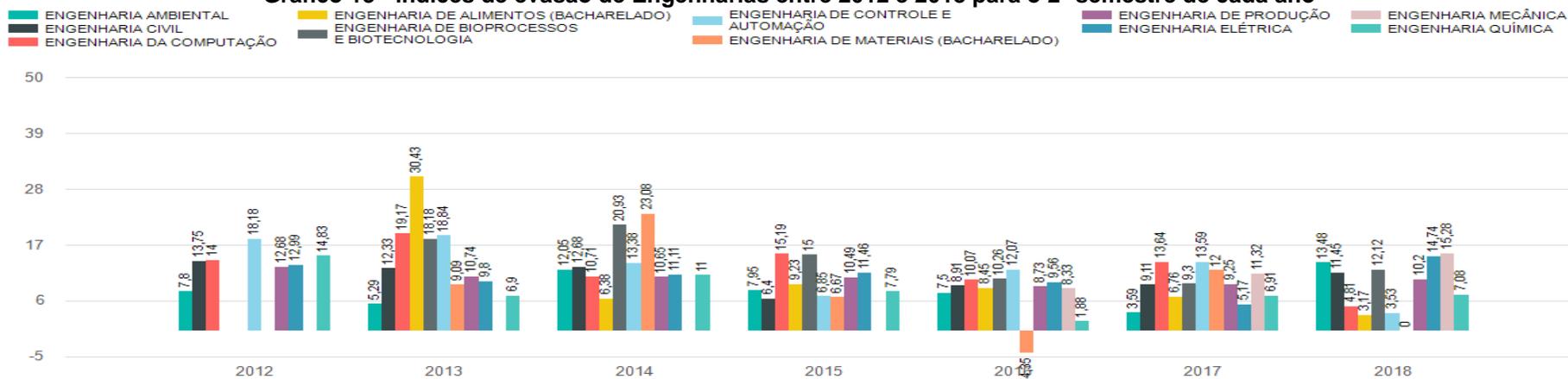
O Gráfico 12 e o Gráfico 13 mostram os índices de evasão dos cursos da grande área de Engenharias, desconsiderando os cursos tecnológicos que serão mostrados separadamente.

Gráfico 12 – Índices de evasão dos cursos de Engenharias entre 2012 e 2018 para o 1º semestre de cada ano



Fonte: Próprio autor.

Gráfico 13 - Índices de evasão de Engenharias entre 2012 e 2018 para o 2º semestre de cada ano



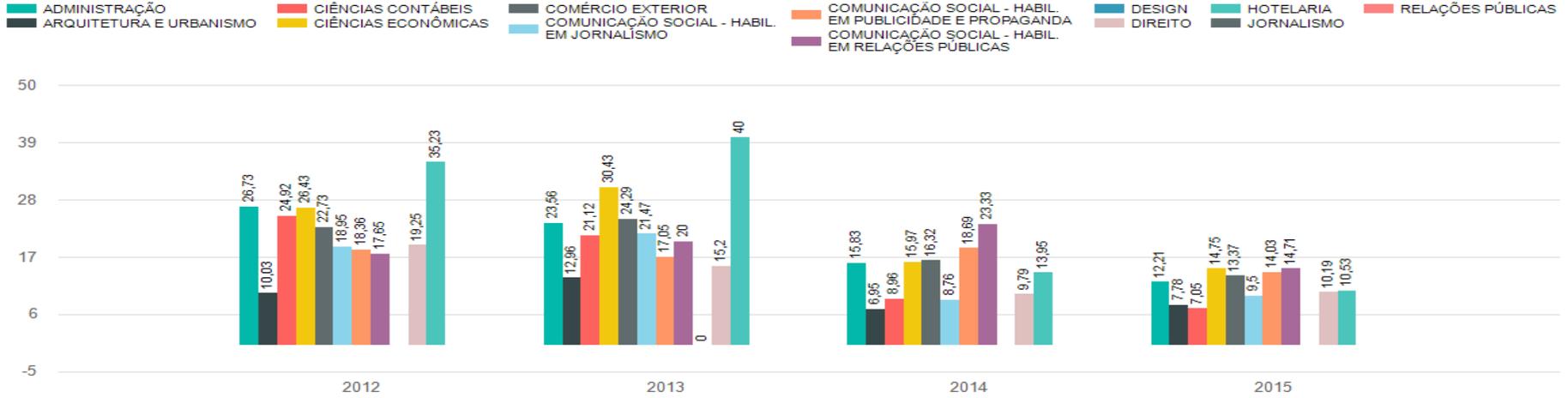
Fonte: Próprio autor.

#### 4.2.2.2 Índices de evasão dos cursos de Ciências Sociais Aplicadas

Para a grande área de Ciências Sociais Aplicadas, a tabela com os dados correspondentes encontra-se no apêndice E, pág.189.

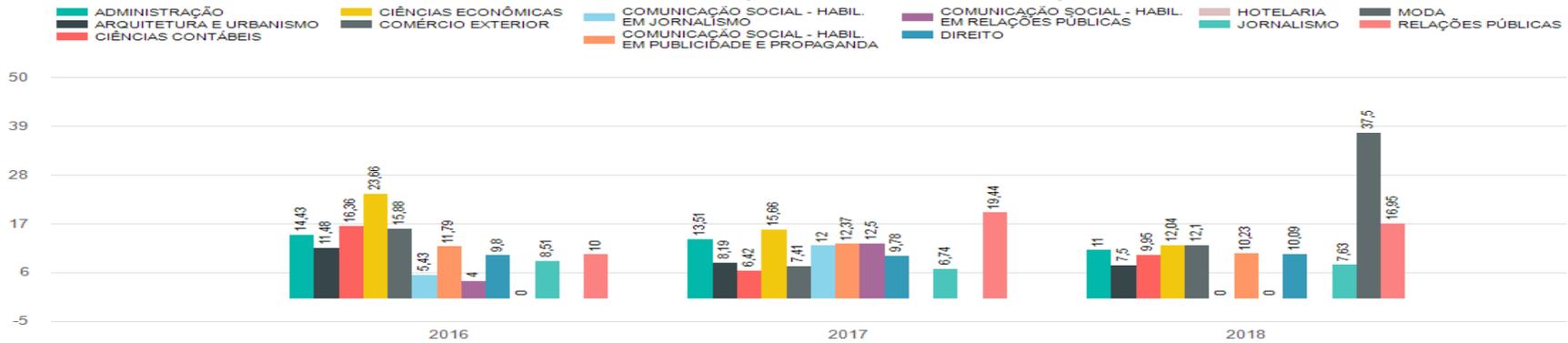
Os cursos tecnológicos dessa área serão mostrados separadamente.

**Gráfico 14 - Índices de evasão de Ciências Sociais Aplicadas entre 2012 e 2015 para o 1º semestre de cada ano**



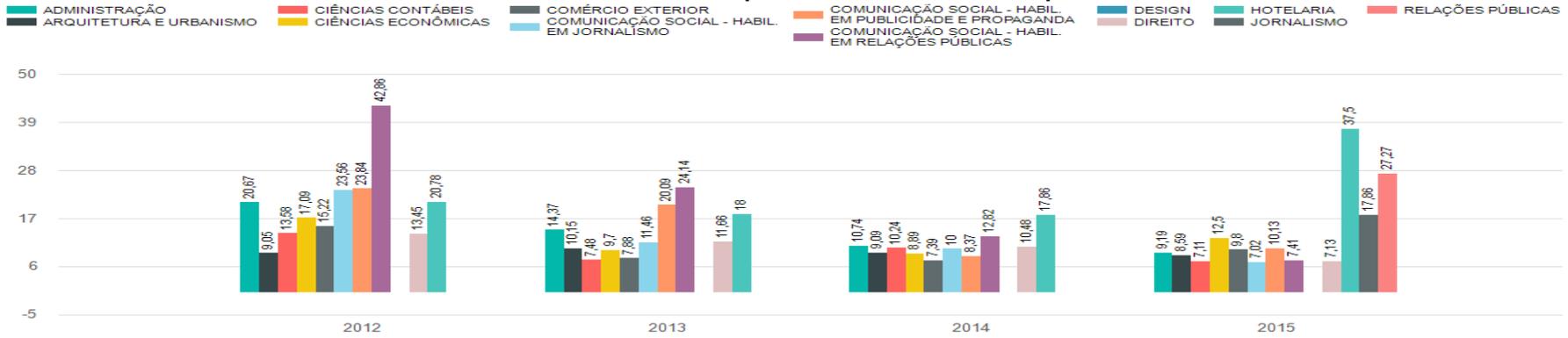
Fonte: Próprio autor

**Gráfico 15 - Índices de evasão de Ciências Sociais Aplicadas entre 2016 e 2018 para o 1º semestre de cada ano**



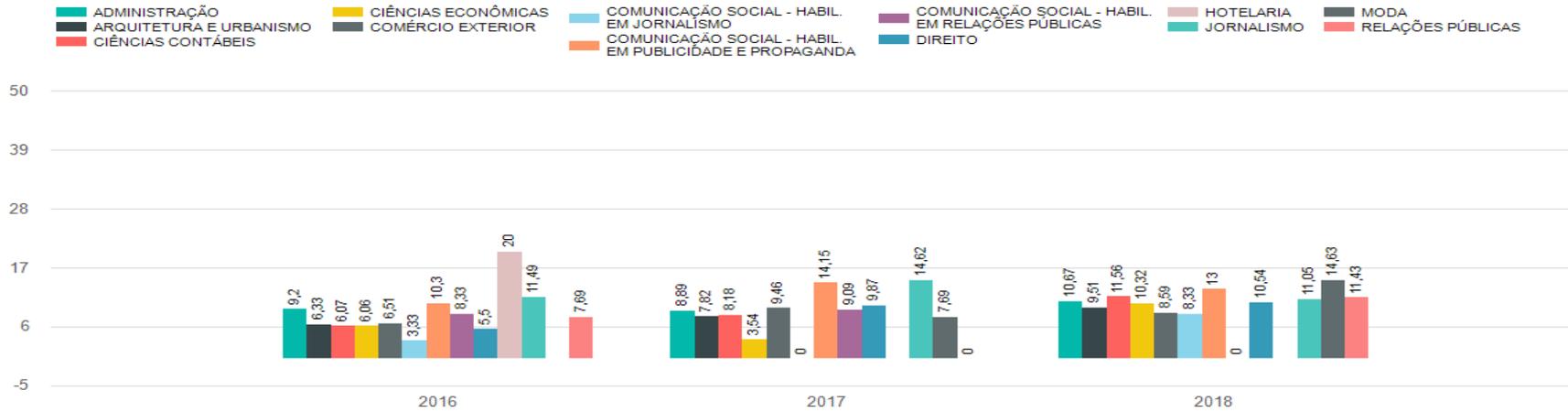
Fonte: Próprio autor

**Gráfico 16 – Índice de evasão de Ciências Sociais Aplicadas entre 2012 e 2015 para o 2º semestre de cada ano**



Fonte: Próprio autor

**Gráfico 17 – Índice de evasão de Ciências Sociais Aplicadas entre 2016 e 2018 para o 2º semestre de cada ano**

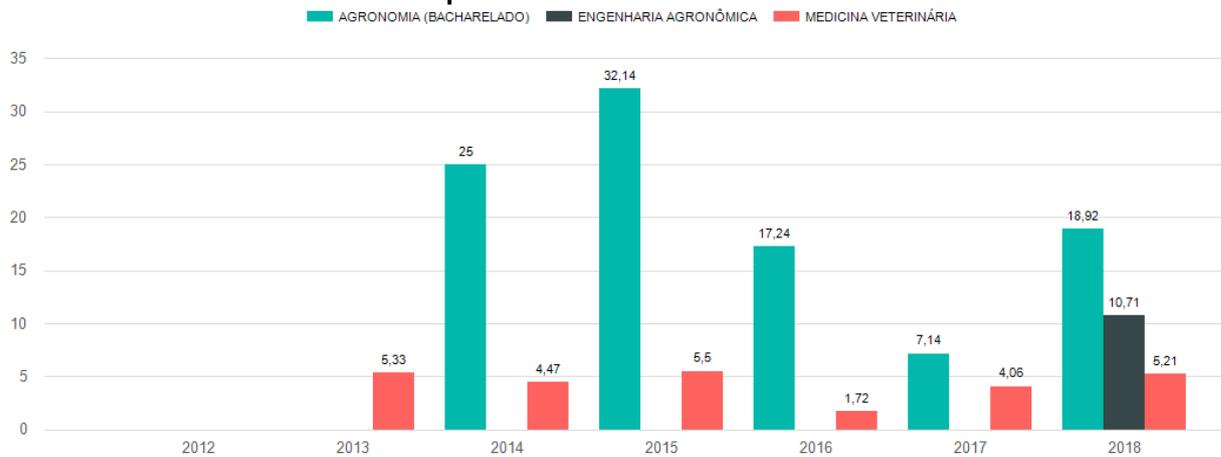


Fonte: Próprio autor.

#### 4.2.2.3 Índices de evasão dos cursos de Ciências Agrárias

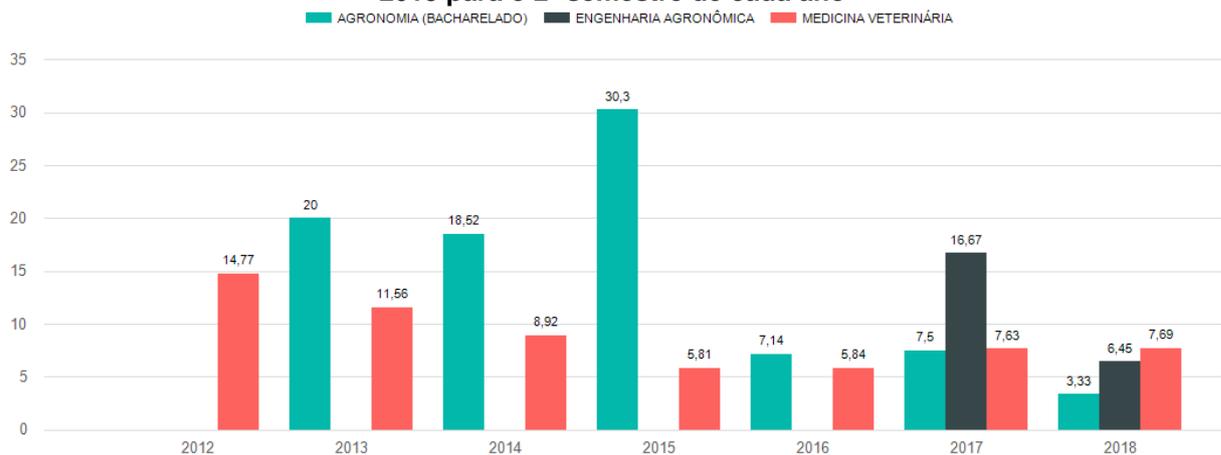
Para a grande área de Ciências Agrárias, a tabela com os dados correspondentes encontra-se no apêndice F, pág. 205.

**Gráfico 18 - Índice de evasão dos cursos da grande área de Ciências Agrárias entre 2012 e 2018 para o 1º semestre de cada ano**



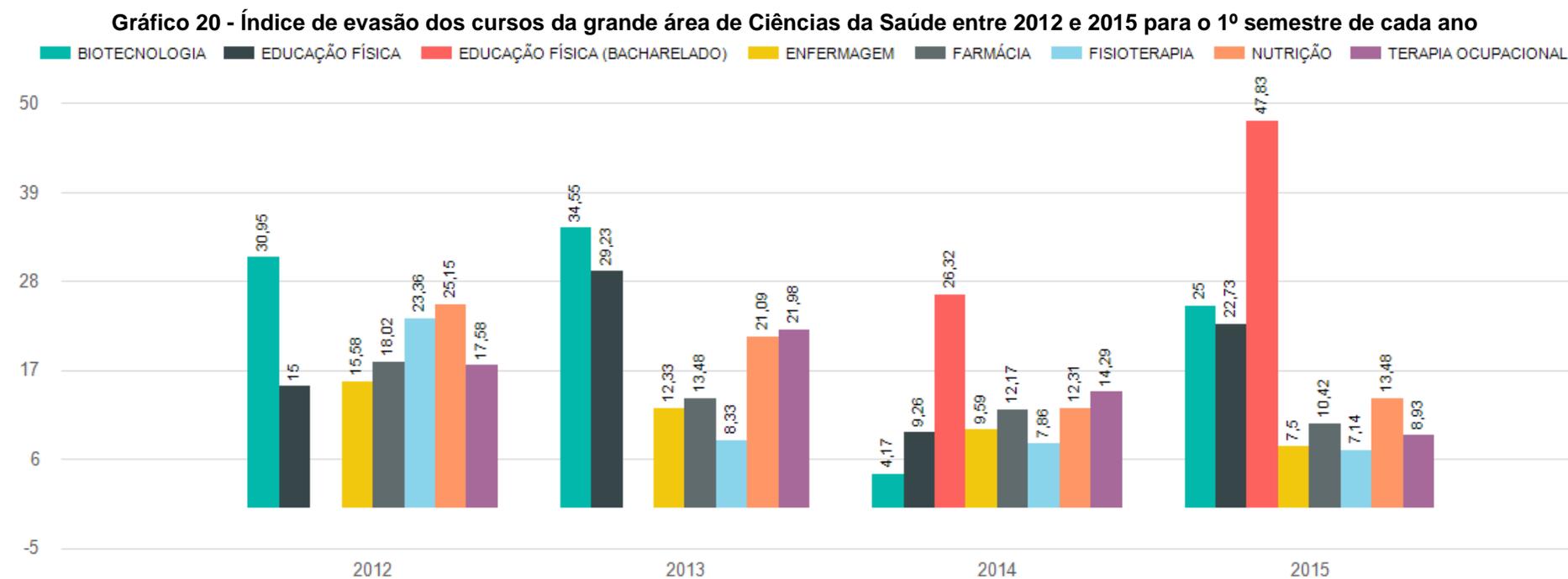
Fonte: Próprio autor

**Gráfico 19 - Índice de evasão dos cursos da grande área de Ciências Agrárias entre 2012 e 2018 para o 2º semestre de cada ano**



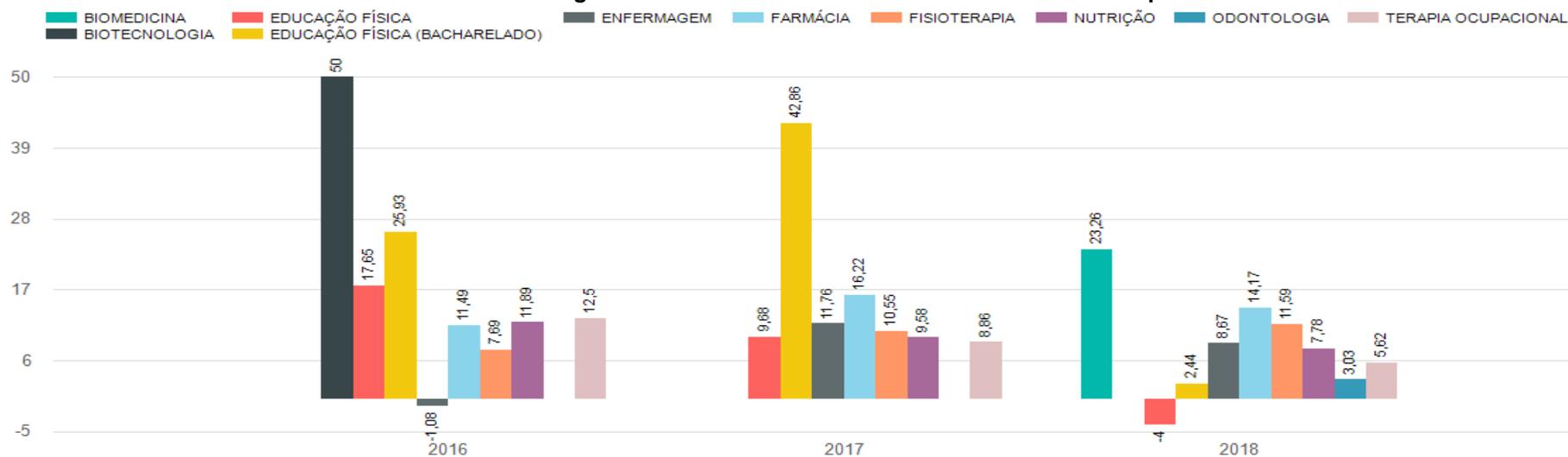
#### 4.2.2.4 Índices de evasão dos cursos de Ciências da Saúde

Para a grande área de Ciências da Saúde, a tabela com os dados correspondentes encontra-se no apêndice G, pág. 209.



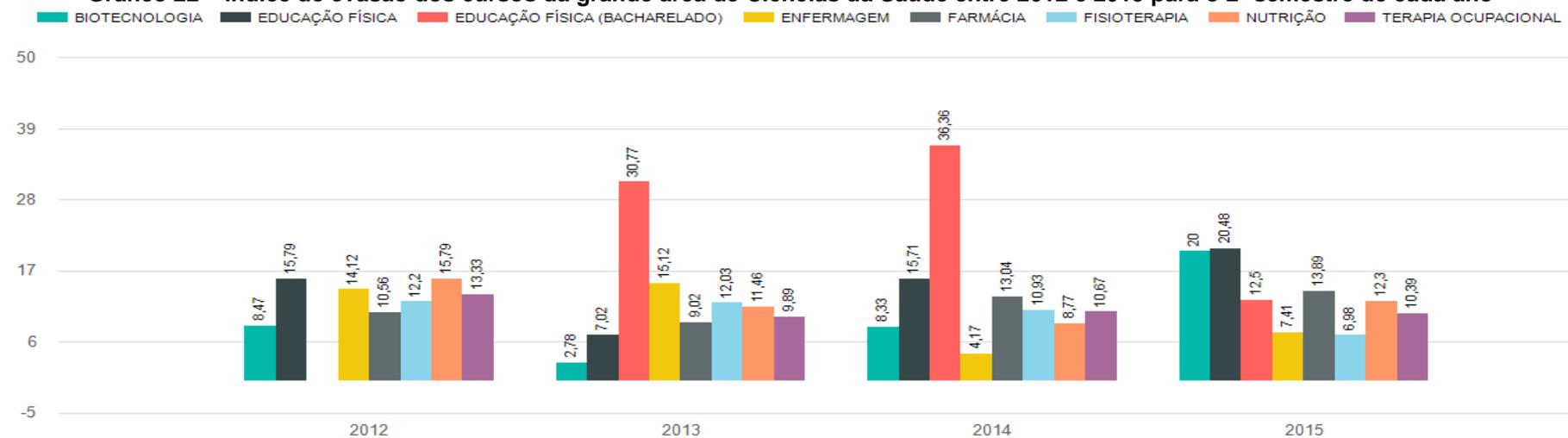
Fonte: Próprio autor

**Gráfico 21 - Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2016 e 2018 para o 1º semestre de cada ano**



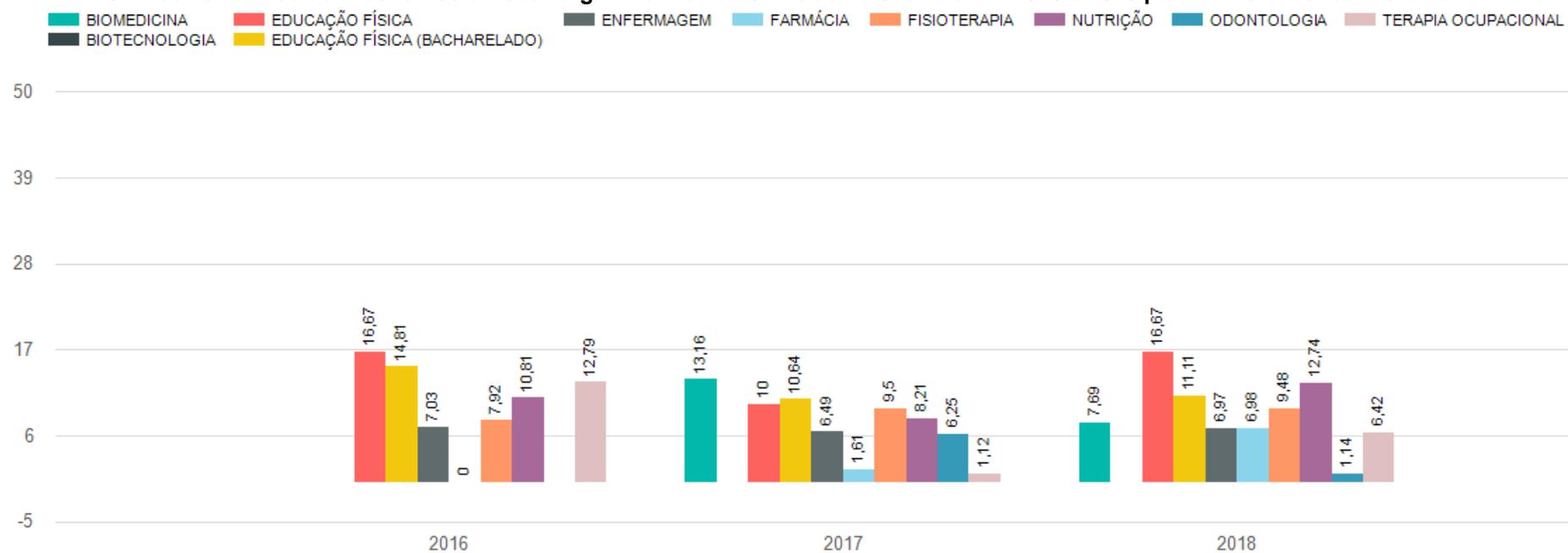
Fonte: Próprio autor

**Gráfico 22 – Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2012 e 2015 para o 2º semestre de cada ano**



Fonte: Próprio autor

**Gráfico 23 - Índice de evasão dos cursos da grande área de Ciências da Saúde entre 2016 e 2018 para o 2º semestre de cada ano**

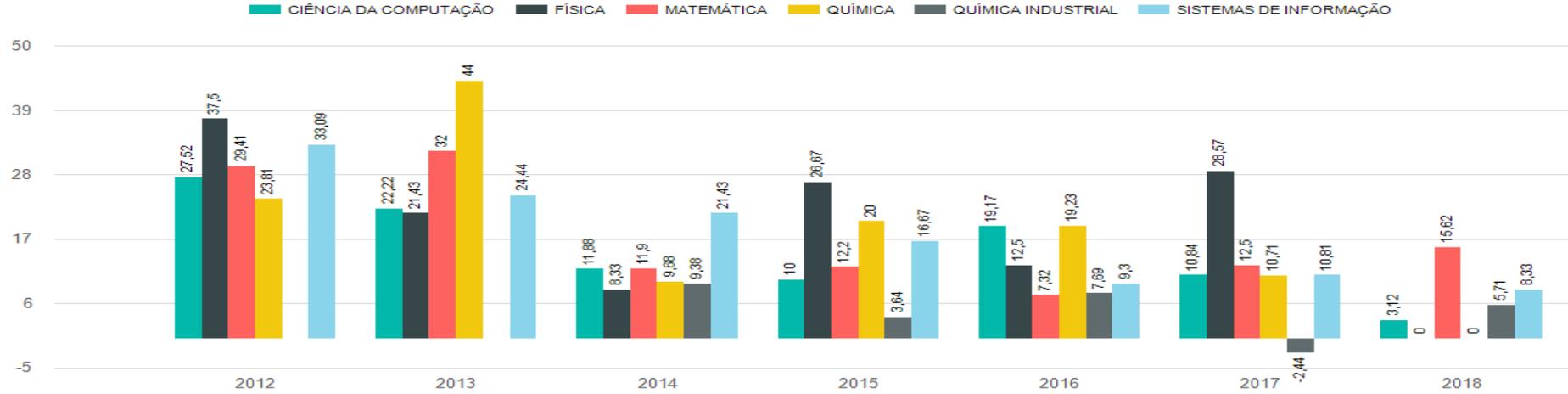


Fonte: Próprio autor

#### 4.2.2.5 Índices de evasão dos cursos de Ciências Exatas e da Terra

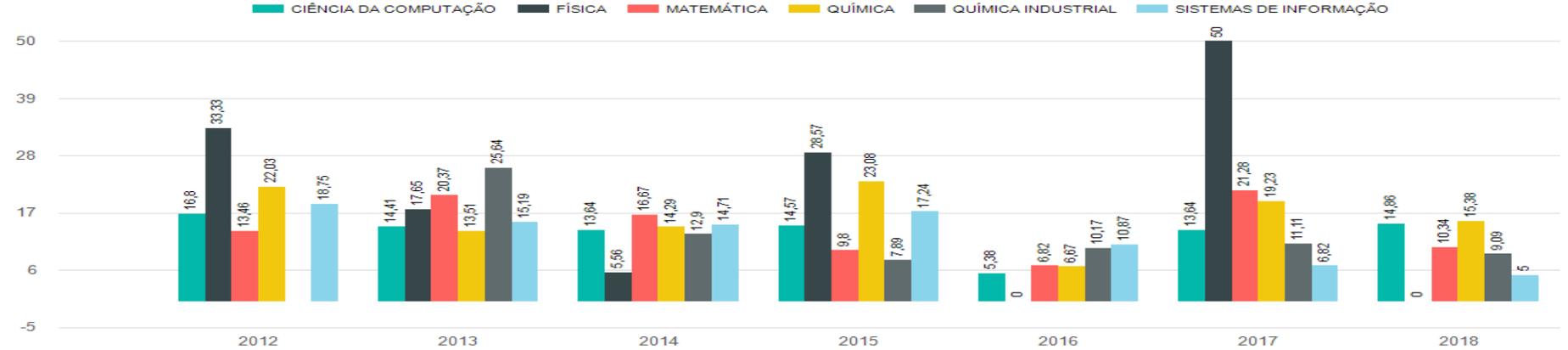
Para a grande área de Ciências Exatas e da Terra, a tabela com os dados correspondentes encontra-se no apêndice H, pág. 216.

**Gráfico 24 - Índice de evasão dos cursos da grande área de Ciências Exatas e da Terra entre 2012 e 2018 para o 1º semestre de cada ano**



Fonte: Próprio autor

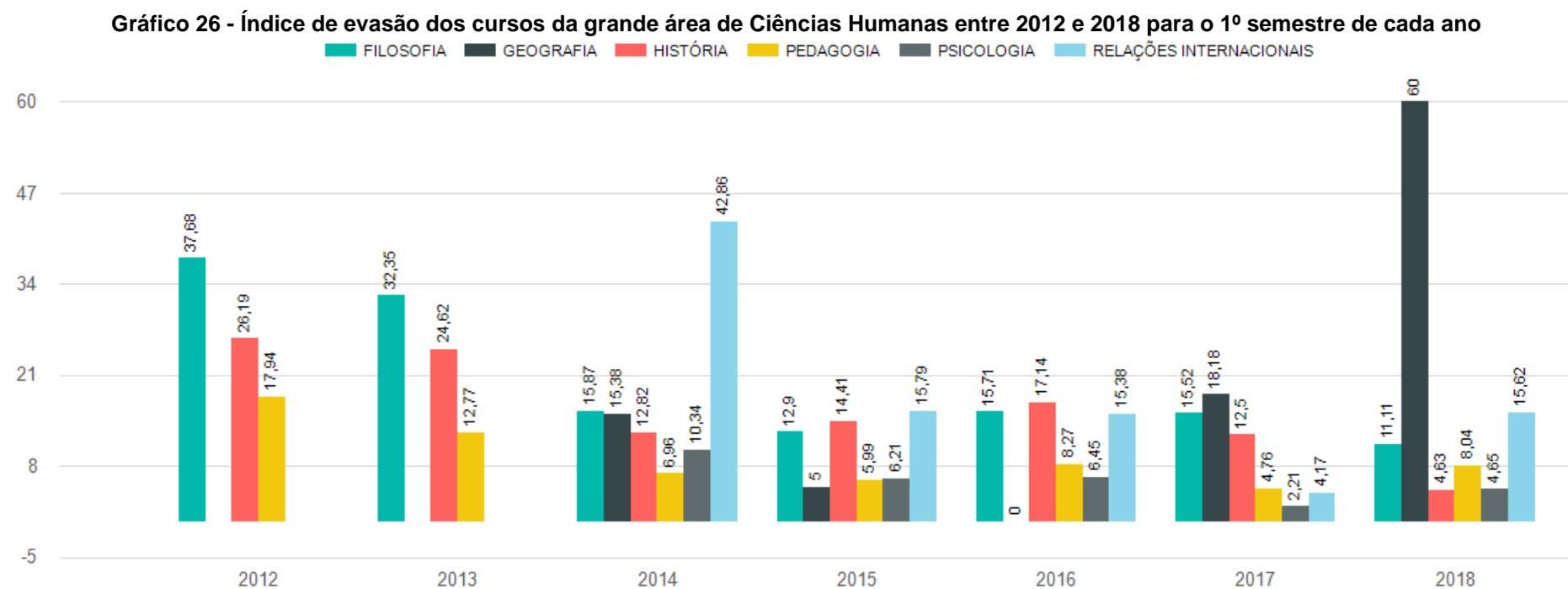
**Gráfico 25 - Índice de evasão dos cursos da grande área de Ciências Exatas e da Terra entre 2012 e 2018 para o 2º semestre de cada ano**



Fonte: Próprio autor

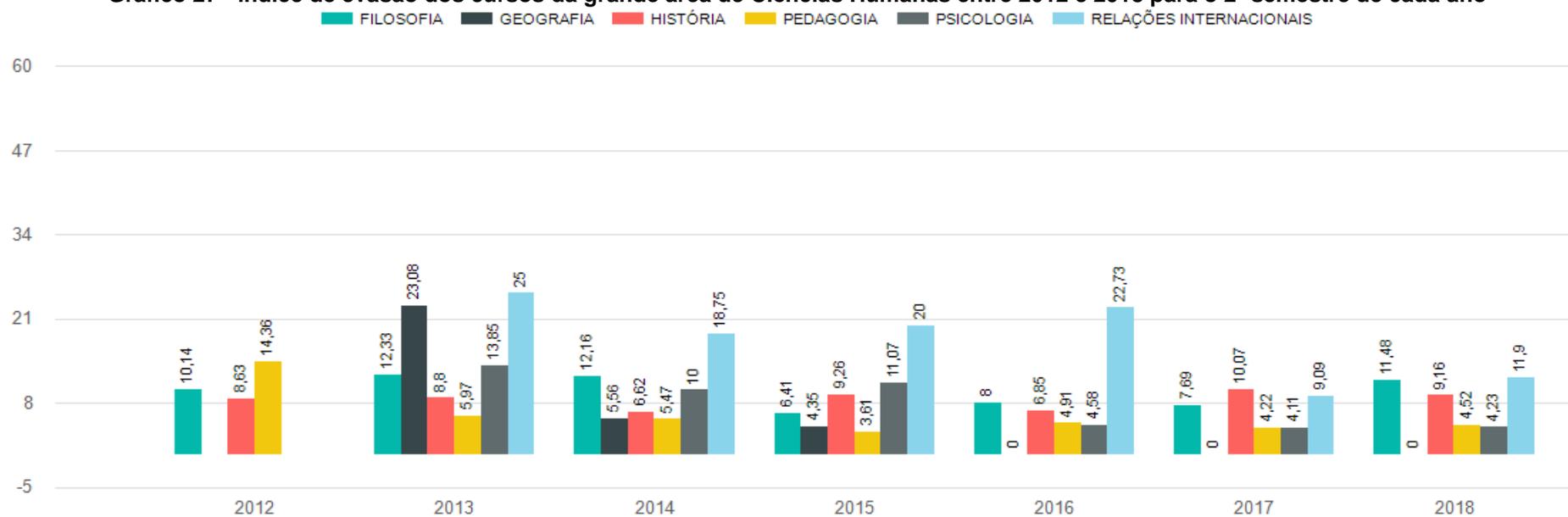
#### 4.2.2.6 Índices de evasão dos cursos de Ciências Humanas

Para a grande área de Ciências Humanas, visualizadas no item 4.2.2.6, a tabela com os dados correspondentes encontra-se no apêndice I, pág. 233.



Fonte: Próprio autor

**Gráfico 27 - Índice de evasão dos cursos da grande área de Ciências Humanas entre 2012 e 2018 para o 2º semestre de cada ano**

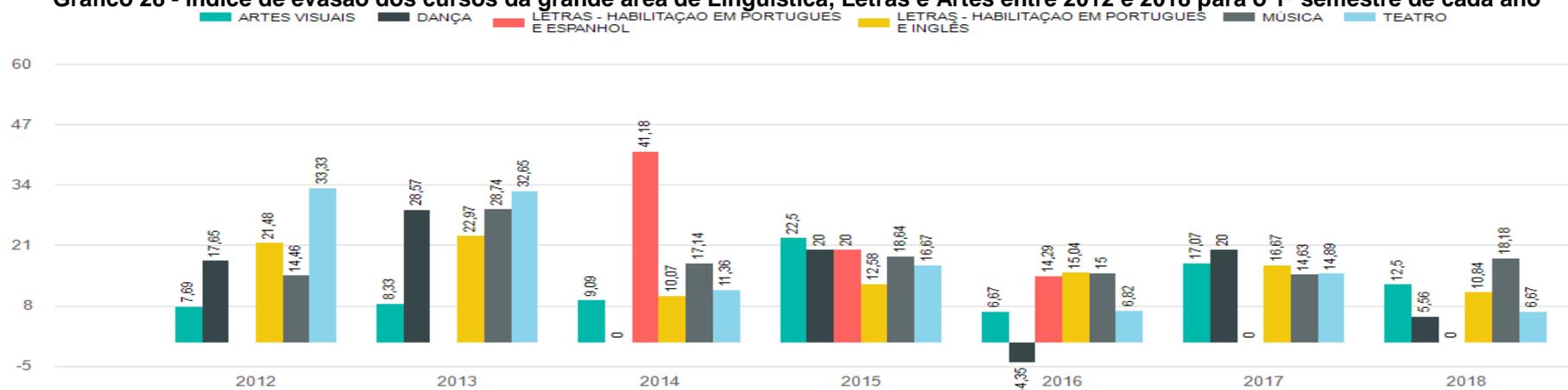


Fonte: Próprio autor

#### 4.2.2.7 Índices de evasão dos cursos de Linguística, Letras e Artes

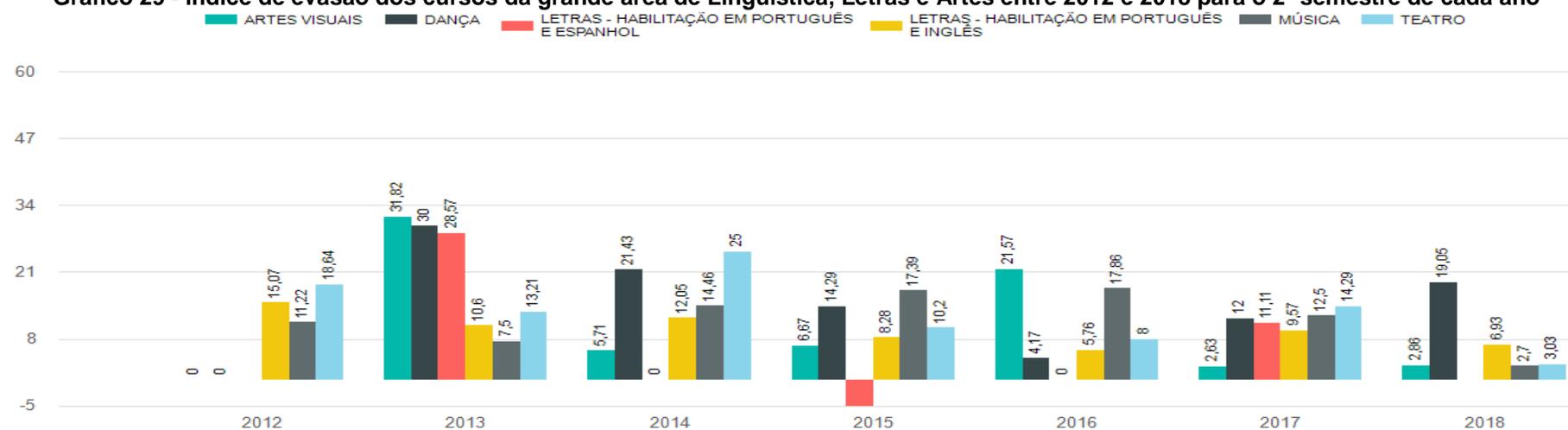
Para a grande área de Linguística Letras e Artes, a tabela com os dados correspondentes encontra-se no apêndice J, pág. 241.

**Gráfico 28 - Índice de evasão dos cursos da grande área de Linguística, Letras e Artes entre 2012 e 2018 para o 1º semestre de cada ano**



Fonte: Próprio autor

**Gráfico 29 - Índice de evasão dos cursos da grande área de Linguística, Letras e Artes entre 2012 e 2018 para o 2º semestre de cada ano**

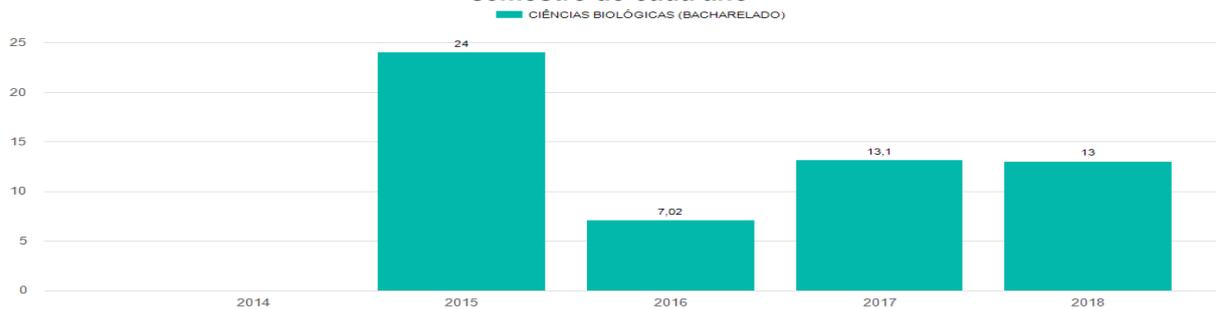


Fonte: Próprio autor

#### 4.2.2.8 Índices de evasão dos cursos de Ciências Biológicas

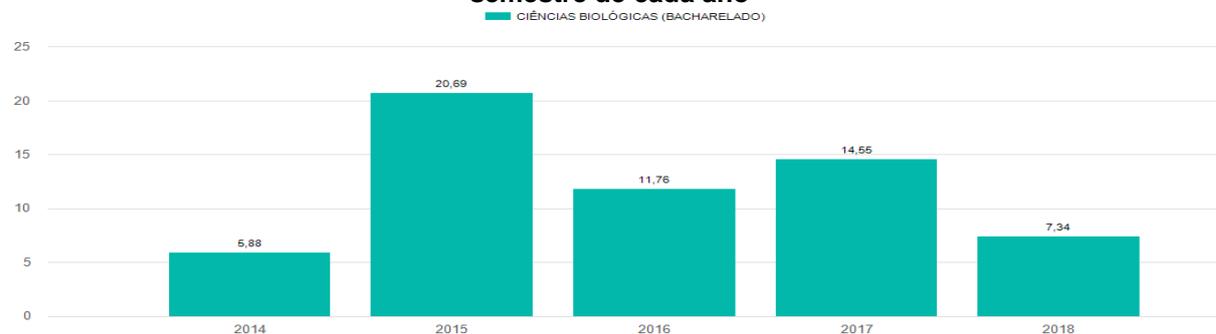
Para a grande área de Ciências Biológicas, a tabela com os dados correspondentes encontra-se no apêndice K, pág. 245.

**Gráfico 30 - Índice de evasão para o curso de Ciências Biológicas entre 2012 e 2018 para o 1º semestre de cada ano**



Fonte: Próprio autor

**Gráfico 31 - Índice de evasão para o curso de Ciências Biológicas entre 2012 e 2018 para o 2º semestre de cada ano**

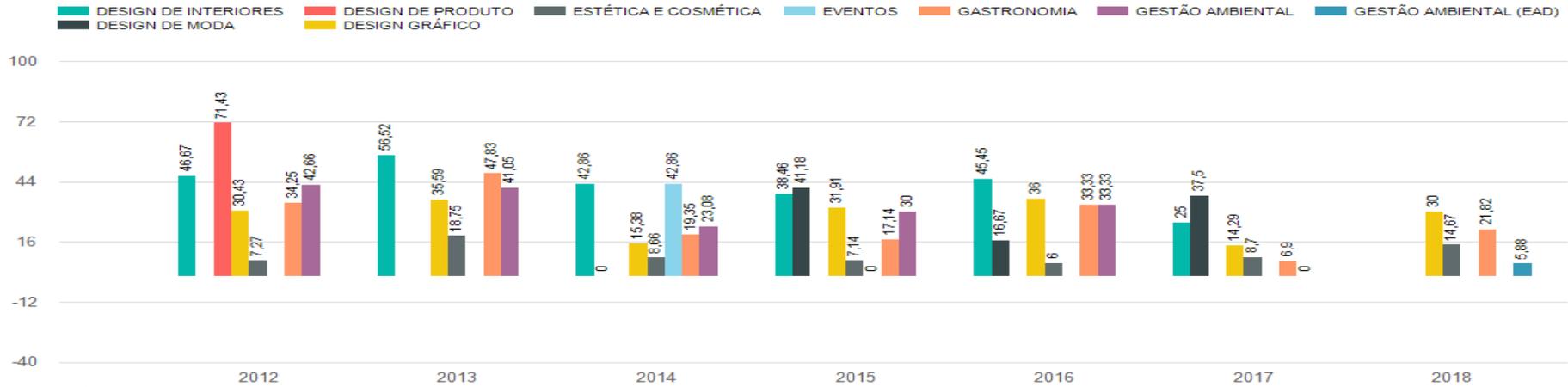


Fonte: Próprio autor

#### 4.2.2.9 Índices de evasão dos cursos tecnológicos

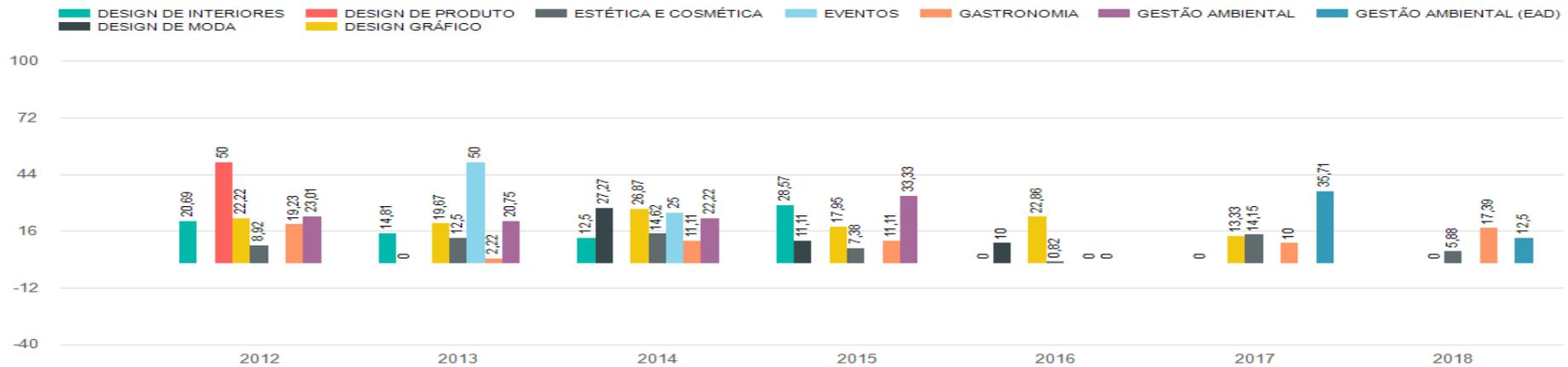
Os cursos tecnológicos foram mostrados separadamente para facilitar a comparação e os dados utilizados para os cálculos encontram-se no apêndice D ao J.

**Gráfico 32 – Índice de evasão dos cursos tecnológicos de Design de Interiores a Gestão Ambiental (EAD) entre 2012 a 2018 para os 1º semestres de cada ano**



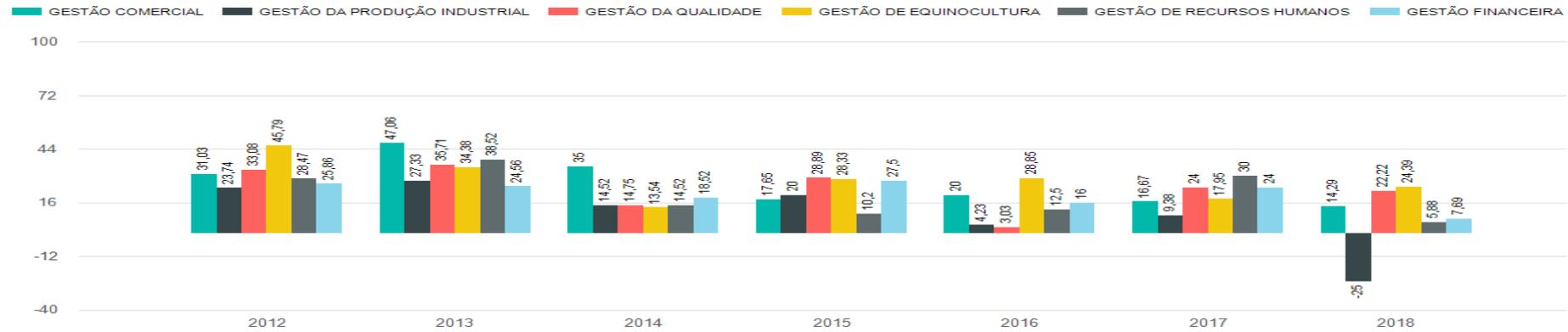
Fonte: Próprio autor

**Gráfico 33 - Índice de evasão dos cursos tecnológicos de Design de Interiores a Gestão Ambiental (EAD) entre 2012 a 2018 para os 2º semestres de cada ano**



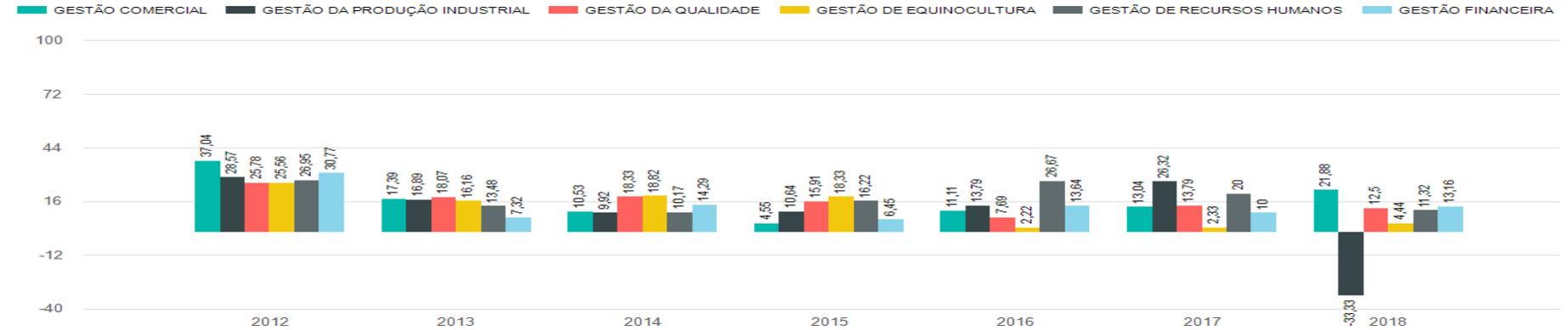
Fonte: Próprio autor

**Gráfico 34 - Índice de evasão dos cursos tecnológicos de Gestão Comercial a Gestão Financeira entre 2012 a 2018 para os 1º semestres de cada ano**



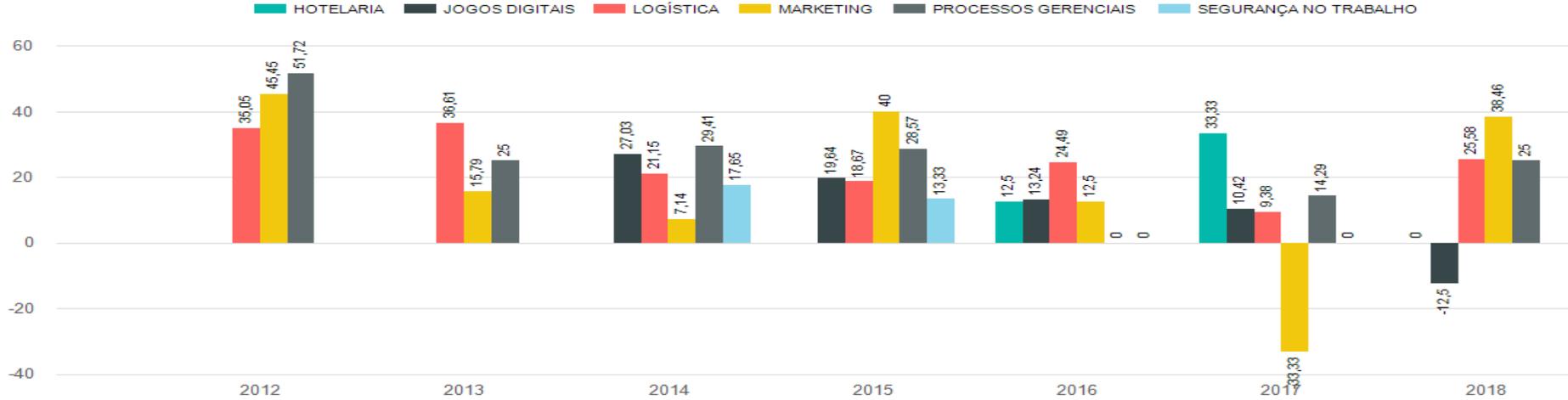
Fonte: Próprio autor

**Gráfico 35 - Índice de evasão dos cursos tecnológicos de Gestão Comercial a Gestão Financeira entre 2012 a 2018 para os 2º semestres de cada ano**



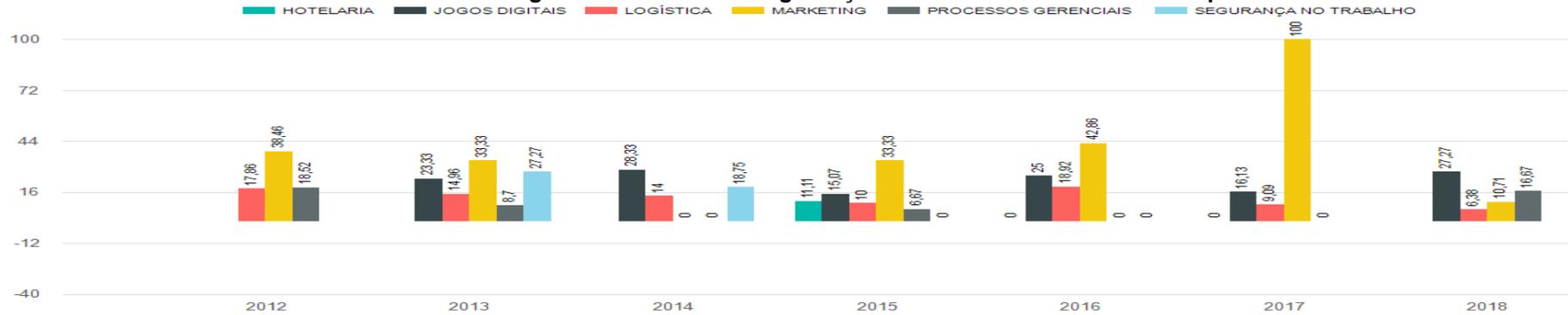
Fonte: Próprio autor

**Gráfico 36 - Índice de evasão dos cursos tecnológicos de Hotelaria a Segurança no Trabalho entre 2012 a 2018 para os 1º semestres de cada ano**



Fonte: Próprio autor

**Gráfico 37 - Índice de evasão dos cursos tecnológicos de Hotelaria a Segurança no Trabalho entre 2012 a 2018 para os 2º semestres de cada ano**



Fonte: Próprio autor

### 4.3 Mineração dos dados

O processo de mineração dos dados foi realizado em duas etapas, onde na primeira, consistia em minerar os dados de todo o Data Mart, ou seja, da população total, independente do aluno ser ou não evadido.

Trabalhando desta forma, o montante dos fatos da evasão representa apenas 7,1% do todo, ou seja, de uma população de 98275 fatos registrados no *Data Mart*, que representam todas as mudanças de status (no período letivo) para todos os alunos na faixa de tempo de 2012 a 2018, apenas 7% diz respeito a evasão.

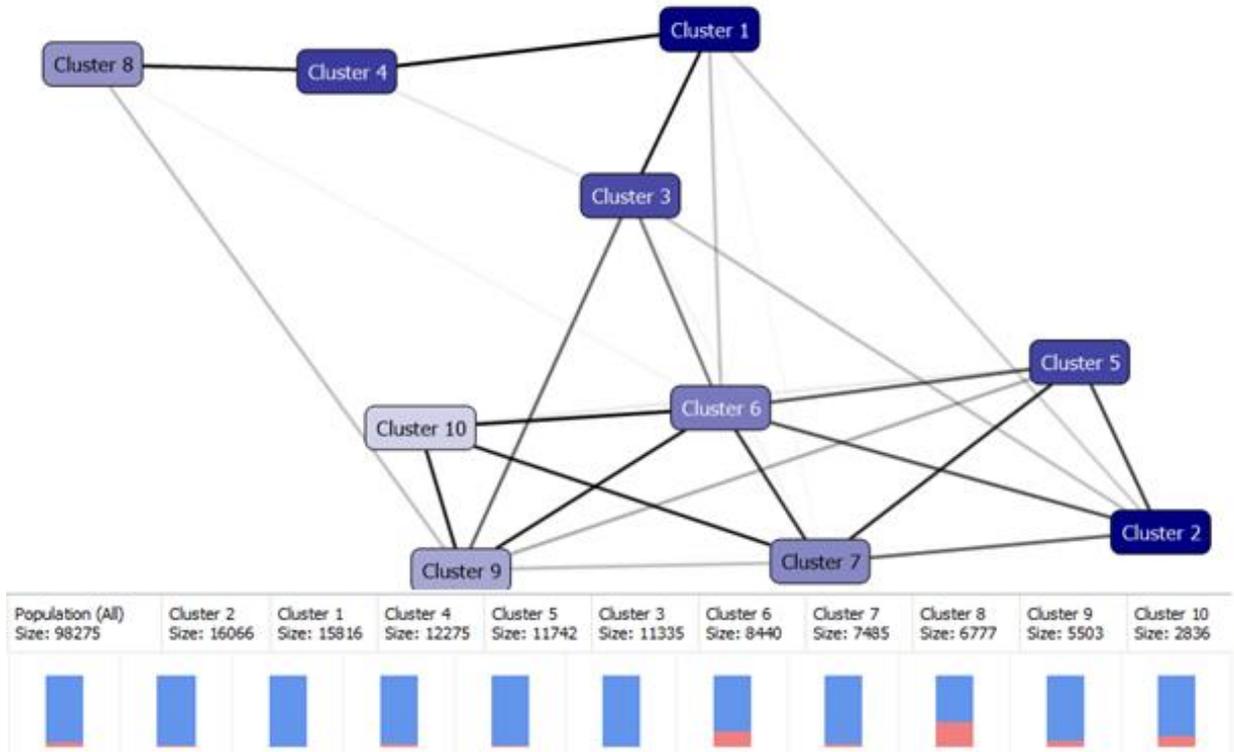
Essa primeira etapa se fez necessário para situar em quais grupos (*clusters*) os fatos de evasão se encontrariam e quais regras o algoritmo de árvore de decisão destacam no contexto geral.

A segunda etapa realiza o mesmo processo de mineração da primeira, porém especificamente com os dados relacionados aos fatos de evasão, para com isso categorizar a evasão já destacada no primeiro processo e identificar as variáveis com maior correlação.

#### 4.3.1 Clustering aplicado a todos os dados

Com o auxílio da ferramenta *Analisis Service* da Microsoft, após a criação do *Data Mart*, detalhado no tópico 3.3.4 e parametrizado, segundo os moldes detalhados no tópico 3.3.5, a aplicação do método de *clustering* é visualizado na Figura 14.

Figura 14 – Visualização do modelo de mineração do algoritmo de *Clustering* para todos os dados



Fonte: Próprio autor

Como mostrado na Figura 14, foram gerados 10 clusters (grupos), onde a maior parte da população encontra-se nos grupos 1, 2, 3, 4 e 5, essa informação pode ser comprovada pelo atributo *size*, abaixo de cada *cluster* e pela cor de cada *cluster* que serve para representar o volume do grupo (quanto mais escura a cor azul, maior a população do grupo).

As quantidades de casos de fatos de evasão são visualizadas na parte inferior da Figura 14, que mostra em vermelho quanto da população do grupo são de casos de evasão.

Do montante total, a porcentagem de casos analisados por grande área, pode ser visualizado na Figura 15.

Figura 15 - Porcentagem de casos por grande área

Value	Cases	Probability	Histogram
<input checked="" type="checkbox"/> Ausente	0	0,00%	
<input checked="" type="checkbox"/> Ciências Agrárias	5033	5,12%	
<input checked="" type="checkbox"/> Ciências Biológicas	1010	1,03%	
<input checked="" type="checkbox"/> Ciências da Saúde	10298	10,48%	
<input checked="" type="checkbox"/> Ciências Exatas e da Terra	3975	4,05%	
<input checked="" type="checkbox"/> Ciências Humanas	7672	7,81%	
<input checked="" type="checkbox"/> Ciências Sociais Aplicadas	41278	42,00%	
<input checked="" type="checkbox"/> Engenharias	25370	25,81%	
<input checked="" type="checkbox"/> Linguística, Letras e Artes	3639	3,70%	

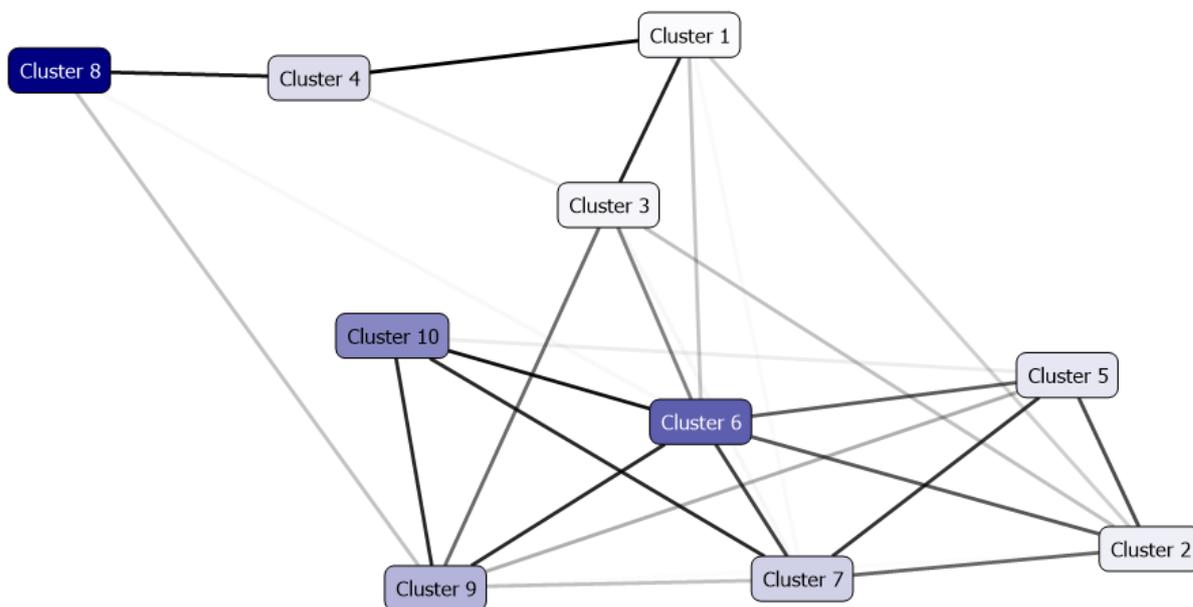
Fonte: Próprio autor

No geral, ou seja, a população como um todo é formado por:

- 85% solteiros;
- 75% da cor branca;
- 69% terminou o 2º grau em escola pública;
- 63% no turno da noite;
- 60% com algum tipo de bolsa de estudo;
- 54% do sexo feminino;
- 42% da área de Ciências Sociais Aplicadas;
- 25% da área de Engenharias;
- 25% com curso de valor médio entre R\$919,90 – R\$1.110,20;
- 25% com curso com valor médio entre R\$729,60- R\$919,20.

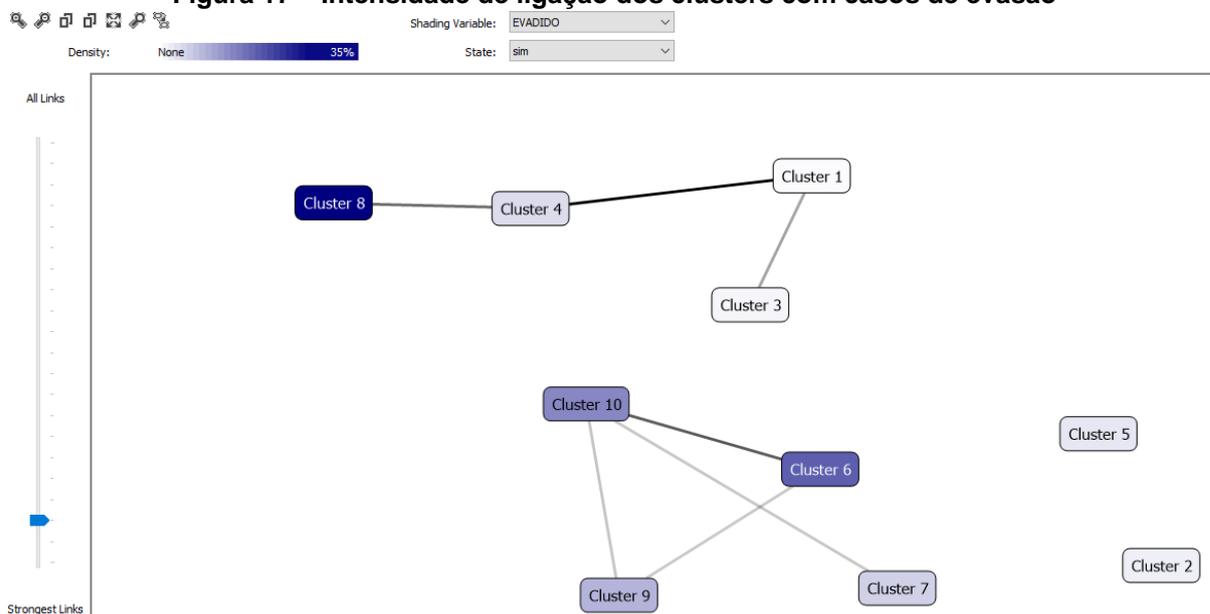
Nos casos de evadidos, a maior parte da população encontra-se nos clusters: 8 com maior relevância, seguido do 10 e 6, como mostrado na Figura 16.

**Figura 16 - Clusters com os casos de alunos evadidos**



Fonte: Próprio autor

**Figura 17 – Intensidade de ligação dos clusters com casos de evasão**



Fonte: Próprio autor

Diminuindo a força de ligação, que representa a afinidade que um *cluster* tem com relação a outro *cluster*, observa-se que o cluster 8 demonstrou maior ligação com o cluster 4, enquanto o cluster 10 e 6 demonstraram maior ligação com o cluster 9.

Dentre as características destacadas pela análise no cluster 8 estão:

- 35% dos casos do cluster 8 são evadidos;
- 84% dos casos são solteiros;
- 75% da cor branca;
- 72% com quantidade de aprovações de 0 -10;
- 72% terminou o segundo grau em escola pública;
- 70% estudo de 11 – 32 meses;
- 64% com número total de matrículas de 1 – 15
- 54% do turno da manhã;
- 51% do sexo masculino;
- 49% está entre o 3 – 4 períodos;
- 47% com faixa de idade entre 21 – 25 anos;
- 40% estão concentrados na área de Ciências Sociais Aplicadas;
- 40% dos casos concentrados entre os períodos de 1 – 2;
- 28% dos casos com valor médio do curso entre R\$729,60 – R\$919,90;
- 27% dos casos com valor médio do curso entre R\$919,90 – R\$1.110,20;
- 19% dos casos com valor médio do curso entre R\$1.110,20 – R\$1766,50.

A ferramenta revelou essas características na Figura 18, sendo que para visualizar as porcentagens, foi necessário entrar em detalhes de cada variável.

**Figura 18 - Características do cluster 8**

Characteristics for Cluster 8

Variables	Values	Probability
ESTADOCIVIL	Solteiro	
CORRACA	Branca	
QTD APROVADAS	0 - 10	
INST2G	Publica	
QUANTIDADEMESES	11 - 32	
MESESENTREDTINICIOEDTFIMPLET...	11 - 32	
EVADIDO	não	
N MATRTOTAL	1 - 15	
TURNO	MANHA	
TEMBOLSA	NÃO	
SEXO	M	
QTD PERIODOS	3 - 4	
SEXO	F	
FAIXA IDADE	21 - 25	
QTD TRANCADAS	2 - 6	
TEMBOLSA	SIM	
QTD CANCELADAS	3 - 7	
QTD DISPENSADAS	0 - 1	
TURNO	NOITE	
QTD REPROVADAS	0 - 2	
DESCSTATUSPLETIVO	Trancado	
STATUSCURSO	Trancado	
AREA	Ciências Sociais Aplicadas	
QTD PERIODOS	1 - 2	
EVADIDO	sim	
N MATRTOTAL	16 - 28	

Fonte: Próprio autor

A descrição dos nomes das variáveis mostradas na Figura 18 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

Dentre as características destacadas no cluster 10 estão:

- 16% dos evadidos estão nesse grupo;
- 81% dos casos são do turno da noite;
- 73% dos casos solteiros;
- 69% dos casos não tem nenhum tipo de bolsa de estudo;
- 60% dos casos do sexo masculino;
- 57% dos casos cursaram de 8 – 15 períodos;
- 52% dos casos possui um total de matrículas de 42 – 86;
- 49% dos casos da grande área de Ciências Sociais aplicadas;
- 46% dos casos terminou o 2º grau em escola privada;
- 44% dos casos possui de 6 – 19 dispensas;
- 42% dos casos cancelaram de 3 – 7 componentes curriculares;

- 38% dos casos tiveram de 6 – 16 reprovadas;
- 36% dos casos tiveram de 2 – 6 trancamentos de componentes curriculares;
- 27% dos casos da grande área de Engenharias.

Características mostradas na Figura 19.

**Figura 19 - Características do cluster 10**

Characteristics for Cluster 10

Variables	Values	Probability
EVADIDO	não	
TURNO	NOITE	
DESCSTATUSPLETIVO	Matriculado	
ESTADOCIVIL	Solteiro	
TEMBOLSA	NÃO	
CORRACA	Branca	
STATUSCURSO	Matriculado	
SEXO	M	
QTD PERIODOS	8 - 15	
INST2G	Publica	
N MATRTOTAL	42 - 86	
AREA	Ciências Sociais Aplicadas	
INST2G	Privada	
QTD DISPENSADAS	6 - 19	
QTD CANCELADAS	3 - 7	
SEXO	F	
QTD REPROVADAS	6 - 16	
QTD APROVADAS	35 - 75	
QTD TRANCADAS	2 - 6	
MESESENTREDTINICIOEDTFIMPLET...	55 - 129	
QUANTIDADEMESES	55 - 129	
FAIXA IDADE	26 - 30	
TEMBOLSA	SIM	
CURSOVALORMEDIO	729,6 - 919,9	
AREA	Engenharias	
DTFIMPLETIVO	30/06/2012 00:00:00 - 30/03/2014 ...	

Fonte: Próprio autor

A descrição dos nomes das variáveis mostradas na Figura 19 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

No cluster 6 demonstrou muita equivalência com o cluster 10, diferenciando-se em:

- 22% dos casos evadidos;
- 76% em terminou o segundo grau em escola pública;
- 46% dos casos com faixa de idade entre 26 – 30 anos;

- 52% dos casos do sexo masculino;

**Figura 20 - Características do cluster 6**

Characteristics for Cluster 6

Variables	Values	Probability
ESTADOCIVIL	Solteiro	
EVADIDO	não	
DESCSTATUSPLETIVO	Matriculado	
INST2G	Publica	
TURNO	NOITE	
CORRACA	Branca	
STATUSCURSO	Matriculado	
TEMBOLSA	NÃO	
SEXO	M	
QTD REPROVADAS	6 - 16	
QTD DISPENSADAS	0 - 1	
SEXO	F	
AREA	Ciências Sociais Aplicadas	
FAIXA IDADE	26 - 30	
QTD REPFALTAS	1 - 2	
TEMBOLSA	SIM	
QTD CANCELADAS	3 - 7	
CURSOVALORMEDIO	729,6 - 919,9	
QUANTIDADEMESES	33 - 54	
MESESENTREDTINICIOEDTFIMPLET...	33 - 54	
QTD PERIODOS	8 - 15	
N MATRTOTAL	42 - 86	
QUANTIDADEMESES	11 - 32	
MESESENTREDTINICIOEDTFIMPLET...	11 - 32	
FAIXA IDADE	21 - 25	
QTD APROVADAS	23 - 34	

Fonte: Próprio autor

A descrição das variáveis mostradas na Figura 20 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

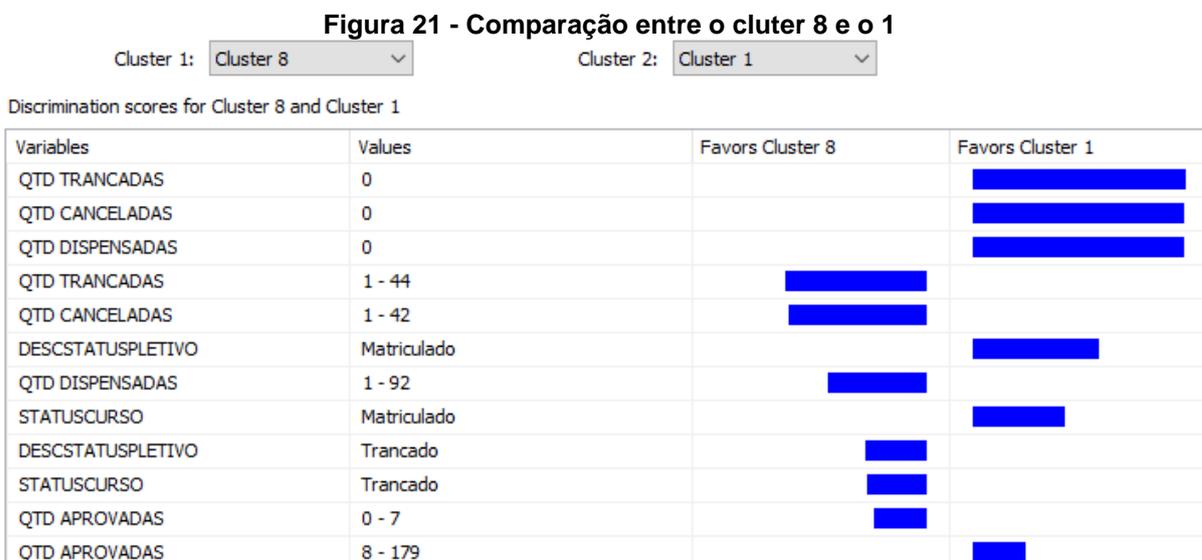
Entre os três *clusters* que concentram a maior população de casos relacionados a evasão, em todos os casos a quantidade de alunos do “sexo masculino” era superior a quantidade de alunos do “sexo feminino”, entretanto foi mostrado no tópico 4.1 que a quantidade geral de alunos do sexo feminino é superior à do sexo masculino.

Fazendo uma comparação entre *clusters*, com o auxílio de um recurso da ferramenta *Analisis Service* que discrimina as variáveis de um cluster selecionado em comparação com outro, ao selecionar o cluster 8 (que possui maior quantidade de casos evadidos) com o cluster 1 (que possui a maior quantidade de casos no total) o cluster 8 obteve destaque nas seguintes variáveis:

- quantidade de trancamentos entre 1 e 44,

- quantidade de cancelamentos entre 1 e 42,
- quantidade de dispensas entre 1 e 92,
- quantidade de aprovações menor igual a 7.

Pode-se observar esse resultado na Figura 21 abaixo.



Fonte: Próprio autor

A descrição das variáveis mostrada na Figura 21 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

Essas mesmas variáveis destacadas em comparação do cluster 8 com o cluster 1, demonstraram-se relevantes quando comparado com cluster 8 com os cluster 2, 3, 4, 5, que possui a maior parte da população de casos.

O resultado obtido na comparação do cluster 8 com os clusters 2, 3, 4, 5, destacou que o aumento da quantidade de trancamentos, cancelamentos e dispensas com valor baixo na quantidade de componentes curriculares aprovadas é um indício da concentração do conjunto de evadidos no cluster 8.

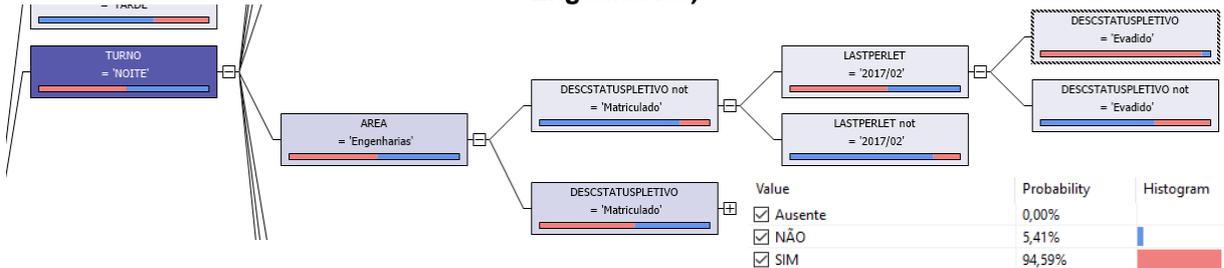
#### 4.3.2 Árvore de decisão aplicado a todos os dados

A árvore de decisão gerada para os mesmos dados, segundo a parametrização definida na Figura 8, pag. 60, não demonstrou diretamente regras para o parâmetro *evadido* na aplicação do método de árvore de decisão, apenas algumas regras onde

esse parâmetro entrava como um nó da árvore, ou seja, como parte de uma condição para atingir um resultado.

Dentre os parâmetros destacados do algoritmo de árvore de decisão, é o parâmetro de bolsa de estudo na categorização da evasão.

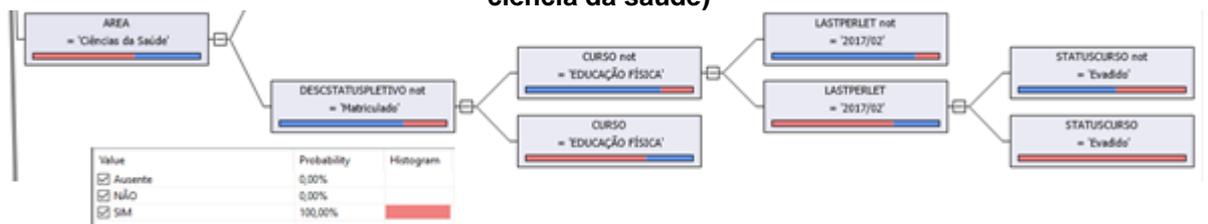
**Figura 22 - Árvore de decisão para previsão de bolsa de estudo (turno da noite, grande área de Engenharias)**



Fonte: Próprio autor

Como mostrado na Figura 22, quando o TURNO é noturno e a GRANDE ÁREA for engenharias e o último PERÍODO LETIVO for o 2º semestre de 2017 e o PERÍODO LETIVO estiver indicando evasão, verifica-se que 94,59% deles, em algum momento de sua vida acadêmica, o aluno já teve algum tipo de bolsa de estudo, ou seja, o algoritmo destacou essa regra devido à alta quantidade de casos de discentes nessas circunstâncias.

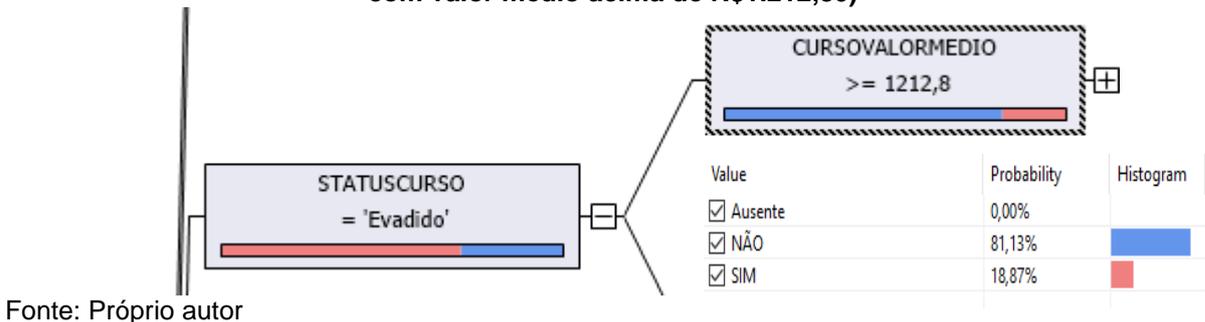
**Figura 23 - Árvore de decisão para previsão de bolsa de estudo (turno da noite, grande área de ciência da saúde)**



Fonte: Próprio autor

Outra associação relacionada a discentes evadidos é mostrada na Figura 23, quando o TURNO for noturno e a GRANDE ÁREA for ciências da saúde e o STATUS do aluno no período letivo não for matriculado e o CURSO não for educação física e o último PERÍODO LETIVO for o 2º semestre de 2017 e o STATUS do aluno no curso for evadido, em 100% dos casos o aluno, em algum momento de sua vida acadêmica, teve algum tipo de bolsa de estudos.

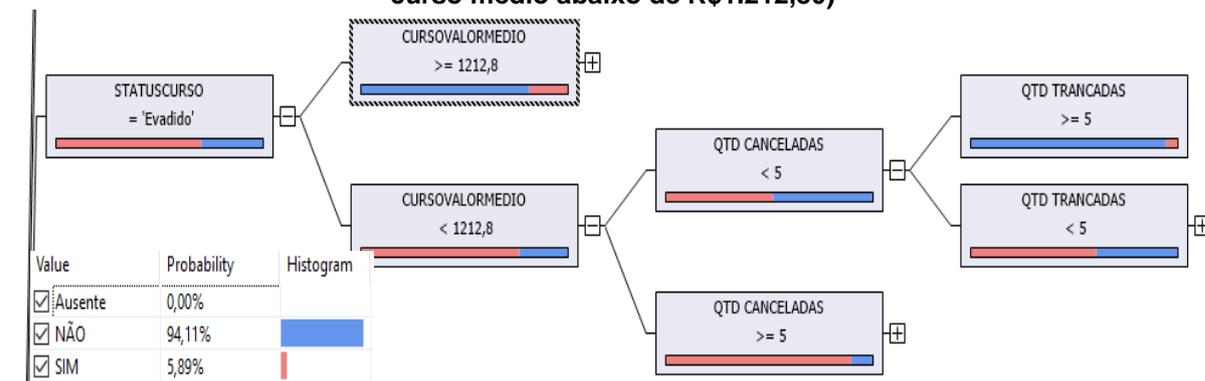
**Figura 24 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, com valor médio acima de R\$1.212,80)**



Fonte: Próprio autor

Na ramificação mostrada na Figura 24, quando o TURNO for manhã e o STATUS do aluno no curso for evadido e o VALOR MÉDIO do curso for maior ou igual a R\$1212,80, em 81,12% dos casos o aluno não teve bolsa de estudo.

**Figura 25 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, curso médio abaixo de R\$1.212,80)**

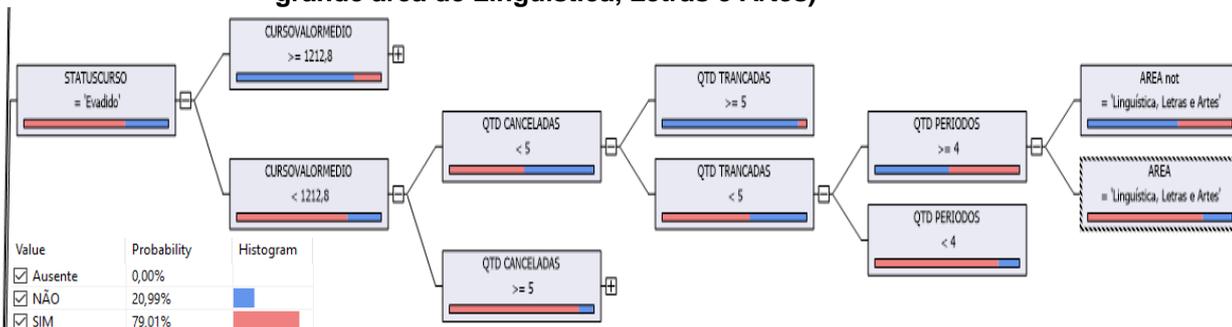


Fonte: Próprio autor

Na Figura 25, o resultado na árvore mostrou que quando o TURNO for manhã e o STATUS no aluno no curso for evadido e o VALOR MÉDIO do curso for menor que R\$1.212,80 e a quantidade de COMPONENTES CURRICULARES canceladas for menor que 5 e a quantidade de COMPONENTES CURRICULARES trancadas for maior ou igual a 5, existe uma probabilidade de 94,11% do aluno não ter tido bolsa de estudo, enquanto em 5,89% dos casos, do aluno já ter usufruído de algum tipo de bolsa de estudo.

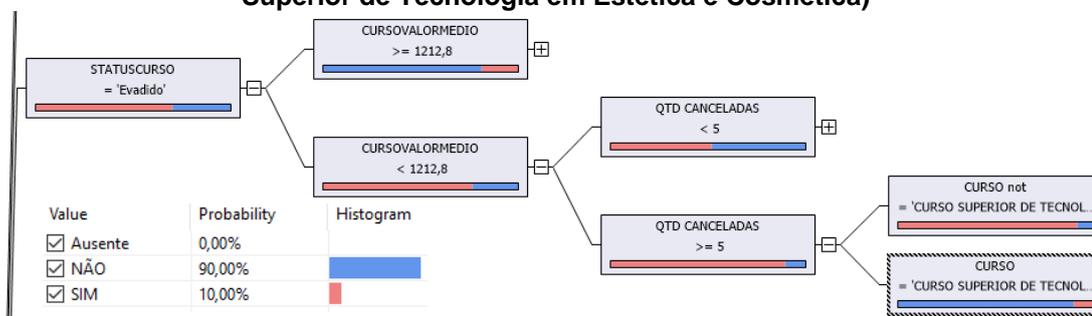
Na Figura 25 quando o TURNO do aluno é de manhã e seu STATUS no curso for evadido e o VALOR MÉDIO DO CURSO for menor que R\$1.212,80 e a quantidade de CANCELAMENTOS for menor que 5 e a quantidade de TRANCAMENTOS for menor que 5 e a quantidade de PERÍODOS estudados pelo aluno for maior ou igual a 4 e a GRANDE ÁREA for linguística, letras e artes, existe uma probabilidade de 79,01% do aluno já ter tido bolsa de estudo e 20,99% de não ter tido bolsa de estudo.

**Figura 26 - Árvore de decisão para o estudo do parâmetro de bolsa de estudo (turno da manhã, grande área de Linguística, Letras e Artes)**



Fonte: Próprio autor

**Figura 27 - Árvore de decisão para previsão de bolsa de estudo (turno da manhã para o Curso Superior de Tecnologia em Estética e Cosmética)**



Fonte: Próprio autor

Na Figura 27, observa-se na árvore que, quando o TURNO do aluno for de manhã e o STATUS no aluno no curso for evadido e o valor médio do aluno no curso for menor que R\$1.212,80 e a quantidade de COMPONENTES CURRICULARES canceladas for maior ou igual a 5 e o CURSO for o curso superior de tecnologia em estética e cosmética, existe uma probabilidade de 90% do aluno não ter tido bolsa de estudo durante o período de vida acadêmica naquele curso.

Apesar do algoritmo aplicado dar relevância a regras relacionadas a bolsa de estudo é necessário levar em consideração que a Uniso é uma universidade filantrópica, ou seja, diferente de instituições de ensino superior privada, não tem fins lucrativos. Por esse motivo possui um alto índice de disponibilização de bolsas de estudo, a fim de prestar serviços à sociedade e principalmente às pessoas mais carentes.

Outro fator a ser considerado com relação a variável bolsa de estudo diz respeito a estratégias aplicadas pela Uniso para melhor alocar os espaços físicos, dentre elas, a aplicação de bolsa para cursos matutinos, onde o aluno ganha bolsa nos primeiros períodos do curso e depois de um determinado tempo estipulado em

contrato passa para o turno da noite sem bolsa de estudo, como nos casos das licenciaturas.

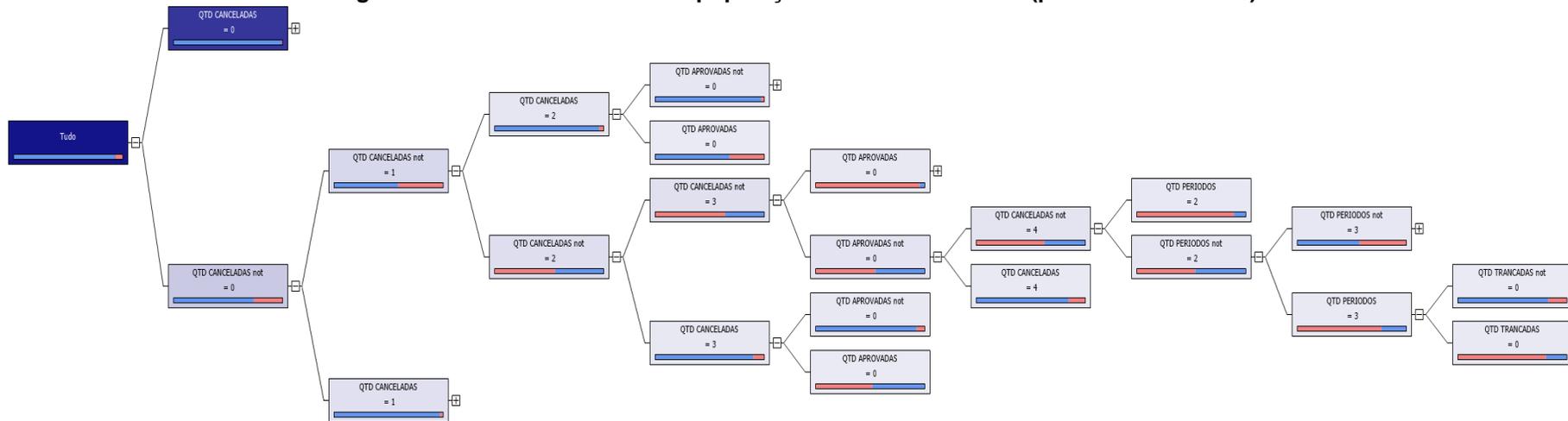
Em um estudo dessa variável em nível geral, ou seja, os cursos como um todo, esse alto volume de bolsas disponibilizadas em estratégias específicas pode trazer conclusões erradas quando considerado para o entendimento da evasão.

A técnica de mineração de dados aplicado a todos os dados do *Data Mart* também encontrou padrões relacionados ao parâmetro total de matrículas adicionados ao Apêndice I.

#### 4.3.2.1 Árvore de decisão aplicado a todos os dados separados por turnos

Separando a população estudada em turno da manhã e noite e configurando a variável evasão como previsão no turno da manhã foi gerada a seguinte árvore (Figura 28).

Figura 28 - Árvore de decisão da população do turno da manhã (previsão de evasão)



Fonte: Próprio autor

Quanto maior a quantidade de CANCELAMENTOS, maior a quantidade de evasões no período da manhã,

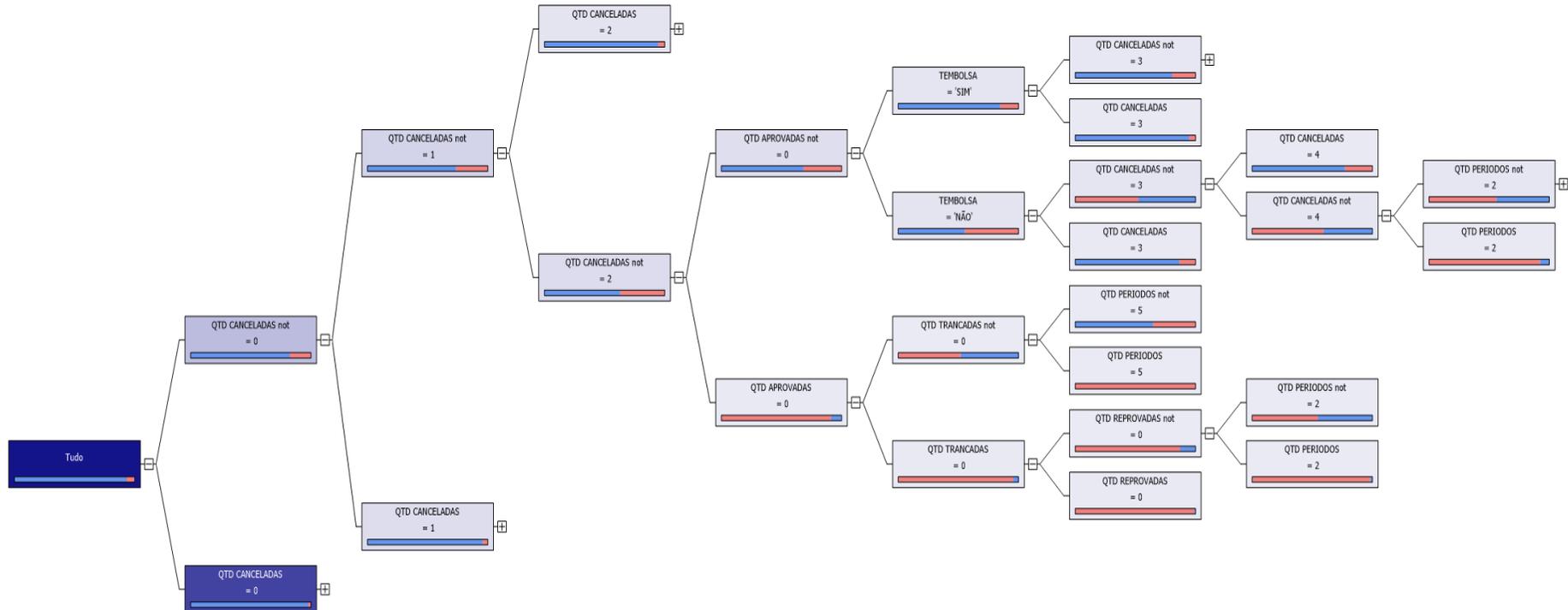
Na árvore acima configurada para prever os evadidos do PERÍODO da manhã, pode-se observar abaixo de cada nó uma faixa de cores vermelha e azul, quanto maior a cor vermelha maiores são as chances de evasão, onde a árvore demonstra:

- QTD\_CANCELADAS (quantidade de disciplinas canceladas) igual a 1 tem um valor baixo de evadidos (probabilidade de 3,06% de chance de evasão).
- QTD\_CANCELADAS (quantidade de disciplinas canceladas) igual a 2 a chance subiu para 4,68%.
- QTD\_CANCELADAS (quantidade de disciplinas canceladas) igual a 3 sobe para 10,23% de chance de evasão, onde dessa população 51,31% dos casos evadidos não tinha nenhuma disciplina aprovada.
- QTD\_CANCELADAS (quantidade de disciplinas canceladas) igual a 4 sobe para 15,53% de chance de evasão.
- QTD\_CANCELADAS (quantidade de disciplinas canceladas) maior que 4 e QTD\_PERIODOS igual a 2, as chances de evasão sobem para 88,49%.

A última regra mostra o ponto mais alto de índice de evasão no período da manhã.

Gerando a árvore de decisão com a população da noite e configurando a variável evasão como previsão gerou-se a seguinte árvore.

**Figura 29 - Árvore de decisão do turno da noite (previsão de evasão)**



Fonte: Próprio autor

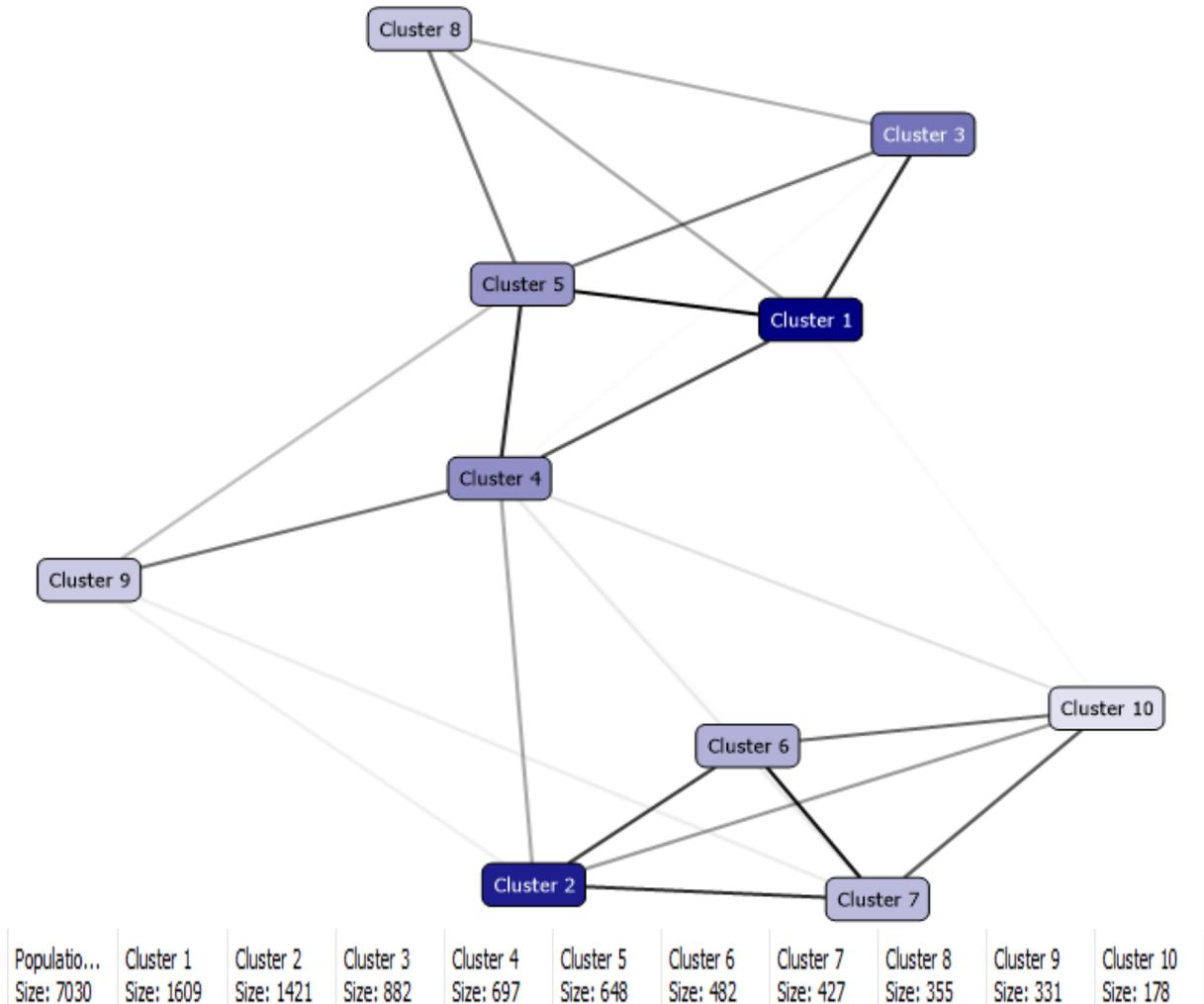
Ao observar o comportamento da árvore de decisão para a população do TURNO da noite, configurando para prever a probabilidade de evasão destacado no modelo em vermelho, fica claro que o aumento das quantidades de CANCELAMENTOS

prediz o aumento da evasão na população da noite apontando para o 2º período e o 5º período letivo com aumento na probabilidade de evasão.

### 4.3.3 Clustering aplicado aos dados dos evadidos

Com o auxílio da ferramenta *Analysis Service*, ao executar o algoritmo de *Clustering*, a ferramenta mostrou o seguinte resultado visto no diagrama apresentado na Figura 30.

Figura 30 - Visualização do modelo de mineração do algoritmo de Clustering para todos os fatos categorizados como evadido



Fonte: Próprio autor

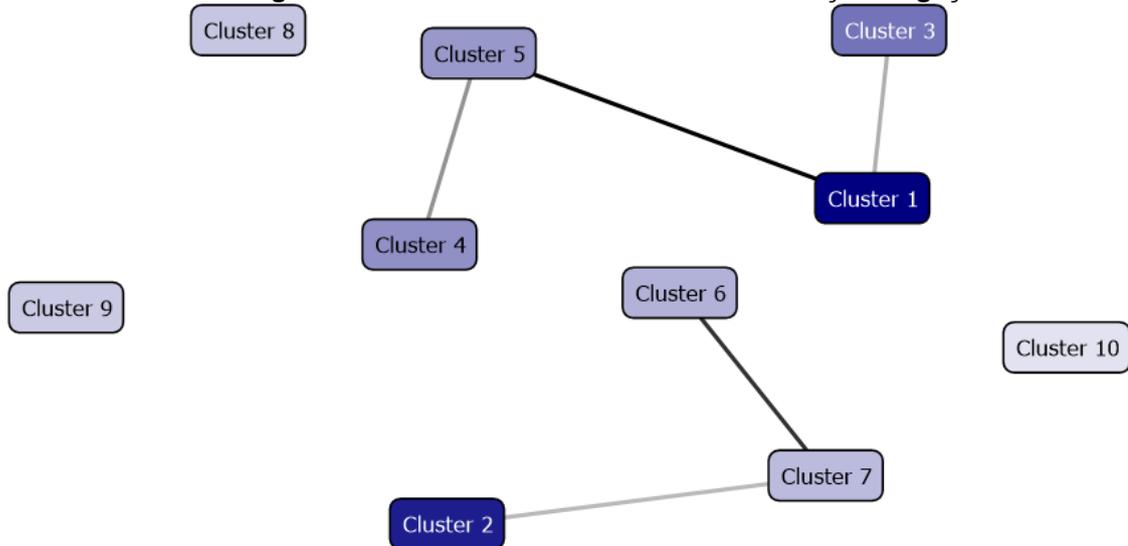
Como pode-se observar na Figura 30, os grupos (*cluster*), que apresentaram maior população são o *cluster 1* e o *cluster 2*, destacado na cor azul mais escura (quando mais escura a cor, maior o tamanho da população no *cluster*).

Outra característica é que o *cluster 1* teve mais forte sua ligação com o Cluster 5, enquanto o *cluster 2* teve mais forte sua ligação com o *cluster 7*. Detalhes que podem ser observados com o realce do traço que faz a ligação entre os *clusters*

(quanto maior for o realce da linha de ligação, maior será o grau de semelhança entre os *clusters*). Lembrando que um cluster reúne casos parecidos, em seu interior.

Essa característica também pode ser observada quando diminuimos a força da ligação na ferramenta, onde o mesmo modelo se apresenta segundo a Figura 31.

**Figura 31 - Visualização do modelo de mineração do algoritmo de Clustering para todos os fatos categorizados como evadido e diminuindo a força de ligação**



Fonte: Próprio autor

As características gerais destacadas na população analisada estão relacionadas abaixo em ordem de ocorrência:

- 83,86% solteiro,
- 69,82% terminou o 2º grau em escola pública,
- 62% dos casos do turno da noite,
- 61,54% não teve bolsa de estudo,
- 51,1% do sexo masculino,
- 45,16% da grande área de Ciências Sociais Aplicadas,
- 37,14% dos casos do turno da manhã,
- 33,52% entre a faixa de idade de 21 – 25 anos,
- 32,32% entre a faixa de idade de 26 – 30 anos,
- 24,99% com valor de curso médio de R\$715,6 – R\$888,70,
- 24,98% cursaram entre 6 e 7 períodos até o fato da evasão,
- 24,98% tinham entre 3 a 6 dispensas,
- 24,98% tinham entre 5 e 6 componentes curriculares canceladas,

- 24,98% tinham 1 reprova por falta,
- 24,98% tinham 5 a 15 componentes curriculares aprovadas,
- 24,98% tinham de 4 a 7 reprovadas,
- 24,98% tinham 1 componente curricular trancado,
- 24,98% cursaram de 3 a 5 períodos letivos,
- 24,98% com número de matrículas total de 14 a 26 componentes curriculares,
- 24,98% com número de matrículas total de 27 a 38 componentes curriculares,
- 24,83% dos casos da grande área de Engenharias,
- 24% dos casos com 0 a 3 reprovações,
- 21,92% dos casos com o valor médio do curso de R\$409,0 à R\$715,60,
- 14,32% com 0 a 2 componentes curriculares dispensados,
- 12,36% dos casos com estado civil casado,
- 9,94% dos casos da grande área de Ciências da Saúde,
- 9,74% dos casos ingressantes no 1º período de 2014,
- 8,86% dos casos do curso de Administração,
- 8,35% dos casos do curso de Direito,
- 7,71% dos casos com faixa etária entre 36 e 40 anos,
- 6,37% do curso de Engenharia Civil,
- 6,19% do curso de Engenharia da Produção,
- 5,70% dos casos da área de Ciências Humanas,
- 5,69% do curso de Arquitetura e Urbanismo.

Detalhes abaixo de 5% não foram citados.

As características destacadas acima na população geral e especificamente em cada cluster pode ser visualizada no Quadro 1. Os cabeçalhos das colunas estão sombreados em intensidades proporcionais de acordo com a proximidade (similaridade) dos clusters.

**Quadro 1 - Resultado do algoritmo de Clustering aplicado aos casos evadidos entre o 1º Sem. de 2012 e o 1º Sem. de 2018**

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	
<b>Grande Área</b>	<p>45,16% da grande área de Ciências Sociais Aplicadas;</p> <p>24,83% dos casos da grande área de Engenharias;</p> <p>9,94% dos casos da grande área de Ciências da Saúde;</p> <p>5,70% dos casos da área de Ciências Humanas</p>	<p>48,31% da grande área de Ciências Sociais Aplicadas;</p> <p>29% da grande área de Engenharias</p>	<p>54,52% da grande área de Ciências Sociais Aplicadas;</p> <p>13,22% da grande área de Engenharias</p> <p>10,61% da grande área de Ciências da Saúde</p>	<p>39,77% da grande área de Ciências Sociais Aplicadas;</p> <p>24,54% da grande área de Engenharias ;</p> <p>12,67% da grande área de Ciências da Saúde;</p> <p>10% da grande área de Ciências Humanas</p>			<p>53,66% da grande área de Ciências Sociais Aplicadas;</p> <p>25,20% da grande área de Engenharias</p>	<p>50,78% da grande área de Ciências Sociais Aplicadas;</p> <p>29,50% da grande área de Engenharias</p>	<p>53,37% da grande área de Ciências Sociais Aplicadas</p>	<p>35,17% da grande área de Ciências Agrárias;</p> <p>20,05% da grande área de ciências da saúde</p>	<p>40,43% da grande área de engenharias</p>	<p>49,91% da grande área de Ciências Sociais Aplicadas</p>

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Curso</b>	<p>8,86% dos casos do curso de Administração;</p> <p>8,35% dos casos do curso de Direito;</p> <p>6,37% do curso de Engenharia Civil;</p> <p>6,19% do curso de Engenharia da Produção;</p> <p>5,69% do curso de Arquitetura e Urbanismo</p>		11,49% dos cursos de Administração				18,05% do curso de Administração	<p>16,72% do curso de Administração;</p> <p>16,33% do curso de Direito</p>	24,83% do curso de medicina veterinária	<p>19,53% do curso de engenharia civil;</p> <p>16,65% do curso de medicina veterinária</p>	

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Valor médio do curso</b>	<p>24,99% com valor de curso médio de R\$715,6 – R\$888,70;</p> <p>21,92% dos casos com o valor médio do curso de R\$409,0 à R\$715,60</p>	<p>37,44% com valor médio de curso de R\$715,60 a R\$888,70;</p> <p>32,66% com valor médio de curso de R\$888,70 a R\$1.061,80;</p> <p>17,96% com valor médio de curso de R\$409,00 a R\$715,60;</p> <p>11,66% com valor médio de curso de R\$1.061,8 a 1.658,70</p>	<p>46,96% com valor médio de curso de R\$409,00 a R\$715,60;</p> <p>17,76% dos casos com valor médio do curso de R\$888,70 a R\$1.061,80</p>	<p>27,76% com valor médio de curso de R\$888,70 a 1.064,80;</p> <p>26,77% com valor médio de curso de R\$715,60 a R\$888,70;</p> <p>24,17% com valor médio de curso de R\$1.061,80 a 1.658,70</p>	<p>33,89% com valor médio de curso entre R\$715,60 e R\$888,70</p>	<p>36,57% dos casos com valor médio do curso entre R\$1.061,80 e 1.658,70</p>		<p>45,52% com valor médio de curso de R\$715,60 a R\$888,70</p>	<p>43,63% com valor médio de curso entre R\$1.061,80 a R\$1.658,70</p>	<p>49,07% com valor médio de curso entre R\$1.061,80 e 1.658,70</p>	<p>36,66% com valor médio de curso de R\$715,60 a R\$888,70</p>
<b>Estado Civil</b>	<p>83,86% solteiro;</p> <p>12,36% casado</p>	<p>88,69% dos casos solteiro</p>	<p>84,60% apresentam estado civil solteiro</p>	<p>85,58% com estado civil solteiro</p>	<p>83,21% de estado civil solteiro</p>	<p>80,38% dos casos possuem estado civil solteiro</p>	<p>73,66% apresentam estado civil solteiro</p>	<p>75,92% com estado civil solteiro</p>	<p>85,02% com estado civil solteiro</p>	<p>86,54% com estado civil solteiro</p>	<p>77,95% com estado civil solteiro</p>

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Faixa de Idade</b>	33,52% entre a faixa de idade de 21 – 25 anos; 32,32% entre a faixa de idade de 26 – 30 anos; 7,71% dos casos com faixa etária entre 36 e 40 anos	47,25% entre faixa etária de 21 a 25 anos; 27,94% com idade de 26 a 30 anos	49,75% com faixa etária entre 26 e 30 anos; 14,78% com faixa etária de 21 a 25 anos	51,16% com idade de 21 a 25 anos	45,31% com faixa de idade de 21 a 25 anos	32,58% com faixa de idade entre 21 e 25 anos	34,62% com faixa de idade de 31 a 35 anos	37,29% com faixa de idade entre 26 e 30 anos	51,70% com faixa de idade de 21 a 25 anos	44,44% com faixa de idade de 21 a 25 anos; 33,81% com faixa de idade de 26 a 30 anos	45,75% de faixa de idade de 26 a 30 anos
<b>Tipo de Instituição de 2º Grau</b>	69,82% terminou o 2º grau em escola pública	75,48% dos casos terminaram o 2º grau em escola pública	75,08% concluíram o 2º grau em escola pública	66,86% terminou o 2º grau em escola pública	77,06% concluíram o 2º grau em escola pública	60,33% concluíram o 2º grau em escola pública	56,66% concluíram o 2º grau em escola privada	62,04% concluíram o 2º grau em escola pública	73,70% concluíram o 2º grau em escola pública	74,20% concluíram o 2º grau em escola pública	65,65% terminou o 2º grau em escola pública

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Número Total de Matrículas</b>	24,98% com número de matrículas total de 14 a 26 disciplinas; 24,98% com número de matrículas total de 27 a 38 disciplinas	51,17% dos casos tiveram entre 1 a 13 matrículas de disciplinas no total; 47,55% com número total de matrículas entre 14 e 26	65,55% com número de matrículas total de 39 a 81	99,70% dos casos com número de matrículas abaixo de 13	40,58% com número total de matrículas entre 14 e 26						
<b>Quantidade de disciplinas aprovadas</b>	24,98% tinham 5 a 15 disciplinas aprovadas	60,74% com 5 a 15 disciplinas aprovadas; 27,66% com 0 a 4 disciplinas aprovadas	78,21% com 27 a 65 disciplinas aprovadas	50% com aprovações abaixo de 4 disciplinas	37,96% entre 16 a 26 disciplinas aprovadas				51,31% com 5 a 15 disciplinas aprovadas		36,84% com 5 a 15 disciplinas aprovadas

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Quantidade de disciplinas canceladas</b>	24,98% tinham entre 5 e 6 disciplinas canceladas	38,24% com 5 a 6 disciplinas canceladas;  28,38% entre 3 a 4 disciplinas canceladas;  24,28% com 7 a 15 disciplinas canceladas	25,85% com 3 a 4 disciplinas canceladas		24,47% entre 7 e 15 disciplinas canceladas;  24,09% com 3 a 4 disciplinas canceladas;  23,37 com 5 a 6 disciplinas canceladas		31,25% com 3 a 4 cancelamentos de disciplinas				
<b>Quantidade de disciplinas dispensadas</b>	24,98% tinham entre 3 a 6 dispensas;  14,32% com 0 a 2 disciplinas dispensadas	48,55% apresentaram de 0 a 2 dispensas de disciplinas	48,55% com 0 a 2 dispensas de disciplina			45,05% apresentaram de 5 a 15 dispensas de disciplina;  27,99% com 5 a 6 dispensas de disciplina	44,06% apresentaram de 7 a 21 dispensas em disciplinas	47,66% com 7 a 21 disciplinas dispensadas	45,49% com 7 a 15 disciplinas dispensadas		

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
<b>Quantidade de períodos letivos cursados</b>	24,98% cursaram entre 6 e 7 períodos até o fato da evasão; 24,98% cursaram de 3 a 5 períodos letivos	49,22% cursaram de 3 a 5 períodos letivos; 48,50% cursaram de 1 a 2 períodos letivos		44,81% cursaram de 1 a 2 períodos letivos		50,24% cursaram de 3 a 5 períodos letivos			58,73% cursaram de 1 a 2 períodos letivos		
<b>Quantidade de disciplinas reprovadas por falta</b>	24,98% tinham 1 reprova por falta	33,84% com 1 reprova por faltas				25,56% com uma reprova por falta		34,78% entre 2 a 4 disciplinas reprovadas por falta	25,56% com 1 reprova por falta	37,06% com uma reprova por faltas	33,69% com 2 a 4 reprovadas por falta
<b>Quantidade de disciplinas reprovadas</b>	24,98% tinham de 4 a 7 reprovadas; 24% dos casos com 0 a 3 reprovações	47,57% dos apresentaram de 0 a 3 reprovadas; 33,76% entre 4 a 7 disciplinas reprovadas	27,57% com 4 a 7 disciplinas reprovadas; 25,77% com 8 a 20 disciplinas reprovadas		39,12% entre 0 a 3 reprovadas; 33,15% entre 4 a 7 disciplinas reprovadas	29,90% com 0 a 3 reprovadas; 29,59 com 4 a 7 reprovadas	72,38% com quantidades de reprovadas de 1 a 3	46,68% apresentaram de 8 a 20 disciplinas reprovadas		43,20% apresentaram de 8 a 20 reprovações	41,72% com 8 a 20 disciplinas reprovadas
<b>Quantidade de disciplinas trancadas</b>	24,98% tinham 1 disciplina trancada	26,19% com 1 disciplina trancada	26,19% com 1 disciplina trancada		45,76% com 2 a 6 disciplina trancadas	26,19% com 1 disciplina trancada		45,67% tiveram de 2 a 6 trancamentos em disciplinas	26,19% com 1 disciplina trancada	37,89% com 2 a 6 disciplinas trancadas	31,56% com 1 disciplina trancada

Variáveis	População Total	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Quantidade de acordos financeiros realizados					38,64% entre 0 a 1 acordo						
Sexo	51,1% do sexo masculino	53,84 do sexo masculino		53,14% do sexo feminino	51,58% do sexo masculino	51,71% do sexo masculino	52,87% do sexo masculino	59,87% do sexo masculino	54,70% do sexo feminino	51,79% do sexo feminino	57,81% do sexo masculino
Tem ou teve bolsa de estudos	61,54% não teve bolsa de estudo	56,39% <b>tiveram</b> algum tipo de bolsa de estudo	68,52% não tiveram nenhum tipo de bolsa de estudo	54,19% não tiveram nenhum tipo de bolsa de estudo	79,73% não tiveram bolsa de estudo	55,24% não tiveram nenhum tipo de bolsa de estudo	66,25% não tiveram nenhum tipo de bolsa	80,15% não tiveram nenhum tipo de bolsa de estudo	59,06% não tiveram nenhum tipo de bolsa	79,15% não tiveram nenhum tipo de bolsa de estudo	68,90% não apresentar am nenhum tipo de bolsa de estudo
Turno	62% dos casos do turno da noite	57,5% do turno da manhã; 42,26% no turno da noite	85,82% no turno da noite	52,79% cursaram no turno da manhã	63,95% dos casos no turno da noite	60,27% dos casos no turno da noite	87,71% do turno da noite	86,17% do turno da noite	66,70% do turno da manhã	66,26% dos casos no turno da noite	65,95% do turno noturno

Fazendo uma comparação entre o cluster 1 e o cluster 2 mostrado na Figura 32, os dois grupos com maior população, a ferramenta mostrou que:

- No cluster 1 estão os alunos com quantidade de aprovações em componentes curriculares menor que 16, enquanto no cluster 2 com aprovações maiores que 16 componentes curriculares,
- no cluster 1 encontram-se os alunos evadidos que estudaram de 1 a 4 períodos letivos, enquanto no cluster 2, valores acima de 5 períodos,
- ressalta que a maioria dos casos do cluster 1 encontram-se no turno da manhã, enquanto no cluster 2 no período da noite.

Ou seja, as evasões do período da manhã possuem uma tendência maior de acontecer já nos primeiros anos do curso, se comparado com as evasões do período da noite.

**Figura 32 - Comparação entre o cluster 1 e 2 da população de evadidos**

Cluster 1:  Cluster 2:

Discrimination scores for Cluster 1 and Cluster 2

Variables	Values	Favors Cluster 1	Favors Cluster 2
QTD APROVADAS	0 - 15		
QTD APROVADAS	16 - 110		
QTD PERIODOS	1 - 4		
QTD PERIODOS	5 - 26		
N MATRTOTAL	25 - 138		
N MATRTOTAL	1 - 24		
TURNO	NOITE		
TURNO	MANHA		
QTD CANCELADAS	3 - 12		
QTD CANCELADAS	0 - 2		
QTD REPROVADAS	0 - 6		
PERLET INGRESSO	2010/01		
FAIXA IDADE	21 - 25		
PERLET INGRESSO	2009/01		
QTD REPROVADAS	7 - 88		
PERLET INGRESSO	2011/01		
TEMBOLSA	NÃO		
TEMBOLSA	SIM		
CURSOVALORMEDIO	409,0 - 811,2		
CURSOVALORMEDIO	811,2 - 2.418,5		

Fonte: Próprio autor

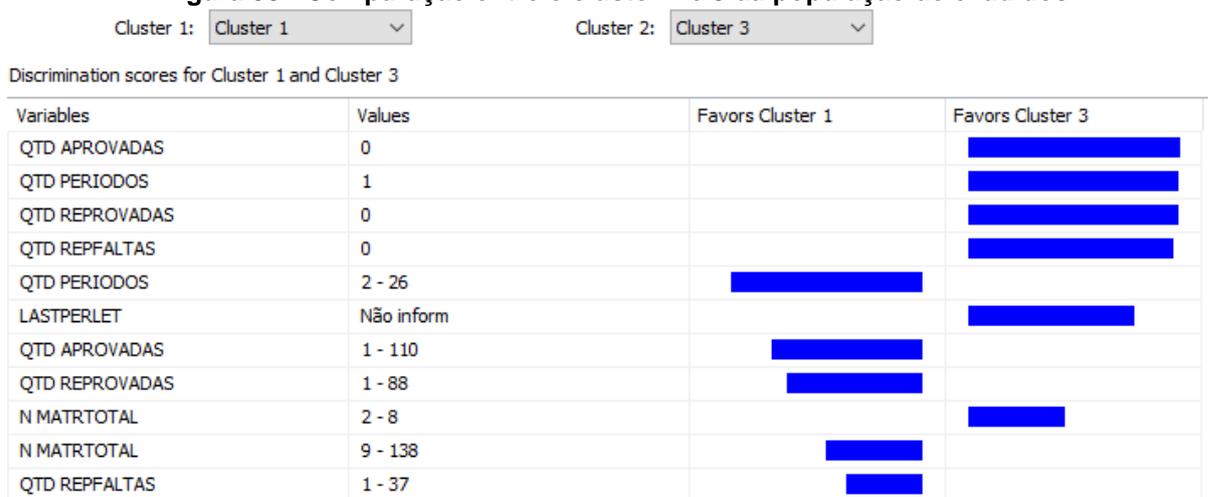
A descrição das variáveis mostradas na Figura 32 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

Comparando o *cluster 1* com o *cluster 3* presente na Figura 33, fica nítido que no *cluster 3* estão os evadidos recém ingressantes, pois apresenta uma população

sem aprovações em componentes curriculares, que estudaram apenas o 1º período, não apresentaram reprovações, com número de matrículas baixo, especificamente entre dois a oito componentes curriculares. Ou seja, evadiram antes de concluir o 1º semestre do curso.

Como no *cluster 3*, ficou claro que estão os grupos dos evadidos recém ingressantes e dentre todos os *clusters* ele representa o terceiro lugar em número de casos, constata-se que os evadidos ingressantes em sua maioria (*cluster 1*) terminam o 1º semestre antes de evadir.

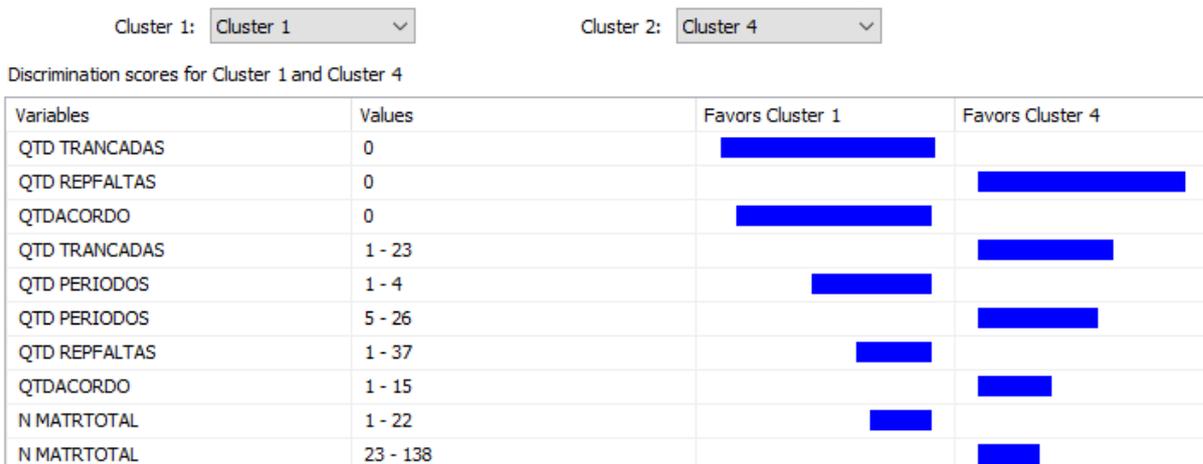
**Figura 33 - Comparação entre o cluster 1 e 3 da população de evadidos**



Fonte: Próprio autor

A descrição das variáveis mostradas na Figura 33 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

Comparando o *cluster 1* com o 4, presente na Figura 34, fica destacado que no *cluster 4*, estão os casos de alunos que tiveram um tempo maior de estudo antes de evadir, pois demonstra maior quantidade de trancamentos em componentes curriculares (entre 1 a 23), maior quantidade de períodos estudado (acima de 5 períodos), maior quantidade de matrículas em componentes curriculares em comparação com o cluster 1, além de apresentar acordos financeiros realizados.

**Figura 34 - Comparação entre o cluster 1 e 4 da população de evadidos**

Fonte: Próprio autor

A descrição das variáveis mostrada na Figura 34 pode ser visualizada no Quadro 4 que se encontra na pág. 245.

#### 4.3.3.1 *Clustering* aplicado aos dados dos evadidos por turnos

Ao aplicar o método de *clustering* foi mostrado na Figura 32 - Comparação entre o *cluster* 1 e 2 da população de evadidos - da página 130 uma forte tendência de evasão no início dos cursos no turno da manhã e alguns períodos mais à frente no turno da noite.

Ao executar o método de *clustering* no turno da manhã, a ferramenta de mineração destacou dos dez grupos criados de uma população de 2.581 fatos de evasão do período da manhã as seguintes características:

- *Cluster 1* (417 casos): Em 96% dos casos formados pelos discentes que não concluíram o primeiro semestre, em sua maioria sem nota ou mesmo reprovadas, ou seja, desistentes.
- *Cluster 2* (365 casos): Em 78% dos casos evadidos do segundo semestre, em 92% dos casos com bolsa de estudo, em média com 11 matrículas em componentes curriculares no total, onde cinco são canceladas, ou seja, não concluiu o 2º semestre.
- *Cluster 3* (324 casos): Em 77,5% dos casos evadidos do terceiro semestre, entre 16 a 18 matrículas em componentes curriculares, dentre

elas 8 a 11 aprovações, com tendência de ter bolsa de estudos e 5 cancelamentos de componentes curriculares.

- *Cluster 4* (263 casos): Predominantemente para o curso de medicina veterinária com 43% dos casos, onde em 50% são da faixa de idade de 21 a 25 anos, sem bolsa de estudo, com 72% da população do sexo feminino.
- *Cluster 5* (263 casos): Evadidos que estudaram entre 4 e 5 períodos, em média com 25 a 27 matrículas em componentes curriculares no total, em 40% dos casos da área de Ciências Sociais Aplicadas, destacando o curso de Direito com 20% das situações.

Ao executar o método de *clustering* no turno da noite da população analisada de 4.384 casos, a ferramenta de mineração destacou os seguintes resultados:

- *Cluster 1* (755 casos): Evadidos que estudaram de 2 a 3 períodos, 38% dos casos da área de Ciências Sociais Aplicadas e 18% de Ciências Humanas.
- *Cluster 2* (527 casos): Desistentes iniciais em 80% dos casos.
- *Cluster 3* (461 casos): Veteranos acima de 5 períodos letivos, onde em 66% dos casos são da faixa de idade de 21 a 25 anos.
- *Cluster 4* (423 casos): Veteranos que não tiveram bolsa em sua vida acadêmica onde em 66% dos casos com faixa de idade dentre 26 a 30 anos, 60% do sexo feminino, 60% da área de Ciências Sociais Aplicadas e 29% das situações do curso de Administração.
- *Cluster 5* (453 casos): Veteranos com tendência aos cursos de Engenharia Civil (21,26% da população do *cluster*) e Arquitetura e Urbanismo (19% dos casos).
- *Cluster 6* (464 casos): Veteranos, de faixa de idade de 26 a 30 anos (42%) sem bolsa de estudo (96%), do sexo masculino, com tendência aos cursos de Engenharia (44% dos casos).
- *Cluster 7* (391 casos): Veteranos, com quantidade média de 5 períodos letivos cursados (32%), da grande área de Ciências Sociais Aplicadas (70%), com tendência ao curso de logística (17%) e Gestão de Recursos Humanos (13%).

#### 4.3.4 Árvore de decisão aplicada aos dados dos evadidos

Considerando a relevância dos perfis encontrados nos dados gerais e dos perfis mais relevantes nos dados especificamente sobre casos de evasão, com o auxílio de árvores de decisão foram estudadas as variáveis que demonstraram maior concentração nos casos de alunos evadidos.

A variável estado civil, configurada na ferramenta *Analysis Service* (Figura 8) como campo a ser analisado na mineração dos casos de evasão destacou as seguintes regras (Figura 35):

- Nos casos onde o aluno evadido tem a idade entre 21 e 25 anos em 96,35% são solteiros, apresentando apenas 1,91% casados.
- Nos casos onde o aluno evadido tem a idade entre 26 e 30 anos manteve-se as mesmas características, onde 92,08% são solteiros, enquanto 5,77% são casados.
- Em casos onde o aluno evadido tem a idade entre 31 e 35 anos, destaca-se o estado civil solteiro, apresentando uma probabilidade de 73,70% de solteiros e 21,28% de casados. Outros casos de estado civil estão abaixo dos 3%.

Ou seja, a evasão apresentou relevância para os casos de alunos solteiros independentes da faixa de idade.

A maioria dos casos são de alunos com faixa etária entre 21 e 25 anos (2.357 casos), por faixa etária entre 26 e 30 anos (2.272 casos), seguido por faixa etária entre 31 e 35 anos (1.090 casos), como mostrado na Figura 35, onde a ferramenta destaca os nós com mais casos aumentando o realce do azul.

**Figura 35 - Árvore de decisão da análise do estado civil dos casos evadidos**



Fonte: Próprio autor

Não foram apresentados os valores de faixa de idade abaixo de 20 anos, por considerar muito remoto os casos de casamento e faixas acima dos 35 anos nem foram mostrados na árvore de decisão.

Utilizando a faixa de idade, variável configurada segundo a Figura 8, pág. 60, para tentar entender os casos dos evadidos que não são solteiros o algoritmo de árvore destacou que esse grupo se encontra na faixa de idade entre 31 e 40 anos, porém com menor quantidade de casos do que a faixa de idade alvo dos cursos de graduação que é de 21 a 30 anos, mostrando um total de 1.135 casos com a distribuição mostrada na Tabela 16.

**Tabela 16 - Distribuição dos casos de evadidos não solteiros**

16 - 20	5	0,52%	
21 - 25	85	7,48%	
26 - 30	179	15,66%	
31 - 35	286	24,97%	
36 - 40	258	22,54%	
41 - 45	143	12,53%	
46 - 50	82	7,22%	
51 - 55	53	4,70%	
56 - 60	24	2,18%	
6 - 10	0	0,09%	
61 - 65	12	1,13%	
66 - 70	5	0,52%	
76 - 80	2	0,26%	

Fonte: Próprio autor

Ainda trabalhando com a faixa de idade e estado civil dos alunos evadidos, a árvore gerada destacou que a maior quantidade de casos se encontra entre os alunos recém ingressantes, onde chegou nas seguintes regras:

- Alunos com estado civil solteiro e quantidade de períodos letivos cursados abaixo de 4 tiveram a maior quantidade de casos com 2.669 casos, sendo sua distribuição mostrada na Tabela 17.

**Tabela 17 - Distribuição dos casos dos alunos evadidos solteiros abaixo de 4 períodos letivos**

16 - 20	263	9,85%	
21 - 25	1355	50,77%	
26 - 30	657	24,62%	
31 - 35	244	9,14%	
36 - 40	97	3,63%	
41 - 45	31	1,16%	

Fonte: Próprio autor

- Alunos com estado civil solteiro e quantidade de períodos letivos maior ou igual a 4 e menor de 10, apresentaram um total de 2.315 casos e sua distribuição com relação a faixa de idade é mostrada na Tabela 18.

**Tabela 18 - Distribuição de casos de alunos solteiros entre 4 e 9 períodos letivos**

16 - 20	18	0,78%	
21 - 25	858	37,06%	
26 - 30	969	41,85%	
31 - 35	312	13,48%	
36 - 40	105	4,54%	
41 - 45	33	1,43%	
46 - 50	13	0,56%	

Fonte: Próprio autor

- Alunos evadidos com número superior a 9 períodos letivos não estão apresentando tantos casos como nos dois tópicos acima (apenas 911), mas destacou um fator que deve ser considerado: quanto maior a quantidade de períodos cursados, maior a faixa etária dos alunos evadidos, aumentando a probabilidade acima de 50% na faixa de idade entre 26 e 30 anos e um valor 4 vezes maior na faixa de 31 a 35 anos se comparado com a faixa de 21 a 25 anos mostrado no primeiro tópico, essa distribuição pode ser visualizada na Tabela 19.

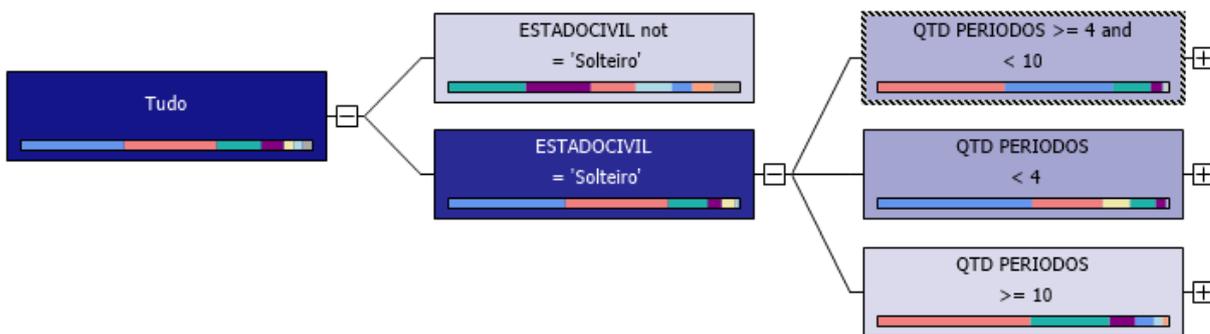
**Tabela 19 - Distribuição dos casos de alunos solteiros acima de 9 períodos letivos**

16 - 20	0	0,00%	
21 - 25	59	6,48%	
26 - 30	467	51,23%	
31 - 35	248	27,21%	
36 - 40	82	9,00%	
41 - 45	32	3,51%	
46 - 50	17	1,87%	
51 - 55	3	0,33%	

Fonte: Próprio autor

A mineração dos dados de alunos evadidos segundo a faixa de idade foi originada segundo a árvore mostrada na Figura 36.

Figura 36 - Árvore de decisão para análise de evadidos segundo a faixa de idade



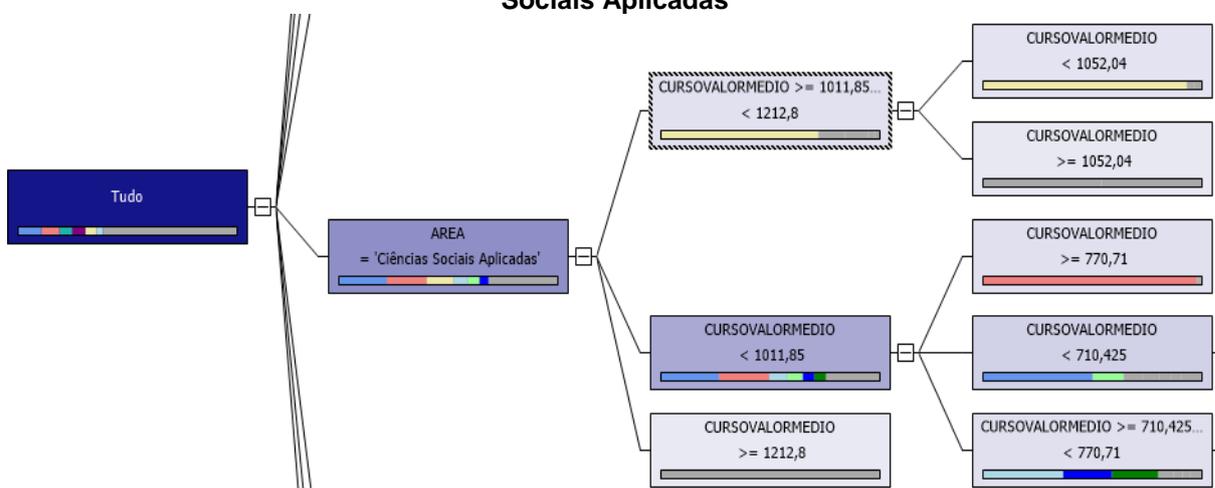
Fonte: Próprio autor

Outra variável que apareceu com frequência no *clustering* sobre os casos de evasão é o da grande área – citada na ferramenta *Analysis Service* com o nome área (Figura 8, pág. 60), trouxe grande relevância para Ciências Sociais Aplicadas e Engenharias. Fato mostrado no tópico 4.1.2 (Índices de evasão das grandes áreas de graduação da UNISO) e 4.2.1 (Discriminação do perfil dos alunos evadidos), pág.79.

Analisando-se essas grandes áreas, o algoritmo de árvore de decisão reforçou essa constatação ao mostrar que a área de Ciências Sociais Aplicadas possui mais casos do que todas do que as áreas juntas: Linguística, Letras e Artes, Ciências Agrárias, Ciências Humanas, Ciências da Saúde, Ciências Exatas e da Terra e Ciências Biológicas. Porém, não significa que esse alto índice de evasão está relacionado a todos os cursos dessa grande área, mas apenas alguns cursos específicos como: Administração, Direito e Arquitetura e Urbanismo.

A árvore de decisão dos casos de evasão configurada para destacar o curso de Administração, mostrou a árvore da Figura 37.

Figura 37 - Árvore de decisão para os casos evadidos predizendo cursos da área de Ciências Sociais Aplicadas



Fonte: Próprio autor

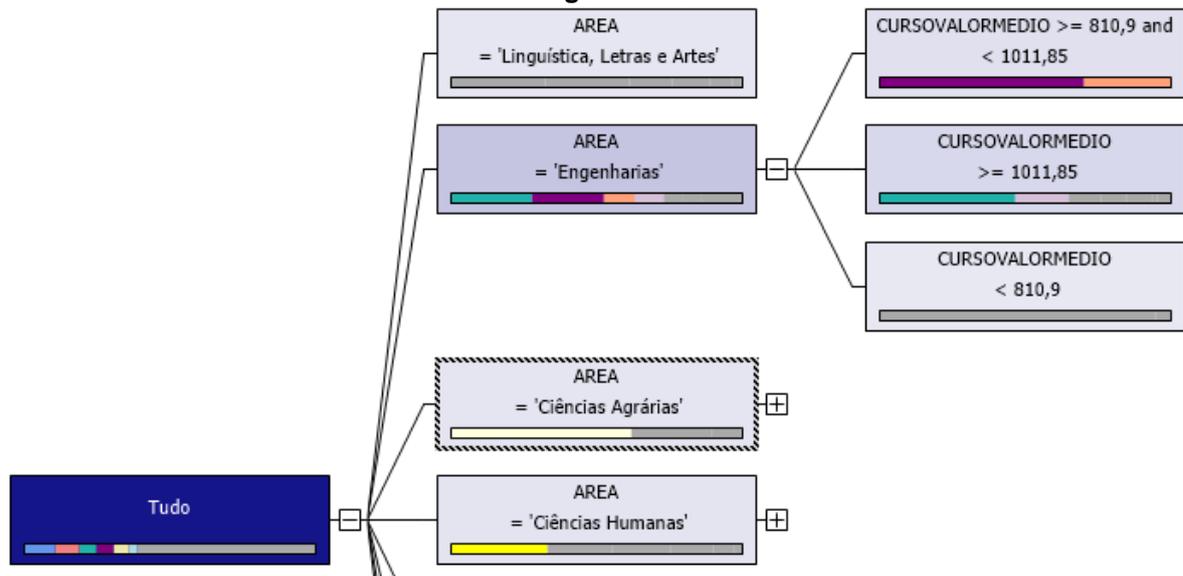
Ao analisar a árvore gerada no nó mais destacado da imagem, chegou-se nas seguintes regras explicativas:

- Quando a grande área for Ciências Sociais Aplicadas e o valor médio do curso for menor que R\$710,43, obteve-se um total de 1293 casos, onde em 48,18% dos casos a evasão é do curso de Administração, 14,31% é do curso de Ciências Contábeis, 7,81% dos casos para o curso de Ciências Econômicas, 9,13% para o curso de Logística, 5,57% para os cursos de Gestão de Recursos Humanos, 5,10% para o curso de Gestão da Qualidade e 3,09% para o curso de Hotelaria e Marketing.
- Quando a grande área for Ciências Sociais Aplicadas e o valor médio do curso estiver entre R\$1.011,85 e R\$1.212,80, foi obtido um total de 569 casos, onde o curso de Arquitetura e Urbanismo aparece com 70,29% dos evadidos, o curso superior de Tecnologia em Design Gráfico tem um total de 11,07% dos casos, design com 12,83% dos casos, Design de Interiores com 2,99% dos casos e Design de Moda com 2,81% dos casos
- Quando a grande área for Ciências Sociais Aplicadas e o valor médio do curso for maior ou igual a R\$1.212,80, com 66 casos computados, aparece o curso superior de Tecnologia em Gastronomia com 99,21% dos casos.

Fazendo uma comparação com o índice de evasão do curso de Administração apresentado no tópico 4.2.2.2, presente na página 88 e considerando os números de matrículas, ingressantes e concluintes apresentados na Tabela 29, pode-se observar que, mesmo alguns pontos acima do gráfico geral (Gráfico 8 e Gráfico 9), ainda assim está dentro da tendência geral, tendo como principal motivo de tantos casos o fato de ser um curso com muita procura.

Estudando a árvore gerada no caso da grande área de Engenharias, chegou-se no resultado mostrado na Figura 38.

**Figura 38 - Árvore de decisão para os casos evadidos predizendo cursos da área de Engenharias**



Fonte: Próprio autor

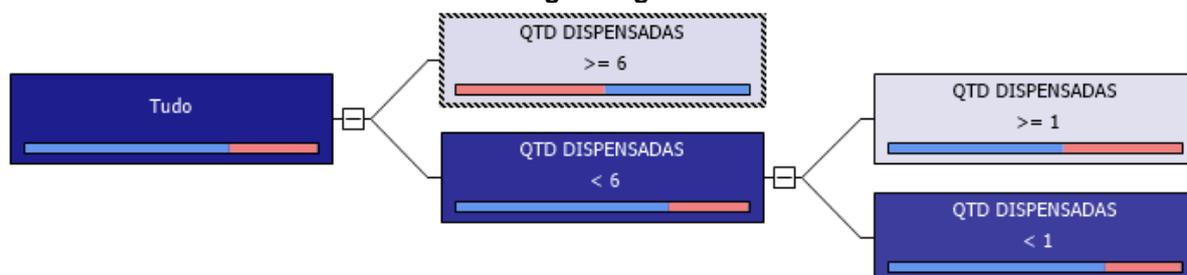
De onde obteve-se as seguintes regras nos 1.746 casos:

- Quando a grande área é de Engenharias e o valor médio do curso for menor que R\$810,90, 94,49% dos casos são do curso de Gestão da Produção Industrial e 5,38% dos casos para o curso de Segurança no Trabalho, tendo um total de 130 casos computados.
- Quando a grande área é de Engenharias e o valor médio do curso for maior ou igual a R\$810,90 e menor que R\$1.011,85, 69,93% dos casos são de Engenharia de Produção e 30,06% para o curso de Engenharia Ambiental, com um total de 622 casos computados.
- Quando a grande área é de Engenharias e o valor médio do curso é maior ou igual a R\$1.011,85, 45,04% dos casos estão em Engenharia Civil, 18,50% em Engenharia Química, 11,56% em Engenharia da Computação, 9,25% em Engenharia de Controle e Automação, 8,85% em Engenharia Elétrica, 2,01% em Engenharia de Bioprocessos e Biotecnologia, 1,81% em Engenharia de Alimentos, 1,71 em Engenharia de Materiais e 1,21% em Engenharia Mecânica, com um total de 994 casos computados.

A mineração dos casos evadidos quando analisando a variável que indica qual instituição de segundo grau o aluno concluiu gerou a árvore de decisão mostrada na Figura 39.

A variável correspondente aos dados relacionados a instituição de segundo grau que o discente concluiu foi configurada com o nome INST2G como mostrado na Figura 8, pág. 60.

**Figura 39 - Árvore de decisão para os casos de evasão predizendo a variável de instituição de segundo grau**



Fonte: Próprio autor

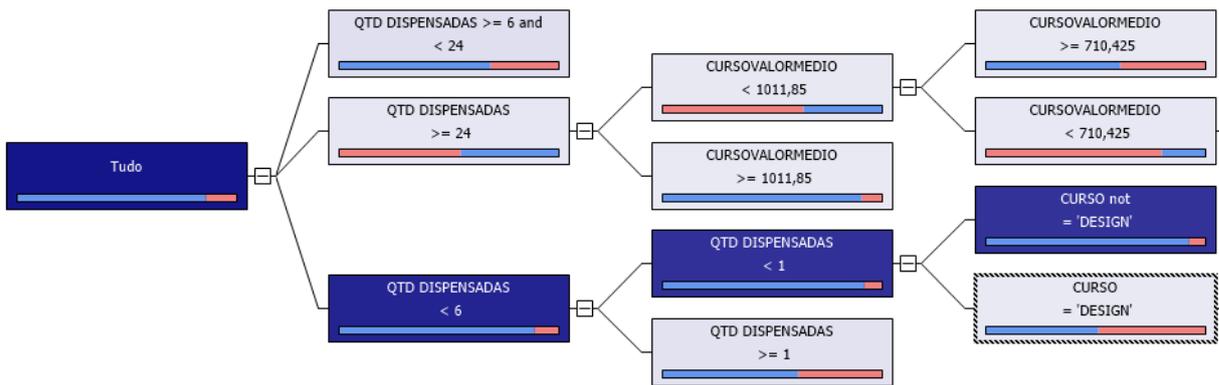
A árvore gerada indica as seguintes regras:

- Quando a quantidade de componentes curriculares dispensados é igual a zero em 72,98% dos casos o aluno concluiu o segundo grau em escola pública, 26,17% em escola privada e 0,85% dos casos não informado, sendo computado nessa análise um total de 5.648 casos.
- Quando a quantidade de componentes curriculares dispensados é maior ou igual a 1 e menor que 6 em 58,82% dos casos o aluno concluiu o segundo grau em escola pública, 40,60% em escola privada e 0,58% dos casos não foram informados, nessa análise foi computado um total de 532 casos.
- Quando a quantidade de componentes curriculares dispensados é maior ou igual a 6 em 49,98% dos casos o aluno concluiu o segundo grau em escola privada, 49,28% em escola pública e 0,75% não informaram, tendo computado um total de 850 casos.

Essa análise mostra a seguinte tendência: quanto maior a quantidade de dispensas de componentes curriculares, maior a probabilidade do aluno evadido ter concluído o segundo grau em escola privada.

Ao configurar a ferramenta de análise para prever a variável que indica outro curso (Figura 8), responsável por indicar se o aluno em semestres anteriores estava cursando um curso diferente do atual, foi gerado a seguinte árvore de decisão mostrada na Figura 40.

**Figura 40 - Árvore de decisão dos casos evadidos configurada para previsão de outro curso**



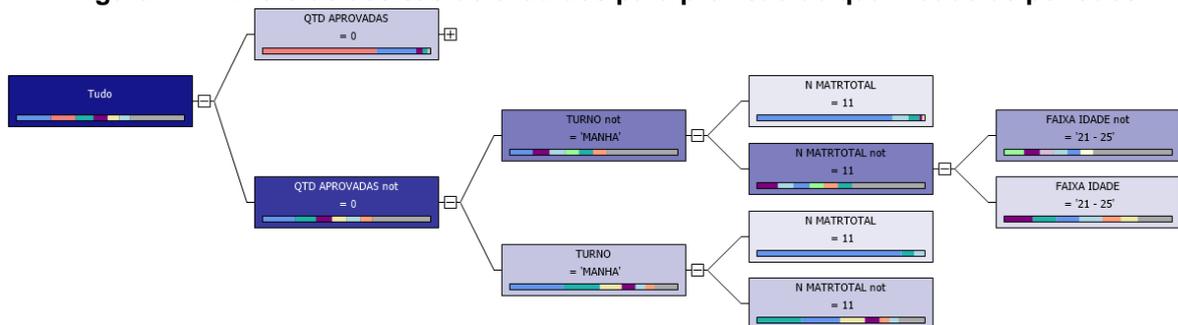
Fonte: Próprio autor

A árvore gerada destacou as seguintes regras:

- Quando a quantidade de dispensas for igual a zero e o curso for Design em 49,32% dos casos o aluno já havia cursado outro curso, nessa situação foi computado um total de 73 casos.
- Quando a quantidade de dispensas for maior ou igual a 24 e o valor médio do curso for menor que R\$710,43 em 79,73% o aluno já havia cursado outro curso, nessa situação foi computado um total de 94 casos.

Ao gerar uma árvore com todos os dados discretos, ou seja, por categorização, configurando para identificar a quantidade de evadidos por quantidade de períodos letivos gerou-se a seguinte árvore.

**Figura 41 - Árvore de decisão de evadidos para previsão de quantidade de períodos**



Fonte: Próprio autor

A árvore gerada deixou explícito que os evadidos do turno da manhã se concentram no 2º período letivo (com 84,56% dos casos) cursado enquanto no turno da noite estão distribuídos, uma parte no 2º período, outra parte períodos acima do 5º período letivo cursado, conforme mostrado na imagem abaixo.

**Figura 42 - Probabilidade de evasão por quantidade de períodos letivos no turno da noite**

Value	Cases	Probabi...	Histogram
<input checked="" type="checkbox"/> 1	4	0,13%	
<input checked="" type="checkbox"/> 10	218	6,07%	
<input checked="" type="checkbox"/> 2	341	9,49%	
<input checked="" type="checkbox"/> 3	307	8,55%	
<input checked="" type="checkbox"/> 4	260	7,24%	
<input checked="" type="checkbox"/> 5	385	10,72%	
<input checked="" type="checkbox"/> 6	346	9,63%	
<input checked="" type="checkbox"/> 7	314	8,74%	
<input checked="" type="checkbox"/> 8	295	8,21%	
<input checked="" type="checkbox"/> 9	325	9,05%	

Fonte: Próprio autor

Essa conclusão justifica os resultados mostrados na Figura 28 - Árvore de decisão da população do turno da manhã (previsão de evasão), pág. 116, Figura 29 - Árvore de decisão do turno da noite (previsão de evasão), pág. 117 com a Figura 32 - Comparação entre o cluster 1 e 2 da população de evadidos.

Outras regras não foram consideradas relevantes por apresentar que o aluno não veio de outro curso, que é o padrão previsível para essa variável.

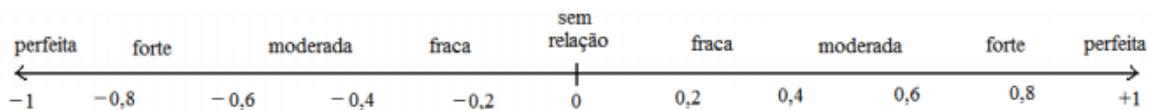
A mineração de dados com aplicação de árvore de decisão aplicada aos casos evadidos também demonstrou influência na variável quantidade de aprovações mostradas no apêndice J.

#### 4.4 MATRIZ DE CORRELAÇÃO DE PEARSON E DENDOGRAMA

A matriz de correlação<sup>8</sup> ficou muito extensa e por esse motivo foi gerado um link de acesso na nota de rodapé para visualizá-la na íntegra.

A matriz de correlação seguiu a seguinte referência para o entendimento dos resultados obtidos:

**Figura 43 - Faixa de força da correlação**



Fonte: FILHO, A. C. R. D. S. **Laboratório de Sistemas Neurais (SisNe)**. Disponível em <<http://sisne.org/ Disciplinas/Grad/ProbEstat2/aula13.pdf>>. Acesso em: 10 Janeiro 2020.

Dentre as correlações encontradas, muitas delas não colocadas pelo fato de serem óbvias ou irrelevantes para a pesquisa, destaca-se:

- Correlação forte: A variável quantidade de disciplinas reprovadas aumenta junto com a variável quantidade de faltas.
- Correlação moderada:
  - o ESTADO CIVIL solteiro diminui com o aumento da IDADE e com o aumento da IDADE aumenta a QUANTIDADE DE ALUNOS casados;
  - com o aumento da quantidade de CANCELAMENTOS de disciplinas aumenta a quantidade de EVADIDOS.
- Correlação fraca:
  - a CARGA HORÁRIA cursada tem correlação positiva com o TURNO da noite e negativa com o TURNO da manhã;
  - a MÉDIA DE NOTAS diminui com o aumento da quantidade de FALTAS e a quantidade de FALTAS diminui para o TURNO da manhã (correlação negativa) e aumenta para o TURNO da noite (correlação positiva);

<sup>8</sup> FEITOSA, E. Matriz de Correlação, 2020. Disponível em: <[https://drive.google.com/open?id=1\\_J5Zt5UiV5fJdA\\_HDShEPmeo6-0D9R8l](https://drive.google.com/open?id=1_J5Zt5UiV5fJdA_HDShEPmeo6-0D9R8l)>. Acesso em: 04 fevereiro 2020.

- com a diminuição da MÉDIA DE NOTAS há um aumento da quantidade de EVADIDOS, CANCELAMENTOS e TRANCAMENTOS de disciplinas;
- o aumento da variável TURNO da noite aumenta a quantidade de disciplinas APROVADAS, PERÍODOS cursados, diminuição na quantidade de alunos que possuem ou possuíram BOLSA de estudo e quando o TURNO é da manhã a correlação é negativa, ou seja, diminuição da quantidade de disciplinas APROVADAS, PERÍODOS cursados e aumento na quantidade de alunos que já tiveram BOLSA ou ainda possuem;
- Com o aumento da variável IDADE há o aumento da variável quantidade de disciplinas APROVADAS e DISPENSADAS;
- A variável ÁREA de Engenharias tem correlação positiva com a variável SEXO masculino e negativa para SEXO feminino.

Esse método aplicado reforça alguns resultados já identificados como o fato da variável evasão estar correlacionada com as variáveis trancamentos e cancelamentos de disciplinas mostrados no tópico 4.3.1, p. 111 e nos tópicos 4.3.2.1, na Figura 28, p. 116, responsável por mostrar o aumento da probabilidade de evasão cruzando com a variável cancelamento; mostra a influência da variável do turno na quantidade de carga horária cursada pelo discente evadido, onde reforça que o turno da noite possui uma tendência a ter alunos mais velhos e com uma pré-disposição a demorar mais para evadir que os alunos do período da manhã, fato já mostrado na Figura 32, pág. 130; reforça a concentração de alunos do sexo masculino na escolha de cursos de Engenharia, fato discutido e complementado nos tópicos 4.1.1 e 4.1.2, p. 74. e mostra que a diminuição da média de notas é um ponteiro afetado pelo aumento da quantidade de faltas, cancelamentos, trancamentos e posteriormente evasão.

Os resultados obtidos foram confirmados com o auxílio do dendrograma gerado, que pode ser visualizado no APÊNDICE L – Dendrograma, pág. 246, ou pelo link na nota de rodapé.

---

9 FEITOSA, E. Dendrograma, 2020. Disponível em:  
<<https://drive.google.com/open?id=1JPqiV6V12wg4ZQTfYzBjkQldDQLLBwDb>>.  
Acesso em: 04 fevereiro 2020.

## 4.5 CONCLUSÃO DO CAPÍTULO

No capítulo 4 foi realizado uma análise dos casos de evasão na Uniso entre o 1º semestre de 2012 ao 1º semestre de 2018, com o auxílio dos métodos: consultas *ad hoc* com o objetivo de traçar o perfil discente dos alunos na faixa de tempo estudada, desenvolvimento de índices de evasão no contexto geral, por grande área e curso, aplicação de métodos de *Clustering* no âmbito geral e especificamente nos casos relacionados a evasão, aplicação de métodos de Árvores de Decisão, a fim de descobrir padrões ocultos nos casos dentro do montante geral dos discentes e especificamente nos casos evadidos, matriz de correlação e dendrograma para confirmação e validação dos resultados.

## 5 CONSIDERAÇÕES FINAIS

O presente estudo buscou lançar luz sobre o fenômeno da evasão dos cursos de graduação da Universidade de Sorocaba, procurando ampliar o entendimento conceitual sobre os fatos já registrados.

Não teve como pretensão utilizar questionários, ou mudar os processos já existentes relacionados a evasão do aluno, e sim, trabalhar com os dados já existentes definindo uma faixa de tempo entre o 1º semestre de 2012 ao 1º semestre de 2018.

Procurando com isso demonstrar que a Uniso possui muitas informações valiosas sobre o entendimento da evasão que não são usadas por estarem ocultas no grande volume de dados existente.

Para tal, o problema foi abordado em duas fases: a) realizando um estudo bibliográfico para entender a evasão que é trabalhada e discutida no âmbito nacional, a fim de identificar a melhor maneira de quantificá-la, b) ampliar o entendimento, procurando extrair o conhecimento oculto do histórico de casos definido com o auxílio de métodos de mineração de dados.

A partir do estudo inicial sobre os fatores que caracterizam a evasão e as técnicas utilizadas para identificar e administrar sua evolução, foi possível definir na Uniso relatórios que demonstrem o comportamento do fenômeno no tempo, categorizando o fato em: índices de evasão geral, cujo objetivo é de identificar o quanto de evasão existe na faixa de tempo a ser definida para os cursos como um todo, índices de evasão por grande área, delimitando os valores encontrados para uma grande área definida, a fim de identificar se o comportamento geral reflete ou não na grande área definida e índice de evasão por curso, para um entendimento e acompanhamento mais individualizado da evasão no curso específico.

Com o auxílio de métodos de mineração de dados, desenvolvidos com base no processo de extração de conhecimento conhecido como KDD, extraiu conhecimento oculto sobre o comportamento dos discentes evadidos dos cursos de graduação entre 2012 e 2018 e as conclusões mais relevantes, evidenciadas pelos dados, são as seguintes:

O perfil do público alvo da Uniso no intervalo de tempo estudado é formado de discentes em sua maioria do sexo feminino, cuja faixa etária preponderante encontra-se entre 21 e 30 anos, com estado civil solteiro, tendo a maioria terminada o segundo grau em escola pública.

O público de discentes masculinos, do ponto de vista da grande área só supera o público feminino nas grandes áreas de Ciências Exatas e da Terra e Engenharias, sendo que dentre todas as áreas a maior concentração de discentes, independente do sexo, está na área de Ciências Sociais Aplicadas.

Mesmo a quantidade de discentes do sexo masculino sendo menor, o volume de evasões é maior do que do sexo feminino, destacando alta concentração de evasão para a grande área de Engenharias pelo sexo masculino e alta para ambos os sexos para a grande área de Ciências Sociais Aplicadas.

Os índices de evasão dos cursos demonstraram-se peculiares, ou seja, de difícil comparação, porém no geral estabilizando a partir do ano de 2015 na faixa de 10% de evasão ao ano.

Dentre os comportamentos apresentados pelos evadidos estão o aumento de TRANCAMENTOS, de CANCELAMENTOS, DISPENSAS e baixo número de APROVAÇÕES.

Na separação de turnos para os dois maiores grupos de evadidos identificados, os discentes do TURNO da manhã apresentaram relevância na quantidade de APROVAÇÕES em componentes curriculares menor que 16, cursando em média de 1 a 4 PERÍODOS LETIVOS, enquanto no turno da noite apresenta quantidade de APROVAÇÕES maiores de 16 componentes curriculares e cursaram número maior ou igual a 5 PERÍODOS LETIVOS.

O trabalho dos métodos de KDD especificamente na população de evadidos para o TURNO da manhã, deixou claro que a evasão acontece no início do curso, porém a população que conclui o 1º período é maior que a população que não conclui, com 85% dos casos.

Ainda no TURNO da manhã as chances sobem para 89% quando a quantidade de CANCELAMENTOS de disciplinas é maior que 4 no 2º SEMESTRE.

Os métodos de KDD aplicados na população do turno na noite, também demonstraram alta probabilidade de a evasão acontecer no início do curso, com relevância no 2º período, mas apresentou uma tendência de a evasão acontecer no 5º período.

Os discentes evadidos do turno da noite do 5º período para frente, demonstraram mais trancamentos de disciplinas.

O estudo apresentou relevância no ato da evasão para os casos de discentes solteiros independente da faixa de idade.

Identificou uma tendência de quanto maior a quantidade de PERÍODOS cursados, maior a FAIXA ETÁRIA dos discentes evadidos, aumentando a probabilidade acima de 50% na FAIXA DE IDADE entre 26 e 30 anos em casos entre 4 e 9 PERÍODOS LETIVOS e um valor 4 vezes maior na faixa de 31 a 35 quando a quantidade de PERÍODOS LETIVOS excede a 9 períodos.

Os métodos mostraram também que quanto maior a quantidade de DISPENSAS do aluno evadido, maior é a probabilidade do discente ter concluído o segundo grau em escola privada.

Identificaram que quando a quantidade de DISPENSAS for maior ou igual a 24 e o VALOR MÉDIO do curso for menor que R\$710,43 em 80% dos casos o discente evadido já havia cursado OUTRO CURSO.

Dentre as diversas variáveis estudadas especificamente na população de discentes evadidos, as que mostraram correlação linear foram:

- Quantidade de REPROVAS aumenta com a quantidade de FALTAS.
- A quantidade de EVASÕES aumenta com a quantidade de CANCELAMENTOS.
- A CARGA HORÁRIA cursada tem correlação positiva (aumenta) com o TURNO da noite e negativa (diminui) com o TURNO da manhã, ou seja, tendência de a evasão noturna não acontecer no começo se comparado ao TURNO da manhã.
- A MÉDIA DE NOTAS diminui com o aumento da quantidade de FALTAS. E a quantidade de FALTAS diminui para o TURNO da manhã (correlação negativa) e aumenta para o TURNO da noite (correlação positiva). Ou seja, a MÉDIA DE NOTAS é sensível com a totalização das FALTAS e os discentes evadidos do TURNO da noite acumulam mais faltas (no decorrer do curso) antes de evadir em comparação com os discentes evadidos do TURNO da manhã.
- Com o decréscimo da MÉDIA DE NOTAS há um aumento da quantidade de discentes EVADIDOS, CANCELAMENTOS e TRANCAMENTOS de disciplinas.
- Com o aumento da variável quantidade de alunos EVADIDOS do TURNO da noite, aumenta a quantidade de disciplinas APROVADAS, PERÍODOS cursados e diminui na quantidade de alunos que possuem

ou possuíram BOLSA DE ESTUDO. E quando a mesma variável é analisada no TURNO da manhã a correlação é negativa, ou seja, diminuição da quantidade de disciplinas APROVADAS, PERÍODOS cursados e aumento na quantidade de alunos que já tiveram BOLSA ou ainda possuem. Ou seja, ressalta novamente o fato da evasão noturno acontecer depois da evasão diurna e acrescenta o detalhe de alunos diurnos terem mais bolsas de estudo do que os alunos noturnos.

- A variável ÁREA de Engenharias tem correlação positiva com a variável SEXO masculino e negativa para SEXO feminino.

Dentre os perfis de evasão identificados na pesquisa um dos mais relevantes é do discente do sexo masculino entre 21 e 25 anos, da grande área de Engenharias que evade no 2º ou 5º semestre, com padrão de comportamentos apontados nas variáveis relacionadas a cancelamentos, trancamentos e média de notas.

Com isso, a pesquisa demonstrou-se útil em enriquecer o conhecimento sobre o fenômeno da evasão na universidade de Sorocaba para os cursos de graduação.

Devido à grande variabilidade dos dados, não foi possível extrair padrões mais gerais que caracterizam a evasão como um todo na Uniso. Existem muitas peculiaridades que variam por curso, por área e por período.

Assim, recomenda-se analisar os casos de evasão (a fim de recomendar ações para diminuí-la) separadamente para cada segmento apontado (curso, área, período, semestre).

As árvores de decisão, que poderiam mostrar regras lógicas que caracterizam os parâmetros influenciadores da evasão, são muito grandes e dispersas, evidenciando a variabilidade das combinações destes parâmetros. Não foi nem mesmo possível apresentar a árvore completa impressa, devido a seu tamanho e complexidade.

A pesquisa contribuiu para levantar muitas outras hipóteses a serem estudadas como: a bolsa de estudos pode servir de estímulo a evasão? Ou visto de outra forma, é possível ser mais assertivo na concessão de bolsas de estudo? Trancamentos e cancelamentos de matrículas, pode-se hipotetizar que estes fenômenos indicam dificuldade de pagamento de mensalidade e culminam com a evasão do discente?

Para trabalhos futuros considerando a complexidade na aplicação do conjunto de regras criadas sobre a evasão nesta pesquisa, outras pesquisas poderiam ser

realizadas para ampliar os resultados desse trabalho ou mesmo utilizar o conhecimento já adquirido.

O trabalho atual teve como dimensão temporal o semestre letivo, contudo poderia ser categorizada a um nível diário, o que acarretaria na possibilidade de identificar se existe períodos de sazonalidade do ato de evasão, mas como o sistema acadêmico/financeiro atual não registra essas informações, de maneira a facilitar a extração desses dados, um possível trabalho a ser desenvolvido seria uma estrutura de atualização de informações acadêmicas e financeiras diárias para ser utilizado pelos métodos de mineração de dados, ou seja, um *Data Mart* específico para esse problema, a fim de aprofundar ainda mais sobre o tema evasão na Uniso.

Pode-se citar também a utilização do *Data Mart* definido nesta pesquisa para o processo de mineração de dados com o auxílio de métodos não trabalhados, como por exemplo redes neurais, regressão logística, análise de componentes principais (PCA) e outros.

Outro fruto que poderia surgir como consequência do trabalho em questão, seria a criação de ferramentas de análise que utilizam as regras criadas para identificar probabilidades de evasão nos alunos atualmente matriculados, onde poder-se-ia criar uma série de ações a serem tomadas antes da concretização do fato, e estudado essas ações para identificar quais são as que trazem melhores resultados e com isso, ensinar ao sistema qual é a melhor maneira de tratar o problema da evasão.

## 6 REFERÊNCIAS

- ABRAHAM SILVERSCHATZ, H. F. K. S. S. **Sistema de Banco de Dados**. 5. ed. Rio de Janeiro: Elsevier, 2010.
- BRASILEIRAS, C. E. D. E. S. A. E. N. U. P. **Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas**. ANDIFES/ABRUEM/SESu/MEC. [S.l.], p. 35. 1997.
- BUENO, J. L. O., Agosto 1993. Disponível em: <<http://dx.doi.org/10.1590/S0103-863X1993000200002>>. Acesso em: 07 Novembro 2017.
- CARVALHO, I. C. **Métodos de mineração de dados (Data Mining) como suporte a tomada de decisão**. Instituto Tecnológico de Aeronáutica. São José dos Campos, p. 98. 2002.
- CARVALHO, L. A. V. D. **DATAMINING - A mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração**. 2. ed. São Paulo: Érica, 2001.
- CORADINE, L. C.; LOPES, R. V. V.; MACIEL, A. F. Mineração de Dados: Uma Introdução. **Journal of the Brazilian Neural Network Society**, Alagoas, v. 9, p. 168-184, 2011.
- DEVEDZIC, V. **Knowledge Discovery e Data Mining in Databases**, in **Handbook of Software Engineering e Knowledge**. [S.l.]: [s.n.], 2001.
- DINIZ, C. A. R.; LOUZADA, F. Data Mining: uma introdução. **Associação Brasileira de Estatística**, São Paulo, p. 123, 2000.
- FAYYAD, U.; SHAPIRO, G. P.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, p. 54, Julho 1996.
- FEITOSA, E. **Dendograma**, 2020. Disponível em: <<https://drive.google.com/open?id=1JPqiV6V12wg4ZQTfYzBjkQIdDQLLBwDb>>. Acesso em: 04 fevereiro 2020.
- FEITOSA, E. **Matriz de Correlação**, 2020. Disponível em: <[https://drive.google.com/open?id=1\\_J5Zt5UiV5fJdA\\_HDShEPmeo6-0D9R8l](https://drive.google.com/open?id=1_J5Zt5UiV5fJdA_HDShEPmeo6-0D9R8l)>. Acesso em: 04 fevereiro 2020.
- FILHO, A. C. R. D. S. **Laboratório de Sistemas Neurais (SisNe)**. Disponível em: <<http://sisne.org/Disciplinas/Grad/ProbEstat2/aula13.pdf>>. Acesso em: 10 Janeiro 2020.
- FILHO, J. P. S. **As reprovações em disciplinas nos cursos de graduação da Universidade Federal do Ceará (UFC) no período de 2000 a 2008 e suas implicações na evasão discente**. Universidade Federal do Ceará. Ceará. 2009.
- GRANATYR, J. et al. Mineração de dados com árvores de decisão. **DevMedia**. Disponível em: <<https://www.devmedia.com.br/mineracao-de-dados-com-arvores-de-decisao/31397>>. Acesso em: 05 Fevereiro 2019.
- HAN, J.; KAMBER, M.; PEI, J. **DATA MINING Concepts and Techniques**. 3. ed. San Francisco: Morgan Kaufmann, 2012.
- HÄRDLE, W.; SIMAR, L. **Applied Multivariate Statistical Analysis**. Berlin: Springer-Verlag, 2003.
- HOFFMANN, I. L. **Metodologia para identificação de fatores estratégicos para acompanhamento sistemático da evasão em cursos de graduação**. Universidade Federal de Santa Maria. Rio Grande do Sul, p. 111. 2016.
- INMON, W. H. **Como construir um Data Warehouse**. 2ª. ed. Rio de Janeiro: Campus, 1997.
- JAIN, A. K.; DUBES, R. C. **Algorithms for Clustering Data**. New Jersey: Prentice Hall, 1988.

- LARSON, R.; FARBER, E. **Estatística aplicada**. 4. ed. [S.l.]: Pearson, 2013.
- LOBO, R. L. et al. A Evasão no Ensino Superior Brasileiro. **Cadernos de Pesquisa**, v. 9, p. 641-659, set/dez 2007.
- MAIMON, O.; ROKACH, L. **DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK**. 2ª. ed. Israel: Springer, 2010.
- MICROSOFT Docs. **Microsoft**, 2017. Disponível em: <<https://docs.microsoft.com/pt-br/sql/analysis-services/data-mining/microsoft-clustering-algorithm-technical-reference?view=sql-server-2017>>. Acesso em: 25 Janeiro 2019.
- MONTEIRO, CARLOS A. ET AL. **O desafio de colocar 10 milhões de estudantes no ensino superior: estudo e projeções: parorama e proposta**. CM Consultoria em Administração. Marília. 2011.
- NERY, F. R. M. **Tecnologia e Projeto de Data Warehouse**. 6ª. ed. São Paulo: Érica, 2013.
- NORVIG, P.; RUSSELL, S. **Inteligência artificial**. 3ª. ed. Rio de Janeiro: Elsevier Editora Ltda, 2013.
- PALESTINO, C. B. **BI-business intelligence: modelagem e tecnologia**. Rio de Janeiro: Axcel Books do Brasil, 2001.
- PAREDES, A. S. **A evasão do terceiro grau em Curitiba**. NUPES Núcleo de Pesquisas sobre Ensino Superior Universidade de São Paulo. Curitiba, p. 28. 1994.
- QUINLAN, J. R. Discovering rules by induction from large collections of examples. **Expert System in the Microeletronic Age**, Edinburgh, UK, p. 168-201, 1979.
- QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1993.
- RAMOS, M. N. **Quadro da Evasão na UFPE: Metodologia, Causas e Ações**. UFPE. Pernambuco. 1995.
- RELATÓRIO Social 2017. **Uniso**, 2017. Disponível em: <[http://uniso.br/publicacoes/pdfs/Relatorio-Social/relatorio\\_social\\_2017.pdf](http://uniso.br/publicacoes/pdfs/Relatorio-Social/relatorio_social_2017.pdf)>. Acesso em: 31 Agosto 2018.
- RICCO, R. M. **Graphes D"Induction**. Université Claude Bernard. Lyon - França, p. 315. 1997.
- ROBERTO LEAL LOBO, M. B. D. C. M. L. Como a mudança na metodologia do INEP altera o cálculo da evasão. **Instituto Lobo**, p. 11, 2012. Disponível em: <<http://www.institutolobo.org.br/paginas/artigos.php?v=1>>. Acesso em: 23 Março 2018.
- ROMÃO, W.; FREITAS, A. A.; PACHECO, R. D. S. Uma revisão de abordagens genético-difusas para descoberta de conhecimento em banco de dados. **Acta Scientiarum. Technology**, Paraná, p. 1347-1359, 2000. ISSN 1415-6814.
- SILVA, W. V. et al. Avaliação da Escolha de um fornecedor sob condição de riscos a partir do método de árvore de decisão. **Revista de Gestão USP**, São Paulo, v. 15, p. 77-94, julho-setembro 2008.
- VANNUCCHI, A. **A caminho da Uniso: história: casos e causos**. Sorocaba: EDUniso, 2012.

## APÊNDICE A – SCRIPT SQL DO DATA MART DA EVASÃO

--Dimensões

CREATE TABLE DIMENSAOPERIODOLETIVO

```
(
    IDPERLET INT PRIMARY KEY,
    CODPERLET VARCHAR(10),
    DTINICIO DATETIME,
    DTFIM DATETIME
```

)

GO

CREATE TABLE DIMENSAOALUNO

```
(
    RA VARCHAR(20) PRIMARY KEY,
    NOME VARCHAR(120),
    INST2G VARCHAR(50),
    ESTADOCIVIL VARCHAR(100),
    SEXO VARCHAR(1),
    CORRACA VARCHAR(50),
    DTNASCIMENTO DATETIME
```

)

GO

CREATE TABLE DIMENSAOCURSO

```
(
    CODCURSO VARCHAR(10) PRIMARY KEY,
    NOMECURSO VARCHAR(60),
    CURSOVALORMEDIO NUMERIC(15,2)
```

)

GO

CREATE TABLE DIMENSAOTURNO

```
(
    CODTURNO INT PRIMARY KEY,
    DESCRICAO VARCHAR(15)
```

)

GO

CREATE TABLE DIMENSAOGRADE

```
(
    CODGRADE VARCHAR(10) PRIMARY KEY,
    DESCRICAO VARCHAR(255)
```

```
)  
GO  
CREATE TABLE DIMENSAOHABILITACAOFILIAL  
(  
    IDHABILITACAOFILIAL INT,  
    CODTURNO INT,  
    CODCURSO VARCHAR(10),  
    CODGRADE VARCHAR(10),  
    PRIMARY KEY(IDHABILITACAOFILIAL, CODTURNO, CODCURSO, CODGRADE)  
)  
GO  
CREATE TABLE DIMENSAOSTATUS  
(  
    CODSTATUS INT PRIMARY KEY,  
    DESCRICAO VARCHAR(30),  
    CUINDICATRANSF VARCHAR(1)  
)  
GO  
  
CREATE TABLE DIMENSAOGRANDEAREA  
(  
    CODAREA INT PRIMARY KEY,  
    AREA VARCHAR(60)  
)  
GO  
  
CREATE TABLE DIMENSAOCONTRATO  
(  
    CODCONTRATO VARCHAR(20) NOT NULL,  
    RA VARCHAR(20) NOT NULL,  
    IDPERLET INT NOT NULL,  
    IDHABILITACAOFILIAL INT,  
    DTCONTRATO DATETIME,  
    DTASSINATURA DATETIME,  
    PRIMARY KEY(RA, CODCONTRATO, IDPERLET)  
)  
GO  
--FATO  
CREATE TABLE FATOEVASAO  
(
```

```

id VARCHAR(110) PRIMARY KEY,
IDPERLET INT REFERENCES DIMENSAOPERIODOLETIVO(IDPERLET),
RA VARCHAR(20) REFERENCES DIMENSAOALUNO(RA),
CODCURSO VARCHAR(10) REFERENCES DIMENSAOCURSO(CODCURSO),
STATUSCURSO INT REFERENCES DIMENSAOSTATUS(CODSTATUS),
CODTURNO INT REFERENCES DIMENSAOTURNO(CODTURNO),
CODGRADE VARCHAR(10) REFERENCES DIMENSAOGRADE(CODGRADE),
IDHABILITACAOFILIAL INT,
STATUSPLETIVO INT REFERENCES DIMENSAOSTATUS(CODSTATUS),
DTINICIOPLETIVO DATETIME,
DTFIMPLETIVO DATETIME,
PERLET_INGRESSO INT REFERENCES DIMENSAOPERIODOLETIVO(IDPERLET),
LASTPERLET INT REFERENCES DIMENSAOPERIODOLETIVO(IDPERLET),
CODAREA INT REFERENCES DIMENSAOGRANDEAREA(CODAREA),
CODCONTRATO VARCHAR(20),
idade INT,
MESESENTREDTINICIOEDTFIMPLETIVO INT,
OUTROCURSO VARCHAR(1),
CHTOTAL INT,
CH_CURSADA INT,
PERCENTFEITO NUMERIC(10,2),
PERCENTFALTAS NUMERIC(10,2),
MEDIANOTA NUMERIC(10,2),
QTD_FALTAS INT,
QTD_APROVADAS INT,
QTD_REPROVADAS INT,
QTD_REPFALTAS INT,
QTD_DISPENSADAS INT,
QTD_CANCELADAS INT,
QTD_TRANCADAS INT,
QTD_INCONCLUIDO INT,
N_MATRTOTAL INT,
QTD_PERIODOS INT,
TEMBOLSA VARCHAR(1),
QTDACORDO INT,
QUANTIDADEDIAS INT,
QUANTIDADEMESES INT,
EVADIDO VARCHAR(1),
CONSTRAINT PK_HABILITACAOFILIAL FOREIGN KEY(IDHABILITACAOFILIAL,
CODTURNO, CODCURSO, CODGRADE)

```

**REFERENCES** DIMENSAOHABILITACAOFILIAL (IDHABILITACAOFILIAL,  
CODTURNO, CODCURSO, CODGRADE),

**CONSTRAINT** PK\_CONTRATO **FOREIGN KEY**(RA, CODCONTRATO, IDPERLET)

**REFERENCES** DIMENSAOCONTRATO (RA, CODCONTRATO, IDPERLET)

)

## APÊNDICE B – CURSOS DE GRADUAÇÃO ANALISADOS

Quadro 2 - Cursos analisados por grande área

<b>Ciências Agrárias</b>
AGRONOMIA (BACHARELADO)
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA
ENGENHARIA AGRONÔMICA
MEDICINA VETERINÁRIA
<b>Ciências Biológicas</b>
CIÊNCIAS BIOLÓGICAS (BACHARELADO)
CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO AMBIENTAL (EAD)
GESTÃO AMBIENTAL
<b>Ciências da Saúde</b>
BIOMEDICINA
BIOTECNOLOGIA
CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA
EDUCAÇÃO FÍSICA
EDUCAÇÃO FÍSICA (BACHARELADO)
ENFERMAGEM
FARMÁCIA
FISIOTERAPIA
NUTRIÇÃO
ODONTOLOGIA
TERAPIA OCUPACIONAL
<b>Ciências Exatas e da Terra</b>
CIÊNCIA DA COMPUTAÇÃO
FÍSICA
JOGOS DIGITAIS
MATEMÁTICA
QUÍMICA
QUÍMICA INDUSTRIAL
SISTEMAS DE INFORMAÇÃO
<b>Ciências Humanas</b>
FILOSOFIA
GEOGRAFIA
HISTÓRIA
PEDAGOGIA
PSICOLOGIA
RELAÇÕES INTERNACIONAIS
<b>Ciências Sociais Aplicadas</b>
ADMINISTRAÇÃO
ARQUITETURA E URBANISMO
CIÊNCIAS CONTÁBEIS
CIÊNCIAS ECONÔMICAS

COMÉRCIO EXTERIOR
COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO
COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA
COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS
CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO
CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA
Curso Superior de Tecnologia em Hotelaria
CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS
DESIGN
DESIGN DE INTERIORES
DESIGN DE MODA
DIREITO
EVENTOS
GESTÃO COMERCIAL
GESTÃO DA QUALIDADE
GESTÃO DE RECURSOS HUMANOS
GESTÃO FINANCEIRA
HOTELARIA
JORNALISMO
LOGÍSTICA
MARKETING
MODA
RELAÇÕES PÚBLICAS
<b>Engenharias</b>
ENGENHARIA AMBIENTAL
ENGENHARIA CIVIL
ENGENHARIA DA COMPUTAÇÃO
ENGENHARIA DE ALIMENTOS (BACHARELADO)
ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA
ENGENHARIA DE CONTROLE E AUTOMAÇÃO
ENGENHARIA DE MATERIAIS (BACHARELADO)
ENGENHARIA DE PRODUÇÃO
ENGENHARIA ELÉTRICA
ENGENHARIA MECÂNICA
ENGENHARIA QUÍMICA
GESTÃO DA PRODUÇÃO INDUSTRIAL
SEGURANÇA NO TRABALHO
<b>Linguística, Letras e Artes</b>
ARTES VISUAIS
DANÇA
LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL
LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS
MÚSICA
TEATRO

**Quadro 3 - Status dos discentes nos cursos**

Cancelado
Concluinte
Evadido
Ingressante
Matriculado
Trancado
Transferência de Curso
Transferência interna
Transferido de curso

Fonte: Próprio autor

## APÊNDICE C – STATUS DOS ALUNOS ENTRE O 1º SEMESTRE DE 2012 AO 1º SEMESTRE DE 2018

**Tabela 20 - Status dos alunos da área de Ciências Agrárias dos semestres de 2012 1º Semestre  
a 2018 1º Semestre**

<b>Ciências Agrárias</b>	<b>8956</b>
<b>2012/01</b>	<b>286</b>
Concluinte	4
Ingressante	125
Matriculado	124
Trancado	33
<b>2012/02</b>	<b>274</b>
Cancelado	2
Concluinte	32
Evadido	19
Ingressante	34
Matriculado	173
Trancado	14
<b>2013/01</b>	<b>517</b>
Cancelado	25
Concluinte	13
Evadido	10
Ingressante	225
Matriculado	208
Trancado	36
<b>2013/02</b>	<b>444</b>
Cancelado	3
Concluinte	14
Evadido	31
Ingressante	45
Matriculado	326
Trancado	25
<b>2014/01</b>	<b>648</b>
Cancelado	12
Concluinte	35
Evadido	22
Ingressante	220
Matriculado	322
Trancado	37
<b>2014/02</b>	<b>561</b>
Cancelado	6
Concluinte	32
Evadido	35
Ingressante	32
Matriculado	433

Trancado	20
Transferido de curso	3
<b>2015/01</b>	<b>765</b>
Cancelado	19
Concluinte	7
Evadido	19
Ingressante	248
Matriculado	411
Trancado	61
<b>2015/02</b>	<b>706</b>
Cancelado	4
Concluinte	12
Evadido	48
Ingressante	25
Matriculado	587
Trancado	28
Transferido de curso	2
<b>2016/01</b>	<b>934</b>
Cancelado	19
Concluinte	9
Evadido	21
Ingressante	223
Matriculado	594
Trancado	67
Transferido de curso	1
<b>2016/02</b>	<b>979</b>
Cancelado	12
Concluinte	55
Evadido	20
Ingressante	20
Matriculado	743
Trancado	126
Transferência interna	1
Transferido de curso	2
<b>2017/01</b>	<b>973</b>
Cancelado	13
Concluinte	14
Evadido	26
Ingressante	186
Matriculado	694
Trancado	38
Transferido de curso	2
<b>2017/02</b>	<b>905</b>
Cancelado	7
Concluinte	78
Evadido	38

Ingressante	23
Matriculado	736
Trancado	17
Transferência de Curso	1
Transferido de curso	5
<b>2018/01</b>	<b>964</b>
Cancelado	15
Concluinte	18
Evadido	21
Ingressante	171
Matriculado	699
Trancado	32
Transferência de Curso	1
Transferência interna	3
Transferido de curso	4

Fonte: Próprio autor

**Tabela 21 - Status dos alunos da área de Ciências Biológicas dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Ciências Biológicas</b>	<b>1684</b>
<b>2012/01</b>	<b>176</b>
Concluinte	4
Ingressante	27
Matriculado	131
Trancado	14
<b>2012/02</b>	<b>141</b>
Cancelado	4
Concluinte	34
Evadido	15
Ingressante	1
Matriculado	82
Trancado	5
<b>2013/01</b>	<b>124</b>
Cancelado	6
Concluinte	24
Evadido	9
Ingressante	27
Matriculado	48
Trancado	10
<b>2013/02</b>	<b>98</b>
Cancelado	2
Concluinte	19
Evadido	20
Ingressante	10
Matriculado	43

Trancado	3
Transferência interna	1
<b>2014/01</b>	<b>81</b>
Cancelado	1
Concluinte	7
Evadido	5
Ingressante	21
Matriculado	38
Trancado	9
<b>2014/02</b>	<b>69</b>
Cancelado	1
Concluinte	6
Evadido	7
Ingressante	11
Matriculado	40
Trancado	4
<b>2015/01</b>	<b>109</b>
Cancelado	2
Concluinte	12
Evadido	6
Ingressante	41
Matriculado	32
Trancado	16
<b>2015/02</b>	<b>89</b>
Cancelado	5
Concluinte	1
Evadido	6
Ingressante	10
Matriculado	59
Trancado	7
Transferido de curso	1
<b>2016/01</b>	<b>124</b>
Cancelado	6
Concluinte	1
Evadido	3
Ingressante	34
Matriculado	70
Trancado	8
Transferido de curso	2
<b>2016/02</b>	<b>152</b>
Evadido	6
Ingressante	6
Matriculado	110
Trancado	29
Transferido de curso	1
<b>2017/01</b>	<b>185</b>

Cancelado	7
Concluinte	1
Evadido	3
Ingressante	70
Matriculado	84
Trancado	17
Transferido de curso	3
<b>2017/02</b>	<b>170</b>
Cancelado	8
Concluinte	5
Evadido	14
Ingressante	12
Matriculado	119
Trancado	9
Transferência de Curso	1
Transferido de curso	2
<b>2018/01</b>	<b>166</b>
Cancelado	3
Evadido	4
Ingressante	27
Matriculado	114
Trancado	16
Transferido de curso	2

Fonte: Próprio autor

**Tabela 22 - Status dos alunos da área de Ciências da Saúde dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Ciências da Saúde</b>	<b>17084</b>
<b>2012/01</b>	<b>1224</b>
Concluinte	10
Ingressante	245
Matriculado	900
Trancado	64
Transferência interna	5
<b>2012/02</b>	<b>1127</b>
Cancelado	23
Concluinte	98
Evadido	51
Ingressante	23
Matriculado	883
Trancado	45
Transferência interna	1
Transferido de curso	3
<b>2013/01</b>	<b>1310</b>
Cancelado	44

Concluinte	49
Evadido	31
Ingressante	318
Matriculado	786
Trancado	68
Transferência interna	1
Transferido de curso	13
<b>2013/02</b>	<b>1148</b>
Cancelado	10
Concluinte	136
Evadido	94
Ingressante	19
Matriculado	843
Trancado	34
Transferência de Curso	1
Transferência interna	1
Transferido de curso	10
<b>2014/01</b>	<b>1360</b>
Cancelado	25
Concluinte	48
Evadido	128
Ingressante	336
Matriculado	736
Trancado	78
Transferido de curso	9
<b>2014/02</b>	<b>1102</b>
Cancelado	17
Concluinte	119
Evadido	60
Ingressante	33
Matriculado	819
Trancado	41
Transferência de Curso	1
Transferido de curso	12
<b>2015/01</b>	<b>1397</b>
Cancelado	37
Concluinte	57
Evadido	51
Ingressante	355
Matriculado	774
Trancado	118
Transferência de Curso	2
Transferido de curso	3
<b>2015/02</b>	<b>1191</b>
Cancelado	25
Concluinte	152

Evadido	75
Ingressante	14
Matriculado	856
Trancado	56
Transferido de curso	13
<b>2016/01</b>	<b>1335</b>
Cancelado	20
Concluinte	72
Evadido	46
Ingressante	333
Matriculado	774
Trancado	84
Transferido de curso	6
<b>2016/02</b>	<b>1428</b>
Cancelado	21
Concluinte	152
Evadido	34
Ingressante	78
Matriculado	873
Trancado	239
Transferido de curso	31
<b>2017/01</b>	<b>1460</b>
Cancelado	36
Concluinte	31
Evadido	52
Ingressante	374
Matriculado	861
Trancado	92
Transferência de Curso	2
Transferido de curso	12
<b>2017/02</b>	<b>1356</b>
Cancelado	15
Concluinte	146
Evadido	48
Ingressante	93
Matriculado	990
Trancado	49
Transferido de curso	15
<b>2018/01</b>	<b>1646</b>
Cancelado	43
Concluinte	33
Evadido	53
Ingressante	436
Matriculado	980
Trancado	85
Transferência de Curso	2

Transferido de curso 14  
 Fonte: Próprio autor

**Tabela 23 - Status dos alunos da área de Ciências Exatas e da Terra dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Ciências Exatas e da Terra</b>	<b>6760</b>
<b>2012/01</b>	<b>553</b>
Concluinte	14
Ingressante	85
Matriculado	395
Trancado	53
Transferência interna	5
Transferido de curso	1
<b>2012/02</b>	<b>440</b>
Cancelado	16
Concluinte	47
Evadido	27
Ingressante	5
Matriculado	326
Trancado	17
Transferência interna	1
Transferido de curso	1
<b>2013/01</b>	<b>581</b>
Cancelado	37
Concluinte	16
Evadido	13
Ingressante	167
Matriculado	290
Trancado	49
Transferido de curso	9
<b>2013/02</b>	<b>484</b>
Cancelado	6
Concluinte	23
Evadido	45
Ingressante	28
Matriculado	355
Trancado	23
Transferido de curso	4
<b>2014/01</b>	<b>656</b>
Cancelado	23
Concluinte	18
Evadido	42
Ingressante	171
Matriculado	335
Trancado	60

Transferido de curso	7
<b>2014/02</b>	<b>553</b>
Cancelado	14
Concluinte	28
Evadido	35
Ingressante	32
Matriculado	399
Trancado	39
Transferência de Curso	1
Transferido de curso	5
<b>2015/01</b>	<b>677</b>
Cancelado	11
Concluinte	21
Evadido	28
Ingressante	152
Matriculado	392
Trancado	70
Transferido de curso	3
<b>2015/02</b>	<b>609</b>
Cancelado	14
Concluinte	71
Evadido	47
Ingressante	32
Matriculado	396
Trancado	46
Transferido de curso	3
<b>2016/01</b>	<b>531</b>
Cancelado	18
Concluinte	32
Evadido	36
Ingressante	48
Matriculado	355
Trancado	41
Transferido de curso	1
<b>2016/02</b>	<b>511</b>
Cancelado	3
Concluinte	31
Evadido	19
Ingressante	4
Matriculado	338
Trancado	115
Transferido de curso	1
<b>2017/01</b>	<b>455</b>
Cancelado	11
Concluinte	48
Evadido	16

Ingressante	82
Matriculado	270
Trancado	28
<b>2017/02</b>	<b>356</b>
Cancelado	6
Concluinte	50
Evadido	25
Ingressante	4
Matriculado	242
Trancado	24
Transferido de curso	5
<b>2018/01</b>	<b>354</b>
Cancelado	7
Concluinte	31
Evadido	12
Ingressante	81
Matriculado	202
Trancado	21

Fonte: Próprio autor

**Tabela 24 - Status dos alunos da área de Ciências Humanas dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Ciências Humanas</b>	<b>12324</b>
<b>2012/01</b>	<b>656</b>
Concluinte	20
Ingressante	130
Matriculado	450
Trancado	46
Transferência interna	10
<b>2012/02</b>	<b>642</b>
Cancelado	16
Concluinte	71
Evadido	29
Ingressante	46
Matriculado	456
Trancado	19
Transferência interna	1
Transferido de curso	4
<b>2013/01</b>	<b>861</b>
Cancelado	47
Concluinte	45
Evadido	17
Ingressante	256
Matriculado	442
Trancado	51

Transferência interna	1
Transferido de curso	2
<b>2013/02</b>	<b>766</b>
Cancelado	5
Concluinte	60
Evadido	47
Ingressante	64
Matriculado	553
Trancado	36
Transferência interna	1
<b>2014/01</b>	<b>974</b>
Cancelado	23
Concluinte	63
Evadido	38
Ingressante	257
Matriculado	513
Trancado	79
Transferido de curso	1
<b>2014/02</b>	<b>885</b>
Cancelado	11
Concluinte	75
Evadido	43
Ingressante	88
Matriculado	621
Trancado	45
Transferência interna	1
Transferido de curso	1
<b>2015/01</b>	<b>1146</b>
Cancelado	31
Concluinte	128
Evadido	24
Ingressante	295
Matriculado	592
Trancado	73
Transferido de curso	3
<b>2015/02</b>	<b>1014</b>
Cancelado	12
Concluinte	61
Evadido	70
Ingressante	54
Matriculado	772
Trancado	41
Transferido de curso	4
<b>2016/01</b>	<b>1097</b>
Cancelado	12
Concluinte	69

Evadido	46
Ingressante	214
Matriculado	681
Trancado	71
Transferido de curso	4
<b>2016/02</b>	<b>1116</b>
Cancelado	2
Concluinte	83
Evadido	32
Ingressante	44
Matriculado	810
Trancado	142
Transferido de curso	3
<b>2017/01</b>	<b>1161</b>
Cancelado	13
Concluinte	78
Evadido	27
Ingressante	194
Matriculado	794
Trancado	55
<b>2017/02</b>	<b>969</b>
Cancelado	2
Concluinte	110
Evadido	22
Ingressante	11
Matriculado	797
Trancado	26
Transferido de curso	1
<b>2018/01</b>	<b>1037</b>
Cancelado	14
Concluinte	70
Evadido	25
Ingressante	190
Matriculado	693
Trancado	44
Transferido de curso	1

Fonte: Próprio autor

**Tabela 25 - Status dos alunos da área de Ciências Sociais Aplicadas dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Ciências Sociais Aplicadas</b>	<b>72068</b>
<b>2012/01</b>	<b>6223</b>
Concluinte	317
Evadido	1
Ingressante	1253

Matriculado	4267
Trancado	353
Transferência interna	12
Transferido de curso	20
<b>2012/02</b>	<b>5796</b>
Cancelado	400
Concluinte	506
Evadido	195
Ingressante	436
Matriculado	3997
Trancado	216
Transferência interna	18
Transferido de curso	28
<b>2013/01</b>	<b>6237</b>
Cancelado	271
Concluinte	375
Evadido	234
Ingressante	1291
Matriculado	3653
Trancado	352
Transferência interna	3
Transferido de curso	58
<b>2013/02</b>	<b>5534</b>
Cancelado	65
Concluinte	473
Evadido	479
Ingressante	374
Matriculado	3860
Trancado	242
Transferência de Curso	3
Transferência interna	2
Transferido de curso	36
<b>2014/01</b>	<b>5993</b>
Cancelado	92
Concluinte	257
Evadido	453
Ingressante	1117
Matriculado	3662
Trancado	358
Transferência interna	1
Transferido de curso	53
<b>2014/02</b>	<b>5351</b>
Cancelado	65
Concluinte	531
Evadido	283
Ingressante	367

Matriculado	3812
Trancado	232
Transferência de Curso	2
Transferência interna	2
Transferido de curso	57
<b>2015/01</b>	<b>5818</b>
Cancelado	96
Concluinte	347
Evadido	287
Ingressante	1045
Matriculado	3607
Trancado	393
Transferência de Curso	5
Transferência interna	1
Transferido de curso	37
<b>2015/02</b>	<b>5375</b>
Cancelado	61
Concluinte	545
Evadido	352
Ingressante	241
Matriculado	3925
Trancado	211
Transferência de Curso	4
Transferido de curso	36
<b>2016/01</b>	<b>5319</b>
Cancelado	66
Concluinte	277
Evadido	223
Ingressante	830
Matriculado	3559
Trancado	326
Transferência de Curso	12
Transferido de curso	26
<b>2016/02</b>	<b>5445</b>
Cancelado	32
Concluinte	576
Evadido	163
Ingressante	247
Matriculado	3641
Trancado	747
Transferência de Curso	3
Transferência interna	1
Transferido de curso	35
<b>2017/01</b>	<b>5295</b>
Cancelado	78
Concluinte	341

Evadido	191
Ingressante	984
Matriculado	3384
Trancado	273
Transferência de Curso	9
Transferido de curso	35
<b>2017/02</b>	<b>4776</b>
Cancelado	56
Concluinte	520
Evadido	160
Ingressante	341
Matriculado	3440
Trancado	187
Transferência de Curso	7
Transferido de curso	65
<b>2018/01</b>	<b>4906</b>
Cancelado	85
Concluinte	259
Evadido	190
Ingressante	907
Matriculado	3165
Trancado	253
Transferência de Curso	8
Transferência interna	1
Transferido de curso	38

Fonte: Próprio autor

**Tabela 26 - Status dos alunos da área de Engenharias dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Engenharias</b>	<b>43015</b>
<b>2012/01</b>	<b>3306</b>
Cancelado	4
Concluinte	4
Ingressante	710
Matriculado	2325
Trancado	234
Transferência interna	4
Transferido de curso	25
<b>2012/02</b>	<b>3168</b>
Cancelado	83
Concluinte	137
Evadido	158
Ingressante	245
Matriculado	2357
Trancado	146

Transferência interna	7
Transferido de curso	35
<b>2013/01</b>	<b>3741</b>
Cancelado	188
Concluinte	88
Evadido	118
Ingressante	795
Matriculado	2283
Trancado	214
Transferência interna	7
Transferido de curso	48
<b>2013/02</b>	<b>3489</b>
Cancelado	60
Concluinte	154
Evadido	216
Ingressante	351
Matriculado	2520
Trancado	148
Transferência de Curso	1
Transferência interna	3
Transferido de curso	36
<b>2014/01</b>	<b>3950</b>
Cancelado	56
Concluinte	83
Evadido	247
Ingressante	685
Matriculado	2579
Trancado	264
Transferência de Curso	4
Transferido de curso	32
<b>2014/02</b>	<b>3544</b>
Cancelado	57
Concluinte	192
Evadido	186
Ingressante	209
Matriculado	2679
Trancado	185
Transferência de Curso	1
Transferido de curso	35
<b>2015/01</b>	<b>3781</b>
Cancelado	76
Concluinte	127
Evadido	129
Ingressante	570
Matriculado	2574
Trancado	284

Transferência de Curso	1
Transferido de curso	20
<b>2015/02</b>	<b>3596</b>
Cancelado	45
Concluinte	232
Evadido	231
Ingressante	103
Matriculado	2769
Trancado	188
Transferência de Curso	2
Transferido de curso	26
<b>2016/01</b>	<b>3404</b>
Cancelado	43
Concluinte	206
Evadido	131
Ingressante	304
Matriculado	2488
Trancado	206
Transferido de curso	26
<b>2016/02</b>	<b>3445</b>
Cancelado	26
Concluinte	294
Evadido	108
Ingressante	96
Matriculado	2395
Trancado	510
Transferido de curso	16
<b>2017/01</b>	<b>2921</b>
Cancelado	42
Concluinte	209
Evadido	98
Ingressante	305
Matriculado	2101
Trancado	139
Transferência de Curso	3
Transferido de curso	24
<b>2017/02</b>	<b>2502</b>
Cancelado	22
Concluinte	299
Evadido	97
Ingressante	93
Matriculado	1877
Trancado	89
Transferência de Curso	4
Transferido de curso	21
<b>2018/01</b>	<b>2168</b>

Cancelado	27
Concluinte	199
Evadido	80
Ingressante	181
Matriculado	1542
Trancado	107
Transferência de Curso	1
Transferência interna	2
Transferido de curso	29

Fonte: Próprio autor

**Tabela 27 - Status dos alunos da área de Linguística, Letras e Artes dos semestres de 2012 1º Semestre a 2018 1º Semestre**

<b>Linguística, Letras e Artes</b>	<b>5641</b>
<b>2012/01</b>	<b>418</b>
Concluinte	6
Ingressante	76
Matriculado	305
Trancado	31
<b>2012/02</b>	<b>399</b>
Cancelado	8
Concluinte	38
Evadido	24
Ingressante	21
Matriculado	282
Trancado	23
Transferido de curso	3
<b>2013/01</b>	<b>479</b>
Cancelado	21
Concluinte	22
Evadido	12
Ingressante	112
Matriculado	261
Trancado	48
Transferido de curso	3
<b>2013/02</b>	<b>465</b>
Cancelado	8
Concluinte	55
Evadido	47
Ingressante	59
Matriculado	273
Trancado	22
Transferido de curso	1
<b>2014/01</b>	<b>519</b>
Cancelado	7

Concluinte	14
Evadido	57
Ingressante	111
Matriculado	294
Trancado	31
Transferência de Curso	3
Transferido de curso	2
<b>2014/02</b>	<b>446</b>
Cancelado	3
Concluinte	38
Evadido	23
Ingressante	38
Matriculado	311
Trancado	29
Transferido de curso	4
<b>2015/01</b>	<b>493</b>
Cancelado	10
Concluinte	21
Evadido	38
Ingressante	92
Matriculado	287
Trancado	44
Transferido de curso	1
<b>2015/02</b>	<b>468</b>
Cancelado	6
Concluinte	42
Evadido	40
Ingressante	41
Matriculado	314
Trancado	25
<b>2016/01</b>	<b>445</b>
Cancelado	3
Concluinte	24
Evadido	26
Ingressante	66
Matriculado	298
Trancado	26
Transferido de curso	2
<b>2016/02</b>	<b>472</b>
Cancelado	3
Concluinte	43
Evadido	14
Ingressante	26
Matriculado	294
Trancado	89
Transferido de curso	3

<b>2017/01</b>	<b>395</b>
Cancelado	3
Concluinte	33
Evadido	19
Ingressante	66
Matriculado	244
Trancado	28
Transferido de curso	2
<b>2017/02</b>	<b>308</b>
Cancelado	3
Concluinte	68
Evadido	13
Ingressante	1
Matriculado	206
Trancado	13
Transferência de Curso	1
Transferido de curso	3
<b>2018/01</b>	<b>334</b>
Cancelado	3
Concluinte	37
Evadido	6
Ingressante	100
Matriculado	162
Trancado	25
Transferido de curso	1

Fonte: Próprio autor

## APÊNDICE D – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE ENGENHARIAS

**Tabela 28 - Índice de evasão por período letivo para os cursos da grande área de Engenharias**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	203	52	198	0	23,74
2012/01	ENGENHARIA DE PRODUÇÃO	702	200	599	0	16,19
2012/01	ENGENHARIA AMBIENTAL	372	75	351	0	15,38
2012/01	ENGENHARIA CIVIL	509	227	339	0	16,81
2012/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	132	49	95	0	12,63
2012/01	ENGENHARIA ELÉTRICA	77	38	47	0	17,02
2012/01	ENGENHARIA QUÍMICA	290	120	187	0	9,09
2012/01	ENGENHARIA DA COMPUTAÇÃO	100	41	74	0	20,27
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	161	16	203	0	28,57
2012/02	ENGENHARIA DE PRODUÇÃO	693	80	702	0	12,68
2012/02	ENGENHARIA AMBIENTAL	362	19	372	0	7,8
2012/02	ENGENHARIA CIVIL	525	86	509	0	13,75
2012/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	113	5	132	0	18,18
2012/02	ENGENHARIA ELÉTRICA	71	4	77	0	12,99

2012/02	ENGENHARIA QUÍMICA	266	19	290	0	14,83
2012/02	ENGENHARIA DA COMPUTAÇÃO	89	3	100	0	14
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	164	47	161	0	27,33
2013/01	ENGENHARIA DE PRODUÇÃO	684	122	693	0	18,9
2013/01	ENGENHARIA AMBIENTAL	353	66	362	0	20,72
2013/01	ENGENHARIA CIVIL	665	218	525	0	14,86
2013/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	138	43	113	0	15,93
2013/01	ENGENHARIA ELÉTRICA	102	45	71	0	19,72
2013/01	ENGENHARIA QUÍMICA	348	100	266	0	6,77
2013/01	ENGENHARIA DA COMPUTAÇÃO	120	42	89	0	12,36
2013/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	154	31	164	16	16,89
2013/02	ENGENHARIA DE PRODUÇÃO	668	86	684	32	10,74
2013/02	ENGENHARIA AMBIENTAL	345	23	353	13	5,29
2013/02	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	28	1	33	0	18,18
2013/02	ENGENHARIA CIVIL	675	92	665	0	12,33
2013/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	137	25	138	0	18,84
2013/02	ENGENHARIA ELÉTRICA	118	26	102	0	9,8
2013/02	ENGENHARIA QUÍMICA	361	37	348	0	6,9

2013/02	ENGENHARIA DA COMPUTAÇÃO	124	27	120	0	19,17
2013/02	SEGURANÇA NO TRABALHO	17	9	11	0	27,27
2013/02	ENGENHARIA DE ALIMENTOS (BACHARELADO)	17	1	23	0	30,43
2013/02	ENGENHARIA DE MATERIAIS (BACHARELADO)	11	1	11	0	9,09
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	135	29	154	30	14,52
2014/01	ENGENHARIA DE PRODUÇÃO	649	119	668	64	12,25
2014/01	ENGENHARIA AMBIENTAL	327	55	345	36	11,97
2014/01	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	43	20	28	0	17,86
2014/01	ENGENHARIA CIVIL	781	184	675	0	11,56
2014/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	157	36	137	0	11,68
2014/01	ENGENHARIA ELÉTRICA	144	37	118	0	9,32
2014/01	ENGENHARIA QUÍMICA	409	79	361	0	8,59
2014/01	ENGENHARIA DA COMPUTAÇÃO	140	40	124	0	19,35
2014/01	SEGURANÇA NO TRABALHO	16	2	17	0	17,65
2014/01	ENGENHARIA DE ALIMENTOS (BACHARELADO)	47	31	17	0	5,88
2014/01	ENGENHARIA DE MATERIAIS (BACHARELADO)	26	17	11	0	18,18

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	123	14	135	14	9,92
2014/02	ENGENHARIA DE PRODUÇÃO	593	39	649	29	10,65
2014/02	ENGENHARIA AMBIENTAL	285	15	327	20	12,05
2014/02	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	34	0	43	0	20,93
2014/02	ENGENHARIA CIVIL	744	62	781	0	12,68
2014/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	152	16	157	0	13,38
2014/02	ENGENHARIA ELÉTRICA	146	18	144	0	11,11
2014/02	ENGENHARIA QUÍMICA	391	27	409	0	11
2014/02	ENGENHARIA DA COMPUTAÇÃO	146	21	140	0	10,71
2014/02	SEGURANÇA NO TRABALHO	15	2	16	0	18,75
2014/02	ENGENHARIA DE ALIMENTOS (BACHARELADO)	44	0	47	0	6,38
2014/02	ENGENHARIA DE MATERIAIS (BACHARELADO)	23	3	26	0	23,08
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	109	25	123	18	20
2015/01	ENGENHARIA DE PRODUÇÃO	588	91	593	41	9,96
2015/01	ENGENHARIA AMBIENTAL	276	47	285	36	8,03

2015/01	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	40	11	34	0	14,71
2015/01	ENGENHARIA CIVIL	827	168	744	18	9,23
2015/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	150	31	152	12	15
2015/01	ENGENHARIA ELÉTRICA	165	32	146	5	5,67
2015/01	ENGENHARIA QUÍMICA	391	66	391	38	7,93
2015/01	ENGENHARIA DA COMPUTAÇÃO	159	37	146	4	14,08
2015/01	SEGURANÇA NO TRABALHO	13	0	15	0	13,33
2015/01	ENGENHARIA DE ALIMENTOS (BACHARELADO)	65	26	44	0	11,36
2015/01	ENGENHARIA DE MATERIAIS (BACHARELADO)	30	12	23	0	21,74
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	85	1	109	15	10,64
2015/02	ENGENHARIA DE PRODUÇÃO	542	47	588	35	10,49
2015/02	ENGENHARIA AMBIENTAL	247	4	276	12	7,95
2015/02	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	34	0	40	0	15
2015/02	ENGENHARIA CIVIL	808	48	827	15	6,4
2015/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	140	4	150	4	6,85
2015/02	ENGENHARIA ELÉTRICA	141	2	165	8	11,46
2015/02	ENGENHARIA QUÍMICA	378	23	391	6	7,79

2015/02	ENGENHARIA DA COMPUTAÇÃO	134	0	159	1	15,19
2015/02	SEGURANÇA NO TRABALHO	13	0	13	0	0
2015/02	ENGENHARIA DE ALIMENTOS (BACHARELADO)	60	1	65	0	9,23
2015/02	ENGENHARIA DE MATERIAIS (BACHARELADO)	28	0	30	0	6,67
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	70	2	85	14	4,23
2016/01	ENGENHARIA DE PRODUÇÃO	494	58	542	46	12,1
2016/01	ENGENHARIA AMBIENTAL	231	27	247	32	5,12
2016/01	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	39	9	34	0	11,76
2016/01	ENGENHARIA CIVIL	779	88	808	41	9,91
2016/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	124	15	140	14	13,49
2016/01	ENGENHARIA ELÉTRICA	139	26	141	13	11,72
2016/01	ENGENHARIA QUÍMICA	355	53	378	32	12,72
2016/01	ENGENHARIA DA COMPUTAÇÃO	141	26	134	7	9,45
2016/01	SEGURANÇA NO TRABALHO	11	0	13	2	0
2016/01	ENGENHARIA DE ALIMENTOS (BACHARELADO)	71	17	60	0	10
2016/01	ENGENHARIA DE MATERIAIS (BACHARELADO)	23	1	28	0	21,43

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	52	2	70	12	13,79
2016/02	ENGENHARIA DE PRODUÇÃO	437	19	494	35	8,93
2016/02	ENGENHARIA AMBIENTAL	195	10	231	31	7,5
2016/02	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	35	0	39	0	10,26
2016/02	ENGENHARIA MECÂNICA	25	14	12	0	8,33
2016/02	ENGENHARIA CIVIL	703	28	779	38	8,91
2016/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	103	1	124	8	12,07
2016/02	ENGENHARIA ELÉTRICA	127	4	139	3	9,56
2016/02	ENGENHARIA QUÍMICA	337	23	355	35	1,88
2016/02	ENGENHARIA DA COMPUTAÇÃO	133	8	141	2	10,07
2016/02	SEGURANÇA NO TRABALHO	10	0	11	1	0
2016/02	ENGENHARIA DE ALIMENTOS (BACHARELADO)	65	0	71	0	8,45
2016/02	ENGENHARIA DE MATERIAIS (BACHARELADO)	27	3	23	0	-4,35
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	30	1	52	20	9,38
2017/01	ENGENHARIA DE PRODUÇÃO	392	45	437	62	7,73

2017/01	ENGENHARIA AMBIENTAL	187	30	195	28	5,99
2017/01	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	43	10	35	0	5,71
2017/01	ENGENHARIA MECÂNICA	53	32	25	0	16
2017/01	ENGENHARIA CIVIL	639	54	703	60	9,02
2017/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	108	22	103	9	8,51
2017/01	ENGENHARIA ELÉTRICA	123	20	127	11	11,21
2017/01	ENGENHARIA QUÍMICA	290	40	337	52	12,28
2017/01	ENGENHARIA DA COMPUTAÇÃO	135	27	133	10	12,2
2017/01	SEGURANÇA NO TRABALHO	5	0	10	5	0
2017/01	ENGENHARIA DE ALIMENTOS (BACHARELADO)	74	11	65	0	3,08
2017/01	ENGENHARIA DE MATERIAIS (BACHARELADO)	25	2	27	0	14,81
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/02	GESTÃO DA PRODUÇÃO INDUSTRIAL	14	0	30	11	26,32
2017/02	ENGENHARIA DE PRODUÇÃO	322	18	392	57	9,25
2017/02	ENGENHARIA AMBIENTAL	167	6	187	20	3,59
2017/02	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	39	0	43	0	9,3
2017/02	ENGENHARIA MECÂNICA	59	12	53	0	11,32
2017/02	ENGENHARIA CIVIL	575	36	639	46	9,11

2017/02	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	97	8	108	5	13,59
2017/02	ENGENHARIA ELÉTRICA	119	9	123	7	5,17
2017/02	ENGENHARIA QUÍMICA	268	12	290	15	6,91
2017/02	ENGENHARIA DA COMPUTAÇÃO	130	16	135	3	13,64
2017/02	ENGENHARIA DE ALIMENTOS (BACHARELADO)	69	0	74	0	6,76
2017/02	ENGENHARIA DE MATERIAIS (BACHARELADO)	22	0	25	0	12
2018/01	GESTÃO DA PRODUÇÃO INDUSTRIAL	5	0	14	10	-25
2018/01	ENGENHARIA DE PRODUÇÃO	285	25	322	38	8,45
2018/01	ENGENHARIA AMBIENTAL	148	19	167	20	12,24
2018/01	ENGENHARIA DE BIOPROCESSOS E BIOTECNOLOGIA	37	4	39	6	0
2018/01	ENGENHARIA MECÂNICA	72	18	59	0	8,47
2018/01	ENGENHARIA CIVIL	480	38	575	83	10,16
2018/01	ENGENHARIA DE CONTROLE E AUTOMAÇÃO	94	10	97	10	3,45
2018/01	ENGENHARIA ELÉTRICA	110	12	119	14	6,67
2018/01	ENGENHARIA QUÍMICA	249	36	268	40	6,58
2018/01	ENGENHARIA DA COMPUTAÇÃO	117	18	130	12	16,1
2018/01	ENGENHARIA DE ALIMENTOS (BACHARELADO)	66	7	69	9	1,67
2018/01	ENGENHARIA DE MATERIAIS (BACHARELADO)	11	0	22	11	0

Fonte: Próprio autor

## APÊNDICE E – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS SOCIAIS APLICADAS

**Tabela 29 - Índice de evasão por período letivo para os cursos da grande área de Ciências Sociais Aplicadas**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	225	71	190	0	18,95
2012/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	281	112	207	0	18,36
2012/01	DIREITO	766	208	693	2	19,25
2012/01	CIÊNCIAS CONTÁBEIS	346	105	323	2	24,92
2012/01	CIÊNCIAS ECONÔMICAS	158	55	140	0	26,43
2012/01	HOTELARIA	77	20	88	0	35,23
2012/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	49	7	51	0	17,65
2012/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	78	30	73	0	34,25
2012/01	GESTÃO DE RECURSOS HUMANOS	141	38	145	1	28,47
2012/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	72	40	46	0	30,43
2012/01	GESTÃO FINANCEIRA	78	35	59	1	25,86
2012/01	ADMINISTRAÇÃO	879	276	828	5	26,73
2012/01	COMÉRCIO EXTERIOR	231	61	220	0	22,73
2012/01	MARKETING	26	8	33	0	45,45
2012/01	LOGÍSTICA	196	57	214	0	35,05
2012/01	DESIGN DE INTERIORES	29	21	15	0	46,67

2012/01	GESTÃO DA QUALIDADE	128	41	132	2	33,08
2012/01	DESIGN DE PRODUTO	2	0	7	0	71,43
2012/01	GESTÃO COMERCIAL	27	7	29	0	31,03
2012/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	27	13	29	0	51,72
2012/01	ARQUITETURA E URBANISMO	475	179	329	0	10,03
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	177	5	225	0	23,56
2012/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	219	5	281	0	23,84
2012/02	DIREITO	750	87	766	0	13,45
2012/02	CIÊNCIAS CONTÁBEIS	323	24	346	0	13,58
2012/02	CIÊNCIAS ECONÔMICAS	138	7	158	0	17,09
2012/02	HOTELARIA	65	4	77	0	20,78
2012/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	30	2	49	0	42,86
2012/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	69	6	78	0	19,23
2012/02	GESTÃO DE RECURSOS HUMANOS	122	19	141	0	26,95
2012/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	62	6	72	0	22,22
2012/02	GESTÃO FINANCEIRA	57	3	78	0	30,77
2012/02	ADMINISTRAÇÃO	776	93	879	18	20,67
2012/02	COMÉRCIO EXTERIOR	211	16	231	1	15,22

2012/02	MARKETING	19	3	26	0	38,46
2012/02	LOGÍSTICA	184	23	196	0	17,86
2012/02	DESIGN DE INTERIORES	25	2	29	0	20,69
2012/02	GESTÃO DA QUALIDADE	112	17	128	0	25,78
2012/02	DESIGN DE PRODUTO	1	0	2	0	50
2012/02	GESTÃO COMERCIAL	17	0	27	0	37,04
2012/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	24	2	27	0	18,52
2012/02	ARQUITETURA E URBANISMO	494	62	475	0	9,05
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	211	72	177	0	21,47
2013/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	257	77	219	2	17,05
2013/01	DIREITO	856	220	750	0	15,2
2013/01	CIÊNCIAS CONTÁBEIS	313	59	323	1	21,12
2013/01	CIÊNCIAS ECONÔMICAS	135	39	138	0	30,43
2013/01	HOTELARIA	57	18	65	0	40
2013/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	41	17	30	0	20
2013/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	57	21	69	0	47,83
2013/01	GESTÃO DE RECURSOS HUMANOS	108	33	122	0	38,52
2013/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	69	31	62	3	35,59

2013/01	GESTÃO FINANCEIRA	55	12	57	0	24,56
2013/01	ADMINISTRAÇÃO	758	174	776	12	23,56
2013/01	COMÉRCIO EXTERIOR	215	56	211	1	24,29
2013/01	MARKETING	28	12	19	0	15,79
2013/01	LOGÍSTICA	170	54	184	1	36,61
2013/01	DESIGN DE INTERIORES	27	17	25	2	56,52
2013/01	GESTÃO DA QUALIDADE	103	31	112	0	35,71
2013/01	DESIGN	97	28	107	22	18,82
2013/01	GESTÃO COMERCIAL	25	16	17	0	47,06
2013/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	29	11	24	0	25
2013/01	ARQUITETURA E URBANISMO	670	240	494	0	12,96
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	198	28	211	19	11,46
2013/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	210	27	257	28	20,09
2013/02	DIREITO	754	19	856	24	11,66
2013/02	CIÊNCIAS CONTÁBEIS	308	36	313	19	7,48
2013/02	CIÊNCIAS ECONÔMICAS	131	10	135	1	9,7
2013/02	HOTELARIA	49	8	57	7	18
2013/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	38	16	41	12	24,14
2013/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	52	8	57	12	2,22
2013/02	GESTÃO DE RECURSOS HUMANOS	92	15	108	19	13,48

2013/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	71	22	69	8	19,67
2013/02	GESTÃO FINANCEIRA	41	3	55	14	7,32
2013/02	ADMINISTRAÇÃO	703	95	758	48	14,37
2013/02	COMÉRCIO EXTERIOR	221	34	215	12	7,88
2013/02	MARKETING	19	3	28	4	33,33
2013/02	LOGÍSTICA	130	22	170	43	14,96
2013/02	DESIGN DE MODA	12	6	6	0	0
2013/02	DESIGN DE INTERIORES	27	4	27	0	14,81
2013/02	GESTÃO DA QUALIDADE	81	13	103	20	18,07
2013/02	DESIGN	86	9	97	5	16,3
2013/02	GESTÃO COMERCIAL	22	3	25	2	17,39
2013/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	25	4	29	6	8,7
2013/02	EVENTOS	7	4	6	0	50
2013/02	ARQUITETURA E URBANISMO	662	60	670	0	10,15
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	225	48	198	4	8,76
2014/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	244	83	210	12	18,69
2014/01	DIREITO	877	232	754	39	9,79
2014/01	CIÊNCIAS CONTÁBEIS	300	56	308	40	8,96
2014/01	CIÊNCIAS ECONÔMICAS	140	40	131	12	15,97
2014/01	HOTELARIA	38	1	49	6	13,95
2014/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	41	18	38	8	23,33

2014/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	56	31	52	21	19,35
2014/01	GESTÃO DE RECURSOS HUMANOS	69	16	92	30	14,52
2014/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	79	35	71	19	15,38
2014/01	GESTÃO FINANCEIRA	40	18	41	13	21,43
2014/01	ADMINISTRAÇÃO	674	137	703	65	15,83
2014/01	COMÉRCIO EXTERIOR	212	53	221	31	16,32
2014/01	MARKETING	20	7	19	5	7,14
2014/01	LOGÍSTICA	120	38	130	26	21,15
2014/01	DESIGN DE MODA	22	10	12	0	0
2014/01	DESIGN DE INTERIORES	17	5	27	6	42,86
2014/01	GESTÃO DA QUALIDADE	76	24	81	20	14,75
2014/01	DESIGN	82	21	86	12	17,57
2014/01	GESTÃO COMERCIAL	21	8	22	2	35
2014/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	21	9	25	8	29,41
2014/01	EVENTOS	4	0	7	0	42,86
2014/01	ARQUITETURA E URBANISMO	847	231	662	0	6,95
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	230	32	225	5	10
2014/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	242	23	244	5	8,37
2014/02	DIREITO	780	20	877	28	10,48
2014/02	CIÊNCIAS CONTÁBEIS	289	26	300	7	10,24

2014/02	CIÊNCIAS ECONÔMICAS	138	15	140	5	8,89
2014/02	HOTELARIA	25	2	38	10	17,86
2014/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	36	2	41	2	12,82
2014/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	48	0	56	2	11,11
2014/02	GESTÃO DE RECURSOS HUMANOS	66	13	69	10	10,17
2014/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	64	15	79	12	26,87
2014/02	GESTÃO FINANCEIRA	42	12	40	5	14,29
2014/02	ADMINISTRAÇÃO	639	74	674	41	10,74
2014/02	COMÉRCIO EXTERIOR	208	20	212	9	7,39
2014/02	MARKETING	28	9	20	1	0
2014/02	LOGÍSTICA	102	16	120	20	14
2014/02	DESIGN DE MODA	18	2	22	0	27,27
2014/02	DESIGN DE INTERIORES	18	4	17	1	12,5
2014/02	GESTÃO DA QUALIDADE	64	15	76	16	18,33
2014/02	DESIGN	77	20	82	12	18,57
2014/02	GESTÃO COMERCIAL	21	4	21	2	10,53
2014/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	25	4	21	0	0
2014/02	EVENTOS	3	0	4	0	25
2014/02	ARQUITETURA E URBANISMO	856	86	847	0	9,09
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	190	9	230	30	9,5

2015/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	250	60	242	21	14,03
2015/01	DIREITO	854	202	780	54	10,19
2015/01	CIÊNCIAS CONTÁBEIS	290	66	289	48	7,05
2015/01	CIÊNCIAS ECONÔMICAS	125	21	138	16	14,75
2015/01	HOTELARIA	20	3	25	6	10,53
2015/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	30	1	36	2	14,71
2015/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	46	17	48	13	17,14
2015/01	GESTÃO DE RECURSOS HUMANOS	52	8	66	17	10,2
2015/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	50	18	64	17	31,91
2015/01	GESTÃO FINANCEIRA	37	8	42	2	27,5
2015/01	ADMINISTRAÇÃO	627	131	639	74	12,21
2015/01	COMÉRCIO EXTERIOR	221	59	208	21	13,37
2015/01	MARKETING	12	0	28	8	40
2015/01	LOGÍSTICA	84	23	102	27	18,67
2015/01	DESIGN DE MODA	21	11	18	1	41,18
2015/01	DESIGN DE INTERIORES	17	9	18	5	38,46
2015/01	GESTÃO DA QUALIDADE	46	14	64	19	28,89
2015/01	DESIGN	67	20	77	17	21,67
2015/01	GESTÃO COMERCIAL	22	8	21	4	17,65
2015/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	19	4	25	4	28,57
2015/01	EVENTOS	1	0	3	2	0
2015/01	ARQUITETURA E URBANISMO	979	220	856	33	7,78

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	164	5	190	19	7,02
2015/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	242	38	250	23	10,13
2015/02	DIREITO	806	38	854	27	7,13
2015/02	CIÊNCIAS CONTÁBEIS	262	27	290	37	7,11
2015/02	CIÊNCIAS ECONÔMICAS	108	3	125	5	12,5
2015/02	HOTELARIA	13	3	20	4	37,5
2015/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	25	0	30	3	7,41
2015/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	42	2	46	1	11,11
2015/02	GESTÃO DE RECURSOS HUMANOS	34	3	52	15	16,22
2015/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	35	3	50	11	17,95
2015/02	GESTÃO FINANCEIRA	30	1	37	6	6,45
2015/02	ADMINISTRAÇÃO	567	63	627	72	9,19
2015/02	COMÉRCIO EXTERIOR	190	6	221	17	9,8
2015/02	MARKETING	9	1	12	0	33,33
2015/02	LOGÍSTICA	67	4	84	14	10
2015/02	DESIGN DE MODA	16	0	21	3	11,11
2015/02	DESIGN DE INTERIORES	15	5	17	3	28,57
2015/02	GESTÃO DA QUALIDADE	40	3	46	2	15,91
2015/02	DESIGN	71	10	67	0	8,96
2015/02	Curso Superior de Tecnologia em Hotelaria	8	0	9	0	11,11

2015/02	JORNALISMO	47	1	56	0	17,86
2015/02	RELAÇÕES PÚBLICAS	10	2	11	0	27,27
2015/02	GESTÃO COMERCIAL	24	3	22	0	4,55
2015/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	14	0	19	4	6,67
2015/02	ARQUITETURA E URBANISMO	951	78	979	24	8,59
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	123	1	164	35	5,43
2016/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	243	56	242	30	11,79
2016/01	DIREITO	849	186	806	71	9,8
2016/01	CIÊNCIAS CONTÁBEIS	231	52	262	48	16,36
2016/01	CIÊNCIAS ECONÔMICAS	103	32	108	15	23,66
2016/01	HOTELARIA	7	0	13	6	0
2016/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	24	0	25	0	4
2016/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	40	24	42	18	33,33
2016/01	GESTÃO DE RECURSOS HUMANOS	22	1	34	10	12,5
2016/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	37	21	35	10	36
2016/01	GESTÃO FINANCEIRA	34	13	30	5	16
2016/01	ADMINISTRAÇÃO	543	128	567	82	14,43
2016/01	COMÉRCIO EXTERIOR	182	39	190	20	15,88
2016/01	MARKETING	7	0	9	1	12,5

2016/01	LOGÍSTICA	47	10	67	18	24,49
2016/01	DESIGN DE MODA	10	0	16	4	16,67
2016/01	DESIGN DE INTERIORES	9	3	15	4	45,45
2016/01	GESTÃO DA QUALIDADE	33	1	40	7	3,03
2016/01	DESIGN	65	20	71	14	21,05
2016/01	Curso Superior de Tecnologia em Hotelaria	7	0	8	0	12,5
2016/01	JORNALISMO	87	44	47	0	8,51
2016/01	RELAÇÕES PÚBLICAS	26	17	10	0	10
2016/01	GESTÃO COMERCIAL	12	0	24	9	20
2016/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	6	0	14	8	0
2016/01	ARQUITETURA E URBANISMO	979	169	951	36	11,48
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	118	2	123	3	3,33
2016/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	210	1	243	10	10,3
2016/02	DIREITO	798	42	849	49	5,5
2016/02	CIÊNCIAS CONTÁBEIS	217	16	231	17	6,07
2016/02	CIÊNCIAS ECONÔMICAS	103	10	103	4	6,06
2016/02	HOTELARIA	4	0	7	2	20
2016/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	23	1	24	0	8,33
2016/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	38	0	40	2	0

2016/02	GESTÃO DE RECURSOS HUMANOS	24	13	22	7	26,67
2016/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	29	2	37	2	22,86
2016/02	GESTÃO FINANCEIRA	32	13	34	12	13,64
2016/02	ADMINISTRAÇÃO	490	36	543	43	9,2
2016/02	COMÉRCIO EXTERIOR	162	4	182	13	6,51
2016/02	MARKETING	4	0	7	0	42,86
2016/02	LOGÍSTICA	44	14	47	10	18,92
2016/02	DESIGN DE MODA	9	0	10	0	10
2016/02	DESIGN DE INTERIORES	6	0	9	3	0
2016/02	GESTÃO DA QUALIDADE	38	14	33	7	7,69
2016/02	DESIGN	59	3	65	6	5,08
2016/02	Curso Superior de Tecnologia em Hotelaria	8	1	7	0	0
2016/02	JORNALISMO	89	12	87	0	11,49
2016/02	RELAÇÕES PÚBLICAS	36	12	26	0	7,69
2016/02	GESTÃO COMERCIAL	17	9	12	3	11,11
2016/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	8	5	6	3	0
2016/02	ARQUITETURA E URBANISMO	936	48	979	31	6,33
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	68	2	118	43	12
2017/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	224	61	210	24	12,37
2017/01	DIREITO	828	164	798	62	9,78

2017/01	CIÊNCIAS CONTÁBEIS	244	69	217	30	6,42
2017/01	CIÊNCIAS ECONÔMICAS	113	43	103	20	15,66
2017/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	15	1	23	7	12,5
2017/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	62	35	38	9	6,9
2017/01	GESTÃO DE RECURSOS HUMANOS	37	23	24	4	30
2017/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	20	2	29	8	14,29
2017/01	GESTÃO FINANCEIRA	31	12	32	7	24
2017/01	ADMINISTRAÇÃO	495	130	490	68	13,51
2017/01	COMÉRCIO EXTERIOR	161	36	162	27	7,41
2017/01	MARKETING	4	0	4	1	-33,33
2017/01	LOGÍSTICA	49	20	44	12	9,38
2017/01	DESIGN DE MODA	5	0	9	1	37,5
2017/01	DESIGN DE INTERIORES	3	0	6	2	25
2017/01	GESTÃO DA QUALIDADE	34	15	38	13	24
2017/01	DESIGN	31	1	59	18	26,83
2017/01	Curso Superior de Tecnologia em Hotelaria	4	2	8	5	33,33
2017/01	JORNALISMO	130	47	89	0	6,74
2017/01	RELAÇÕES PÚBLICAS	52	23	36	0	19,44
2017/01	GESTÃO COMERCIAL	26	16	17	5	16,67
2017/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	16	10	8	1	14,29
2017/01	ARQUITETURA E URBANISMO	956	149	936	58	8,09

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/02	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	58	3	68	13	0
2017/02	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	209	27	224	12	14,15
2017/02	DIREITO	754	51	828	48	9,87
2017/02	CIÊNCIAS CONTÁBEIS	230	28	244	24	8,18
2017/02	CIÊNCIAS ECONÔMICAS	124	15	113	0	3,54
2017/02	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	10	0	15	4	9,09
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	61	7	62	2	10
2017/02	GESTÃO DE RECURSOS HUMANOS	37	9	37	2	20
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	14	1	20	5	13,33
2017/02	GESTÃO FINANCEIRA	36	9	31	1	10
2017/02	ADMINISTRAÇÃO	460	50	495	45	8,89
2017/02	COMÉRCIO EXTERIOR	144	10	161	13	9,46
2017/02	MARKETING	13	13	4	2	100
2017/02	LOGÍSTICA	51	11	49	5	9,09
2017/02	DESIGN DE MODA	4	0	5	1	0
2017/02	GESTÃO DA QUALIDADE	30	5	34	5	13,79
2017/02	DESIGN	23	3	31	6	20
2017/02	Curso Superior de Tecnologia em Hotelaria	4	0	4	0	0
2017/02	JORNALISMO	131	20	130	0	14,62
2017/02	RELAÇÕES PÚBLICAS	59	7	52	0	0

2017/02	MODA	24	0	26	0	7,69
2017/02	GESTÃO COMERCIAL	22	2	26	3	13,04
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	29	13	16	0	0
2017/02	ARQUITETURA E URBANISMO	884	47	956	48	7,82
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2018/01	COMUNICAÇÃO SOCIAL - HABIL. EM JORNALISMO	24	0	58	34	0
2018/01	COMUNICAÇÃO SOCIAL - HABIL. EM PUBLICIDADE E PROPAGANDA	220	62	209	33	10,23
2018/01	DIREITO	803	188	754	70	10,09
2018/01	CIÊNCIAS CONTÁBEIS	238	57	230	29	9,95
2018/01	CIÊNCIAS ECONÔMICAS	131	36	124	15	12,84
2018/01	COMUNICAÇÃO SOCIAL - HABIL. EM RELAÇÕES PÚBLICAS	3	1	10	8	0
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM GASTRONOMIA	75	32	61	6	21,82
2018/01	GESTÃO DE RECURSOS HUMANOS	53	21	37	3	5,88
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM DESIGN GRÁFICO	10	3	14	4	30
2018/01	GESTÃO FINANCEIRA	38	14	36	10	7,69
2018/01	ADMINISTRAÇÃO	468	104	460	51	11
2018/01	COMÉRCIO EXTERIOR	139	30	144	20	12,1
2018/01	MARKETING	28	20	13	0	38,46
2018/01	LOGÍSTICA	48	16	51	8	25,58
2018/01	GESTÃO DA QUALIDADE	32	11	30	3	22,22

2018/01	DESIGN	8	1	23	9	50
2018/01	Curso Superior de Tecnologia em Hotelaria	2	1	4	3	0
2018/01	JORNALISMO	173	52	131	0	7,63
2018/01	RELAÇÕES PÚBLICAS	70	21	59	0	16,95
2018/01	MODA	41	26	24	0	37,5
2018/01	GESTÃO COMERCIAL	32	14	22	1	14,29
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM PROCESSOS GERENCIAIS	36	15	29	1	25
2018/01	ARQUITETURA E URBANISMO	819	104	884	111	7,5

Fonte: Próprio autor

## APÊNDICE F – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS AGRÁRIAS

Tabela 30 - Índice de evasão por período letivo para os cursos da grande área de Ciências Agrárias

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº concluintes	Taxa de Evasão
2012/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	90	32	108	1	45,79
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº concluintes	Taxa de Evasão
2012/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	96	29	90	0	25,56
2012/02	MEDICINA VETERINÁRIA	75	0	88	0	14,77
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº concluintes	Taxa de Evasão
2013/01	MEDICINA VETERINÁRIA	199	128	75	0	5,33
2013/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	111	48	96	0	34,38
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº concluintes	Taxa de Evasão
2013/02	AGRONOMIA (BACHARELADO)	8	4	5	0	20

2013/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	107	24	111	12	16,16
2013/02	MEDICINA VETERINÁRIA	179	3	199	0	11,56
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2014/01	MEDICINA VETERINÁRIA	314	143	179	0	4,47
2014/01	AGRONOMIA (BACHARELADO)	27	21	8	0	25
2014/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	110	27	107	11	13,54
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2014/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	83	14	110	25	18,82
2014/02	MEDICINA VETERINÁRIA	291	5	314	0	8,92
2014/02	AGRONOMIA (BACHARELADO)	28	6	27	0	18,52
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2015/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	66	23	83	23	28,33

2015/01	MEDICINA VETERINÁRIA	430	155	291	0	5,5
2015/01	AGRONOMIA (BACHARELADO)	33	14	28	0	32,14
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2015/02	AGRONOMIA (BACHARELADO)	29	6	33	0	30,3
2015/02	MEDICINA VETERINÁRIA	407	2	430	0	5,81
2015/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	60	11	66	6	18,33
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2016/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	54	17	60	8	28,85
2016/01	MEDICINA VETERINÁRIA	548	148	407	0	1,72
2016/01	AGRONOMIA (BACHARELADO)	42	18	29	0	17,24
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2016/02	AGRONOMIA (BACHARELADO)	42	3	42	0	7,14
2016/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	51	7	54	9	2,22
2016/02	MEDICINA VETERINÁRIA	523	7	548	0	5,84

<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2017/01	MEDICINA VETERINÁRIA	593	120	523	30	4,06
2017/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	51	19	51	12	17,95
2017/01	AGRONOMIA (BACHARELADO)	40	1	42	0	7,14
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2017/02	MEDICINA VETERINÁRIA	551	6	593	3	7,63
2017/02	AGRONOMIA (BACHARELADO)	37	0	40	0	7,5
2017/02	ENGENHARIA AGRONÔMICA	28	3	30	0	16,67
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	57	15	51	8	2,33
<b>Período Letivo</b>	<b>CURSO</b>	<b>Nº Matrículas</b>	<b>Nº Ingressantes Atuais</b>	<b>Nº Matrículas (Per. Anterior)</b>	<b>Nº concluintes</b>	<b>Taxa de Evasão</b>
2018/01	ENGENHARIA AGRONÔMICA	62	37	28	0	10,71
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO DE EQUINOCULTURA	47	16	57	16	24,39
2018/01	MEDICINA VETERINÁRIA	569	96	551	52	5,21
2018/01	AGRONOMIA (BACHARELADO)	31	1	37	0	18,92

Fonte: Próprio autor

## APÊNDICE G – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS DA SAÚDE

**Tabela 31 - Índice de evasão por período letivo para os cursos da grande área de Ciências da Saúde**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	FARMÁCIA	161	20	172	0	18,02
2012/01	TERAPIA OCUPACIONAL	105	30	91	0	17,58
2012/01	NUTRIÇÃO	171	49	163	0	25,15
2012/01	BIOTECNOLOGIA	59	1	84	0	30,95
2012/01	EDUCAÇÃO FÍSICA	76	25	60	0	15
2012/01	FISIOTERAPIA	123	41	107	0	23,36
2012/01	ENFERMAGEM	85	20	77	0	15,58
2012/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	157	55	110	0	7,27
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	FARMÁCIA	149	5	161	0	10,56
2012/02	TERAPIA OCUPACIONAL	92	1	105	0	13,33
2012/02	NUTRIÇÃO	147	3	171	0	15,79
2012/02	BIOTECNOLOGIA	55	1	59	0	8,47
2012/02	EDUCAÇÃO FÍSICA	65	1	76	0	15,79
2012/02	FISIOTERAPIA	108	0	123	0	12,2
2012/02	ENFERMAGEM	73	0	85	0	14,12
2012/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	160	17	157	0	8,92

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	FARMÁCIA	162	40	149	8	13,48
2013/01	TERAPIA OCUPACIONAL	94	23	92	1	21,98
2013/01	NUTRIÇÃO	161	45	147	0	21,09
2013/01	BIOTECNOLOGIA	37	1	55	0	34,55
2013/01	EDUCAÇÃO FÍSICA	58	12	65	0	29,23
2013/01	FISIOTERAPIA	158	59	108	0	8,33
2013/01	ENFERMAGEM	86	22	73	0	12,33
2013/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	205	75	160	0	18,75
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/02	FARMÁCIA	126	5	162	29	9,02
2013/02	TERAPIA OCUPACIONAL	82	0	94	3	9,89
2013/02	NUTRIÇÃO	142	3	161	4	11,46
2013/02	BIOTECNOLOGIA	37	2	37	1	2,78
2013/02	EDUCAÇÃO FÍSICA (BACHARELADO)	19	10	13	0	30,77
2013/02	EDUCAÇÃO FÍSICA	59	6	58	1	7,02
2013/02	FISIOTERAPIA	140	1	158	0	12,03
2013/02	ENFERMAGEM	73	0	86	0	15,12
2013/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	176	1	205	5	12,5

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	FARMÁCIA	134	33	126	11	12,17
2014/01	TERAPIA OCUPACIONAL	76	22	82	19	14,29
2014/01	NUTRIÇÃO	176	62	142	12	12,31
2014/01	BIOTECNOLOGIA	24	1	37	13	4,17
2014/01	EDUCAÇÃO FÍSICA (BACHARELADO)	22	8	19	0	26,32
2014/01	EDUCAÇÃO FÍSICA	79	30	59	5	9,26
2014/01	FISIOTERAPIA	183	54	140	0	7,86
2014/01	ENFERMAGEM	96	30	73	0	9,59
2014/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	182	66	176	49	8,66
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	FARMÁCIA	102	2	134	19	13,04
2014/02	TERAPIA OCUPACIONAL	70	3	76	1	10,67
2014/02	NUTRIÇÃO	161	5	176	5	8,77
2014/02	BIOTECNOLOGIA	23	1	24	0	8,33
2014/02	EDUCAÇÃO FÍSICA (BACHARELADO)	23	9	22	0	36,36
2014/02	EDUCAÇÃO FÍSICA	75	16	79	9	15,71
2014/02	FISIOTERAPIA	165	2	183	0	10,93
2014/02	ENFERMAGEM	93	1	96	0	4,17

2014/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	149	3	182	11	14,62
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	FARMÁCIA	121	35	102	6	10,42
2015/01	TERAPIA OCUPACIONAL	78	27	70	14	8,93
2015/01	NUTRIÇÃO	201	79	161	20	13,48
2015/01	BIOTECNOLOGIA	11	2	23	11	25
2015/01	EDUCAÇÃO FÍSICA (BACHARELADO)	24	12	23	0	47,83
2015/01	EDUCAÇÃO FÍSICA	85	34	75	9	22,73
2015/01	FISIOTERAPIA	217	74	165	11	7,14
2015/01	ENFERMAGEM	110	36	93	13	7,5
2015/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	155	38	149	22	7,87
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	FARMÁCIA	97	4	121	13	13,89
2015/02	TERAPIA OCUPACIONAL	73	4	78	1	10,39
2015/02	NUTRIÇÃO	164	0	201	14	12,3
2015/02	BIOTECNOLOGIA	7	3	11	6	20
2015/02	EDUCAÇÃO FÍSICA (BACHARELADO)	27	6	24	0	12,5
2015/02	EDUCAÇÃO FÍSICA	77	11	85	2	20,48

2015/02	FISIOTERAPIA	200	0	217	2	6,98
2015/02	ENFERMAGEM	100	0	110	2	7,41
2015/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	138	0	155	6	7,38
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	FARMÁCIA	123	46	97	10	11,49
2016/01	TERAPIA OCUPACIONAL	86	37	73	17	12,5
2016/01	NUTRIÇÃO	193	67	164	21	11,89
2016/01	BIOTECNOLOGIA	1	0	7	5	50
2016/01	EDUCAÇÃO FÍSICA (BACHARELADO)	27	7	27	0	25,93
2016/01	EDUCAÇÃO FÍSICA	58	2	77	9	17,65
2016/01	FISIOTERAPIA	253	85	200	18	7,69
2016/01	ENFERMAGEM	141	47	100	7	-1,08
2016/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	131	37	138	38	6
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	FARMÁCIA	118	6	123	11	0
2016/02	TERAPIA OCUPACIONAL	85	10	86	0	12,79
2016/02	NUTRIÇÃO	186	21	193	8	10,81
2016/02	EDUCAÇÃO FÍSICA (BACHARELADO)	26	3	27	0	14,81
2016/02	EDUCAÇÃO FÍSICA	45	0	58	4	16,67

2016/02	FISIOTERAPIA	238	17	253	13	7,03
2016/02	ENFERMAGEM	128	9	141	13	7,03
2016/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	133	12	131	9	0,82
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	FARMÁCIA	127	34	118	7	16,22
2017/01	TERAPIA OCUPACIONAL	92	20	85	6	8,86
2017/01	NUTRIÇÃO	198	47	186	19	9,58
2017/01	EDUCAÇÃO FÍSICA (BACHARELADO)	50	38	26	5	42,86
2017/01	EDUCAÇÃO FÍSICA	30	2	45	14	9,68
2017/01	FISIOTERAPIA	247	52	238	20	10,55
2017/01	ENFERMAGEM	155	50	128	9	11,76
2017/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	111	27	133	41	8,7
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/02	FARMÁCIA	131	9	127	3	1,61
2017/02	TERAPIA OCUPACIONAL	97	9	92	3	1,12
2017/02	NUTRIÇÃO	199	20	198	3	8,21
2017/02	EDUCAÇÃO FÍSICA (BACHARELADO)	48	6	50	3	10,64
2017/02	BIOMEDICINA	43	10	38	0	13,16
2017/02	EDUCAÇÃO FÍSICA	31	4	30	0	10
2017/02	FISIOTERAPIA	231	12	247	5	9,5

2017/02	ENFERMAGEM	159	15	155	1	6,49
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	99	8	111	5	14,15
2017/02	ODONTOLOGIA	33	3	32	0	6,25
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2018/01	FARMÁCIA	133	30	131	11	14,17
2018/01	TERAPIA OCUPACIONAL	110	26	97	8	5,62
2018/01	NUTRIÇÃO	213	59	199	32	7,78
2018/01	EDUCAÇÃO FÍSICA (BACHARELADO)	67	27	48	7	2,44
2018/01	BIOMEDICINA	91	58	43	0	23,26
2018/01	EDUCAÇÃO FÍSICA	43	17	31	6	-4
2018/01	FISIOTERAPIA	235	52	231	24	11,59
2018/01	ENFERMAGEM	201	64	159	9	8,67
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM ESTÉTICA E COSMÉTICA	93	29	99	24	14,67
2018/01	ODONTOLOGIA	88	56	33	0	3,03

Fonte: Próprio autor

## APÊNDICE H – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS EXATAS E DA TERRA

**Tabela 32 – Índice de evasão por período letivo para os cursos da grande área de Ciências Exatas e da Terra**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	MATEMÁTICA	52	16	51	0	29,41
2012/01	QUÍMICA	59	11	63	0	23,81
2012/01	SISTEMAS DE INFORMAÇÃO	112	21	136	0	33,09
2012/01	FÍSICA	18	3	24	0	37,5
2012/01	CIÊNCIA DA COMPUTAÇÃO	125	46	109	0	27,52
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	MATEMÁTICA	51	6	52	0	13,46
2012/02	QUÍMICA	50	4	59	0	22,03
2012/02	SISTEMAS DE INFORMAÇÃO	93	2	112	0	18,75
2012/02	FÍSICA	14	2	18	0	33,33
2012/02	CIÊNCIA DA COMPUTAÇÃO	110	6	125	0	16,8
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	MATEMÁTICA	56	22	51	1	32
2013/01	QUÍMICA	40	12	50	0	44
2013/01	SISTEMAS DE INFORMAÇÃO	81	13	93	3	24,44
2013/01	FÍSICA	18	7	14	0	21,43
2013/01	CIÊNCIA DA COMPUTAÇÃO	123	39	110	2	22,22
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão

2013/02	MATEMÁTICA	44	1	56	2	20,37
2013/02	QUÍMICA	35	3	40	3	13,51
2013/02	SISTEMAS DE INFORMAÇÃO	80	13	81	2	15,19
2013/02	FÍSICA	15	1	18	1	17,65
2013/02	CIÊNCIA DA COMPUTAÇÃO	103	2	123	5	14,41
2013/02	JOGOS DIGITAIS	37	14	30	0	23,33
2013/02	QUÍMICA INDUSTRIAL	32	3	39	0	25,64
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	MATEMÁTICA	56	19	44	2	11,9
2014/01	QUÍMICA	45	17	35	4	9,68
2014/01	SISTEMAS DE INFORMAÇÃO	72	17	80	10	21,43
2014/01	FÍSICA	19	8	15	3	8,33
2014/01	CIÊNCIA DA COMPUTAÇÃO	136	47	103	2	11,88
2014/01	JOGOS DIGITAIS	60	33	37	0	27,03
2014/01	QUÍMICA INDUSTRIAL	62	33	32	0	9,38
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	MATEMÁTICA	46	1	56	2	16,67
2014/02	QUÍMICA	39	3	45	3	14,29
2014/02	SISTEMAS DE INFORMAÇÃO	67	9	72	4	14,71
2014/02	FÍSICA	18	1	19	1	5,56
2014/02	CIÊNCIA DA COMPUTAÇÃO	128	14	136	4	13,64
2014/02	JOGOS DIGITAIS	56	13	60	0	28,33
2014/02	QUÍMICA INDUSTRIAL	55	1	62	0	12,9

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	MATEMÁTICA	54	18	46	5	12,2
2015/01	QUÍMICA	39	11	39	4	20
2015/01	SISTEMAS DE INFORMAÇÃO	63	13	67	7	16,67
2015/01	FÍSICA	14	3	18	3	26,67
2015/01	CIÊNCIA DA COMPUTAÇÃO	162	54	128	8	10
2015/01	JOGOS DIGITAIS	74	29	56	0	19,64
2015/01	QUÍMICA INDUSTRIAL	76	23	55	0	3,64
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	MATEMÁTICA	47	1	54	3	9,8
2015/02	QUÍMICA	32	2	39	0	23,08
2015/02	SISTEMAS DE INFORMAÇÃO	55	7	63	5	17,24
2015/02	FÍSICA	10	0	14	0	28,57
2015/02	CIÊNCIA DA COMPUTAÇÃO	139	10	162	11	14,57
2015/02	JOGOS DIGITAIS	74	12	74	1	15,07
2015/02	QUÍMICA INDUSTRIAL	71	1	76	0	7,89
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	MATEMÁTICA	46	8	47	6	7,32
2016/01	QUÍMICA	32	11	32	6	19,23
2016/01	SISTEMAS DE INFORMAÇÃO	53	14	55	12	9,3
2016/01	FÍSICA	7	0	10	2	12,5
2016/01	CIÊNCIA DA COMPUTAÇÃO	97	0	139	19	19,17
2016/01	JOGOS DIGITAIS	72	13	74	6	13,24

2016/01	QUÍMICA INDUSTRIAL	68	8	71	6	7,69
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	MATEMÁTICA	42	1	46	2	6,82
2016/02	QUÍMICA	30	2	32	2	6,67
2016/02	SISTEMAS DE INFORMAÇÃO	41	0	53	7	10,87
2016/02	FÍSICA	8	1	7	0	0
2016/02	CIÊNCIA DA COMPUTAÇÃO	88	0	97	4	5,38
2016/02	JOGOS DIGITAIS	51	0	72	4	25
2016/02	QUÍMICA INDUSTRIAL	54	1	68	9	10,17
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	MATEMÁTICA	50	15	42	2	12,5
2017/01	QUÍMICA	33	8	30	2	10,71
2017/01	SISTEMAS DE INFORMAÇÃO	45	12	41	4	10,81
2017/01	FÍSICA	5	0	8	1	28,57
2017/01	CIÊNCIA DA COMPUTAÇÃO	91	17	88	5	10,84
2017/01	JOGOS DIGITAIS	44	1	51	3	10,42
2017/01	QUÍMICA INDUSTRIAL	61	19	54	13	-2,44
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/02	MATEMÁTICA	40	3	50	3	21,28
2017/02	QUÍMICA	22	1	33	7	19,23
2017/02	SISTEMAS DE INFORMAÇÃO	41	0	45	1	6,82
2017/02	FÍSICA	2	0	5	1	50
2017/02	CIÊNCIA DA COMPUTAÇÃO	77	1	91	3	13,64

2017/02	JOGOS DIGITAIS	26	0	44	13	16,13
2017/02	QUÍMICA INDUSTRIAL	42	2	61	16	11,11
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2018/01	MATEMÁTICA	35	8	40	8	15,62
2018/01	QUÍMICA	28	9	22	3	0
2018/01	SISTEMAS DE INFORMAÇÃO	42	9	41	5	8,33
2018/01	FÍSICA	2	0	2	0	0
2018/01	CIÊNCIA DA COMPUTAÇÃO	82	20	77	13	3,12
2018/01	JOGOS DIGITAIS	39	21	26	10	-12,5
2018/01	QUÍMICA INDUSTRIAL	47	14	42	7	5,71

Fonte: Próprio autor

## APÊNDICE I – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS HUMANAS

**Tabela 33 - Índice de evasão por período letivo para os cursos da grande área de Ciências Humanas**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	HISTÓRIA	139	46	126	0	26,19
2012/01	FILOSOFIA	69	26	69	0	37,68
2012/01	PEDAGOGIA	404	189	262	0	17,94
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	HISTÓRIA	132	5	139	0	8,63
2012/02	FILOSOFIA	70	8	69	0	10,14
2012/02	PEDAGOGIA	382	36	404	0	14,36
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	HISTÓRIA	131	33	132	2	24,62
2013/01	FILOSOFIA	74	28	70	2	32,35
2013/01	PEDAGOGIA	414	86	382	6	12,77
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/02	HISTÓRIA	130	16	131	6	8,8
2013/02	GEOGRAFIA	13	3	13	0	23,08
2013/02	FILOSOFIA	74	10	74	1	12,33
2013/02	PEDAGOGIA	390	28	414	29	5,97
2013/02	PSICOLOGIA	58	2	65	0	13,85
2013/02	RELAÇÕES INTERNACIONAIS	7	4	4	0	25
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	HISTÓRIA	145	43	130	13	12,82

2014/01	GEOGRAFIA	18	7	13	0	15,38
2014/01	FILOSOFIA	77	24	74	11	15,87
2014/01	PEDAGOGIA	390	69	390	45	6,96
2014/01	PSICOLOGIA	130	78	58	0	10,34
2014/01	RELAÇÕES INTERNACIONAIS	16	12	7	0	42,86
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	HISTÓRIA	139	12	145	9	6,62
2014/02	GEOGRAFIA	20	3	18	0	5,56
2014/02	FILOSOFIA	73	8	77	3	12,16
2014/02	PEDAGOGIA	327	33	390	79	5,47
2014/02	PSICOLOGIA	145	28	130	0	10
2014/02	RELAÇÕES INTERNACIONAIS	19	6	16	0	18,75
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	HISTÓRIA	167	66	139	21	14,41
2015/01	GEOGRAFIA	24	5	20	0	5
2015/01	FILOSOFIA	84	30	73	11	12,9
2015/01	PEDAGOGIA	332	65	327	43	5,99
2015/01	PSICOLOGIA	244	108	145	0	6,21
2015/01	RELAÇÕES INTERNACIONAIS	25	9	19	0	15,79
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	HISTÓRIA	159	12	167	5	9,26
2015/02	GEOGRAFIA	22	0	24	1	4,35
2015/02	FILOSOFIA	78	5	84	6	6,41
2015/02	PEDAGOGIA	302	35	332	56	3,26
2015/02	PSICOLOGIA	217	0	244	0	11,07
2015/02	RELAÇÕES INTERNACIONAIS	26	6	25	0	20

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	HISTÓRIA	153	37	159	19	17,14
2016/01	GEOGRAFIA	18	0	22	4	0
2016/01	FILOSOFIA	80	21	78	8	15,71
2016/01	PEDAGOGIA	313	58	302	24	8,27
2016/01	PSICOLOGIA	284	81	217	0	6,45
2016/01	RELAÇÕES INTERNACIONAIS	22	0	26	0	15,38
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	HISTÓRIA	152	16	153	7	6,85
2016/02	GEOGRAFIA	14	0	18	4	0
2016/02	FILOSOFIA	74	5	80	5	8
2016/02	PEDAGOGIA	273	21	313	48	4,91
2016/02	PSICOLOGIA	272	1	284	0	4,58
2016/02	RELAÇÕES INTERNACIONAIS	26	9	22	0	22,73
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	HISTÓRIA	150	31	152	16	12,5
2017/01	GEOGRAFIA	9	0	14	3	18,18
2017/01	FILOSOFIA	71	22	74	16	15,52
2017/01	PEDAGOGIA	278	38	273	21	4,76
2017/01	PSICOLOGIA	341	75	272	0	2,21
2017/01	RELAÇÕES INTERNACIONAIS	33	10	26	2	4,17
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/02	HISTÓRIA	129	4	150	11	10,07
2017/02	GEOGRAFIA	7	0	9	2	0
2017/02	FILOSOFIA	62	2	71	6	7,69
2017/02	PEDAGOGIA	227	0	278	41	4,22
2017/02	PSICOLOGIA	331	4	341	0	4,11

2017/02	RELAÇÕES INTERNACIONAIS	36	6	33	0	9,09
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2018/01	HISTÓRIA	135	32	129	21	4,63
2018/01	GEOGRAFIA	2	0	7	2	60
2018/01	FILOSOFIA	65	25	62	17	11,11
2018/01	PEDAGOGIA	225	42	227	28	8,04
2018/01	PSICOLOGIA	357	70	331	30	4,65
2018/01	RELAÇÕES INTERNACIONAIS	42	15	36	4	15,62

Fonte: Próprio autor

## APÊNDICE J – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE LÍNGUÍSTICA, LETRAS E ARTES

**Tabela 34 – Índice de evasão por período letivo para os cursos da grande área de Linguística, Letras e Artes**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	146	40	135	0	21,48
2012/01	TEATRO	59	21	58	1	33,33
2012/01	MÚSICA	98	27	83	0	14,46
2012/01	DANÇA	14	0	17	0	17,65
2012/01	ARTES VISUAIS	12	0	13	0	7,69
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	148	24	146	0	15,07
2012/02	TEATRO	49	1	59	0	18,64
2012/02	MÚSICA	87	0	98	0	11,22
2012/02	DANÇA	14	0	14	0	0
2012/02	ARTES VISUAIS	12	0	12	0	0
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	159	45	148	0	22,97
2013/01	TEATRO	57	24	49	0	32,65
2013/01	MÚSICA	87	25	87	0	28,74
2013/01	DANÇA	10	0	14	0	28,57
2013/01	ARTES VISUAIS	24	13	12	0	8,33

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2013/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	154	19	159	8	10,6
2013/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	17	12	7	0	28,57
2013/02	TEATRO	54	8	57	4	13,21
2013/02	MÚSICA	88	14	87	7	7,5
2013/02	DANÇA	11	4	10	0	30
2013/02	ARTES VISUAIS	23	8	24	2	31,82
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	173	48	154	15	10,07
2014/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	11	1	17	0	41,18
2014/01	TEATRO	58	19	54	10	11,36
2014/01	MÚSICA	87	29	88	18	17,14
2014/01	DANÇA	15	10	11	6	0
2014/01	ARTES VISUAIS	35	15	23	1	9,09
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2014/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	165	19	173	7	12,05
2014/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	15	4	11	0	0

2014/02	TEATRO	47	5	58	2	25
2014/02	MÚSICA	73	2	87	4	14,46
2014/02	DANÇA	16	5	15	1	21,43
2014/02	ARTES VISUAIS	42	9	35	0	5,71
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	158	26	165	14	12,58
2015/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	13	1	15	0	20
2015/01	TEATRO	55	20	47	5	16,67
2015/01	MÚSICA	70	22	73	14	18,64
2015/01	DANÇA	21	9	16	1	20
2015/01	ARTES VISUAIS	45	14	42	2	22,5
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2015/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	153	20	158	13	8,28
2015/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPANHOL	14	0	13	0	-7,69
2015/02	TEATRO	52	8	55	6	10,2
2015/02	MÚSICA	69	12	70	1	17,39
2015/02	DANÇA	23	5	21	0	14,29
2015/02	ARTES VISUAIS	48	6	45	0	6,67
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	150	37	153	20	15,04

2016/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPAÑHOL	12	0	14	0	14,29
2016/01	TEATRO	55	14	52	8	6,82
2016/01	MÚSICA	61	10	69	9	15
2016/01	DANÇA	24	0	23	0	-4,35
2016/01	ARTES VISUAIS	54	12	48	3	6,67
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2016/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	146	15	150	11	5,76
2016/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPAÑHOL	12	0	12	0	0
2016/02	TEATRO	51	5	55	5	8
2016/02	MÚSICA	51	5	61	5	17,86
2016/02	DANÇA	28	5	24	0	4,17
2016/02	ARTES VISUAIS	43	3	54	3	21,57
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	129	24	146	20	16,67
2017/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPAÑHOL	9	0	12	3	0
2017/01	TEATRO	49	9	51	4	14,89
2017/01	MÚSICA	52	17	51	10	14,63
2017/01	DANÇA	26	6	28	3	20
2017/01	ARTES VISUAIS	45	11	43	2	17,07
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão

2017/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	105	1	129	14	9,57
2017/02	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPAÑHOL	8	0	9	0	11,11
2017/02	TEATRO	44	2	49	0	14,29
2017/02	MÚSICA	43	1	52	4	12,5
2017/02	DANÇA	22	0	26	1	12
2017/02	ARTES VISUAIS	38	1	45	7	2,63
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2018/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E INGLÊS	112	38	105	22	10,84
2018/01	LETRAS - HABILITAÇÃO EM PORTUGUÊS E ESPAÑHOL	2	0	8	6	0
2018/01	TEATRO	40	12	44	14	6,67
2018/01	MÚSICA	45	18	43	10	18,18
2018/01	DANÇA	25	8	22	4	5,56
2018/01	ARTES VISUAIS	39	11	38	6	12,5

Fonte: Próprio autor

## APÊNDICE K – ÍNDICE DE EVASÃO POR CURSO DA GRANDE ÁREA DE CIÊNCIAS BIOLÓGICAS

**Tabela 35 - Índice de evasão por período letivo para os cursos da grande área de Ciências Biológicas**

Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2012/01	GESTÃO AMBIENTAL	113	31	143	0	42,66
2012/02	GESTÃO AMBIENTAL	96	9	113	0	23,01
2013/01	GESTÃO AMBIENTAL	77	21	96	0	41,67
2013/02	GESTÃO AMBIENTAL	54	12	77	24	20,75
2014/01	GESTÃO AMBIENTAL	32	2	54	15	23,08
2014/02	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	25	9	17	0	5,88
2014/02	GESTÃO AMBIENTAL	25	4	32	5	22,22

2015/01	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	58	39	25	0	24
2015/01	GESTÃO AMBIENTAL	16	2	25	5	30
2015/02	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	57	11	58	0	20,69
2015/02	GESTÃO AMBIENTAL	4	0	16	10	33,33
2016/01	GESTÃO AMBIENTAL	2	0	4	1	33,33
2016/01	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	85	32	57	0	7,02
2016/02	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	84	9	85	0	11,76
2016/02	GESTÃO AMBIENTAL	1	0	2	1	0

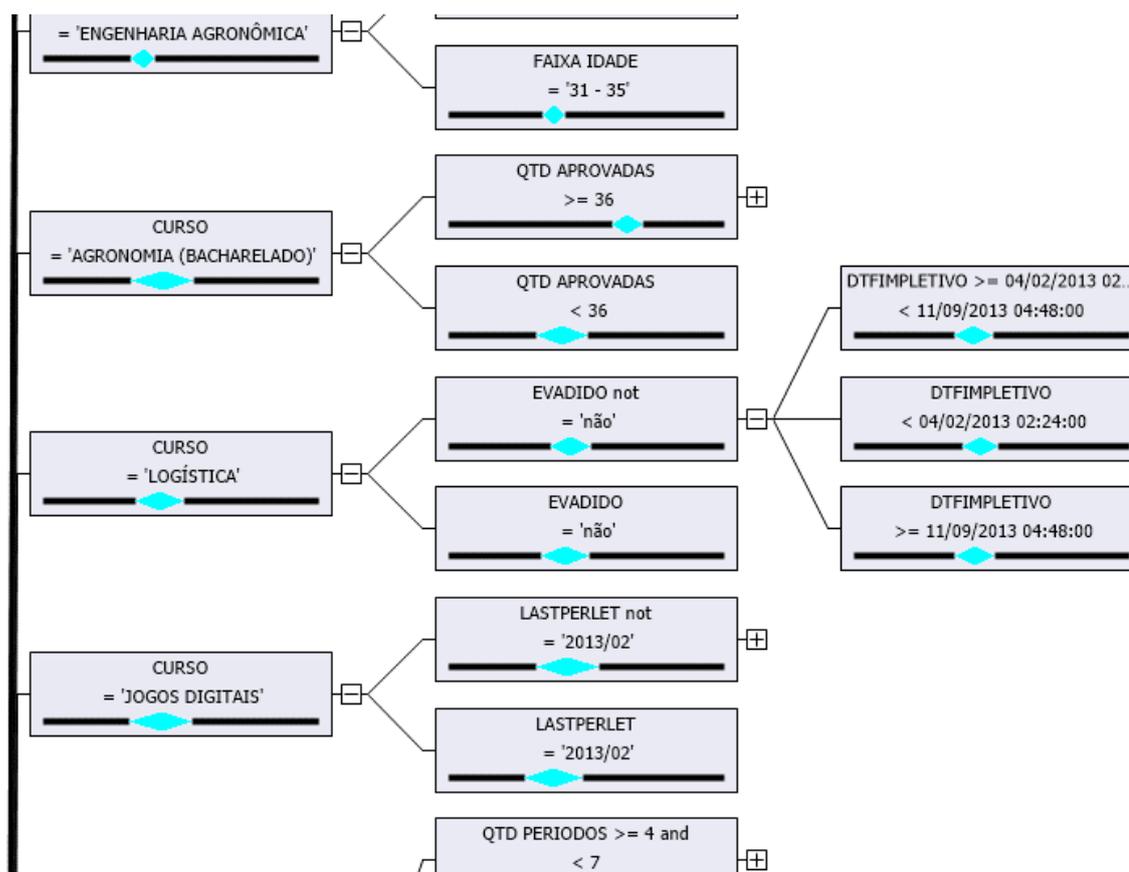
Período Letivo	CURSO	Nº Matrículas	Nº Ingressantes Atuais	Nº Matrículas (Per. Anterior)	Nº Concluintes	Taxa de Evasão
2017/01	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	110	37	84	0	13,1
2017/01	GESTÃO AMBIENTAL	1	0	1	0	0
2017/02	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO AMBIENTAL (EAD)	18	0	28	0	35,71
2017/02	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	105	11	110	0	14,55
2018/01	CIÊNCIAS BIOLÓGICAS (BACHARELADO)	109	22	105	5	13
2018/01	CURSO SUPERIOR DE TECNOLOGIA EM GESTÃO AMBIENTAL (EAD)	16	0	18	1	5,88

Fonte: Próprio autor

## APÊNDICE I – ANÁLISE DO PARÂMETRO TOTAL DE MATRÍCULAS COM A TÉCNICA ÁRVORES DE DECISÃO EM TODOS OS DADOS

Um dos parâmetros que utilizam a evasão como previsão é o número total de matrículas, como mostrado na Figura 44.

Figura 44 - Árvore de decisão para previsão do número total de matrículas (curso de logística)

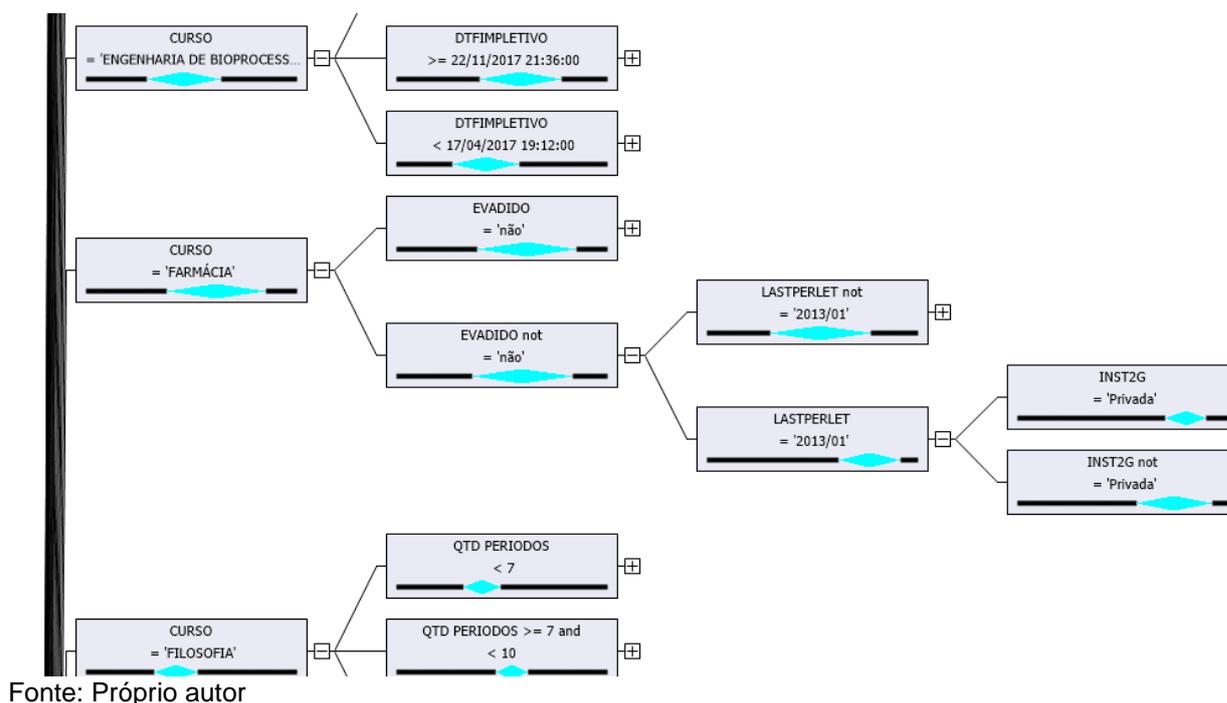


Fonte: Próprio autor

Como mostrado no tópico 4.2.2, para o curso de Logística no 1º semestre de 2013, o índice de evasão gerou uma a seguinte regra na árvore de decisão.

Se o curso for Logística e o aluno estiver evadido e a data final do período letivo for maior que 04/02/2013 e menor que 11/09/2013 o número total de matrículas do aluno equivale a  $21,800 + 0,713 * (\text{quantidade de cancelamentos} - 3,300) + 1,095 * (\text{quantidade de períodos letivos} - 4,200) - 0,028 * (\text{Data final do período letivo} - 41.439,256) + 0,801 * (\text{Quantidade de reprovações} - 4,500) + 0,755 * (\text{quantidade de aprovações} - 12,750) - 2,539 * (\text{quantidade de trancamentos} - 0,150) + (\text{quantidade de dispensas} - 0,750)$ .

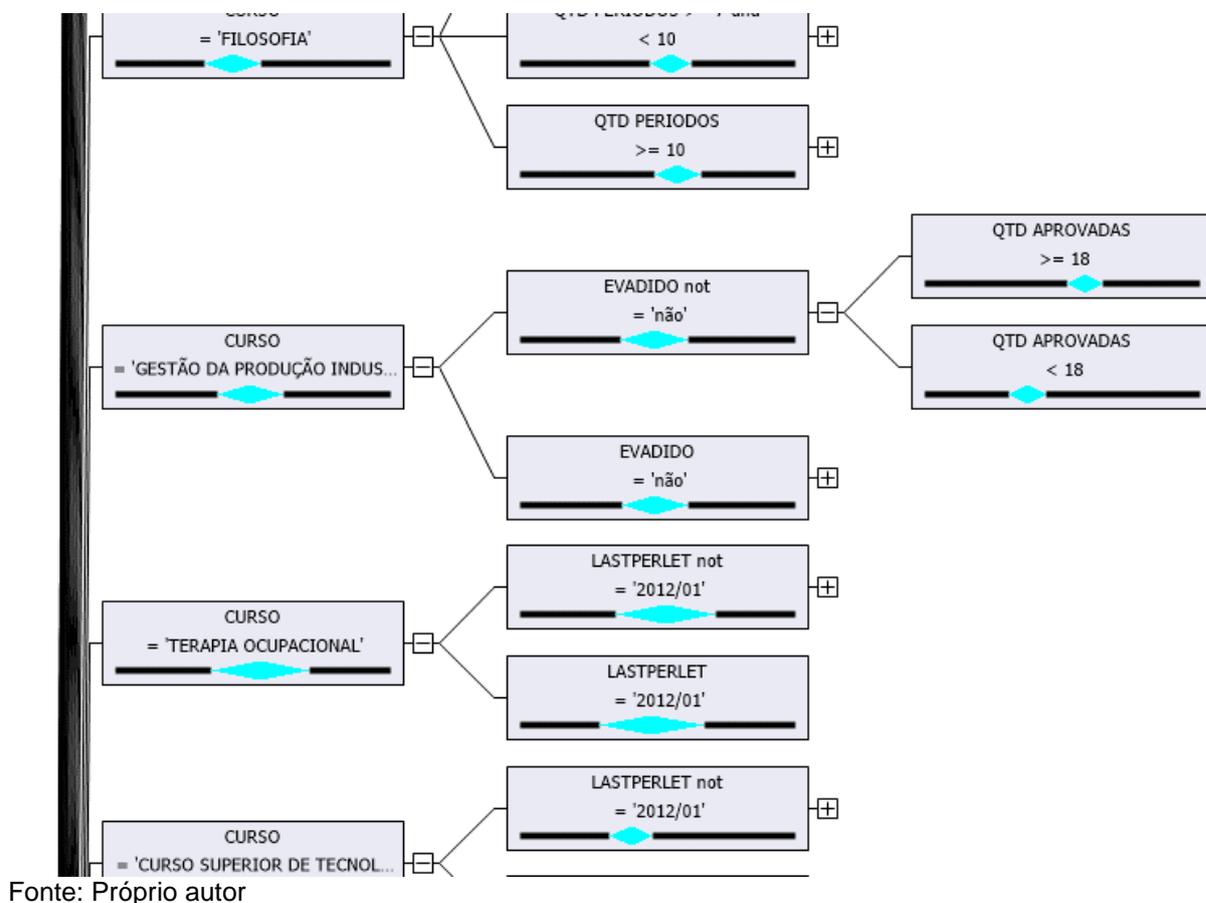
Figura 45 - Árvore de decisão para previsão do número total de matrículas (curso de farmácia)



Ao analisar a árvore, quando o curso for Farmácia, o aluno estiver evadido e o último período letivo do aluno for 1º semestre de 2013 e o aluno concluiu o 2º grau em escola privada o número de matrícula total do aluno equivale a  $63,615 + 0,430 * (\text{quantidade de reprovadas} - 1,154) + 0,775 * (\text{quantidade de aprovações} - 60,615) + 1,106 * (\text{quantidade de cancelamentos} - 1,231) + 0,059 * (\text{data final do período letivo} - 41.629,035)$ .

Outra regra gerada para o curso de farmácia foi que quando o aluno está evadido, o último período letivo cursado foi 1º semestre de 2013, não concluiu o 2º grau em escola privada o número total de matrículas equivale a  $57,250 + 0,986 * (\text{quantidade de disciplinas dispensadas} - 1,583) + 0,135 * (\text{quantidade de disciplinas reprovadas por falta} - 0,417) + 0,946 * (\text{quantidade de disciplinas trancadas} - 1,167) + 0,999 * (\text{quantidade de disciplinas reprovadas} - 3,917) + 0,953 * (\text{quantidade de disciplinas canceladas} - 4,417) - 0,0004 * (\text{data final do período letivo} - 41.760,667) + 0,988 * (\text{quantidade de aprovações} - 46,167)$ .

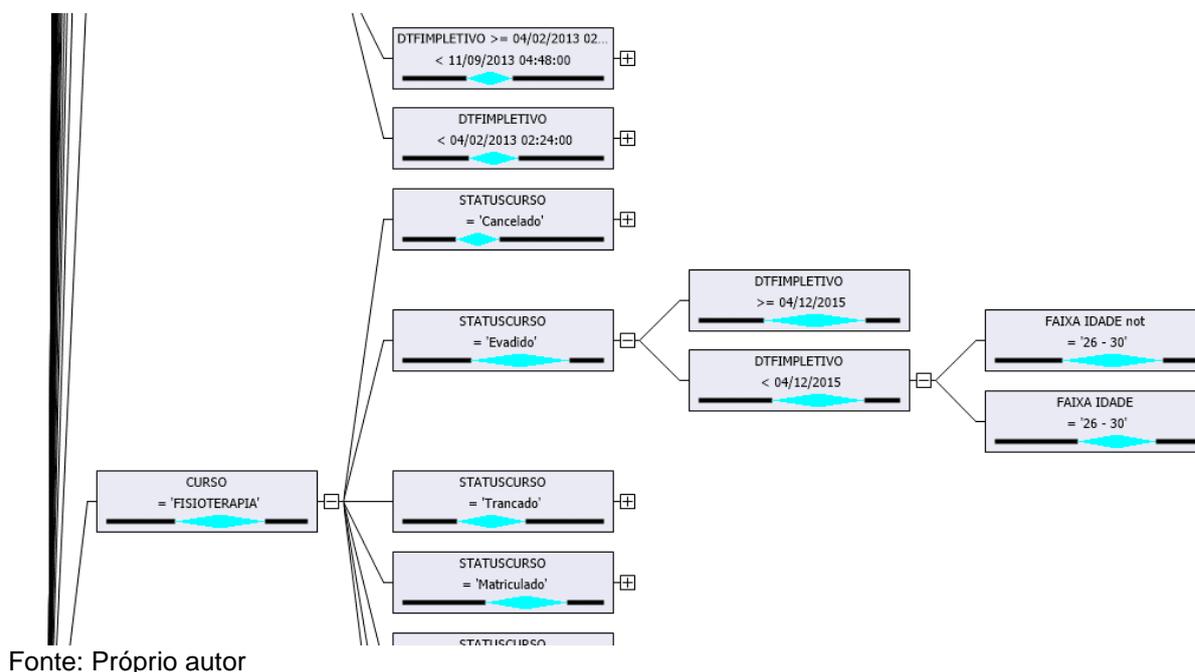
**Figura 46 - Árvore de decisão para previsão do número total de matrículas (curso de gestão de produção industrial)**



No caso da previsão do número de matrícula total para o curso de gestão da produção industrial, mostrado na Figura 46, quando o aluno está evadido e a quantidade de disciplinas aprovadas for maior ou igual a 18, o número de matrículas equivale a  $38,924 + 1,000 * (\text{quantidade de disciplinas reprovadas} - 1,970) + 1,000 * (\text{quantidade de disciplinas aprovadas} - 34,470) + 4,995 * (\text{quantidade de acordos financeiros realizados} - 0,015) + 1,000 * (\text{quantidade de disciplinas trancadas} - 0,682) + 0,999 * (\text{quantidade de disciplinas canceladas} - 1,106) + 1,000 * (\text{quantidade de disciplinas dispensadas} - 0,621)$ .

Outro caso encontrado é quando o curso é Gestão da Produção Industrial, o aluno encontra-se evadido e a quantidade de aprovações for menor que 18, o número de matrículas total equivale a  $14,554 + 0,951 * (\text{quantidade de disciplinas dispensadas} - 1,000) + 0,939 * (\text{quantidade de disciplinas reprovadas} - 2,804) + 0,739 * (\text{quantidade de disciplinas canceladas} - 4,607) + 0,877 * (\text{quantidade de disciplinas trancadas} - 0,607) + 0,960 * (\text{quantidade de disciplinas aprovadas} - 5,071)$ .

**Figura 47 - Árvore de decisão para previsão do número total de matrículas (curso de Fisioterapia)**

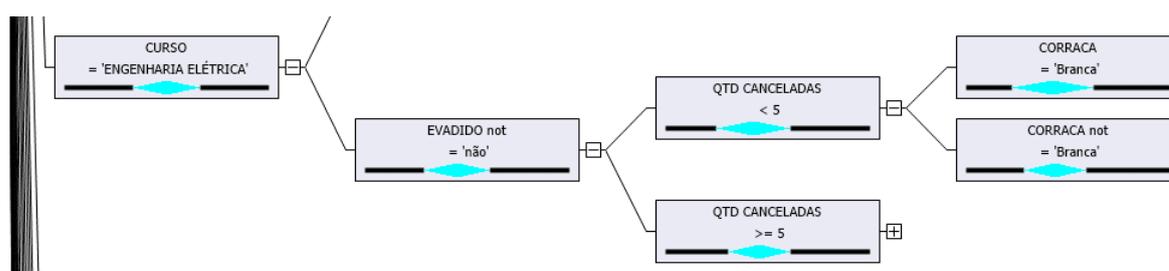


Com relação ao curso de fisioterapia, como demonstrado na Figura 47, o algoritmo mostrou que quando o status do aluno no curso for evadido e o período letivo for maior ou igual a 04/12/2015, o número de matrículas total equivale a  $38,205 + 0,995 * (\text{quantidade de disciplinas reprovadas} - 8,114) + 0,998 * (\text{quantidade de disciplinas dispensadas} - 1,273) + 1,014 * (\text{quantidade de disciplinas trancadas} - 0,455) + 1,000 * (\text{quantidade de disciplinas aprovadas} - 20,614) - 0,012 * (\text{quantidade de meses cursados} - 27,250) + 0,186 * (\text{quantidade de períodos letivos cursados} - 4,636) - 0,012 * (\text{meses entre o início e fim do período letivo} - 27,250) + 0,989 * (\text{quantidade de disciplinas canceladas} - 7,682)$ .

Quando o curso for Fisioterapia e o status do curso for evadido e a data final do período letivo for menor que 04/12/2015 e a faixa de idade não for 26 – 30, o número total de matrículas equivale a  $38,867 + 0,001 * (\text{data final do período letivo} - 41.772,133) + 0,932 * (\text{quantidade de disciplinas reprovadas} - 7,067) + 1,857 * (\text{quantidade de períodos letivos cursados} - 4,000) + 0,954 * (\text{quantidade de disciplinas reprovadas por falta} - 0,533) + 0,945 * (\text{quantidade de disciplinas aprovadas} - 20,533) + 0,965 * (\text{quantidade de disciplinas canceladas} - 9,800) - 0,074 * (\text{quantidade de meses cursados} - 24,333) + 0,448 * (\text{quantidade de disciplinas trancadas} - 0,533) - 0,074 * (\text{meses entre início e fim do período letivo} - 24,333) + 0,919 * (\text{quantidade de disciplinas dispensadas} - 0,400)$ .

Quando o curso for fisioterapia e o status do curso for evadido e a data final do período letivo for menor que 04/12/2015 a faixa de idade for igual a 26 – 30, o número de matrículas total equivale a  $41,400 + 1,264 * (\text{quantidade de disciplinas aprovadas} - 23,500) + 3,727 * (\text{quantidade de períodos cursados} - 5,500) + 0,409 * (\text{quantidade de disciplinas canceladas} - 6,800) - 0,505 * (\text{meses entre o início e fim do período letivo} - 31,800) - 0,505 * (\text{quantidade de meses cursados} - 31,800) + 1,205 * (\text{quantidade de disciplinas reprovadas} - 9,400)$ .

**Figura 48 - Árvore de decisão para previsão do número total de matrículas (curso de Engenharia Elétrica)**



Fonte: Próprio autor

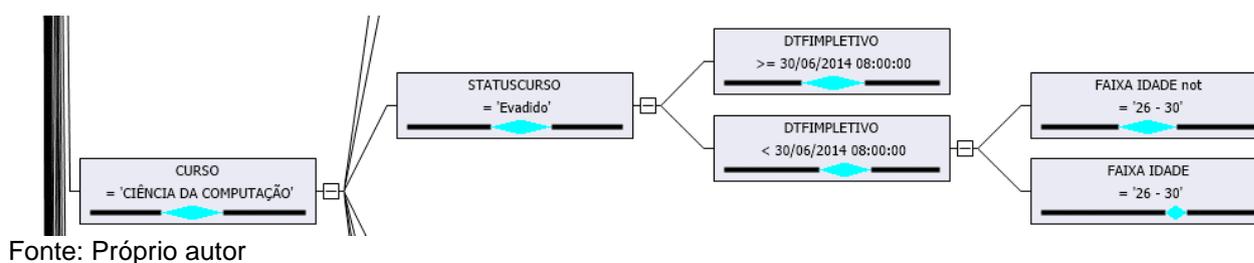
A árvore de decisão quando expandido o nó relacionado ao curso de engenharia elétrica trouxe os seguintes resultados.

Quando o curso for engenharia elétrica e o aluno for evadido e quantidade de cancelamentos for menor que 5 e a cor/raça for branca, o número de matrículas total equivale a  $20,250 + 0,972 * (\text{quantidade de disciplinas dispensadas} - 4,583) - 3,179 * (\text{quantidade de disciplinas reprovadas por falta} - 0,250) + 1,176 * (\text{quantidade de disciplinas aprovadas} - 10,333) + 1,021 * (\text{quantidade de disciplinas trancadas} - 1,000) + 1,395 * (\text{quantidade de disciplinas reprovadas} - 2,500) - 0,823 * (\text{quantidade de períodos letivos cursados} - 4,500)$ .

Quando o curso de engenharia elétrica e o aluno for evadido e a quantidade de cancelamentos for maior ou igual a 5, o número de matrículas total equivale a  $24,279 - 0,009 * (\text{meses entre o início e o fim do período letivo} - 21,765) + 0,981 * (\text{quantidade de disciplinas canceladas} - 7,176) + 1,002 * (\text{quantidade de disciplinas aprovadas} - 9,926) - 0,009 * (\text{quantidade de meses cursados} - 21,765) + 1,011 * (\text{quantidade de}$

disciplinas trancadas - 0,485 ) + 0,138 \* (quantidade de períodos letivos cursados - 3,897 ) + 1,002 \* (quantidade de disciplinas reprovadas - 4,676 ) + 1,000 \* (quantidade de disciplinas dispensadas - 1,971).

**Figura 49 - Árvore de decisão para previsão do número total de matrículas (curso de Ciência da Computação)**



O método de árvore de decisão demonstrou as seguintes regras relacionadas ao número de matrículas em disciplinas quando expandido o nó sobre o curso de ciência da computação.

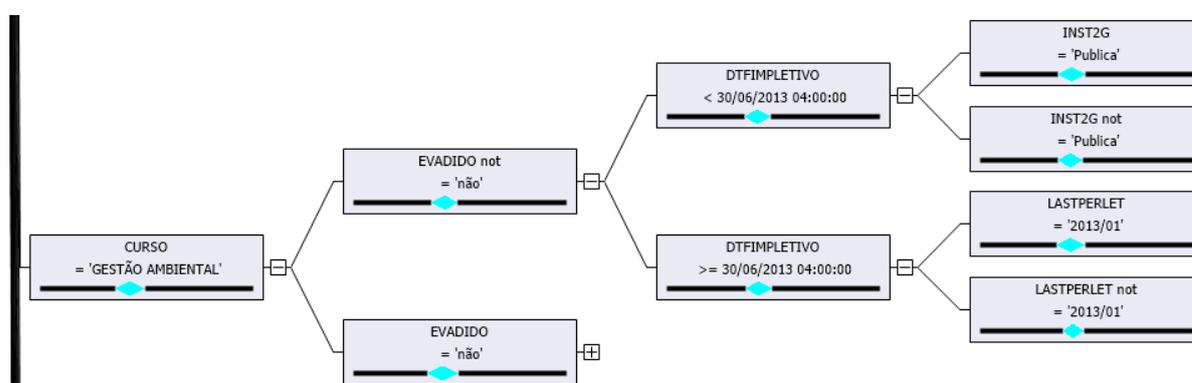
Quando o curso for ciência da computação e o status do aluno no curso for evadido e a data final do período letivo for maior ou igual a 30/06/2014 o número de matrícula total equivale a  $29,833 + 1,233 * (\text{quantidade de disciplinas dispensadas} - 0,400) + 0,999 * (\text{quantidade de disciplinas aprovadas} - 16,433) + 1,006 * (\text{quantidade de disciplinas reprovadas} - 7,633) + 1,025 * (\text{quantidade de disciplinas trancadas} - 0,100) + 0,985 * (\text{quantidade de disciplinas canceladas} - 5,200)$ .

Quando o curso for ciência da computação e o status do aluno no curso for evadido e a data final do período letivo for menor que 30/06/2014 e a faixa de idade não for entre 26 a 30 anos, o número de matrículas total equivale a  $28,300 + 0,539 * (\text{quantidade de disciplinas reprovadas} - 5,800) + 0,588 * (\text{quantidade de disciplinas aprovadas} - 17,400) - 0,195 * (\text{quantidade de disciplinas dispensadas} - 0,400) + 1,241 * (\text{quantidade de períodos cursados} - 6,600) - 0,020 * (\text{quantidade de disciplinas trancadas} - 0,400) + 0,015 * (\text{meses entre o início e fim do período letivo} - 41,600) + 0,015 * (\text{quantidade de meses cursados} - 41,600) + 0,662 * (\text{quantidade de disciplinas canceladas} - 3,500) + 0,00003 * (\text{data de fim do período letivo} - 41.529,200) - 0,102 * (\text{quantidade de disciplinas reprovadas por falta} - 0,800)$ .

Quando o curso for ciência da computação e o status do aluno no curso for evadido e a data do período letivo for menor que 30/06/2014 e a faixa de idade for entre 26 e 30 anos, o número de matrículas total equivale a  $43,455 + 0,928 * (\text{quantidade de disciplinas aprovadas} - 36,182) + 0,788 * (\text{quantidade de disciplinas$

canceladas - 0,545 ) + 0,0003 \* (data final do período letivo -41.555,364) + 0,853 \* (quantidade de disciplinas dispensadas - 0,636 ) + 0,027 \* (quantidade de meses cursados - 62,818 ) + 0,942 \* (quantidade de disciplinas reprovadas - 6,091 ) -0,194 \* (quantidade de períodos cursados - 10,273 ) - 0,348 \* (quantidade de disciplinas reprovadas por falta - 0,545 ) + 0,027 \* (meses entre o início e fim do período letivo - 62,818).

**Figura 50 - Árvore de decisão para previsão do número total de matrículas (curso de Gestão Ambiental)**



O nó relacionado ao curso de gestão ambiental mostrado na Figura 50, para a variável número de matrículas total gerou as seguintes regras.

Quando o curso for gestão ambiental e o aluno for evadido e a data final do período letivo for menor que 30/06/2013 e o aluno concluiu o segundo grau em escola pública o número de matrículas total pode ser representado por  $20,929 - 0,017 * (\text{meses entre o início e fim do período letivo} - 30,214) + 0,715 * (\text{quantidade de disciplinas aprovadas} - 16,214) - 0,017 * (\text{quantidade de meses cursadas} - 30,214) - 0,0002 * (\text{data final do período letivo} - 41.330,514) + 0,239 * (\text{quantidade de disciplinas dispensadas} - 0,143) + 0,292 * (\text{quantidade de períodos cursados} - 4,500) + 0,730 * (\text{quantidade de disciplinas reprovadas} - 1,929)$ .

Quando o curso for gestão ambiental e o aluno for evadido e a data final do período letivo for menor que 30/06/2013 e o aluno não concluiu o segundo grau em escola pública, o número de matrículas total pode ser representado por  $20,083 + 0,979 * (\text{quantidade de disciplinas reprovadas} - 2,500) + 0,0006 * (\text{quantidade de meses cursados} - 28,667) + 0,0006 * (\text{meses entre o início e fim do período letivo} - 28,667) + 1,007 * (\text{quantidade de disciplinas dispensadas} - 0,167) + 0,988 * (\text{quantidade de disciplinas trancadas} - 1,417) + 0,977 * (\text{quantidade de disciplinas canceladas} - 1,667)$

) + 0,033 \* (quantidade de períodos cursados - 4,583 ) + 0,983 \* (quantidade de disciplinas aprovadas - 14,333 ) - 0,0002 \* (data final do período letivo - 41.322,037).

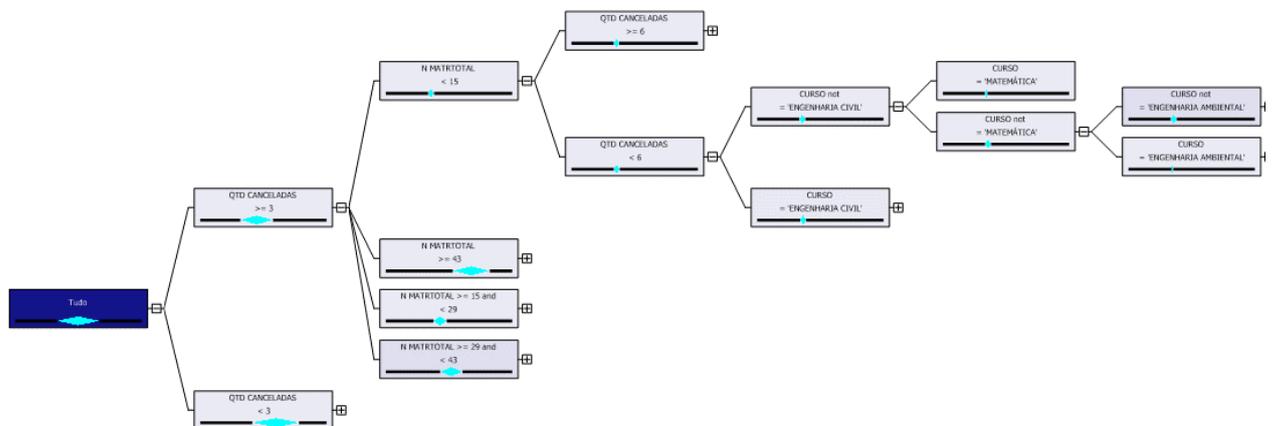
Quando o curso for gestão ambiental e o aluno for evadido e a data final do período letivo for maior ou igual a 30/06/2013 e o último período letivo for 1º semestre de 2013, o número de matrículas total pode ser representado por  $20,357 + 0,0006 * (data\ final\ do\ período\ letivo - 41.691,143 ) + 0,992 * (quantidade\ de\ disciplinas\ trancadas - 1,071 ) + 1,006 * (quantidade\ de\ disciplinas\ dispensadas - 0,500 ) + 0,037 * (quantidade\ de\ períodos\ letivos\ cursados - 5,357 ) + 0,455 * (quantidade\ de\ disciplinas\ canceladas - 0,071 ) + 0,932 * (quantidade\ de\ disciplinas\ reprovadas - 0,214 ) - 0,0007 * (quantidade\ de\ meses\ cursados - 33,643 ) - 0,0007 * (meses\ entre\ o\ início\ e\ fim\ do\ período\ letivo - 33,643 ) + 0,994 * (quantidade\ de\ disciplinas\ aprovadas - 18,500)$ .

Quando o curso for gestão ambiental e aluno for evadido e a data final do período letivo for maior ou igual a 30/06/2013 e o último período letivo não for o 1º semestre de 2013, o número total de matrículas pode ser representado por  $21,933 + 0,990 * (quantidade\ de\ disciplinas\ aprovadas - 18,533 ) + 1,007 * (quantidade\ de\ disciplinas\ canceladas - 1,333 ) + 0,985 * (quantidade\ de\ disciplinas\ trancadas - 0,200 ) - 0,00004 * (data\ final\ do\ período\ letivo - 41.918,284 ) + 0,016 * (quantidade\ de\ disciplinas\ reprovadas\ por\ falta - 0,067 ) + 0,982 * (quantidade\ de\ disciplinas\ reprovadas\ por\ nota - 1,867 ) + 0,051 * (quantidade\ de\ períodos\ cursados - 4,800)$ .

## APÊNDICE J – ANÁLISE DO PARÂMETRO QUANTIDADE DE APROVAÇÕES APLICADA COM TÉCNICAS DE ÁRVORE DE DECISÃO NOS DADOS RELACIONADOS AOS CASOS DE EVASÃO

Ao minerar configurando a ferramenta para prever a quantidade de aprovações dos evadidos e destacando as principais regras geradas, utilizando como critério o maior número de casos foi gerado a árvore de decisão segundo a Figura 51.

Figura 51 - Árvore de decisão gerada para os casos de evasão ao prever a quantidade de aprovações



Fonte: Próprio autor

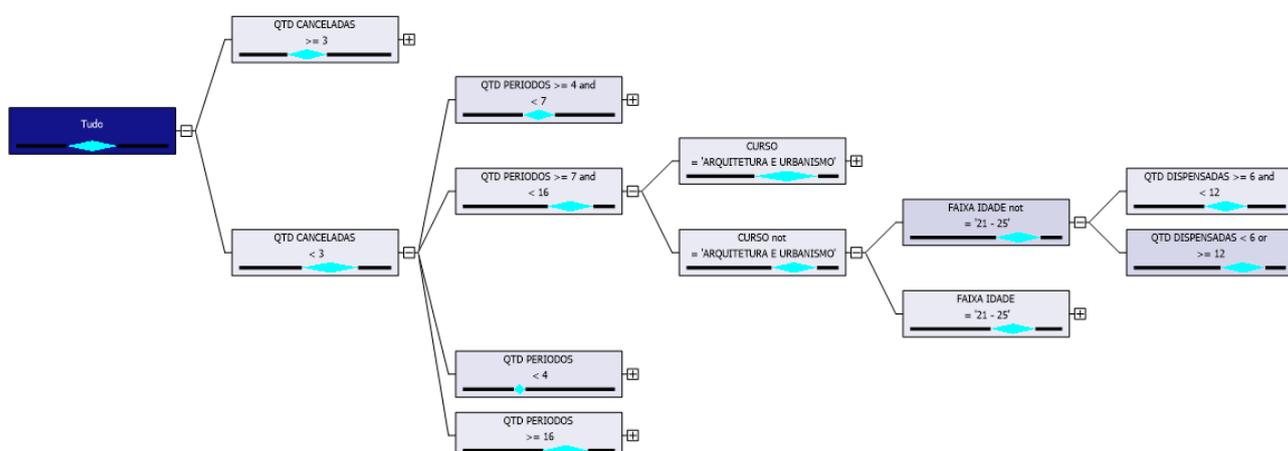
Dentre as regras foi destacada:

- Quando a quantidade de cancelamentos é maior ou igual a 6 e o número total de matrículas for menor que 15 e a área não é ciências biológicas, com 1.134 casos computados, a fórmula para prever a quantidade de aprovações é equivalente a  $1,419 - 0,999 * (\text{quantidade de disciplinas reprovadas} - 1,119) + 0,999 * (\text{número total de matrículas} - 9,439) - 0,999 * (\text{quantidade de disciplinas trancadas} - 0,114) - 0,998 * (\text{quantidade de disciplinas canceladas} - 6,742) - 0,999 * (\text{quantidade de dispensas} - 0,044)$ .
- Quando a quantidade de cancelamentos é maior ou igual a 6 e o número de matrículas total é menor que 15 e a área é ciências biológicas, com 17 casos computados, a quantidade de disciplinas aprovadas equivale a  $1,235 - 1,110 * (\text{quantidade de disciplinas canceladas} - 6,294) - 1,147 * (\text{quantidade de disciplinas reprovadas} - 2,294) + 1,042 * (\text{número de matrículas total} - 9,941)$ .

- Quando a quantidade de cancelamentos é maior ou igual a 3 e menor que 6 e o número de matrículas total é menor que 15 e o curso é engenharia civil, com 71 casos computados, a quantidade de disciplinas aprovadas equivale a  $2,324 - 0,911 * (\text{quantidade de disciplinas reprovadas} - 2,690) + 0,900 * (\text{número total de matrículas} - 10,380) - 0,850 * (\text{quantidade de disciplinas dispensadas} - 0,338) - 0,595 * (\text{quantidade de disciplinas canceladas} - 4,803)$ .
- Quando a quantidade de cancelamentos é maior ou igual a 3 e menor que 6 e o número de matrículas total é menor que 15 e o curso é matemática, com 16 casos computados, a quantidade de disciplinas aprovadas equivale a  $0,375 + 0,282 * (\text{número total de matrículas} - 7,813) + 0,518 * (\text{quantidade de disciplinas canceladas} - 4,688) - 0,082 * (\text{quantidade de disciplinas canceladas} - 1,813) - 0,138 * (\text{quantidade de disciplinas dispensadas} - 0,625)$ .
- Quando a quantidade de cancelamentos é maior ou igual a 3 e menor que 15 e o curso é engenharia ambiental, com 21 casos computados, a quantidade de disciplinas aprovadas equivale a  $0,286 + 0,583 * (\text{quantidade de disciplinas reprovadas por falta} - 0,095) - 0,161 * (\text{quantidade de disciplinas reprovadas} - 1,524) + 0,200 * (\text{número total de matrículas} - 6,524) - 0,267 * (\text{quantidade de disciplinas canceladas} - 4,571)$ .

Outro ramo da árvore que apresentou uma quantidade alta de casos, com relação a previsão da quantidade de aprovações dos alunos evadidos é a apresentada na Figura 52.

Figura 52 - Árvore de decisão dos casos evadidos na previsão da quantidade de aprovações



Fonte: Próprio autor

Da árvore pode-se destacar as seguintes regras:

- Quando a quantidade de cancelamentos é menor que 3 e a quantidade de períodos é maior ou igual a 7 e menor que 16 e o curso for Arquitetura e Urbanismo, com 22 casos computados, a quantidade de disciplinas aprovadas equivale a  $31,273 - 0,973 * (\text{quantidade de disciplinas reprovadas} - 5,455) - 1,321 * (\text{quantidade de disciplinas dispensadas} - 1,545) - 0,978 * (\text{quantidade de disciplinas canceladas} - 0,682) - 0,993 * (\text{quantidade de disciplinas trancadas} - 1,955) + 0,976 * (\text{número total de matrículas} - 41,773)$ .
- Quando a quantidade de cancelamentos é menor que 3 e a quantidade de períodos é maior ou igual a 7 e menor que 16 e o curso não é arquitetura e urbanismo e a faixa de idade não está entre 21 e 25 anos e a quantidade de dispensas é menor que 6 ou maior igual a 12, com 1.110 casos, a quantidade de disciplinas aprovadas equivale a  $37,377 + 0,997 * (\text{número total de matrículas} - 44,503) - 0,997 * (\text{QTD\_REPROVADAS} - 2,656) - 0,998 * (\text{quantidade de disciplinas dispensadas} - 3,646) - 0,993 * (\text{quantidade de disciplinas trancadas} - 0,376) - 1,066 * (\text{quantidade de disciplinas canceladas} - 0,365)$ .

Quando a quantidade de cancelamentos é menor que 3 e a quantidade de períodos é maior ou igual a 7 e menor que 16 e o curso não é arquitetura e urbanismo e a faixa de idade está entre 21 e 25 anos, com 110 casos computados, a quantidade de disciplinas aprovadas equivale a  $33,100 - 1,186 * (\text{quantidade de disciplinas}$

canceladas - 0,573 ) -1,060 \* (quantidade de disciplinas dispensadas - 0,482 ) -1,001 \* (quantidade de disciplinas reprovadas - 3,345 ) - 1,001 \* (quantidade de disciplinas trancadas - 1,609 ) + 0,982 \* (número total de matrículas - 39,545).

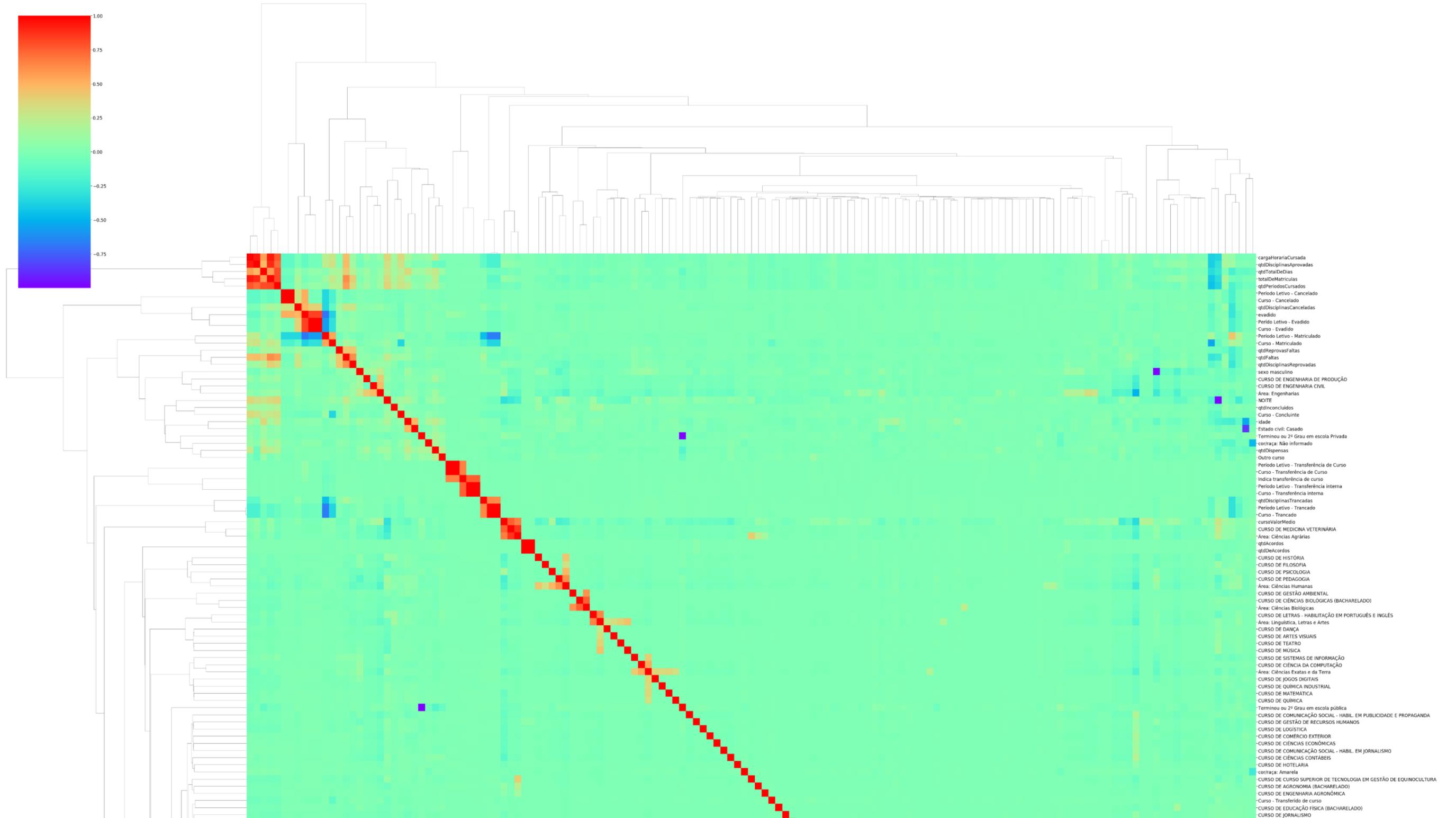
## APÊNDICE K – DESCRIÇÃO DOS ATRIBUTOS USADOS PELA FERRAMENTA DE MINERAÇÃO DE DADOS (ANALISYS SERVICE)

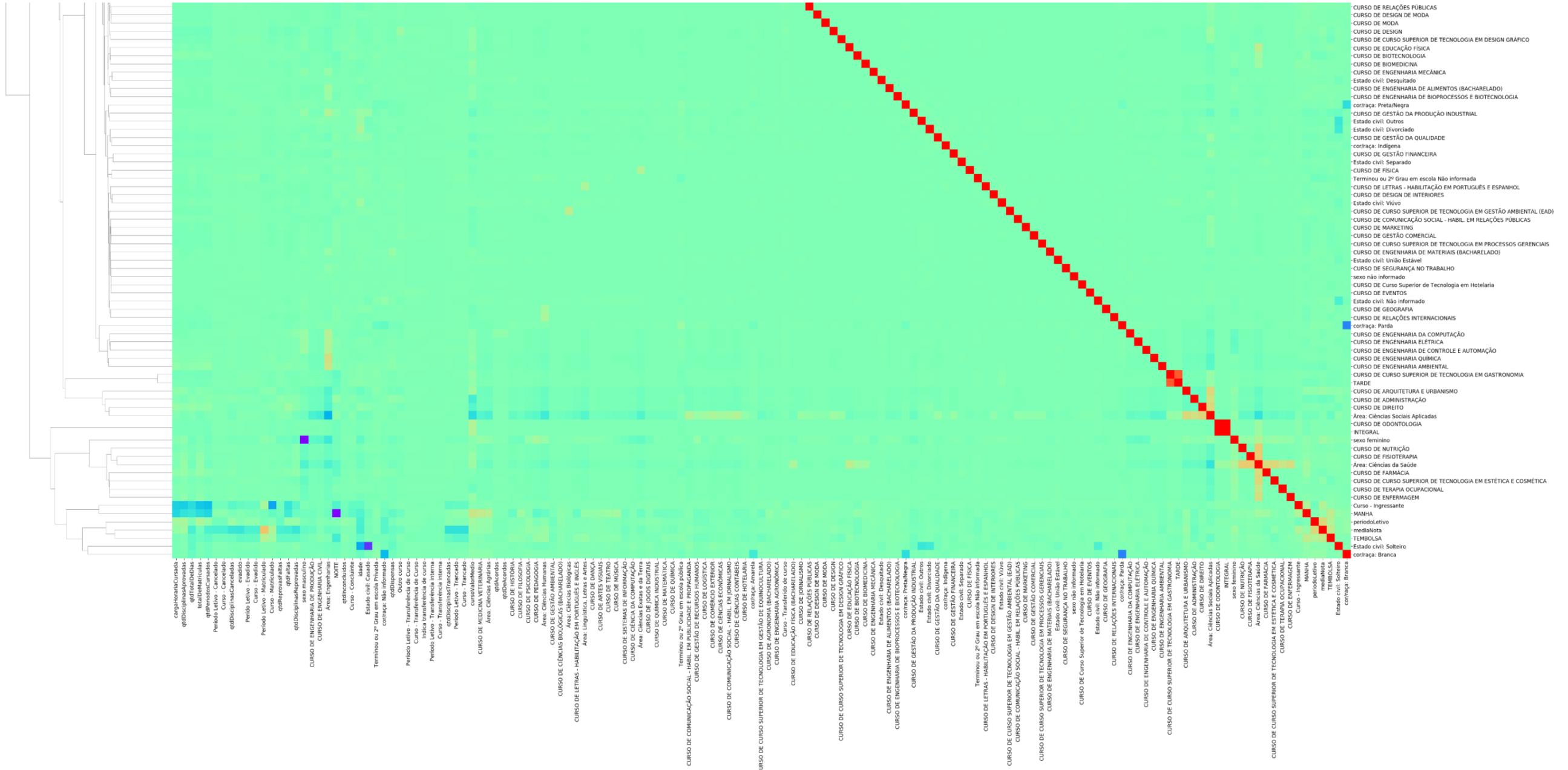
**Quadro 4 - Descrição dos atributos usados na mineração de dados**

Nome usado pelo Analise Service	Descrição do atributo
AREA	Área
CODCURSO	Código do Curso
CORRACA	Cor e Raça
CURSO	Curso
CURSOVALORMEDIO	Valor médio do curso
DESCSTATUSPLETIVO	Descrição do Status no Período Letivo
DTFIMPLETIVO	Data final do período letivo
ESTADOCIVIL	Estado Civil
EVADIDO	Evadido
FAIXA IDADE	Faixa de idade
id	identificação do registro
INST2G	Instituição que concluiu o 2º Grau
LASTPERLET	Último período letivo cursado
MESESENTREDTINICIODTFIM	Meses entre data de início e fim do período letivo
N MATRTOTAL	Número total de matrículas cursadas
PERLET_INGRESSO	Périodo letivo que o discente ingressou
QTD APROVADAS	Quantidade de disciplinas aprovadas
QTD CANCELADAS	Quantidade de disciplinas reprovadas
QTD DISPENSADAS	Quantidade de disciplinas dispensadas
QTD PERIODOS	Quantidade de períodos cursados
QTD REPFALTAS	Quantidade de disciplinas reprovadas por falta
QTD REPROVADAS	Quantidade de disciplinas reprovadas por nota
QTD TRANCADAS	Quantidade de disciplinas trancadas
QTDACORDO	Quantidade de acordos financeiros realizados
QUANTIDADEMESES	Quantidade de meses cursados
SEXO	Sexo do discente
STATUSCURSO	Status do discente no curso
TEMBOLSA	Verificação se já teve bolsa de estudos
TURNO	Turno

# APÊNDICE L – DENDOGRAMA

Figura 53 - Dendograma





Fonte: Próprio autor

## ANEXO A – LEIAUTE DE IMPORTAÇÃO DOS DADOS DO CENSO PARA O INEP (VERSÃO 2017)

Campo	Situação de vínculo do aluno ao curso
Tamanho	1
Fixo/Variável	Fixo
Tipo	Numérico
Preenchimento	Obrigatório
1ª Etapa de Validação	<p>Valores válidos:  2 – Cursando  3 – Matrícula trancada  4 – Desvinculado do curso  5 - Transferido para outro curso da mesma IES  6 – Formado  7 – Falecido</p> <p>Se a Situação do vínculo for informada com algum valor diferente do especificado, deve aparecer a mensagem: <b>"Preencher com um dos valores: 2,3,4,5,6,7"</b></p> <p>Se a situação do vínculo do aluno for "6 - Formado" e o aluno tem idade maior ou igual a 15 anos com relação à data de referência do Censo (31/12/2016), deverá apresentar a mensagem: <b>"Idade do aluno formado não deve ser inferior a 15 anos."</b></p>
2ª Etapa de Validação	<p>Se informada a opção 6 – Formado, quando o curso informado for do tipo Área Básica do curso, deverá apresentar a mensagem: <b>"A opção "Formado" não pode ser informada quando o curso for do tipo Área Básica do curso."</b></p> <p>Se um vínculo em uma IES possuir situação do aluno igual a "5 - transferido para outro curso da mesma IES" e não possuir outro vínculo com o curso de origem preenchido, deverá apresentar a mensagem: <b>"Existe um aluno com situação de vínculo igual a Transferido, portanto é necessário que exista outro vínculo com o curso origem preenchido."</b></p> <p>---Para Universidade Federal:</p> <p><b>"As situações Cursando e Matrícula Trancada não poderão ser informadas no vínculo do aluno no 2º semestre, se o curso estiver com situação de funcionamento extinto no e-MEC no ano de referência do Censo."</b></p> <p>Quando informada a situação de vínculo "Transferido", onde o curso origem for um curso BI/LI, não pode ser vinculado a um curso terminalidade correspondente à origem. Quando o curso origem for terminalidade, não pode ser informado o BI/LI correspondente. Mensagem: <b>"Não pode haver transferência de curso BI/LI para suas terminalidades, nem das terminalidades para o BI/LI."</b></p> <p>Se informado situação de vínculo "Formado" no vínculo no 1º semestre e possui vínculo no segundo semestre do mesmo curso, apresentar a mensagem <b>"Como o aluno possui o vínculo com situação Formado no Primeiro semestre, não é possível ter vínculo com o mesmo curso no Segundo semestre."</b></p> <p>Se informado situação de vínculo "Falecido" em algum vínculo no 1º semestre e possui vínculo no segundo semestre, apresentar a mensagem <b>"Como o aluno possui o vínculo com situação Falecido no Primeiro semestre, não é possível ter vínculo com qualquer curso no Segundo semestre."</b></p> <p>Se o aluno, em uma IES, possuir um vínculo com um curso ABI e um outro com um curso vinculado à esta ABI, e em ambos os vínculos a situação seja 2- Cursando ou 3 - Matrícula Trancada, no mesmo semestre de referência, apresentar mensagem <b>"Aluno não pode estar com mais de um vínculo ativo (na situação de Cursando ou Matrícula Trancada) em ABI e curso vinculado a mesma ABI simultaneamente."</b></p> <p>---Para demais IES: <b>"As situações Cursando e Matrícula trancada não poderão ser informadas quando o curso estiver com situação de funcionamento extinto no e-MEC no ano de referência do Censo."</b></p> <p><b>"As situações Cursando e Matrícula trancada não poderão ser informadas quando o curso estiver com situação de funcionamento extinto no e-MEC no ano de referência do Censo."</b></p>

## ANEXO B – EXEMPLO DE FUNCIONAMENTO DO ALGORITMO DE ÁRVORE DE DECISÃO

Para entender o funcionamento do algoritmo será apresentado a seguir um exemplo que já foi mostrado nos trabalhos de Quinlan (1993), Ricco (1997) e Carvalho (2002).

No exemplo em questão é apresentado um conjunto de dados compostos por 14 registros constituídos de 4 atributos sendo eles o tempo, a temperatura, a umidade relativa e o vento, onde o objetivo da análise é prever o atributo jogar tênis que apresenta duas modalidades {sim, não}, conforme tabela abaixo:

**Tabela 36 - Dados de exemplo para explicação do método de árvore de decisão**

Registro	Tempo	Temperatura (°F)	Umidade Relativa(%)	Vento	Jogar
1	ensolarado	quente	alta	fraco	não
2	ensolarado	quente	alta	forte	não
3	encoberto	quente	alta	fraco	sim
4	chuvoso	suave	alta	fraco	sim
5	chuvoso	frio	normal	fraco	sim
6	chuvoso	frio	normal	forte	não
7	encoberto	frio	normal	forte	sim
8	ensolarado	suave	alta	fraco	não
9	ensolarado	frio	normal	fraco	sim
10	chuvoso	suave	normal	fraco	sim
11	ensolarado	suave	normal	fraco	sim
12	encoberto	suave	alta	forte	sim
13	encoberto	quente	normal	fraco	sim
14	chuvoso	suave	alta	forte	não

Fonte: QUINLAN, J. R. **C4.5**: Programs form Machine Learning. San Mateo, CA, USA: Morgan Kaufmann Publishers, 1993

Primeiramente deve-se dentro dos 14 registros que denominarei de S identificar a quantidade de valores positivos (sim) e valores negativos (não) no atributo a ser previsto, no nosso caso Jogar.

Contabiliza-se 9 exemplos positivos e 5 negativos [9+, 5-], onde a entropia do conjunto S é mostrado abaixo:

$$Entropia ([9+, 5 -]) = -\left(\frac{9}{14} \log_2 \frac{9}{14} - \frac{5}{14} \log_2 \frac{5}{14}\right) = 0.940 \quad (5)$$

Após essa primeira análise, o próximo passo é definir qual atributo deve compor o próximo nó e para isso é necessário calcular o ganho de informação Ganho para cada atributo (Tempo, Temperatura, Umidade, Vento) e selecionar aquele cujo ganho é mais alto.

Para calcular Ganho do atributo vento é preciso identificar quais são seus valores possíveis, no caso (fraco, forte) e identificar quantos casos de vento fraco encontram-se no conjunto de jogar = sim, quantos casos no jogar = não e para o vento forte identificar quantos casos encontram-se no jogar = sim e jogar = não.

Valores possíveis de Vento (fraco, forte).

$S = [9+, 5-]$

Vento fraco =  $[6+, 2-]$

Vento forte =  $[3+, 3-]$

A aplicação da fórmula da Ganho fica assim:

$$\begin{aligned}
 \text{Gain}(S, \text{Vento}) &\equiv \text{Entropia}(S) - \sum_{V \in \{\text{fraco}, \text{forte}\}} \frac{S_v}{S} \text{Entropia}(S_v) & (6) \\
 &= 0.940 - \frac{8}{14} \text{Entropia}(\text{fraco}) - \frac{6}{14} \text{Entropia}(\text{forte}) \\
 &= 0.940 - \frac{8}{14} \left( \frac{6}{8} \log_2 \left( \frac{6}{8} \right) - \frac{-2}{8} \log_2 \left( \frac{-2}{8} \right) \right) - \\
 &\quad \frac{6}{14} \left( \frac{3}{6} \log_2 \left( \frac{3}{6} \right) - \frac{-3}{6} \log_2 \left( \frac{-3}{6} \right) \right) \\
 &= 0.940 - \frac{8}{14} 0.811 - \frac{6}{14} 1.00 = 0,048
 \end{aligned}$$

Aplicando Ganho para todos os campos os resultados são:

$\text{Gain}(S, \text{Vento}) = 0,048$

$\text{Gain}(S, \text{Temperatura}) = 0,029$

$\text{Gain}(S, \text{Umidade}) = 0,151$

$\text{Gain}(S, \text{Tempo}) = 0,246$

Ou seja, o atributo Tempo fornece a melhor previsão para atributo alvo que é Jogar, logo será a raiz da árvore.

Tempo possui 3 resultados possíveis (ensolarado, encoberto, chuvoso) e para os casos de jogar colocado acima como  $S = [9+, 5-]$ , ou seja, 9 sim e 5 não onde temos a seguinte divisão para o atributo tempo:

- Ensolarado  $[2+, 3-] \rightarrow$  (2 sim e 3 não)
  - Umidade alta =  $[0+, 3-] \rightarrow$  (3 não)
  - Umidade normal =  $[2+, 0-] \rightarrow$  (2 sim)
  - Temperatura quente =  $[0+, 2-] \rightarrow$  (2 não)
  - Temperatura suave =  $[1+, 1-] \rightarrow$  (1 sim e 1 não)
  - Temperatura frio =  $[1+, 0-] \rightarrow$  (1 sim)
  - Vento fraco =  $[1+, 2-] \rightarrow$  (1 sim e 2 não)
  - Vento forte =  $[0+, 1-] \rightarrow$  (1 não)
- Encoberto  $[4+, 0-] \rightarrow$  (4 sim)
- Chuvoso  $[3+, 2-] \rightarrow$  (3 sim e 2 não)
  - Umidade alta =  $[1+, 1-] \rightarrow$  (1 sim e 1 não)
  - Umidade normal =  $[2+, 1-] \rightarrow$  (2 sim e 1 não)
  - Temperatura quente =  $[0+, 0-] \rightarrow$  não possui resultado
  - Temperatura suave =  $[2+, 1-] \rightarrow$  (2 sim e 1 não)
  - Temperatura frio =  $[1+, 1-] \rightarrow$  (1 sim e 1 não)
  - Vento fraco =  $[3+, 0-] \rightarrow$  (3 sim)
  - Vento forte =  $[0+, 2-] \rightarrow$  (2 não)

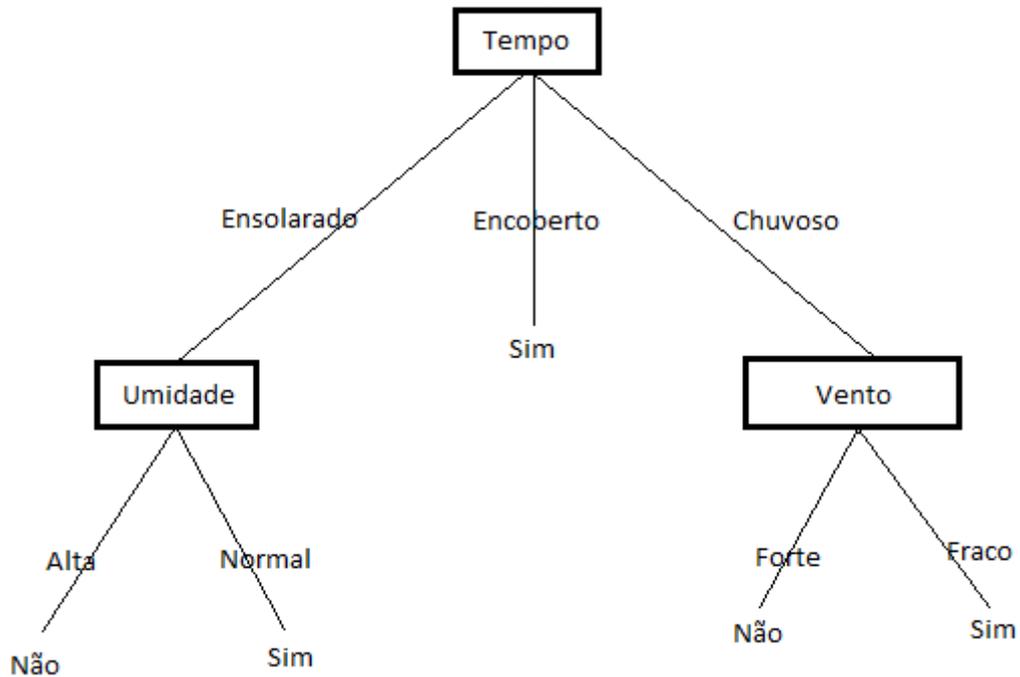
E com esses valores deve ser aplicado a fórmula de Ganho para umidade, temperatura e vento, porém o nó relacionado a encoberto torna-se folha, pois só possui um resultado possível que é sim.

O valores da fórmula de Ganho para ensolarado e chuvoso geram o valor de 0,970, logo ambos são definidos como nó da árvore.

Continuando o processo é necessário definir o próximo nó para ensolarado e chuvoso, então aplicando a fórmula de Ganho para definir qual atributo possui o maior ganho de informação, mas essa fórmula só faz-se necessário para a temperatura, pois como visualizado acima no caso de dia ensolarado com a umidade alta todos os caso gerados foram não jogar tenis, e nos casos de dia ensolarado com a umidade normal em todos os casos foram sim para jogar tênis enquanto para os casos de dia chuvoso onde o vento é fraco em todos os casos foram sim para jogar tênis e não quando o dia for chuvoso e o vento for forte.

Com isso é gerado a seguinte árvore mostrada na Figura 54.

Figura 54 - Árvore de Decisão para exemplo (Jogar Tênis)



Fonte: Elaborado pelo autor

Montando no formato SE-ENTÃO, a mesma árvore seria reescrita da seguinte maneira:

Se ((Tempo = Ensolarado) E (Umidade = Normal)) Ou  
 (Tempo = Encoberto) Ou  
 ((Tempo = Chuvoso) E (Vento = Fraco))  
 Jogar tênis  
 Senão  
 Não jogar tênis