

**UNIVERSIDADE DE SOROCABA
PRÓ-REITORIA ACADÊMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO**

Rafael Fernando de Moraes Moreno

**ANÁLISE DO PERFIL DISCENTE DA UNIVERSIDADE DE
SOROCABA, MEDIANTE A TÉCNICA DE MINERAÇÃO DE DADOS**

**Sorocaba/SP
2014**

Rafael Fernando de Moraes Moreno

**ANÁLISE DO PERFIL DISCENTE DA UNIVERSIDADE DE
SOROCABA, MEDIANTE A TÉCNICA DE MINERAÇÃO DE DADOS**

Dissertação apresentada à Banca Examinadora do Programa de Pós-Graduação em Educação da Universidade de Sorocaba, como exigência parcial para obtenção do título de Mestre em Educação.

Orientador: Prof. Dr. José Dias Sobrinho

**Sorocaba/SP
2014**

Ficha Catalográfica

M842a Moreno, Rafael Fernando de Moraes
Análise do perfil discente da Universidade de Sorocaba,
mediante a técnica de mineração de dados / Rafael Fernando de
Moreaes Moreno. -- 2014.
107 f. : il.

Orientador: Prof. D r. José Dias Sobrinho
Dissertação (Mestrado em Educação) - Universidade de
Sorocaba, Sorocaba, SP, 2014.

1. Estudantes universitários – Sorocaba (SP) – Aspectos sociais.
2. Universidade de Sorocaba - Sorocaba (SP) – Estudantes. 3.
Ensino superior - Avaliação. 4. Mineração de dados (Computação). I.
Dias Sobrinho, José, orient. II. Universidade de Sorocaba. III. Título.

Rafael Fernando de Moraes Moreno

**ANÁLISE DO PERFIL DISCENTE DA UNIVERSIDADE DE
SOROCABA, MEDIANTE A TÉCNICA DE MINERAÇÃO DE DADOS**

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre no Programa de Pós-Graduação em Educação da Universidade de Sorocaba.

Aprovado em:

BANCA EXAMINADORA:

Pres.: Prof. Dr. José Dias Sobrinho,
Universidade de Sorocaba

Prof. Dr. Waldemar Marques
Universidade de Sorocaba

Prof. Dr. Rogério Augusto Profeta
Universidade de Sorocaba

Dedico este trabalho à minha amada
esposa, pais e irmãos sempre presentes
em minha vida.

AGRADECIMENTOS

Primeiramente agradeço a Deus pela saúde e o privilégio de poder alcançar mais esse degrau.

Agradeço à minha esposa, pelo amor, paciência e apoio que tem oferecido nas horas de dificuldade.

Agradeço à minha família e familiares pelo apoio e ajuda que sempre me ofereceram.

Agradeço a todos os professores que contribuíram para minha formação. Sem eles não seria possível tornar este sonho realidade.

Em especial, agradeço ao Prof.^o Dr. José Dias Sobrinho, meu orientador, por ter acreditado e orientado minha pesquisa no decorrer destes anos.

Agradeço a todos os amigos do mestrado e doutorado, que sempre estiveram prontos em compartilhar e ajudar.

Agradeço aos amigos e funcionários da Universidade de Sorocaba que estiveram presentes e acompanharam o desenvolvimento desta pesquisa.

A todos, muito obrigado.

Tenha em mente que tudo que você aprende na escola é trabalho de muitas gerações. Receba essa herança, honre-a, acrescente a ela e, um dia, fielmente, deposite-a nas mãos de seus filhos.

(Albert Einstein)

RESUMO

No cenário complexo que constitui a universidade nos dias atuais, impulsionada e pressionada pela lógica do mercado, cabe a perguntar: quem está frequentando a universidade? O objetivo desta pesquisa é realizar um estudo de caso na Universidade de Sorocaba identificando possíveis padrões e classificações no perfil dos discentes que expliquem a sua escolha por determinado curso. Para este feito será realizado uma análise empírica no banco de dados da Universidade fazendo uso de um software de mineração de dados e utilizando um processo denominado KDD (*Knowledge Discovery in Databases*), que tem a finalidade de descobrir conhecimento em grande volume de dados. Com esta técnica e ferramenta serão analisados os dados socioeconômicos e do questionário informativo dos estudantes ingressantes entre 2006 a 2011, buscando assim, extrair possíveis classificações que expliquem o perfil do estudante que busca determinado curso.

Palavras-chave: Perfil discente. Universidade de Sorocaba. Mineração de dados.

ABSTRACT

In the complex scenario that constitutes the present-day university, driven and pressed by the logic of market, one may wonder: who is attending college? The objective of this research is to conduct a case study at the University of Sorocaba identifying possible standards and classifications in the students' profile to explain their choice for a specific academic major. For this purpose an empirical analysis will be performed on the University's database by making use of a data mining software and using a process called KDD (*Knowledge Discovery in Databases*), which aims to discover knowledge in large volume of data. This technique and tool will be used to analyze socioeconomic and informative survey data of incoming students between 2006 to 2011, thus seeking to extract possible classifications that explain the profile of a student that seeks a certain academic major.

Keywords: Student profile. University of Sorocaba. Data Mining.

LISTA DE FIGURAS

Figura 1 - Etapas do KDD	37
Figura 2 - Exemplo de árvore	43
Figura 3 - Exemplo de cluster.....	44
Figura 4 - Árvore por sexo/faixa etária	52
Figura 5 - Árvore sexo/faixa etária sem Parfor	54
Figura 6 - Todas as relações.....	54
Figura 7 - Somente as relações fortes	55
Figura 8 - Árvore Atividade que desempenha e curso escolhido	70
Figura 9 - Árvore motivo da escolha e curso escolhido.....	71
Figura 10 – Árvore Turno de conclusão do ensino médio e curso escolhido	72
Figura 11 – Árvore Motivo da escolha do curso x modalidade do curso	74
Figura 12 - Árvore múltiplas variáveis	75
Figura 13 - Árvore múltiplas variáveis nó 1	75
Figura 14 - Árvore múltiplas variáveis nó 2	76
Figura 15 - Árvore múltiplas variáveis nó 3 e 4	76
Figura 16 - Árvore múltiplas variáveis por modalidade.....	77
Figura 17 - Árvore múltiplas variáveis por modalidade 1.....	78
Figura 18 - Árvore múltiplas variáveis por modalidade 2.....	78
Figura 19 - Árvore múltiplas variáveis por modalidade 3.....	79
Figura 20 - Árvore múltiplas variáveis por modalidade 4.....	79
Figura 21 - Árvore dados demográficos x processo seletivo.....	81
Figura 22 - Árvore dados demográficos e processo seletivo 2.....	82
Figura 23 - Árvore motivo da escolha do curso	84

LISTA DE TABELAS

Tabela 1 - Quantidade de Habitantes por Sexo/Idade	23
Tabela 2 - Produto Interno Bruto de Sorocaba (2009)	25
Tabela 3 - Produto Interno Bruto de Sorocaba (2010)	26
Tabela 4 - Quantidade de escola por tipo	26
Tabela 5 - Quantidade de ingressantes na UNISO de 2006 - 2011 por Estado	33
Tabela 6 – Vinte primeiras cidades que mais apresentam ingressantes na UNISO de 2006 - 2011	33
Tabela 7 - Dez primeiros cursos com maior número de ingressante de 2006 - 2011	35
Tabela 8 – Curso escolhido, sexo feminino e faixa etária	53
Tabela 9 - Curso escolhido, sexo masculino e faixa etária	53

LISTA DE QUADROS

Quadro 1 - Renda do Candidato (Original/Padronizada).....	58
Quadro 2 - Número de dependentes que vivem com a renda (Original/Padronizada)	59
Quadro 3 – Se o candidato já iniciou algum curso Superior (Original/Padronizada) .	60
Quadro 4 - Atividade que desempenha (Original/Padronizada)	61
Quadro 5 - Meio que utiliza para manter-se informado (Original/Padronizada)	62
Quadro 6 - Candidato possui acesso a internet (Original/Padronizada).....	63
Quadro 7 - Motivo da escolha do curso (Original/Padronizada).....	65
Quadro 8 - Conheceu a UNISO através de qual meio (Original/Padronizada).....	66
Quadro 9 - Quantidade de livros que lê (Original/Padronizada).....	67
Quadro 10 - Exerce alguma atividade remunerada (Original/Padronizada)	68
Quadro 11 - Turno de conclusão ensino médio (Original/Padronizada).....	69

LISTA DE GRÁFICOS

Gráfico 1 - Quantidade de Habitantes por Sexo/Idade.....	23
Gráfico 2 – Número de habitantes de Sorocaba – 1991-2010	24
Gráfico 3 - Pessoas de 10 anos ou mais de idade, estado civil	24
Gráfico 4 - Produto Interno Bruto de Sorocaba (2009).....	25
Gráfico 5 - Produto Interno Bruto de Sorocaba (2010).....	26
Gráfico 6 - Quantidade de escolas por tipo	27
Gráfico 7- Matrículas por série	27
Gráfico 8 - Ingressantes na Uniso por ano	32
Gráfico 9 - Ingressantes na UNISO de Sorocaba/Outras Cidades.....	34
Gráfico 10 - Dez primeiros cursos com maior número de ingressantes de 2006 - 2011	35
Gráfico 11 - Quantidade de estudantes por ano de ingresso/sexo.....	49
Gráfico 12 – Percentual de Estudantes/Faixa etária - 2006 a 2011	50
Gráfico 13 - Percentual de alunos por cor/raça	56
Gráfico 14 - Renda do candidato.....	59
Gráfico 15 - Número de dependentes que vivem com a renda	60
Gráfico 16 – Se o candidato já iniciou algum curso superior.....	61
Gráfico 17 - Atividade que desempenha	62
Gráfico 18 - Meio que utiliza para manter-se informado.....	63
Gráfico 19- Candidato possui acesso a internet.....	64
Gráfico 20 - Motivo da escolha do curso	65
Gráfico 21 - Conheceu a UNISO através de qual meio.....	66
Gráfico 22 - Quantidade de livros que lê	67
Gráfico 23 - Exerce alguma atividade remunerada	68
Gráfico 24 - Turno de conclusão do ensino médio	69

SUMÁRIO

1	INTRODUÇÃO	14
2	GLOBALIZAÇÃO, NEOLIBERALISMO E UNIVERSIDADE	17
3	CIDADE DE SOROCABA E UNIVERSIDADE DE SOROCABA	21
3.1	Cidade de Sorocaba	21
3.1.1	História	21
3.1.2	Atualmente	22
3.2	Universidade de Sorocaba	28
4	DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS	37
4.1	Mineração de dados e suas técnicas	38
4.1.1	Classificação	40
4.1.2	Estimativa	40
4.1.3	Previsão.....	40
4.1.4	Análise de Afinidade.....	41
4.1.5	Análise de Agrupamentos.....	41
4.2	Árvore de decisão	41
4.3	Análise por agrupamento (Cluster Analysis)	43
4.4	Casos de Mineração de Dados	44
4.4.1	Descoberta de Conhecimento sobre o Processo Seletivo da UFPR	44
4.4.2	Aspectos teóricos do data mining e aplicação das redes neurais em previsões de preços agropecuários.....	46
4.4.3	Descoberta de conhecimento aplicado a dados eleitorais.....	47
5	MINERAÇÃO DE DADOS – UNISO	48
5.1	Dados socioeconômicos e demográficos dos estudantes	49
5.1.1	Mineração dos dados socioeconômicos demográficos dos estudantes	51
5.2	Dados do processo seletivo	56
5.2.1	Renda do candidato e curso escolhido.....	70
5.2.2	Número de dependentes que vivem com a renda e curso escolhido	70
5.2.3	Candidato já iniciou algum curso superior e curso escolhido	70
5.2.4	Atividade que desempenha e curso escolhido	70
5.2.5	Meio que utiliza para manter-se informado e curso escolhido.....	71

5.2.6	Acesso à internet e curso escolhido	71
5.2.7	Motivo da escolha do curso e curso escolhido	71
5.2.8	Conheceu a UNISO através de qual meio e curso escolhido	72
5.2.9	Quantidade de livros que lê em média por ano e curso escolhido	72
5.2.10	Exerce alguma atividade remunerada e curso escolhido	72
5.2.11	Turno de conclusão do ensino médio e curso escolhido	72
5.2.12	Múltiplas variáveis	75
5.3	Dados demográficos e processo seletivo	80
6	CONSIDERAÇÕES FINAIS.....	86
	REFERÊNCIAS.....	90
	APÊNDICE A - CURSOS DA GRADUAÇÃO – UNIVERSIDADE DE SOROCABA	94
	APÊNDICE B - CURSOS DA PÓS-GRADUAÇÃO LATO SENSU – UNIVERSIDADE DE SOROCABA	96
	APÊNDICE C – QUANTIDADE DE INGRESSANTES 2006-2011 POR CIDADE DO ESTADO DE SÃO PAULO	97
	APÊNDICE D – QUANTIDADE DE INGRESSANTES POR CURSO.....	101
	ANEXO A – POPULAÇÃO CENSO DEMOGRÁFICO 2010	104
	ANEXO B – EVOLUÇÃO DO NÚMERO DE INSTITUIÇÕES DE EDUCAÇÃO SUPERIOR.....	105
	ANEXO C – EVOLUÇÃO DO NÚMERO DE MATRÍCULAS	106
	ANEXO D – EVOLUÇÃO DAS TAXAS DE ESCOLARIZAÇÃO BRUTA E LÍQUIDA NA EDUCAÇÃO SUPERIOR.....	107

1 INTRODUÇÃO

Hoje ocorre uma forte presença do mercado operando no âmbito universitário e modificando seus valores primordiais. A universidade é produto de um acontecimento histórico que vem desde a baixa Idade Média (Bolonha, 1088), passando por um longo processo de reconstrução social (BERNHEIM, 2010).

Uma Universidade deve exercer o ensino, a pesquisa e a extensão. Cada um destes elementos estruturais emergiu ao longo dos séculos em diferentes modelos universitários. Pode-se destacar alguns: Universidade de Bolonha, Universidade de Paris, e mais tarde, muitas outras, como a Universidade Napoleônica, a Universidade de Berlim (Humboldt) e as Universidades Estadunidenses. Todas exerceram papel fundamental na construção da Universidade dos dias atuais.

A Educação Superior vem passando por grandes transformações de seus valores e perdendo sua força/significado pela opressão do cenário neoliberal que constitui a globalização econômica.

Nesse cenário complexo da universidade dos dias atuais, cabe a pergunta: quem está frequentando a universidade? Existe alguma tendência estatística que define o perfil do aluno mediante a opção do curso escolhido?

Com esses questionamentos, será tomada a Universidade de Sorocaba como um estudo de caso com objetivo de entender e conhecer melhor o público que está inserido na educação superior e assim analisar e delinear possíveis classificações existentes no perfil do aluno sobre o curso escolhido.

Para o desenvolvimento desta pesquisa, será trabalhado o banco de dados da Universidade de Sorocaba, relativos aos ingressantes entre 2006 a 2011, no qual será aplicado a tecnologia denominada mineração de dados (*data mining*) para buscar possíveis regras do perfil discente quando este opta por um curso.

A mineração de dados efetua a junção da estatística com a inteligência computacional, possibilitando a descoberta de comportamentos ocultos em grandes bancos de dados e, a partir daí, a inferência de regras para previsão de comportamento futuro e orientação na tomada de decisões (NAMEN, BORGES e SADALA, 2013 apud HIRJI, 2001).

Turban, Rainer e Potter (2005) apresentam um exemplo clássico da utilização da mineração de dados na rede Wall Mart. Neste exemplo foi identificado que a compra de fraldas estava tendencialmente ligada a compra de cervejas. Chegou-se

à conclusão que se torna impossível a olho nu verificar uma tendência quando há um grande volume de dados.

A tecnologia de mineração de dados foi originalmente utilizada em grandes empresas para tomada de decisão, mas ela serve perfeitamente para a pesquisa em questão, pois, com ela é possível analisar grande volume de dados e mediante a utilização de algoritmos matemáticos garimpar comportamentos que estavam invisíveis ao olhar humano.

Este estudo será de grande importância para melhor compreender o cenário da região de Sorocaba, decifrando quem frequenta a universidade, vista do ponto da Universidade de Sorocaba e se existe alguma tendência no perfil do discente quando este opta pelo curso.

Para realizar esse feito, partimos do pressuposto de que, analisando três hipóteses, será possível alcançar o objetivo da pesquisa. A primeira hipótese é que exista uma regra na escolha do curso de acordo com os dados socioeconômicos/demográficos, ou seja, mediante esses dados do estudante será possível cruzá-los e assim extrair possíveis regras que definam o perfil discente sobre um determinado curso.

A segunda hipótese é que exista uma regra nos dados do processo seletivo. Ou seja, de acordo com o questionário informativo que o candidato preenche no dia do vestibular e com outras variáveis como número de acertos, pontuação e outros dados, será possível traçar a regra que define o perfil discente com relação a um determinado curso.

A terceira hipótese diz respeito à combinação dos dados socioeconômicos/demográficos e do processo seletivo na qual será possível cruzar as informações e chegar a uma regra que defina o perfil discente de acordo com o curso escolhido.

Para o desenvolvimento desta pesquisa foi necessário primeiramente trazer alguns assuntos em discussão. Sendo assim, no capítulo 2, é discutido o cenário da educação superior nos dias de hoje, destacando os temas: globalização e neoliberalismo. Não é possível tratar hoje da educação superior sem abordar esses dois temas fundamentais para conhecer como se encontra o cenário atual que servirá de base para interpretar os resultados identificados na pesquisa.

A pesquisa utiliza como estudo de caso a Universidade de Sorocaba. O capítulo 3 traz um relato histórico sobre a Universidade, baseada nos trabalhos de Neves (1998) e de Vannucchi (2012) e apresenta alguns dados sobre a Universidade. Também é apresentado neste capítulo um breve estudo sobre a cidade de Sorocaba, apontando suas origens e a situação atual, baseada nos dados do censo demográfico de 2010.

No capítulo 4, será apresentada e detalhada a descoberta de conhecimento em banco de dados, denominado *KDD*. Após uma pesquisa bibliográfica, será descrito o que é, como funciona a mineração de dados e apresentadas algumas pesquisas acadêmicas já realizadas utilizando esta técnica.

No capítulo 5 é descrito passo a passo o desenvolvimento da pesquisa, no qual será apresentado detalhadamente a verificação das três hipóteses e apontado como se deu o desenvolvimento e utilização da mineração de dados.

No capítulo 6, considerações finais, é apresentado um relatório dos resultados obtidos na pesquisa, as limitações encontradas e as contribuições trazidas pela pesquisa.

2 GLOBALIZAÇÃO, NEOLIBERALISMO E UNIVERSIDADE

Quando tratamos da Universidade e da Educação de um modo geral, seja ela básica ou superior, não podemos deixar de citar um fenômeno que vem ocorrendo e que trouxe consigo mudanças que causaram grande impacto na educação. Estamos falando da globalização.

Se quisermos entender e decifrar o perfil do candidato que está inserido na Universidade, seja na Universidade de Sorocaba ou em qualquer outra instituição, primeiro precisamos compreender os processos mais amplos, políticos e econômicos que operam e exercem influência sobre a Educação e quais são os efeitos que estes produzem.

Com o cenário competitivo instaurado pela globalização, a diluição das fronteiras territoriais e a supervalorização da economia como motor da sociedade, a educação, em especial a educação superior, é forçada a suprir e alimentar o mercado com mão-de-obra capacitada para produzir e consumir, fomentando assim o fortalecimento da economia e a expansão do mercado.

Vale lembrar que a base da economia de mercado é a oferta e demanda. Estando essas duas bases bem alinhadas e proporcionais, a economia tende a ir bem. Trazendo essa lógica para o ponto de vista educacional, uma pessoa analfabeta não consome e uma pessoa sem qualificação mínima não produz.

Assim, podemos perceber um alinhamento das políticas impostas pelo Banco Mundial com essa lógica de mercado. Uma destas políticas é o fortalecimento de investimentos em educação básica, capacitando a sociedade minimamente para consumir (demanda). Outra política imposta pelo Banco Mundial resulta no aumento das privatizações em educação superior formando mão-de-obra qualificada para produzir o que podemos chamar de oferta.

Conforme cita Mello (2012, p. 232), de “apenas 2% dos empréstimos do banco mundial para educação, na década de 90 passamos para 22,1% do total de recursos emprestados pelo Banco Mundial ao Brasil no final do período de 1991-1994”.

Foi sobretudo na década de 90 que o Banco Mundial impôs fortemente ao Brasil o seu modelo, fazendo com que o governo priorizasse os investimentos em educação básica e aumentasse a privatização na educação superior.

Essas trouxeram um acréscimo ao número de pessoas cursando a educação superior. Mas o que mais inquieta e se torna questionável é a forma como isso está ocorrendo, sem entrar no mérito da qualidade desses cursos, pois aí teríamos que questionar de qual qualidade se está tratando. Estamos hoje formando cidadãos-profissionais ou somente profissionais para o mercado de trabalho?

Não podemos nos esquecer de que educar não é simplesmente ensinar uma profissão aos alunos. Educar é mais amplo do que ensinar, conforme escreve Barata-Moura (2003, p. 31), “educar não é domesticar, nem formatar; educar é eduzir: desenvolver e formar a nossa aptidão de senhorio e de orientação do viver”.

Com o aumento de privatizações na educação superior, investidores internos e externos brigam pela fatia do bolo, que se demonstrou ser muito atrativa e lucrativa. Conforme relata Cunha (2004), de acordo com o censo de 2003, houve um aumento de 209% dos alunos que cursam uma instituição superior, mas o número de universidades públicas ficou estagnado, ou seja, tivemos o aumento de instituições privadas e a criação dos centros universitários.

Podemos observar esse aquecimento no “mercado da educação” com a seguinte passagem de Marques (2013, p. 72) “[...] A matrícula na educação superior [...] de 3.036.113 matriculados em 2001, alcança 6.379.299 em 2010”. Subdividindo esse número em público ou privado, Marques (2013, p. 72) detalha que “[...] as públicas aumentaram de 944.584 para 1.643.298, uma taxa de 42,5%, já as privadas de 2.091.529 para 4.736.001, uma taxa de 55,8%”.

O surgimento da sociedade do conhecimento vem agravando mais a situação, pois, segundo Dias Sobrinho (2005, p. 110) a chamada sociedade do conhecimento, que é fundada na informação e comunicação, trouxe consigo mais assimetrias sociais. O conhecimento tornou-se motor da economia. Hoje os países ricos e desenvolvidos são aqueles que detêm os meios de comunicação/informação gerando mais conhecimento, detendo as pesquisas de ponta e conseqüentemente obtendo riqueza.

Assim, os países detentores da comunicação/informação e de modo geral da tecnologia ditam aos países consumidores sua vontade. Deste modo, a educação acaba entrando nesse meio político/mercantil.

Os avanços tecnológicos e informacionais que movem a economia nos dias atuais fizeram crescer o número de pessoas excluídas. De acordo com Dias Sobrinho (2005, p. 111), cerca de 4,5 bilhões não possuem condições necessárias de acesso às novas tecnologias de comunicação e informação e conseqüentemente a estas serão negados empregos, condições mínimas de vida digna e esperança de futuro.

Uma frase que representa o desafio atual é o que diz Dias Sobrinho (2005, p. 112):

Um dos maiores desafios do mundo de hoje é dar um sentido mais humano ao progresso, isto é, fazer com que os benefícios do desenvolvimento tecnológico e global cheguem ao maior número possível de pessoas em todos os cantos do planeta.

Temos duas forças interagindo na educação: a mercantilização da educação como algo que se possa comprar, equiparando-a com um simples produto/serviço, e a formação de mão-de-obra para o mercado, minimamente qualificada pra executar aquilo que é necessário do ponto de vista empresarial e com necessidades de consumo desenfreado.

Goergen (2013, p. 48) constata que a educação hoje é vista como investimento individual, visando retorno econômico privado na forma de emprego/salário e retorno estatal em termos de impostos e crescimento econômico. Ou seja, a educação é vista como um serviço que se compra para proporcionar ao indivíduo melhor poderio de competitividade visando sua própria ascensão, não se preocupando com os outros ao redor e nem com a sociedade em que está inserido.

De modo geral, apenas ensinar (adestrar) as pessoas para exercerem uma profissão não é suficiente. Goergen (2013, p. 46) expõe a necessidade da formação e não apenas do ensino. Formação é algo mais amplo que ensino, pois inclui dimensões epistêmicas, políticas e éticas dos alunos.

Sem a educação e a formação de cidadãos conscientes e profissionais capacitados, não conseguiremos alterar esse cenário e sempre seremos um país consumidor daquilo que vem de fora. Se não investimos em educação superior, estamos automaticamente não investindo em pesquisa e pesquisadores que poderiam estar trabalhando em prol da nossa sociedade.

Diante desse quadro da educação, seja ela superior ou básica, não podemos deixar de acreditar que esta cumpra seu papel de transformar e formar pessoas, que serão capazes de pensar, agir e propor mudanças em nossa sociedade.

Frente a todas essas dificuldades que encontramos na educação, particularmente falando da educação superior, emerge a pergunta: Qual é a missão da educação superior? Vemos hoje um forte apelo para a formação de mão-de-obra, mas será que sua missão é apenas capacitar para o mercado?

Um trecho que Goergen (2013, p. 49) cita traz a definição de qual seria, segundo a UNESCO, a missão da educação superior, sintetizada nas conclusões da Conferência Mundial do Ensino Superior em Paris em 1998:

[...] A missão do ensino superior consiste em contribuir com o desenvolvimento e a melhoria do conjunto da sociedade, ou seja, educar e formar graduados altamente qualificados, capazes de responder às necessidades de todos os aspectos da atividade humana; promover, gerar e difundir conhecimentos através da pesquisa; interpretar, preservar e promover as culturas no contexto do pluralismo e da diversidade cultural; oferecer possibilidades de aprendizagem ao longo de toda a vida; contribuir com o desenvolvimento e melhoria da educação em todos os níveis; proteger e promover a sociedade civil, formando os jovens de acordo com os valores nos quais se baseia a sociedade civil democrática e que proporcionem perspectivas críticas e independentes no debate sobre as opções estratégicas e no fortalecimento das perspectivas humanísticas.

3 CIDADE DE SOROCABA E UNIVERSIDADE DE SOROCABA

Neste capítulo apresentaremos o histórico da Universidade de Sorocaba junto com o cenário da cidade de Sorocaba. Para tanto, o capítulo foi dividido em dois subcapítulos. No primeiro subcapítulo trataremos da cidade de Sorocaba e, no segundo da Universidade de Sorocaba.

3.1 Cidade de Sorocaba

3.1.1 História

Sorocaba foi fundada oficialmente em 1654, sendo declarado o dia 15 de agosto, dia da padroeira da cidade, Nossa Senhora da Ponte, como o dia da cidade. Vannucchi (2012) conta que Sorocaba surgiu com a vinda do bandeirante Baltasar Fernandes que, ao chegar à região, trouxe os monges Beneditinos que garantiram ao povoado a prática religiosa e o magistério das primeiras letras.

Anterior a Baltasar Fernandes, Sorocaba “era uma encruzilhada aonde convergiam, por onde viajavam e se limitavam os índios tupis do Tietê, os tupiniquins e guaianazes de Piratininga, os carijós de Curitiba, os guaranis do Paranapanema” (VANNUCCHI, 2012, p. 17). O nome surgiu dos nativos, “sorok”, que significava romper-se, rasgar-se, ou seja, terra rasgada.

Vannucchi (2012) destaca que os primeiros forasteiros que chegaram à região foram os portugueses Afonso Sardinha e seu filho, por volta de 1589. Esses estiveram no morro Araçoiaba à procura de ouro, mas foi Baltasar Fernandes que veio para instalar-se e povoar a região.

Com a fundação de Sorocaba por um bandeirante, logo essa se notabilizou pela vinda de outros bandeirantes, como: Pascoal Moreira Cabral, Miguel Sutil, André de Zunega, Antunes Maciel, Antônio Raposo Tavares e outros que conforme diz Vannucchi (2012), alargaram o mapa brasileiro, sertão adentro, em busca de ouro e de esmeralda.

Outro movimento a favor do fortalecimento e crescimento de Sorocaba foi o movimento Tropeiro. Com o passar dos anos, Sorocaba se tornou sede das Feiras de Muares. Conforme destaca Vannucchi (2012, p. 19)

A cidade se tornou rica e famosa, chegando a superar em notoriedade a própria capital da Província. As feiras tiravam os nativos da pacatez provinciana, acordados para vender e enriquecer, recebendo forasteiros de todos os quadrantes, prontos para gastar em compras[...]. Para cá vinham brasileiros de todos os rincões, como também uruguaios e argentinos. Apareciam também companhias de teatro lírico, grandes comerciantes e personalidades de projeção nacional, como o Pe. Feijó, o Duque de Caxias e o próprio D. Pedro II, que aqui esteve quatro vezes. (VANNUCCHI, 2012, p. 19)

Sorocaba também foi pioneira na metalurgia. Nos arredores da cidade funcionava a primeira fábrica de ferro da América Latina, a fábrica São João do Ipanema, constituída por D. João VI e dirigida por um tempo por Francisco Adolfo Varnhagen, o Visconde de Porto Seguro, o pai da História do Brasil.

Vannucchi (2012) ressalta que Sorocaba não foi formada apenas por pessoas da região, como é o caso do húngaro Luiz Matheus Maylasky, o maior comprador de algodão da região, que em 1868 resolveu fundar uma fábrica de tecidos em Sorocaba, cujo prédio mais tarde serviu como oficina da locomoção Sorocabana.

Em 1870 Maylasky liderou a construção de uma via férrea de Sorocaba a São Paulo, que deu origem em 1871 à “Companhia Sorocabana”. Em junho de 1875 a linha férrea foi terminada sem a utilização de mão-de-obra escrava. Por esse feito, em 1880, D. Pedro II conferiu o título de Visconde de Sapucahy a Maylasky, como recompensa aos serviços prestados.

Vannucchi (2012) também ressalta a importância que as fábricas de tecidos tiveram para a cidade. Em 1852 surgiu a primeira fábrica de tecidos e a partir daí a indústria têxtil sorocabana viveu um século de crescimento notável.

3.1.2 Atualmente

Sorocaba possui aproximadamente 600.000 habitantes e está entre as dez principais cidades do Estado e a 21ª do País. De acordo com o censo do IBGE de 2010, Sorocaba e Ribeirão Preto foram as cidades que mais cresceram no Estado de São Paulo dentre as com mais de 300 mil habitantes. Do ano de 2000 a 2010 a população de Sorocaba cresceu 15,5%.

O censo de 2010 também indica que a população da cidade de Sorocaba é de 586.625 habitantes e que a cidade possui uma área de 449 KM². Analisando a pirâmide etária do censo, observamos que atualmente as mulheres representam 51% da população de Sorocaba, conforme os gráficos a seguir ilustram:

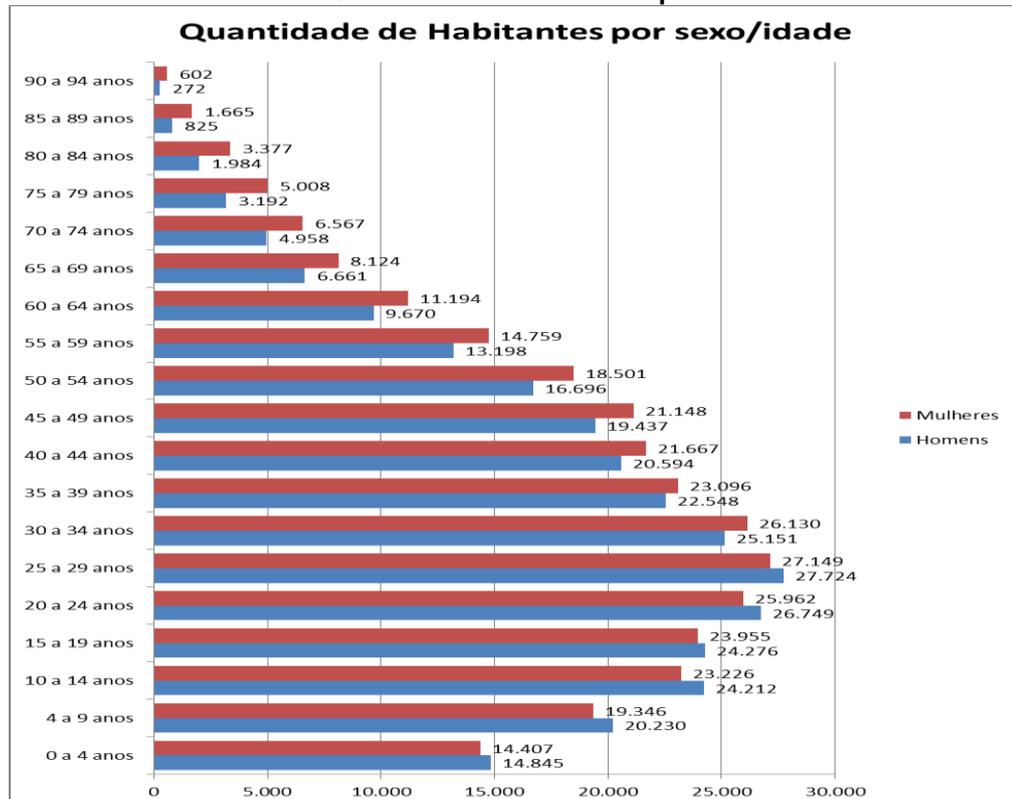
Tabela 1 - Quantidade de Habitantes por Sexo/Idade

Quantidade de Habitantes por sexo/idade					
	Homens	Mulheres	Total	% Homens	% Mulheres
0 a 4 anos	14.845	14.407	29.252	50,75	49,25
4 a 9 anos	20.230	19.346	39.576	51,12	48,88
10 a 14 anos	24.212	23.226	47.438	51,04	48,96
15 a 19 anos	24.276	23.955	48.231	50,33	49,67
20 a 24 anos	26.749	25.962	52.711	50,75	49,25
25 a 29 anos	27.724	27.149	54.873	50,52	49,48
30 a 34 anos	25.151	26.130	51.281	49,05	50,95
35 a 39 anos	22.548	23.096	45.644	49,40	50,60
40 a 44 anos	20.594	21.667	42.261	48,73	51,27
45 a 49 anos	19.437	21.148	40.585	47,89	52,11
50 a 54 anos	16.696	18.501	35.197	47,44	52,56
55 a 59 anos	13.198	14.759	27.957	47,21	52,79
60 a 64 anos	9.670	11.194	20.864	46,35	53,65
65 a 69 anos	6.661	8.124	14.785	45,05	54,95
70 a 74 anos	4.958	6.567	11.525	43,02	56,98
75 a 79 anos	3.192	5.008	8.200	38,93	61,07
80 a 84 anos	1.984	3.377	5.361	37,01	62,99
85 a 89 anos	825	1.665	2.490	33,13	66,87
90 a 94 anos	272	602	874	31,12	68,88
95 a 99 anos	0	157	157	0,00	100,00
Total	283.222	296.040	579.262	48,89	51,11

Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Nota: Tabela elaborada pelo autor

Gráfico 1 - Quantidade de Habitantes por Sexo/Idade

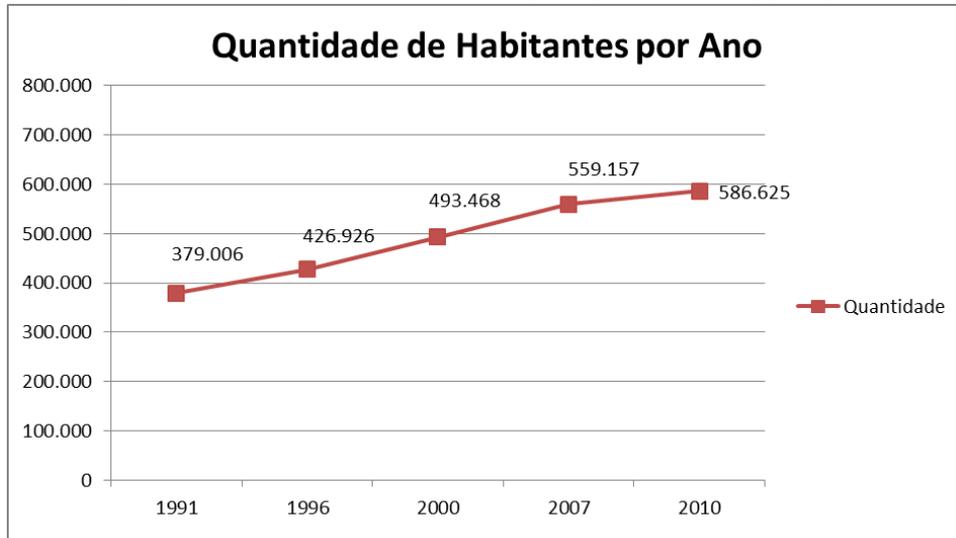


Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Nota: Gráfico elaborado pelo autor

O censo também detalha a evolução do número de habitantes da cidade, como podemos observar no gráfico a seguir. De 379.006 habitantes em 1991 passou para 586.625 em 2010.

Gráfico 2 – Número de habitantes de Sorocaba – 1991-2010

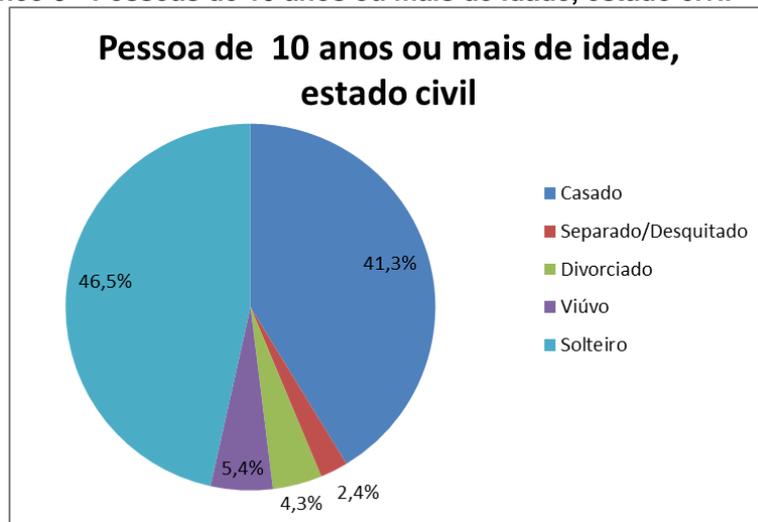


Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Nota: Gráfico elaborado pelo autor

Podemos observar também que a grande massa dos habitantes se divide em solteiros, que representam 46,5% dos habitantes com mais de 10 anos, e casados com 41,3% da população.

Gráfico 3 - Pessoas de 10 anos ou mais de idade, estado civil



Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Nota: Gráfico elaborado pelo autor

Analisando a economia da cidade de Sorocaba, pode ser verificado que a grande concentração do Produto Interno Bruto, ou seja, tudo aquilo que ela produz de bens e serviços, está na prestação de serviços e na indústria.

Conforme descreve Silva (2002), o setor de serviços, denominado setor terciário da economia, é caracterizado por produtos não materiais em que pessoas ou empresas prestam a terceiros, para satisfazer determinadas necessidades. Como por exemplo: educação, comércio, saúde, telecomunicações, turismo, transporte etc. Já a indústria se denomina setor secundário da economia e é responsável por transformar matéria prima em produtos industrializados.

Segundo Vannucchi (2012), Sorocaba conta atualmente com 1.500 indústrias de perfil diversificado, com 1.000 pontos de comércio e 2.000 prestadoras de serviços, nos diversos ramos de negócios.

Podemos identificar, através dos dados do censo, que Sorocaba apresenta um fato econômico comum tanto em São Paulo como no Brasil, na qual em primeiro lugar está o setor terciário (serviços) e em segundo o setor secundário (indústria). As tabelas e os gráficos abaixo detalham os números do Produto Interno Bruto do ano de 2009 e 2010 do município de Sorocaba.

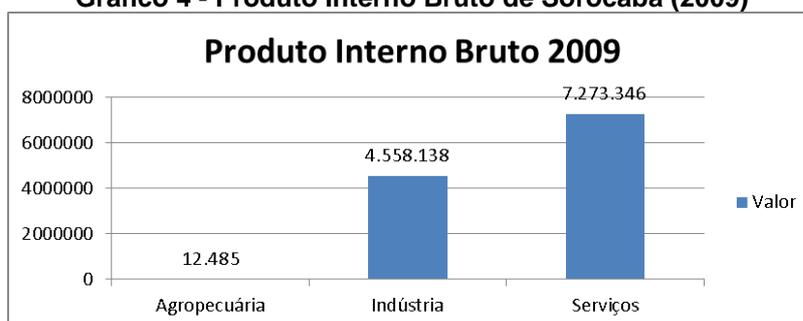
- **Produto interno Bruto (2009)**

Tabela 2 - Produto Interno Bruto de Sorocaba (2009)

Tipo	Valor
Agropecuária	12.485,00
Indústria	4.558.138,00
Serviço	7.273.346,00

Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Gráfico 4 - Produto Interno Bruto de Sorocaba (2009)



Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

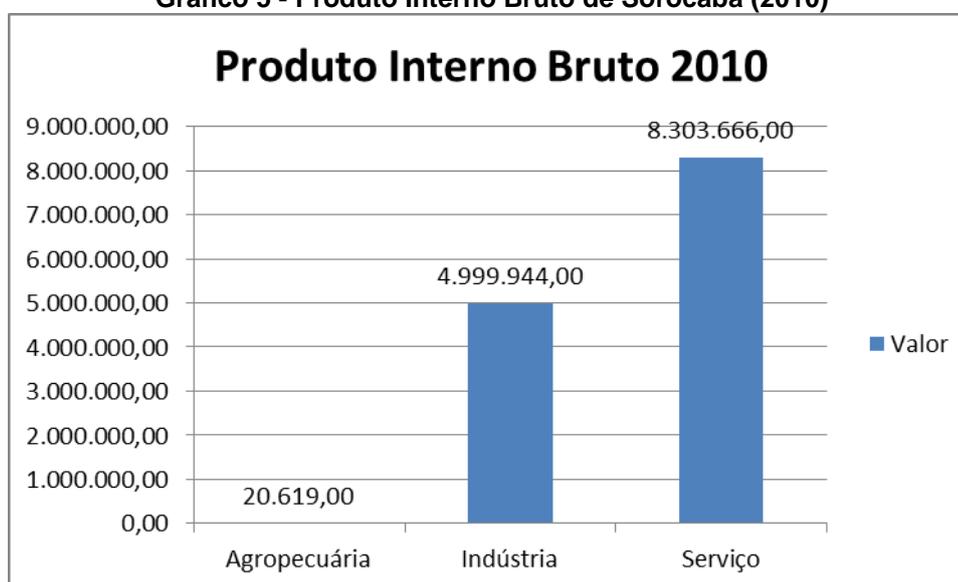
- **Produto Interno Bruto (2010)**

Tabela 3 - Produto Interno Bruto de Sorocaba (2010)

Tipo	Valor
Agropecuária	20.619,00
Indústria	4.999.944,00
Serviço	8.303.666,00

Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Gráfico 5 - Produto Interno Bruto de Sorocaba (2010)



Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Conforme Vannucchi (2012), logo que Baltasar Fernandes fundou a cidade de Sorocaba, trouxe monges Beneditinos para ensinar as primeiras letras na região. Hoje, Sorocaba é uma cidade formada e desenvolvida e que possui 431 escolas, subdividas nas seguintes séries:

Tabela 4 - Quantidade de escola por tipo

Tipo	Quantidade
Fundamental	181
Pré-Escola	161
Médio	89

Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010**. Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

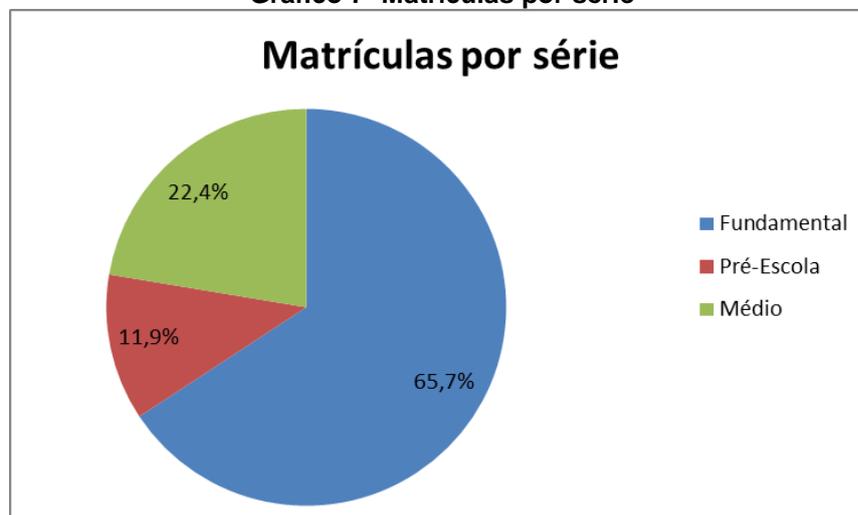
Gráfico 6 - Quantidade de escolas por tipo



Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

De acordo com o censo do IBGE 2010, o município conta com 127.235 mil alunos matriculados, nas séries Pré-escola, fundamental e médio. Sendo 15.172 mil alunos da pré-escola, 83.621 mil do fundamental e 28.442 mil no ensino médio. O gráfico abaixo detalha o percentual da matrícula por série.

Gráfico 7- Matrículas por série



Fonte: Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Vannucchi (2012) destaca a presença de mais de 100 cursos de graduação e outros de pós-graduação que hoje são oferecidos em Sorocaba, espalhados pelas universidades: Universidade de Sorocaba, a Pontifícia Universidade Católica de São Paulo, a Universidade Paulista, a UNESP, a Universidade Federal de São Carlos e

dezenas de outras instituições isoladas que compõem a Educação Superior na cidade de Sorocaba.

Vannucchi (2012) também aponta a importância de Sorocaba para as cidades vizinhas. Hoje Sorocaba é sede de uma Região Administrativa com 78 cidades, a mais extensa do Estado, com aproximadamente 2,5 milhões de habitantes.

3.2 Universidade de Sorocaba

As informações a respeito da Universidade de Sorocaba que se seguem estão baseadas nos trabalhos de Neves (1998) e de Vannucchi (2012), com grande destaque para este último. Vannucchi é personagem de enorme importância para a UNISO, de cuja história participou ativamente desde sua fundação como gestor e historiador.

Com o crescimento da cidade de Sorocaba, nos meados da década de 50, esta estava se tornando um grande centro estudantil e universitário. A cidade possuía 102.000 habitantes e inúmeras escolas estaduais, municipais e particulares, atendendo a 20.210 alunos. (NEVES, 1998, p. 15)

Com esse crescimento a necessidade de cursos superiores logo se alastrou. Vannucchi (2012) relata que em 1949, o bispo de Sorocaba Dom José de Aguirre, que sempre privilegiara a educação para semear valores evangélicos, registra em sua agenda uma visita a um cirurgião na Santa Casa para discutir sobre a possível colaboração no projeto de Faculdade de Medicina.

Com base na agenda do bispo, Vannucchi (2012) destaca que Dom Aguirre trabalhou em prol do surgimento da Faculdade de Medicina de Sorocaba. Vannucchi (2012) aponta a alegria que o bispo teve em 1951 ao benzer o hospital Santa Lucinda que seria um anexo da Faculdade de Medicina.

Em 1951, a prefeitura de Sorocaba solicita à diocese que administre a Faculdade de Filosofia, Ciências e Letras de Sorocaba. Diante do grande desafio, Vannucchi (2012) informa que Dom Aguirre não recusa, mas deixa expresso em ofício que a administração deva ser independente da política, conforme descreve o trecho:

- 1) A Diocese ficará na sua administração alheia a qualquer injunção política ou vínculo de sujeição;
- 2) A Diocese terá toda liberdade na escolha do corpo docente, discente e administrativo da faculdade.

Em 1954 começa a funcionar a Faculdade de Filosofia, Ciências e Letras de Sorocaba. Vannucchi (2012) relata que Dom Aguirre solicitou recursos financeiros a Francisco e Nicolau Scarpa, para reformar e melhorar o prédio cedido pela prefeitura na chácara Trujillo.

Neves (1998) aponta que o primeiro vestibular para a Faculdade de Filosofia, Ciências e Letras de Sorocaba teve 31 candidatos inscritos, sendo 13 para o curso de Pedagogia e 18 para o curso de Letras Neolatinas. Destes, 27 alunos se matricularam.

Visto que a subvenção municipal mostrava-se insuficiente e cada vez mais inviável, Dom Aguirre se dirigiu a família Scarpa para salvar a instituição que passava por um grave problema financeiro. O resultado foi à criação da Fundação Scarpa que subvencionou a faculdade de 1954 a 1956. (VANNUCCHI, 2012)

Neves (1998) relata que a família Scarpa se dispusera a tomar as primeiras providências, como: reforma do prédio, instalação de biblioteca especializada, verba para sua manutenção e construção de novas instalações.

Em 1955, segundo Neves (1998) a Fundação Scarpa se preocupou com a crescente despesa que a Faculdade vinha gerando e, em meio a essa crise financeira junto com problemas internos de administração, a situação ficou ainda pior quando se tornou pública. Alguns professores denunciaram a Faculdade para jornais locais alegando que o salário estava atrasado dois meses. Os alunos também se dividiram em pró e contra o movimento, sendo a Faculdade alvo das manchetes.

Ainda de acordo com Neves (1998), em meio à crise, a Fundação Scarpa utilizou de empréstimos bancários para colocar em dia o salário dos professores e superar os problemas, mas que depois desses acontecimentos, membros da Fundação Scarpa sugeriram sua extinção.

Vannucchi (2012, p. 28) relata esse acontecimento da seguinte maneira: “a Fundação, à vista dos crescentes encargos financeiros, teve vida breve, voltando à prefeitura o patrimônio e a responsabilidade financeira pela faculdade”.

Em meio a crises e problemas, Dom Aguirre sempre esteve à frente da Faculdade. Em 1963, de acordo com Neves (1998) houve uma reunião na qual o Prof^o Aldo Vannucchi apresenta a ideia da criação de uma fundação para manutenção, administração e desenvolvimento da Faculdade de Filosofia, Ciências

e Letras de Sorocaba. Ao término da reunião, por decisão unânime, fora criada a Fundação Dom Aguirre.

Vannucchi (2012) e Neves (1998) relatam que o bispo José Carlos de Aguirre ficou assustado ao ver a fundação com o seu nome e relutou para que fosse colocado outro nome na Fundação, mas cedeu e ficou tranquilizado quando lhe explicaram que, utilizando seu nome, essa teria mais credibilidade e apoio.

Neves (1998, p. 46) afirma que uma,

Fundação pode ser definida como uma instituição de caráter social, com finalidades filantrópicas, educacionais, assistenciais, culturais, científicas ou tecnológicas, tendo como fundamento de sua existência, um patrimônio destinado a esse fim.

Sendo a Fundação Dom Aguirre criada, um projeto de Lei nº 1153 autoriza o Poder Público Municipal de Sorocaba a doar à Fundação Dom Aguirre o imóvel de propriedade da prefeitura, localizado na Avenida General Osório nº 35.

Neves (1998) destaca que os primeiros anos de vida da Fundação não foram fáceis. Esta lutou contra problemas econômicos/financeiros, de espaço físico e administrativo.

Vannucchi (2012) informa que em 1956 fora criada a Faculdade de Direito e que essa também funcionou no mesmo prédio da Faculdade de Filosofia, Ciências e Letras, na chácara Trujillo.

Em 1965, segundo relata Vannunchi (2012) o professor Arthur Fonseca em reunião com o Conselho Superior da Fundação Dom Aguirre, sugere que a Fundação assumira mais uma instituição, a fim de atender à necessidade da sociedade sorocabana.

Assim, o Conselho Superior aprova a ideia e começa os preparativos para construção da Faculdade de Ciências Econômicas (FACCAS). Vannucchi (2012) escreve que a ideia era criar três cursos, na seguinte ordem: Administração de Empresas, Ciências Contábeis e Ciências Econômicas.

Um ano depois, em 1966, iniciaram-se as aulas no primeiro curso, Administração de Empresas, com 69 alunos no período noturno. Vannucchi (2012, p. 40) salienta que muito esforço fora feito para que esses cursos funcionassem

É de justiça anotar que o fundador e primeiro diretor da Faccas, Profº Arthur Fonseca, abriu mão de seu salário, em todo o tempo de sua gestão, que foi

de agosto de 1966 até setembro de 1970, objetivando, assim, aliviar os custos da Instituição. (VANNUCCHI, 2012, p. 40)

Vannucchi (2012) também relata que o segundo curso, Ciências Contábeis, iniciou a primeira turma em 1975 e o terceiro e último curso, Ciências Econômicas, iniciou em 1987, completando assim, o planejamento dos cursos que seriam oferecidos na Faculdade de Ciências Econômicas (FACCAS).

No começo a convivência das duas faculdades no mesmo local fora fácil, mas, com o crescimento das duas instituições a situação foi piorando. Vannucchi (2012, p. 41) diz: “[...] as duas Faculdades irmãs continuavam, fisicamente, juntas e, cordialmente, separadas”.

Vannucchi (2012) conta uma história que comprova essa desavença entre as duas instituições. Conta ele que, em 1975, quando estava em construção o prédio da FACCAS, faltou verba e a construção iria parar. Diante disso, o diretor da FACCAS conversou com o diretor da Faculdade de Filosofia, verificando a possibilidade de utilizar um andar do prédio dessa faculdade e que para isso iria pagar um aluguel simbólico. Vannucchi (2012, p. 42) comenta: “Como se vê, era irmã cobrando aluguel de irmã!”.

Segundo Vannucchi (2012), em 1977, quando Dom Amaury Castanho, assumiu a presidência da Fundação Dom Aguirre, o processo de unificação das duas Faculdades ganhou força e assim as primeiras mudanças começaram a ocorrer, como foi o caso da unificação dos setores administrativos. Porém, a oposição era intensa.

O esforço para unificação das duas Faculdades foi redobrado quando se lançou o projeto UNISO em 1988. Vannucchi (2012) descreve que sem essa unificação não haveria Universidade.

Após uma longa batalha política, com boicotes e rejeições por parte do corpo docente e administrativo da FACCAS, em 1993 após a conclusão do Regimento Unificado aprovado pela Portaria Ministerial, as duas Faculdades da Fundação Dom Aguirre passaram a ser uma única instituição, denominada: Faculdades Integradas Dom Aguirre (FIDA).

Em 1988, o Prof^o Aldo Vannucchi foi nomeado pelo Conselho Superior da Fundação e pela prefeitura de Sorocaba, coordenador do projeto que visava à construção de uma Universidade para a cidade de Sorocaba.

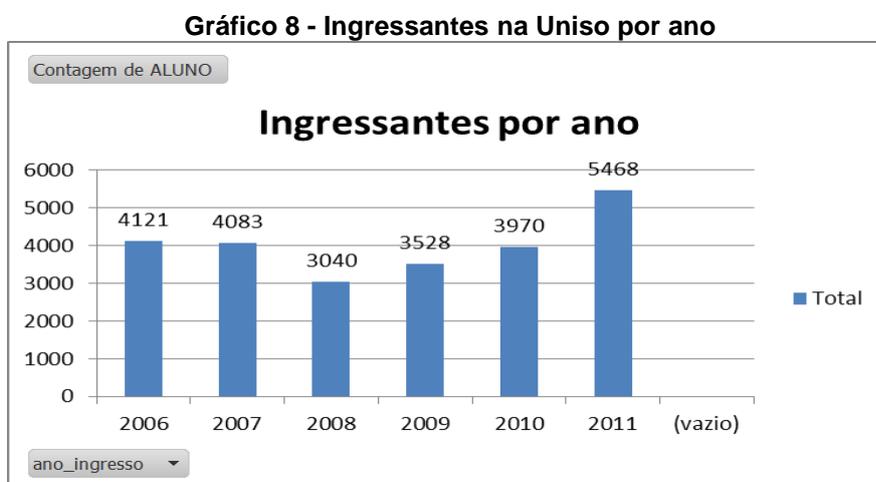
Vannucchi (2012) relata que o primeiro nome pensado para a Universidade foi Universo, ou seja, **Universidade de Sorocaba**. Porém no decorrer do projeto descobriu-se que esse nome já era utilizado por outra instituição e a Universidade acabou sendo batizada de UNISO.

Em seu livro, Vannucchi (2012), detalha os passos que foram necessários percorrer e as dificuldades a superar para que a Universidade de Sorocaba (UNISO) pudesse sair do papel e se concretizar. Vannucchi (2012) relata que visitou muitas Universidades, prefeituras da região e empresas a fim de conseguir conhecer as expectativas, as necessidades e arrecadar doações. Com o trabalho árduo e com ajuda de muitos benfeitores, em 1994 a Universidade de Sorocaba foi oficialmente criada.

Em 1999, com o grande número de alunos e cursos, foi inaugurado o terceiro campus da Universidade, o campus Cidade Universitária. Recentemente, (Setembro de 2013), em homenagem ao coordenador do projeto de implantação da Universidade e seu primeiro Reitor, o campus ganhou o nome de “Cidade Universitária Professor Aldo Vannucchi”.

Atualmente a Universidade conta com 68 cursos de graduação, 4 de pós-graduação *Stricto Sensu*, sendo três mestrados (educação, farmácia e comunicação), um doutorado (educação) e 13 cursos de pós-graduação *Lato Sensu*. A relação dos cursos da graduação e pós-graduação *Lato Sensu* está nos apêndices A e B.

Entre o ano de 2006 a 2011 a Universidade obteve 24.210 novos ingressantes, sendo esses ingressantes divididos nos seguintes anos:



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Hoje a Universidade não atende apenas à cidade de Sorocaba; podem ser encontrados alunos de inúmeras cidades vizinhas como também alguns de outros estados, conforme a tabela a seguir.

Tabela 5 - Quantidade de ingressantes na UNISO de 2006 - 2011 por Estado

UF	Quantidade
BA	1
CE	1
DF	3
ES	2
GO	4
MA	1
MG	42
MS	12
MT	3
PA	2
PB	3
PR	30
RJ	12
RS	10
SC	10
SP	24072
TO	2

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

No anexo G é possível visualizar a quantidade de ingressantes de 2006 a 2011 por cidade do estado de SP, porém vale destacar as 20 primeiras cidades que mais apresentam alunos na instituição.

Tabela 6 – Vinte primeiras cidades que mais apresentam ingressantes na UNISO de 2006 - 2011

CIDADE	QUANTIDADE
SOROCABA	15151
VOTORANTIM	1973
TATUI	651
SALTO DE PIRAPORA	480
PIEDADE	476
ARACOIABA DA SERRA	427
SAO ROQUE	404
ITAPETININGA	373
PORTO FELIZ	370
MAIRINQUE	369
ITU	337
IBIUNA	319
BOITUVA	310

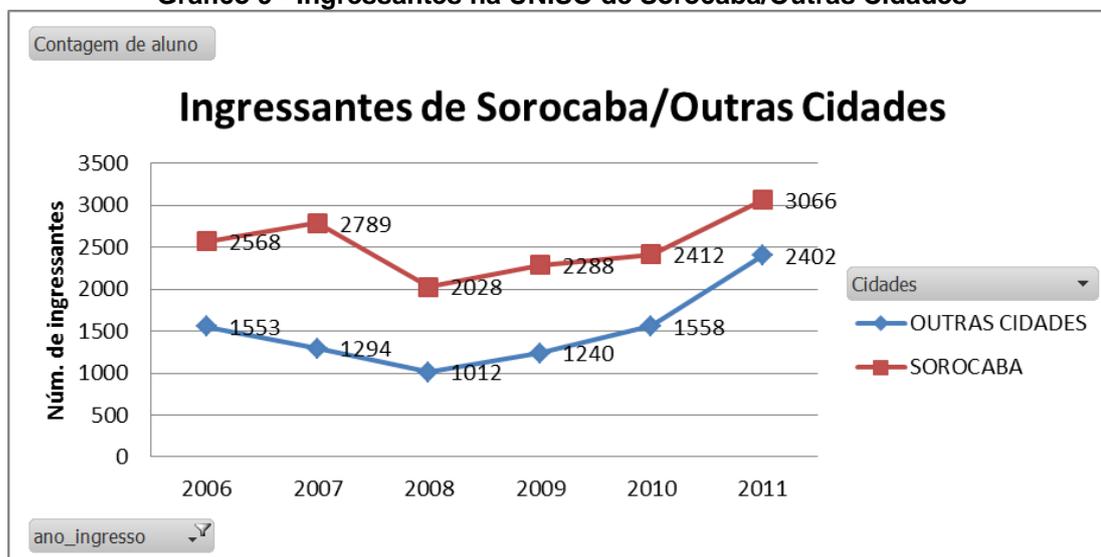
PILAR DO SUL	302
ALUMINIO	265
CERQUILHO	237
SALTO	204
TIETE	141
IPERO	132
CAPELA DO ALTO	132

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Conforme é possível analisar, 62,58% dos ingressantes são da cidade de Sorocaba, sendo 37,42% das demais cidades. Porém ao observar os ingressantes de Sorocaba comparado aos das demais cidades percebe-se que essa diferença tende a diminuir com o passar dos anos.

Gráfico 9 - Ingressantes na UNISO de Sorocaba/Outras Cidades



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Como mostra o gráfico, no ano de 2009, 2010 e 2011 o número de ingressantes aumentou, porém a quantidade de ingressantes de outras cidades foi maior do que os da cidade de Sorocaba. De 2009 a 2010 obteve-se um aumento de 318 ingressantes de outras cidades enquanto no mesmo período aumentou 124 ingressantes da cidade de Sorocaba.

De 2010 a 2011 ocorreu um aumento de 844 ingressantes de outras cidades enquanto no mesmo período foram apenas 654 da cidade de Sorocaba. Analisando de maneira geral, o número de ingressantes da cidade de Sorocaba é maior do que o de outras cidades, mas conforme analisado ano a ano, percebemos uma possível inversão no cenário.

Esses ingressantes estão subdivididos em inúmeros cursos que a Universidade de Sorocaba oferece. Assim, no anexo H é apresentada a quantidade de ingressantes por curso de 2006 – 2011. Vale destacar os dez primeiros cursos desta lista, sendo eles os seguintes:

Gráfico 10 - Dez primeiros cursos com maior número de ingressantes de 2006 - 2011



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Podemos detalhar esse gráfico dividindo essa quantidade de ingressantes por ano, conforme ilustra a tabela a seguir.

Tabela 7 - Dez primeiros cursos com maior número de ingressante de 2006 - 2011

Curso	2006	2007	2008	2009	2010	2011	Total Geral
Administração	11	637	336	321	242	430	1977
Pedagogia	84	571	389	337	163	204	1748
Direito	203	171	144	244	242	413	1417
Curso Superior de Tecnologia em Gestão da Produção Industrial	245	329	244	181	100	104	1203
Engenharia de Produção			135	248	225	351	959
Ciências Contábeis	273	151	100	115	83	188	910
Curso Superior de Tecnologia em Logística		202	138	154	121	179	794
Administração de Negócios	706	46	22	7	1		782
Curso Superior de Tecnologia em Gestão de Recursos Humanos	92	229	142	105	86	86	740
Comunicação Social - Habil. em Publicidade e Propaganda	173	112	87	76	82	135	665
Total Geral	1787	2448	1737	1788	1345	2090	11195

Fonte: Banco de dados da UNISO.

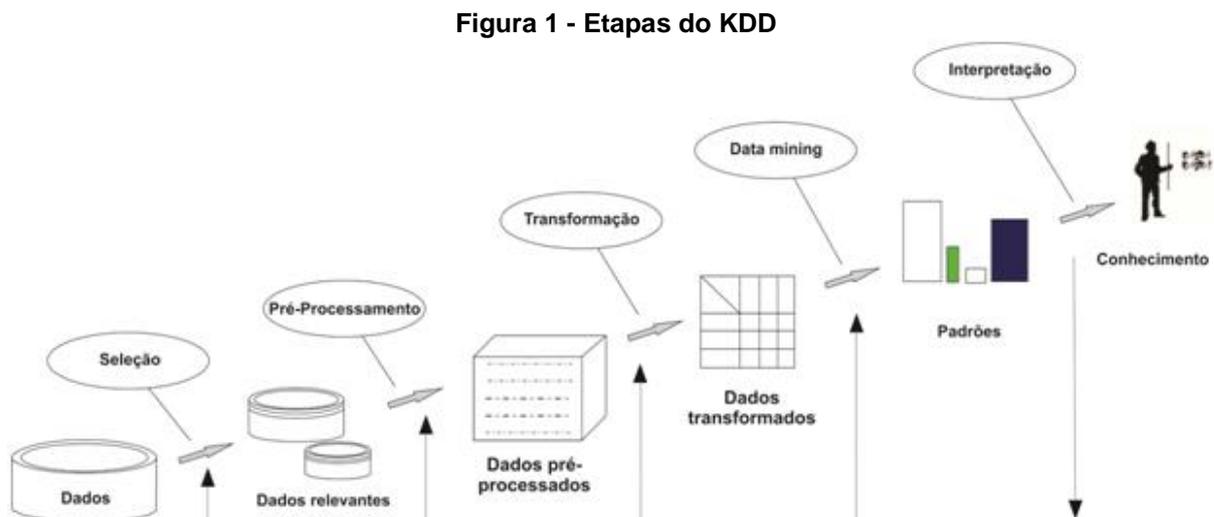
Nota: Elaborado pelo autor.

Observa-se que o curso de Administração anterior ao ano de 2007 se denominava Administração de Negócios é por isso que o ano de 2006 esse curso ficou com apenas 11 ingressantes.

Como pôde ser observado, a Universidade de Sorocaba é atuante em Sorocaba e inúmeras cidades da região, sendo essencial para a formação de seus munícipes. Também, está presente em inúmeras atividades de extensão, como por exemplo, os programas: Terceira Idade, Serviço de Assistência Jurídica, Fundação C.A.S.A, PROEJA, Ler e Aprender, Universidade vai à Escola, Fisioterapia, Enfermagem, Terapia Ocupacional, Núcleo de Esporte e Recreação, Unisaúde, AdministrAÇÃO e outros programas.

4 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

Nesta pesquisa será utilizado um processo denominado KDD (*Knowledge Discovery in Databases*), que, em Português, quer dizer Descoberta de Conhecimento em Banco de Dados. Este processo compreende a transformação de dados em conhecimento. Para que isso seja possível o KDD apresenta as seguintes etapas que a figura abaixo ilustra:



Fonte: FAYYAD, Usama; PIATETSKI-SHAPIRO, Gregory; SMYTH, Padhraic. **The KDD Process for Extracting Useful Knowledge from Volumes of Data.** In: Communications of the ACM, v.39, n.11, p.27-34, nov.1996

De acordo com Fayyad, Piatetski-Shapiro e Smyth (1996), KDD é um processo de várias etapas não trivial, interativo e iterativo, para identificação de padrões compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados.

O objetivo do KDD é descobrir conhecimento útil, válido e relevante através do uso de algoritmos, o que torna necessário passar pelos seguintes passos que a imagem acima ilustra e que serão detalhados abaixo:

- **Seleção** - A primeira etapa do processo da KDD é a seleção. É nesta etapa que se define a segmentação dos dados necessários para a análise de acordo com critérios pré-definidos. No caso desta pesquisa, foi neste passo que foram filtrados somente os dados dos alunos que ingressaram nos anos de 2006 a 2011.
- **Pré-Processamento** - A segunda etapa do KDD diz respeito à limpeza dos dados selecionados na etapa anterior. Com relação à pesquisa, foi

aplicada uma série de procedimentos para eliminar os dados inconsistentes. Esse processo será explicado detalhadamente no decorrer do trabalho.

- **Transformação** – Dependendo de qual informação é preciso trabalhar, é necessário transformar alguns dados. Por exemplo, foi necessário saber a faixa etária dos alunos, dado este que foi transformado da data de nascimento para a faixa etária; ou seja, é nesta etapa que os dados são transformados em um padrão necessário para posteriormente serem aplicados a uma ferramenta de mineração.
- **Mineração dos dados (*Data Mining*)** – É nesta etapa que se fará o uso de uma ferramenta (software) para aplicar as técnicas de mineração (algoritmos) nos dados já definidos das etapas anteriores.
- **Interpretação** – Esta é a última etapa do KDD. É nesta fase que será interpretado o resultado extraído da mineração. Caso o resultado seja irrelevante, será necessário refazer todas as etapas anteriores.

4.1 Mineração de dados e suas técnicas

De acordo com Silberschatz, Korth e Sudarshan (2006), a mineração de dados lida com “descoberta de conhecimento em banco de dados”. O termo mineração de dados refere-se ao processo de escavar grandes bancos de dados com objetivo de encontrar conhecimentos úteis que se encontram escondidos entre os dados.

Existem vários tipos de conhecimento que podem ser extraídos do banco de dados. Alguns desses conhecimentos podem assumir a forma de um conjunto de regras, equações, previsões de resultados futuros, identificação de padrões etc.

Carvalho (2001) define a mineração de dados como o uso de uma técnica automática de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações que não seriam facilmente descobertas a olho nu. O autor também relata que a mineração de dados pode ser utilizada em qualquer ramo que possua uma grande massa de dados.

Quanto maior o volume de dados que temos para minerar, mais preciso será o resultado da mineração, pois os algoritmos utilizados necessitam de grandes volumes para calibrarem e extraírem conclusões confiáveis.

Outra questão que está ligada diretamente ao bom resultado da mineração é a qualidade dos dados da base. As técnicas de mineração de dados necessitam que os bancos de dados estejam armazenados de uma forma consistente, padronizada e organizada. O recurso computacional disponível também é uma questão crucial, pois, as ferramentas de data mining necessitam de muito recurso para processar seus algoritmos.

Carvalho (2001) destaca que a mineração de dados pode ser aplicada em três situações:

1. Quando não se tem conhecimento nenhum sobre o fenômeno que está sendo trabalhado, Carvalho (2001) o nomeia de “Descoberta não supervisionada de relações”;
2. Quando se tem algum conhecimento sobre o campo de atuação ou alguma ideia sobre a área trabalhada, denominado Teste de Hipótese.
3. Quando se tem um nível mais elevado de conhecimento na área que deseja estudar, denominado Modelagem matemática dos dados.

Para cada uma destas situações, o autor detalha que a forma de se trabalhar muda. Por exemplo, na descoberta não supervisionada de relações, a pessoa não apresenta nenhum problema específico a ser resolvido, deixando que o computador vasculhe o banco de dados na busca de algo novo. Quando se trabalha com teste de hipótese, o pesquisador já possui alguma hipótese e utiliza a mineração de dados para comprovar se aquilo é verdade ou não. Já na modelagem matemática dos dados, o pesquisador tem o conhecimento extraído anteriormente por alguma das outras maneiras, mas necessita identificar algo específico, ou seja, a mineração de dados é utilizada para buscar uma informação específica de um conhecimento já extraído.

Independente da situação em que é aplicada a mineração de dados, todas necessitarão de uma técnica para sua realização. Existem várias técnicas que podem ser utilizadas nas ferramentas de mineração de dados. Contudo, cinco técnicas englobam todas as outras formas de representações. Abaixo será detalhada cada uma destas técnicas.

4.1.1 Classificação

Uma das tarefas cognitivas humanas mais realizadas no auxílio da compreensão do ambiente é a classificação. Carvalho (2001) afirma que essa técnica é uma das mais utilizadas nas ferramentas de *data mining*. Nós seres humanos estamos sempre classificando coisas à nossa volta. Conforme ilustra o autor, o ser humano cria classes¹ de suas relações como: colegas de trabalho, amigos, familiares, classes sociais, preconceitos etc. Para cada uma destas classes uma forma diferente de tratamento é definida. Assim, vamos criando estereótipos construídos com base no passado e prontos para serem utilizados no presente.

Da mesma maneira, os algoritmos de classificação das ferramentas de data mining buscam classificar os dados. Por exemplo, classificar os clientes em baixo, médio ou alto risco de empréstimo bancário; classificar os alunos em grande, médio ou baixo grau de evasão etc.

4.1.2 Estimativa

Carvalho (2001) escreve que estimar é determinar da melhor forma possível um valor baseando-se em situações similares, mas nunca exatamente iguais.

De acordo com Carvalho (2001, p. 22) “estimar algum índice é determinar um valor mais provável diante de dados do passado ou de dados de outros índices semelhantes sobre os quais se tem conhecimento”.

Para ilustrar essa ideia, vamos supor que uma escola deseja determinar o gasto dos alunos do ensino médio com doces na cantina e que para isso existam apenas índices de gastos com doces do ensino médio de outra escola dentro de uma faixa etária. Assim, será necessário criar uma estimativa com base em índices semelhantes.

4.1.3 Previsão

Diferentemente da estimativa, a previsão se preocupa com o valor futuro de um determinado índice, baseando-se em dados do comportamento passado deste mesmo índice. Como o próprio nome diz, não existe uma forma de verificar se o resultado encontrado pela técnica é verdadeiro; a única maneira de verificar o

¹ Estrutura que abstrai um conjunto de objetos com características similares

resultado é aguardar e comparar o quanto a previsão acertou. Carvalho (2001) relata que a previsão é uma das técnicas mais difíceis de serem executadas, não somente em uma ferramenta de data mining, mas também em nossa vida.

4.1.4 Análise de Afinidade

De acordo com Carvalho (2001), a análise de afinidade consiste em verificar as correlações e co-ocorrências que existam nos dados. Isso quer dizer que em uma massa de dados pode ocorrer que alguns itens estejam presentes juntos com certa frequência (correlação) ou que determinados fatos ocorram simultaneamente com uma probabilidade razoável (co-ocorrência).

4.1.5 Análise de Agrupamentos

Agrupar não é uma tarefa fácil de ser realizada, porém, a sua definição é mais simples. De acordo com Carvalho (2001), agrupar seria classificar uma massa de dados em classes desconhecidas. A dificuldade de criar esses agrupamentos está no fato dos grupos ou classes serem construídos com base na semelhança entre os elementos, cabendo ao pesquisador avaliar e encontrar um significado útil dos grupos encontrados.

O que distingue a técnica de agrupamento da técnica de classificação é que na classificação as possíveis classes já são conhecidas, enquanto na de agrupamento o analista será o responsável por defini-las.

Carvalho (2001) cita um exemplo para ilustrar melhor: imaginemos que temos uma massa de dados sobre o consumo no Brasil. Determinar quantas classes ou padrões de comportamento de consumo existam é algo complexo e requer muito trabalho do pesquisador.

Em cada uma das técnicas apresentadas por Carvalho (2001) existem alguns algoritmos matemáticos que poderão ser utilizados. Nesta pesquisa abordaremos dois algoritmos, um de classificação (árvore de decisão) e um de agrupamento (*cluster analysis*).

4.2 Árvore de decisão

Árvore de decisão é um algoritmo derivado da técnica de classificação. Este será o principal algoritmo utilizado nesta pesquisa. Com ele será possível organizar

regras de classificação em um diagrama no formato de uma árvore. Esta técnica ganhou este nome devido ao fato de seu resultado ser demonstrado como se fosse uma árvore, ou seja, existe um primeiro nó, denominado raiz e a partir deste vai se dividindo em outros nós, denominados nós folhas. Conforme descreve Palestino (2001, p. 190):

Árvore de decisão é uma técnica que, a partir de uma massa de dados, cria e organiza regras de classificação e decisão em formato de diagramas de árvore, que irão classificar suas observações ou prever resultados futuros. [...] Começa-se com um único grupo que reúne todos os casos em estudo. Na medida em que a árvore vai se expandindo, esta base é dividida em nódulos que representam categorias das variáveis analisadas.

Como colocado, a ideia desta técnica é, a partir de um grande volume de dados, encontrar uma variável, entre todas as variáveis de entradas fornecidas, que detecte com mais eficácia um indicador de comportamento. Sendo assim, o algoritmo cria uma ramificação a partir desta variável encontrada e começa com base nessa nova ramificação encontrar outras possíveis variáveis que tenham uma grande proporção e representatividade. Com esse mecanismo, serão criadas várias ramificações de acordo com as categorias encontradas e por fim temos uma possível regra que era intrínseca ao dado bruto.

O algoritmo só irá parar de buscar novas categorias nas ramificações quando as evidências forem poucas ou o volume de dados nessas ramificações também for reduzido. Por isso quanto mais dados tivermos, mais ramificações teremos e a chance de classificar, categorizar e descobrir uma regra no nosso dado cresce.

De acordo com Palestino (2001) a árvore de decisão pode ser empregada em vários momentos. Por exemplo, para identificar um grupo através de características comuns, especificar regras e determinar qual dado tem maior influência sobre o processo analisado. No manual da ferramenta de mineração de dados há um exemplo interessante que explica como esse algoritmo se comporta:

Para entender como uma árvore de decisão funciona, considere um exemplo, como o cenário de comprador de bicicleta[...]. Pressupondo que o atributo previsível corresponda às compras do cliente, o algoritmo de árvores de decisão tenta encontrar uma coluna de dados, entre todas as entradas fornecidas, que detecte com mais eficácia os clientes que provavelmente comprarão uma bicicleta e aqueles que provavelmente não comprarão. Por exemplo, talvez o modelo descubra que a idade é o melhor indicador do comportamento de compra. Especificamente, que os clientes com mais de 30 anos têm maior probabilidade de comprar uma bicicleta, e

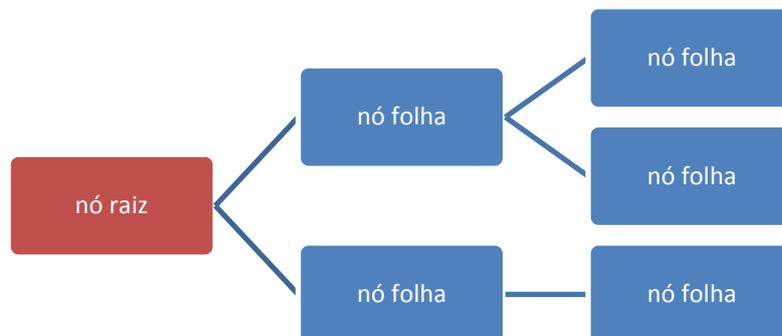
todos os outros clientes provavelmente não farão uma compra. Neste cenário, o modelo cria uma divisão no atributo Idade. Isso significa que a árvore se divide em duas ramificações, uma que contém os clientes com mais de 30 anos e outra com os clientes com menos de 30 anos. [...]

Para cada ramificação, o modelo continua procurando atributos adicionais a serem usados na diferenciação de clientes. Se houver poucas evidências nos dados para continuar criando subgrupos de clientes, o modelo parará de criar a árvore. O modelo também parará de criar a árvore sempre que o número de casos no nó for pequeno demais para continuar, independentemente de quão boa seja a divisão ou de o valor ser nulo ou estar ausente. Ao interromper o crescimento da árvore precocemente, você impede que o modelo treine muito perto de um determinado conjunto de dados.

Cada nó de árvore interior contém nós folha que fornecem uma análise dos resultados, de acordo com os resultados de classificação atuais. Por exemplo, talvez você tenha um nó interior que represente Idade ≥ 30 e Sexo = Masculino. O nó desse grupo mostra quantos clientes nessa categoria compraram ou não algo. [...] (MICROSOFT a)

A seguir é apresentado um exemplo do formato do diagrama da árvore de decisão para melhor explicar e representar o resultado que será obtido com o uso desta técnica. No exemplo podemos observar que existe um nó raiz que é a concentração bruta dos dados. A partir desta, várias ramificações vão surgindo, de acordo com as variáveis que contenham fortes evidências, originando assim os nós folhas.

Figura 2 - Exemplo de árvore



Fonte: Próprio autor.

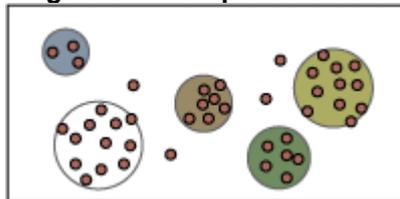
4.3 Análise por agrupamento (Cluster Analysis)

Como um segundo algoritmo a ser utilizado caso o algoritmo de árvore de decisão não venha a encontrar alguma classificação útil, utilizaremos o algoritmo de *Cluster* para assim criar agrupamentos dos dados, buscando identificar um grupo ou uma classe que determine a existência de um padrão no perfil dos alunos com relação ao curso.

De acordo com Palestino (2001, p. 196), “o objetivo desta técnica é identificar a existência de diferentes grupos dentro de um conjunto de dados, com isso conseguiremos agrupar os que possuem características semelhantes”.

O algoritmo de *Cluster* primeiro coloca todos os casos em plano, onde cada caso é representado por um ponto no gráfico de dispersão. Após realizada essa tarefa, o algoritmo começa a criar os *clusters* (grupos) com os pontos que estão mais próximos um do outro. Após definir os primeiros *clusters*, o algoritmo começa a recalcular os *clusters* buscando uma representação mais satisfatoriamente. Quando o mesmo não consegue mais melhorar a redefinição dos clusters esse encerra a execução.

Figura 3 - Exemplo de cluster



Fonte: Microsoft b.

O manual da ferramenta de mineração apresenta um exemplo de quando devemos utilizar o algoritmo de cluster

Considere um grupo de pessoas que compartilha informações demográficas similares e que compram produtos semelhantes da empresa Adventure Works. Esse grupo de pessoas representa um cluster de dados. Podem existir vários desses clusters em um banco de dados. Observando as colunas que formam um cluster, é possível observar claramente como os registros em um conjunto de dados são relacionados entre si. (MICROSOFT b)

4.4 Casos de Mineração de Dados

Abaixo são relatados alguns casos e trabalhos acadêmicos que utilizaram o processo de KDD para descoberta de padrões e comportamentos:

4.4.1 Descoberta de Conhecimento sobre o Processo Seletivo da UFPR

Este trabalho foi desenvolvido para obtenção do grau de Mestre em Ciências do programa de Pós-Graduação em Métodos Numéricos em Engenharia da Universidade Federal do Paraná.

O Objetivo geral da pesquisa foi o de delinear o perfil do candidato ao processo seletivo do ensino superior da Universidade Federal do Paraná. Para isto, o autor aplicou a técnica de mineração de dados (Data Mining) utilizando o algoritmo de Árvore de Decisão.

Ele utilizou um recorte dos vestibulares realizados em dezembro de 2003 para o ano letivo de 2004, no qual coletou dados dos questionários sócio-educacionais preenchidos pelos candidatos no momento da inscrição. O autor também utilizou as notas obtidas pelo candidato nas provas, na redação e no ENEM.

O autor relata, no decorrer da pesquisa, os passos que foram dados para selecionar, limpar e aplicar os dados numa ferramenta de mineração de dados. No final de todo o processo ele relata alguns resultados que conseguiu obter. São eles:

- Nos cursos mais concorridos, os dados socioeconômicos e culturais do candidato são relevantes para o seu bom desempenho. O mesmo não ocorre nos cursos menos concorridos.
- As pontuações obtidas nas provas de outras áreas diferentes daquelas à qual o curso pertence influenciaram na aprovação dos candidatos a cursos das áreas de Exatas e Humanas.
- Percebeu-se, também, através das regras geradas, que a nota em Língua Estrangeira é um atributo importante no sucesso do candidato em praticamente todos os cursos, independente da área.
- Notou-se que a nota de redação, pelo fato de estar presente apenas nos registros de candidatos classificados ou aprovados, prejudicou a análise das regras geradas. (MARTINHAGO,2005)

O autor relata no decorrer da pesquisa outros trabalhos que foram desenvolvidos utilizando a aplicação da mineração de dados:

- Na Universidade Federal de Minas Gerais (UFMG), as técnicas de mineração de dados vem sendo utilizadas na determinação do perfil dos alunos, com base nos dados de pesquisa econômico-social. Essas informações vêm sendo correlacionadas com o desempenho dos alunos no vestibular e mesmo durante o curso de graduação (CESAR,2000)
- O artigo apresentado por Freitas Junior et al. (2001), mostra um estudo de caso no qual foram aplicadas técnicas de mineração, mais especificamente, as regras de associação, sobre uma base de dados do quadro de docentes d Universidade Estadual de Maringá (UEM). O objetivo da mineração foi descobrir o perfil do corpo docente da UEM, extraindo conhecimentos tais como, a relação entre titulação e produção acadêmica, a relação entre regime de trabalho e produção acadêmica. [...]
- A pesquisa de Pansanato e Soares(1999) avaliou a capacidade preditiva do resultado do ENEM em relação ao desempenho dos candidatos ao vestibular da UFMG de 1999. O resultados obtidos demonstram que a utilização dos resultados de ENEM, em substituição à primeira etapa do vestibular, não traria prejuízo para a universidade ou para os alunos envolvidos.

- Em 1998, a pesquisa de Soares e Fonseca teve como principal objetivo analisar o desempenho dos candidatos ao vestibular da UFMG em 1997. Para isto, foram utilizadas várias características socioeconômicas coletadas através do questionário respondido pelos candidatos que se inscreveram ao vestibular. Os resultados verificaram que estes fatores estão, como previsto, fortemente associados com o desempenho dos alunos, embora tenham no conjunto, pequena capacidade preditiva.[...] (MARTINHAGO,2005).

4.4.2 Aspectos teóricos do data mining e aplicação das redes neurais em previsões de preços agropecuários

A pesquisa foi realizada na Universidade Federal de Viçosa. Seu objetivo geral é o de avaliar a qualidade das previsões de preços agropecuários obtidas com o emprego da ferramenta de data mining. Para isso, a pesquisa analisou dois artigos científicos que usaram a ferramenta de mineração de dados. O primeiro artigo realizou a previsão do preço da soja e o segundo a previsão da arroba do boi gordo.

O autor destaca que a análise do comportamento dos preços no ramo agropecuário é fundamental para avaliar os riscos e a viabilidade econômica. Sendo assim, o autor busca analisar se o uso de mineração de dados na previsão do preço poderia proporcionar bons resultados e que de fato esses representassem a realidade do setor.

No decorrer da pesquisa é relato alguns casos de mineração de dados

- a) A Wal-Mart constitui uma das maiores cadeias varejistas dos Estados Unidos. É conhecida tanto por sua política de baixos níveis de estoque e ressurgimento constante de produtos (baixos lotes e alta frequência), como por sua política agressiva com os concorrentes regionais. Utilizando ferramentas de data mining, que auxiliam na precisão de cada item transacionado nas lojas da empresa, essa empresa modificou seus sistemas e ressurgimento automático de produtos. Além disso, identificou padrões de consumo, em cada loja, para a escolha do mix de produtos a ser ofertado.
- b) A ShopKo, rede varejista americana, utilizou ferramentas de data mining para determinar quais produtos eram vendidos por meio da venda indireta de outros produtos[...]
- g) No Vestibular PUC-RJ, utilizando as técnicas de data mining, um programa de obtenção de conhecimento, depois de examinar milhares de alunos, forneceu a seguinte regra: se o candidato é do sexo feminino, trabalha e teve aprovação com boas notas, então não efetiva matrícula. Uma reflexão justifica essa regra: de acordo com os costumes do Rio de Janeiro, uma mulher com idade para realizar o vestibular, se trabalha é porque precisa, e nesse caso deve ter feito, também, inscrição para ingressar na universidade pública gratuita. Se teve boas notas, provavelmente, foi aprovada na universidade pública, onde efetivará a matrícula.[...]

Nas considerações finais o autor conclui que a mineração de dados é um processo poderoso para transformar dado em conhecimento, porém sua utilização possui limitações. O autor cita que a principal limitação é com a análise dos resultados gerados. Isso porque, em geral, não é qualquer pessoa que pode interpretá-los. Faz-se necessária a utilização de profissionais especializados. Ao término ele conclui que os resultados encontrados nos dois artigos foram satisfatórios e que a mineração de dados é uma possível solução para previsão dos preços agropecuários.

4.4.3 Descoberta de conhecimento aplicado a dados eleitorais

O artigo publicado na revista *Gestão e Conhecimento Online* tem como objetivo encontrar uma possível tendência no tocante às pessoas que almejam se candidatar e conseqüentemente se eleger deputado estadual. Para isso, Tavares, Bozza e Kono (2007) selecionou dados das eleições de 1994 e 1998.

Após a limpeza, padronização dos dados e a aplicação de um algoritmo de árvore de decisão, o autor chegou às seguintes regras:

- Homens têm uma representatividade muito maior na câmara dos deputados.
- Mulheres têm um desempenho muito melhor do que os homens, quando estas se elegem.
- Nível de escolaridade dos candidatos eleitos é maior ou igual ao superior completo.
- Os candidatos eleitos estavam na faixa dos 45 aos 55 anos.
- Duas profissões tiveram o maior destaque que é a de comerciante e advogado.

5 MINERAÇÃO DE DADOS – UNISO

Neste capítulo será explicado e detalhado como ocorreu a pesquisa, desde a coleta dos dados passando pela “limpeza”, aplicação das ferramentas de data mining até a chegarmos aos resultados e sua análise.

Vale destacar que ao iniciarmos o trabalho de coleta e limpeza dos dados armazenados no banco de dados da Universidade de Sorocaba observamos uma grande inconsistência nos dados anteriores ao ano 2006. Várias informações que poderiam ser importantes para a pesquisa encontravam-se incompletas, não preenchidas e em alguns casos incorretas, dificultando assim a sua utilização. Com isso, ficou definido que trabalharíamos com os alunos ingressantes nos anos 2006 a 2011, pois, anterior a esse período a instituição utilizava outro sistema de gestão e ao implantar um novo sistema muitas informações acabaram sendo descartadas.

Após definido o corte com que seria trabalhado nesta pesquisa, passamos por uma fase de limpeza dos dados, A partir deste momento foram deixados de lado os dados que apresentavam alguma inconsistência ou aqueles que apresentavam alguma suspeita de anomalia, pois esses poderiam distorcer a pesquisa e gerar um resultado duvidoso.

Para ilustrar o problema de inconsistência/anomalia citaremos a variável data de nascimento. Em alguns casos essa informação era inválida ou por não possuir dado ou por apresentar uma data muito recente que não seria possível à pessoa estar matriculada em um curso superior com uma data nascimento muito recente.

O trabalho de mineração de dados foi dividido de acordo com as hipóteses apresentadas, ou seja, primeiro foram explorados os dados socioeconômicos dos alunos, depois os dados do processo seletivo e por fim o cruzamento dos dados socioeconômicos com os do processo seletivo.

O trabalho está dividido em subcapítulos. Cada subcapítulo apresenta a verificação de uma hipótese, começando pelos dados socioeconômicos, passando para os dados do processo seletivo e, por fim, o cruzamento dos dados socioeconômicos com os do processo seletivo.

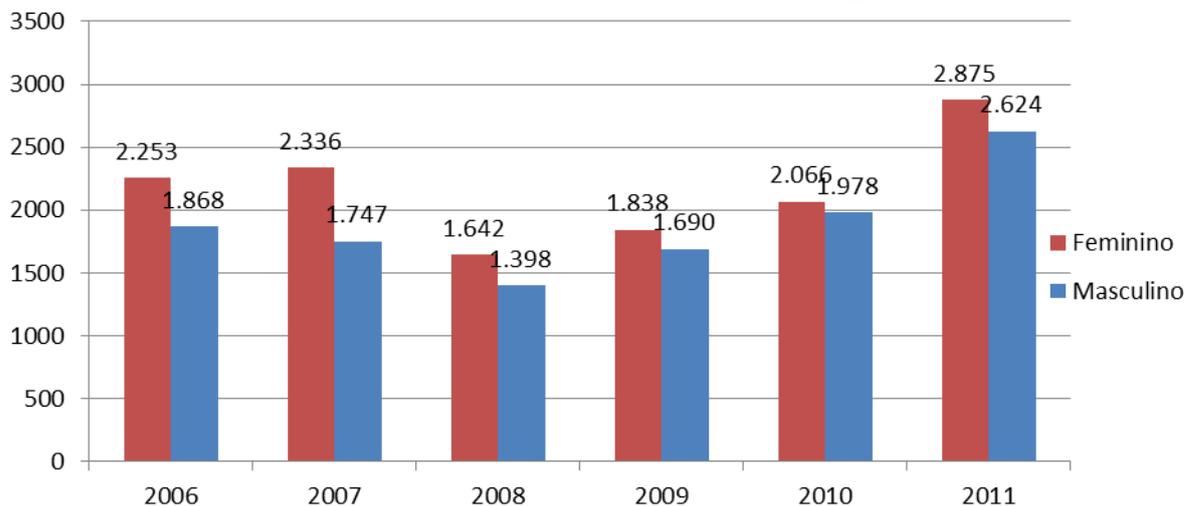
5.1 Dados socioeconômicos e demográficos dos estudantes

Primeiramente são analisados os dados socioeconômicos e demográficos dos estudantes, que são captados pela Universidade no momento da matrícula; ou seja, após passar no vestibular o estudante se dirige à instituição trazendo seus documentos e assim efetivando sua matrícula em um determinado curso. É neste momento que são coletadas informações que serão utilizadas para atingir o objetivo do presente estudo.

Inicialmente trabalhamos com uma forma de consulta direta no banco de dados, buscando algumas informações específicas, sem a utilização de uma ferramenta de mineração de dados. Esse tipo de consulta é utilizado quando sabemos onde está a informação que queremos e temos em mente o que buscar, diferentemente de uma ferramenta de *Data Mining* na qual muitas vezes não sabemos o que estamos procurando ou qual será o resultado extraído.

Desta forma, em um primeiro momento, a identificação é feita de acordo com o gênero dos estudantes matriculados ou egressos da instituição delimitando o período de ingresso entre 2006 a 2011. Após a finalização dessa etapa foram obtidos os seguintes resultados que são apresentados a seguir

Gráfico 11 - Quantidade de estudantes por ano de ingresso/sexo



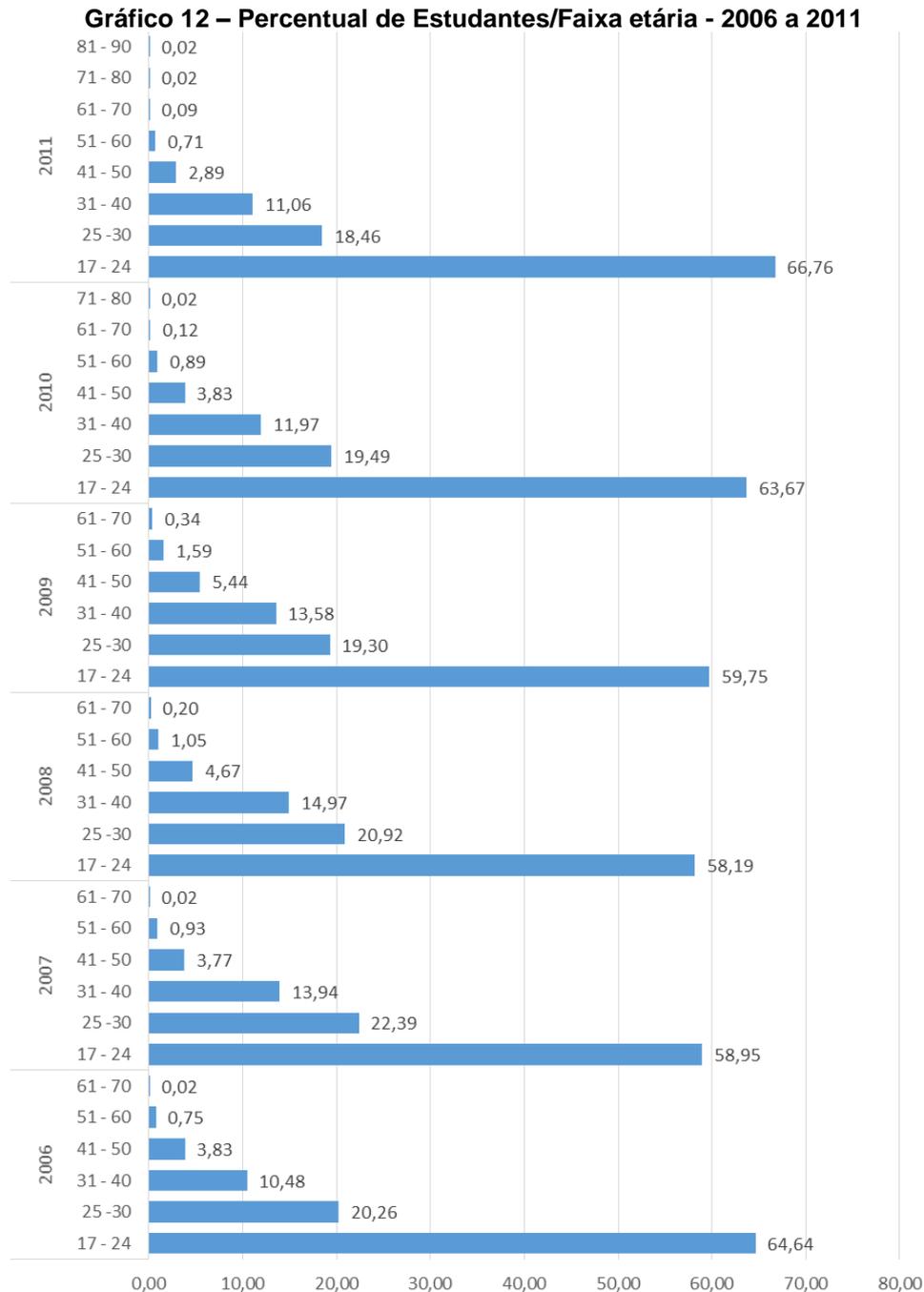
Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Como pode ser apontado no gráfico, desde 2006 temos um maior número de mulheres frequentando a Educação Superior, tendência já apontada por Segre

(2006), relativamente a América Latina, em que 54% dos estudantes são do sexo feminino.

Outro levantamento realizado é sobre a faixa etária destes estudantes. Qual a faixa etária que apresenta maior número de ingressantes? Fazendo a consulta no banco de dados da instituição, obtivemos o seguinte resultado:



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Para melhor ilustrar, foi gerado o gráfico por ano letivo e faixa etária. Esta faixa etária foi criada para conseguirmos melhor agrupar e visualizar os resultados. Podemos observar que a grande maioria dos ingressantes por ano são estudantes com idade inferior a 24 anos.

Analisando essas duas variáveis, sexo e idade, de conjunto, observa-se que o maior número de ingressantes na Universidade concentra-se na faixa de 24 anos ou menos e em sua maioria mulheres.

Cruzar sexo e idade com a opção do curso escolhido se torna inviável e de difícil aplicação utilizando essa forma de consulta diretamente no banco de dados. Sendo assim, o que veremos no capítulo a seguir é a aplicação de uma ferramenta de mineração de dados que tem por finalidade trabalhar com algoritmos matemáticos buscando previsões e relações existentes em grandes volumes de dados.

5.1.1 Mineração dos dados socioeconômicos demográficos dos estudantes

Para trabalharmos com uma ferramenta de mineração dos dados, primeiramente precisamos passar pelas etapas de seleção, transformação e processamento dos dados. É necessário percorrer cada uma das etapas em ordem para que consigamos atingir nosso objetivo. Para tanto, primeiramente, precisam ser selecionados todos os dados para a realização da mineração e transformá-los/armazená-los em um padrão adequado para que possam ser processados por uma ferramenta de mineração de dados.

A primeira tarefa de seleção dos dados necessitou de um conhecimento técnico para explorar o banco de dados da instituição. Mediante a utilização de uma linguagem de manipulação de banco de dados, denominada SQL, conseguimos encontrar todas as informações necessárias para trabalharmos com os dados socioeconômicos.

Após ter selecionado os dados necessários, foram realizadas algumas transformações para criar dados novos ou para adequá-los a nossa necessidade. Foi o caso dos campos:

- Idade: para que consigamos trabalhar com a idade dos alunos, foi necessário transformar a data de nascimento em idade. Para tanto,

verificamos o ano de ingresso do aluno e realizamos o cálculo para se chegar à idade.

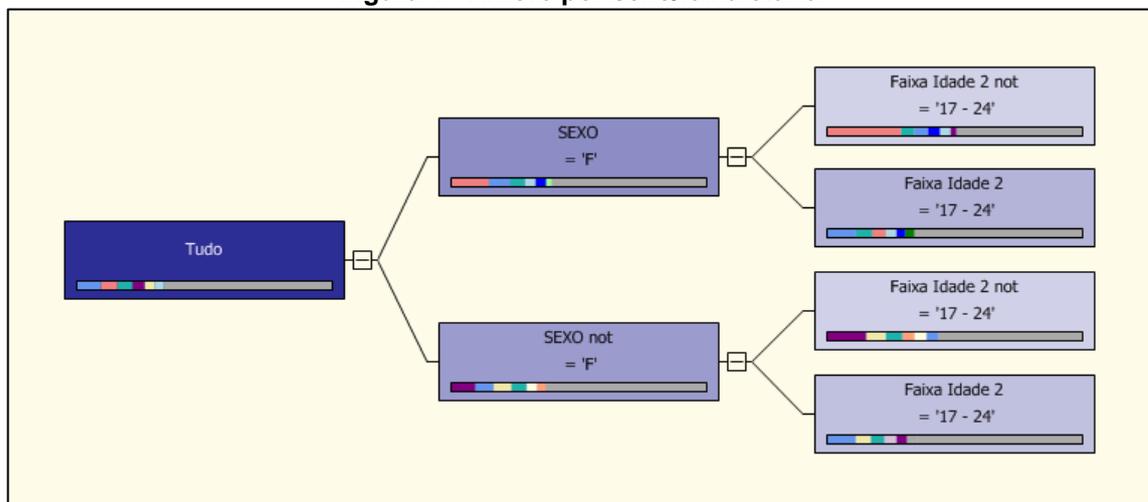
- Faixa etária: tendo a informação idade, definimos algumas faixas de idade, como por exemplo: 17 – 24, 25 – 30 etc.
- Acentuação nos registros: a ferramenta de mineração de dados não permite acento nos registros, por isso, foi necessário retirar toda acentuação. Por exemplo: Estado civil igual a “Viúvo”, tivemos que transformar em “Viuvo”.

Tendo realizada a seleção dos dados e executadas as transformações necessárias, foi criada uma estrutura no banco de dados para armazená-los e assim podermos submetê-los a uma ferramenta de mineração de dados.

Na primeira análise trabalhamos com as variáveis: sexo e idade aplicados a um algoritmo denominado árvore de decisão. A finalidade desse algoritmo que se caracteriza por nós e ramificações é subdividir os dados analisados em busca de regras que existem de acordo com sua proporção e assim poder classificá-las.

Ao executar a ferramenta de mineração de dados configurado para prever o curso do aluno de acordo com seu sexo e idade, chegamos à seguinte árvore:

Figura 4 - Árvore por sexo/faixa etária



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Analisando mais precisamente essa árvore obtivemos a seguinte classificação: Se sexo for feminino e faixa etária maior que 24 anos o curso que tem maior procura é o de pedagogia com 26,90% das matrículas. Já se a faixa etária for até 24 anos, o resultado ficou bastante dividido em três cursos distintos. São eles: administração com 9,88%, direito com 6,62% e pedagogia com 5,56%.

Fica claro com esse resultado que a probabilidade de mulheres com mais de 24 anos optarem pelo curso de pedagogia é muito forte e se destaca completamente das demais probabilidades. Abaixo segue uma tabela que representa essas informações destacadas pela ferramenta de mineração de dados.

Tabela 8 – Curso escolhido, sexo feminino e faixa etária

Curso	Sexo Feminino e Maior que 24 anos	Sexo Feminino e Menor ou igual a 24 anos
Pedagogia	26,90%	5,56%
Administração		9,88%
Direito		6,62%

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Em outra ramificação da árvore podemos encontrar a seguinte classificação: se o sexo for masculino e faixa etária maior que 24 anos o curso que tem maior procura é o de tecnologia em gestão de produção industrial, com 13,96% de probabilidade das matrículas. Caso a idade seja menor ou igual a 24 anos o curso que tem maior procura é o de administração, com 8,97%, acompanhado pelos cursos de engenharia de produção (6,46%) e direito (5,62%).

Tabela 9 - Curso escolhido, sexo masculino e faixa etária

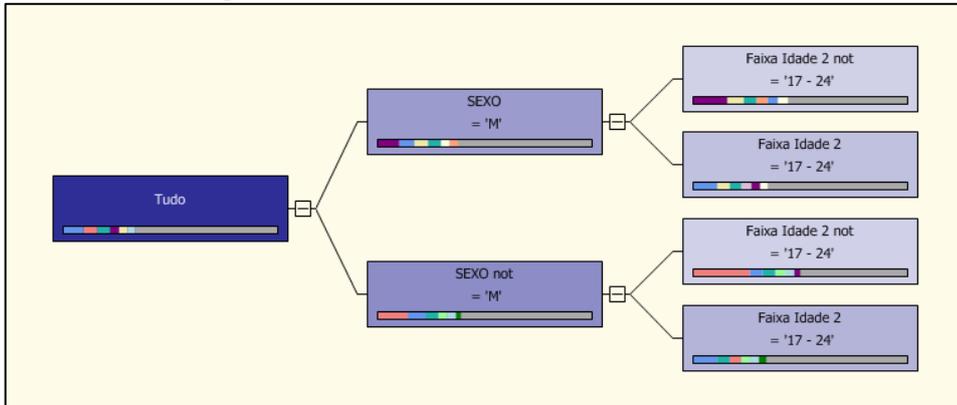
Curso	Sexo Masculino e Maior que 24 anos	Sexo Masculino e Menor ou igual a 24 anos
Tecnologia em gestão de produção industrial	13,96%	
Administração		8,97%
Engenharia de Produção		6,46%
Direito		5,62%

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Após uma análise detalhada das classificações encontradas, foi notado que os alunos do programa PARFOR também estavam sendo considerados nesta primeira mineração, podendo assim distorcer a classificação encontrada. Desta forma, foi necessário voltarmos para a etapa de seleção dos dados para retirar estes alunos. Com esta limpeza realizada, foi executada novamente a mineração que gerou a seguinte árvore:

Figura 5 - Árvore sexo/faixa etária sem Parfor



Fonte: Banco de dados da UNISO.

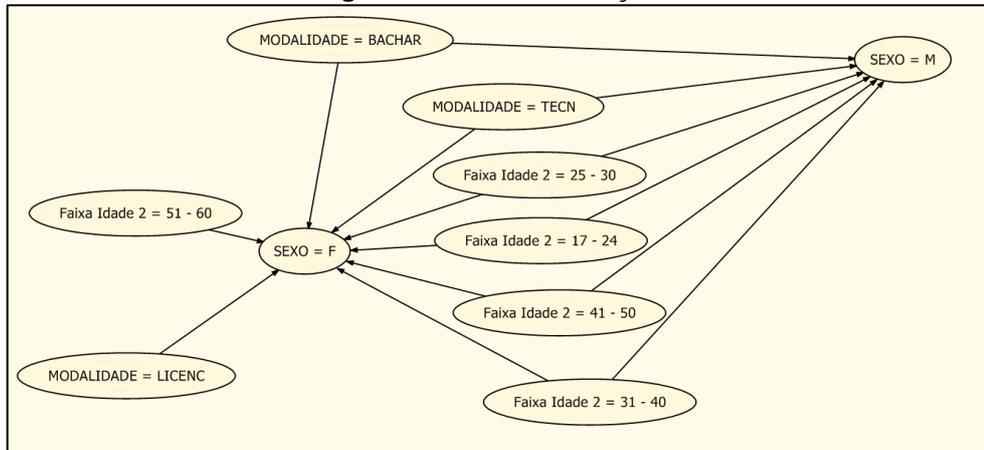
Nota: Elaborado pelo autor.

Como pode ser observado, o resultado obtido foi o mesmo que o anterior, mudando apenas a porcentagem de mulheres que cursa pedagogia. Anteriormente, o resultado apresentado foi de 26,90%, agora, sem os alunos do PARFOR, o número caiu para 25,12%, continuando assim uma grande concentração de mulheres com idade superiores a 24 anos que optam por pedagogia.

Após encontrarmos essa tendência, fizemos uso de outro algoritmo denominado “Regra de Associação” para comprovarmos o resultado obtido. Este algoritmo utiliza as variáveis e dados criando associações que possuem uma forte relação. Ao invés de utilizarmos o curso, utilizamos a modalidade do mesmo, como por exemplo: bacharel, tecnológico e licenciatura.

Na primeira execução obtivemos as seguintes associações:

Figura 6 - Todas as relações

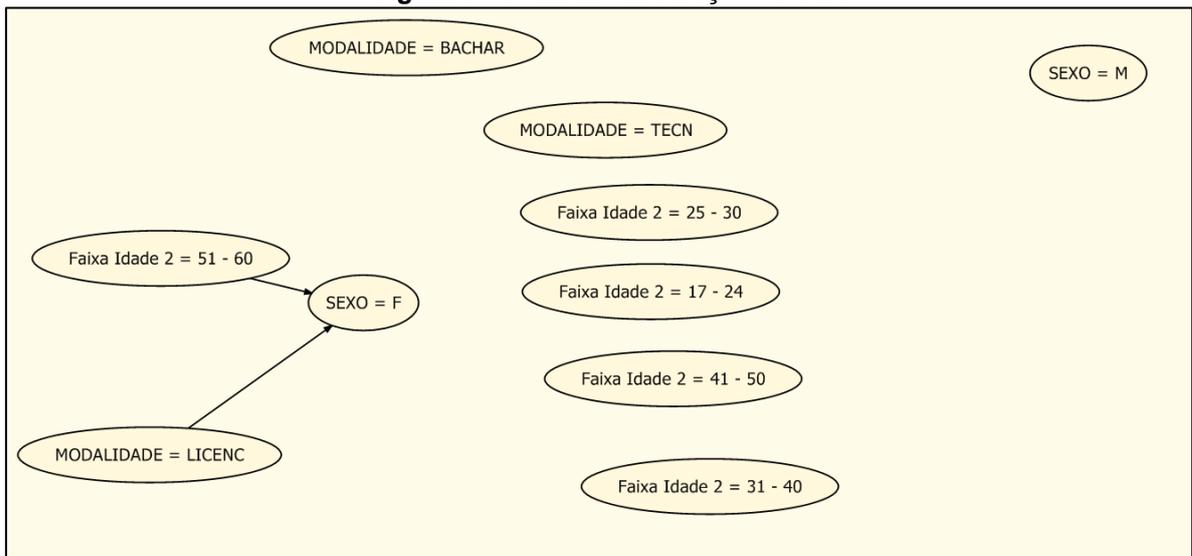


Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Nessa primeira imagem gerada pela ferramenta obtivemos todas as relações que existem. Por exemplo, o sexo feminino se relaciona com as faixas etárias e com as modalidades de licenciatura, bacharel e tecnológica. Até esse ponto, nenhuma novidade, mas quando alteramos a configuração da ferramenta para mostrar apenas as associações que possuem uma forte relação, obtivemos o seguinte resultado:

Figura 7 - Somente as relações fortes



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

É possível observar que, ao configurarmos a ferramenta para mostrar somente os dados com associações fortes, o sexo feminino tem alta relação com a modalidade de licenciatura e faixa etária entre 51 e 60 anos, reforçando assim a regra encontrada com o algoritmo de árvore de decisão que indicou que a probabilidade de mulheres com mais de 24 anos optam pelo curso de pedagogia era de 25,12%.

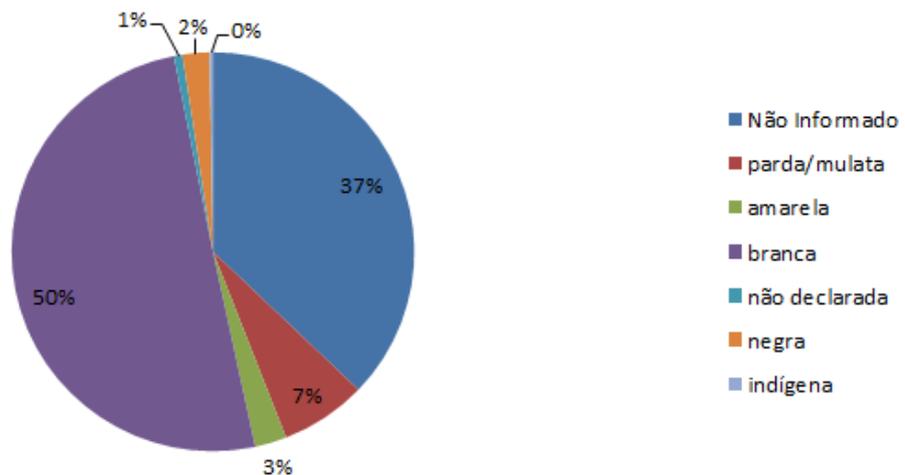
Adicionamos a esse modelo a variável cor/raça para tentarmos verificar a probabilidade da escolha do curso através deste dado. Porém sua adição não gerou um resultado diferente do qual já havíamos obtido. Isso quer dizer que a ferramenta de mineração de dados não conseguiu classificar os dados mediante essa variável e, por tanto, essa não possui uma forte representatividade dentro de um determinado curso, ou seja, para nossa pesquisa, a cor/raça da pessoa não impacta na escolha do curso, pois não há uma divisão clara que realce cor/raça associada ao curso.

Um problema identificado, que pode ter agravado a não utilização deste dado para classificação, é que este não é uma informação obrigatória no sistema de gestão da instituição. Sendo assim, um grande número de estudantes não possuem essa informação na base de dados da Universidade.

Por outro lado, observamos a prevalência da cor/raça branca com relação às demais, conforme demonstra o gráfico:

Gráfico 13 - Percentual de alunos por cor/raça

Percentual de alunos por cor/raça



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Ao todo utilizamos as seguintes variáveis: sexo, idade, estado civil, município, faixa etária e cor/raça. O algoritmo de árvore de decisão só conseguiu encontrar uma classificação para as variáveis sexo e faixa etária, ou seja, as demais não tem representatividade na definição do curso para que a ferramenta criasse uma ramificação na árvore.

5.2 Dados do processo seletivo

Como colocado, a segunda hipótese a ser explorada são os dados do processo seletivo. Desta forma, iniciamos o trabalho levantando essas informações junto ao banco de dados da instituição.

Nossa primeira tarefa foi a de limpeza dos dados para podermos tabulá-los e aplicá-los ao processo de mineração de dados. Neste trabalho de limpeza, encontramos nossa primeira dificuldade, pois, os questionários informativos que o

aluno preenche no dia da prova do vestibular não possuíam um padrão. Ou seja, no passar dos anos, a UNISO foi alterando/reformulando suas perguntas ou adicionando novas e removendo outras. Assim, para cada processo seletivo tínhamos um modelo de questionário diferente.

Foi Identificado que o questionário informativo de 2011 possui apenas 13 questões, enquanto os questionários anteriores possuíam 30 questões. Sendo assim só poderemos trabalhar com questões que são comuns aos processos seletivos, de forma que os candidatos analisados tenham respondido à mesma questão.

Outra dificuldade encontrada foi que a ordem das questões não era a mesma. Da mesma forma, as alternativas de respostas variavam de acordo com o processo seletivo. Vale destacar também que em um semestre podemos ter muitos processos seletivos com datas/horários de provas diferentes, o que pode caracterizar um novo questionário informativo. Ao todo, foram analisados 48 processos seletivos buscando padronizar as questões e as alternativas de resposta.

Ao término desta análise, conseguimos identificar as questões comuns a todos os processos seletivos. São elas:

1. Renda do candidato;
2. Número de dependentes que vivem com a renda;
3. Se o candidato já iniciou algum curso superior;
4. Atividade que desempenha;
5. Meio que utiliza para manter-se informado;
6. Se o candidato possui acesso a internet;
7. Motivos da escolha do curso;
8. Conheceu a UNISO através de qual meio;
9. Quantidade de livros que lê em média por ano;
10. Exerce alguma atividade remunerada;
11. Turno de conclusão do ensino médio;

Fora essas questões que o candidato responde no dia do processo seletivo, há outras que ele preenche no momento da sua inscrição, como: curso a que está se candidatando, turno do curso, data de nascimento (conseguimos identificar através deste dado a idade do candidato na época em que prestou o concurso), sexo, estado civil, município em que reside, classificação no curso, nota obtida na prova e se efetivou ou não a matrícula.

Como nosso objetivo é identificar um padrão quando o aluno opta por um curso, consideramos apenas os candidatos que foram classificados (os desclassificados não efetivam sua matrícula) para que adiante consigamos cruzar os dados socioeconômicos coletados no momento da matrícula com os do processo seletivo.

Depois de realizado o processo de limpeza das questões a serem trabalhadas e restringidos os candidatos somente aos classificados, realizamos um processo de limpeza das respostas e das questões do questionário informativo.

Foi identificado que as alternativas de respostas mudavam de um processo seletivo para outro. Dessa maneira, foi necessária uma tabulação dessas respostas para melhorar a precisão da mineração de dados.

Abaixo serão detalhadas as possíveis respostas encontradas nos processos seletivos, como ficou sua padronização e a quantidade de resposta para cada pergunta:

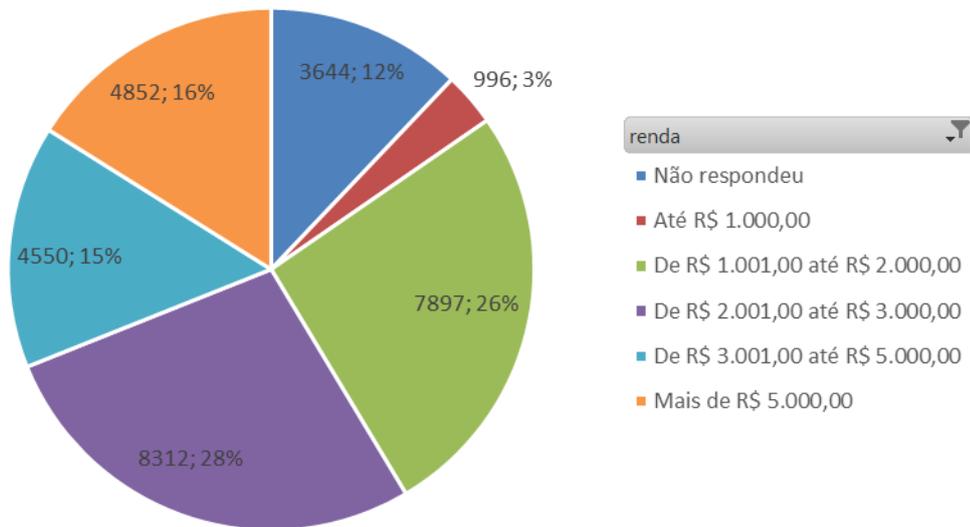
- Renda do Candidato

Quadro 1 - Renda do Candidato (Original/Padronizada)

Resposta Original	Resposta Padronizada
Até R\$ 1.000,00	Até R\$ 1.000,00
De 1 até 3 Salários Mínimos	De R\$ 1.001,00 até R\$ 2.000,00
De R\$ 1.001,00 até R\$ 2.000,00	
Mais de 3 até 5 Salários Mínimos	De R\$ 2.001,00 até R\$ 3.000,00
De R\$ 2.001,00 até R\$ 3.000,00	
Mais de 5 até 7 Salários Mínimos	De R\$ 3.001,00 até R\$ 5.000,00
De R\$ 3.001,00 até R\$ 5.000,00	
Mais de 15 até 20 Salários Mínimos	Mais de R\$ 5.000,00
Mais de 20 Salário Mínimos	
Mais de 20 salários mínimos	
Mais de R\$ 5.000,00	
Mais de 10 até 15 Salários Mínimos	
Mais de 7 até 10 Salários Mínimos	Não respondeu
Não respondeu	
Não Informada	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 14 - Renda do candidato

Fonte: Banco de dados da UNISO

Nota: Elaborado pelo autor.

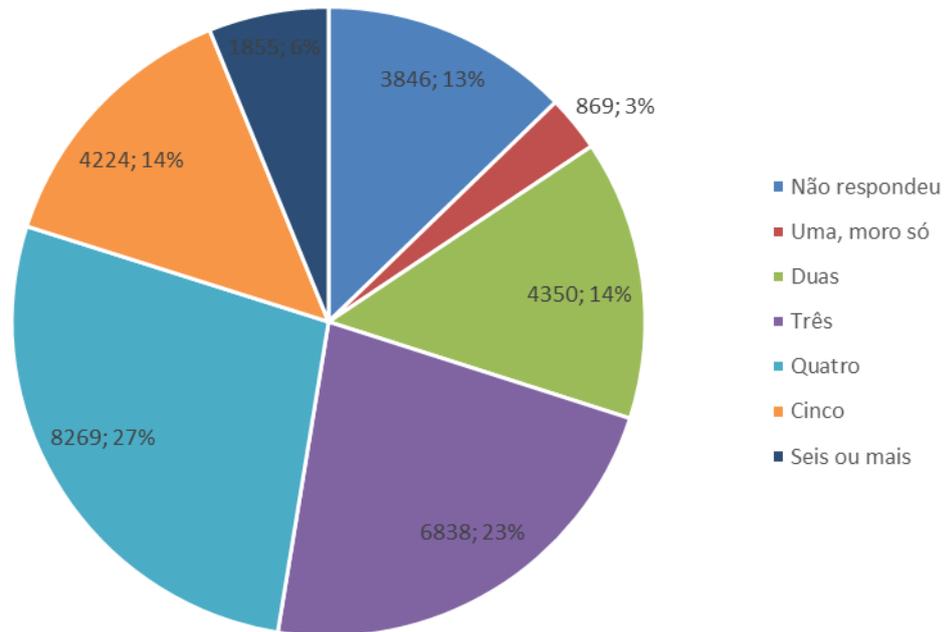
- Número de dependentes que vivem com a renda

Quadro 2 - Número de dependentes que vivem com a renda (Original/Padronizada)

Resposta Original	Resposta Padronizada
Uma, moro só	Uma, moro só
Uma (moro só)	
Duas	Duas
Três	Três
Quatro	Quatro
Cinco	Cinco
Seis ou mais	Seis ou mais
Não respondeu	Não respondeu
Não Informada	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 15 - Número de dependentes que vivem com a renda

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Se o candidato já iniciou algum curso superior

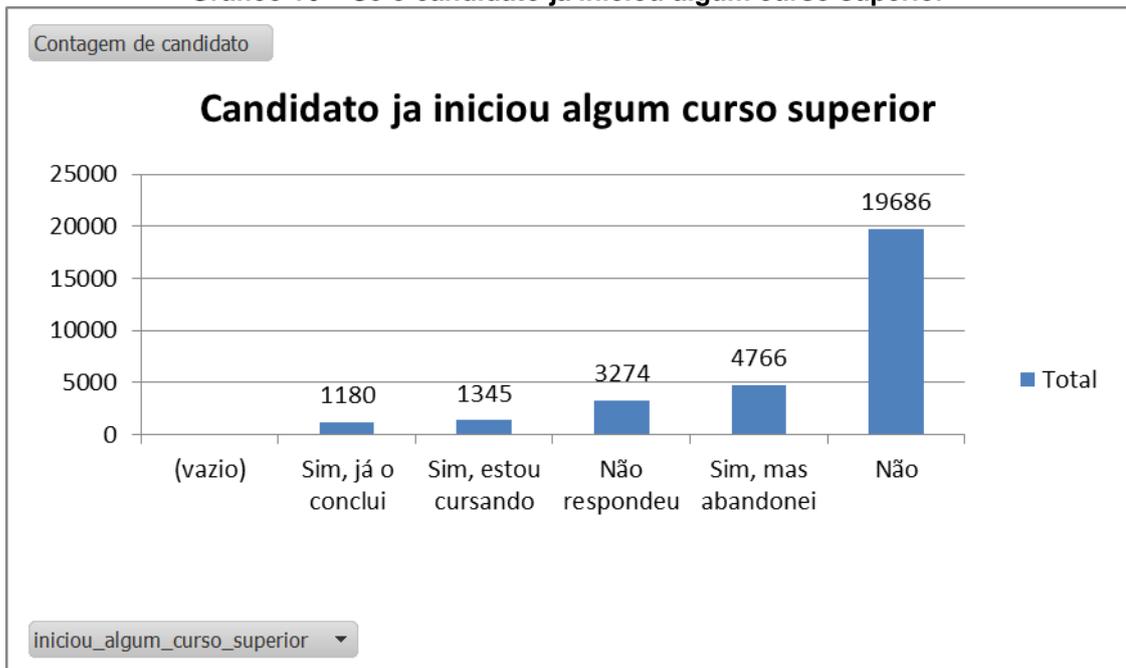
Quadro 3 – Se o candidato já iniciou algum curso Superior (Original/Padronizada)

Resposta Original	Resposta Padronizada
Não	Não
Sim, estou cursando	Sim, estou cursando
Sim, mas abandonei	Sim, mas abandonei
Sim, já o conclui	Sim, já o concluí
Sim, já o concluí	
Sim, já conclui	
Não respondeu	Não respondeu
Não Informada	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 16 – Se o candidato já iniciou algum curso superior



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Atividade que desempenha

Quadro 4 - Atividade que desempenha (Original/Padronizada)

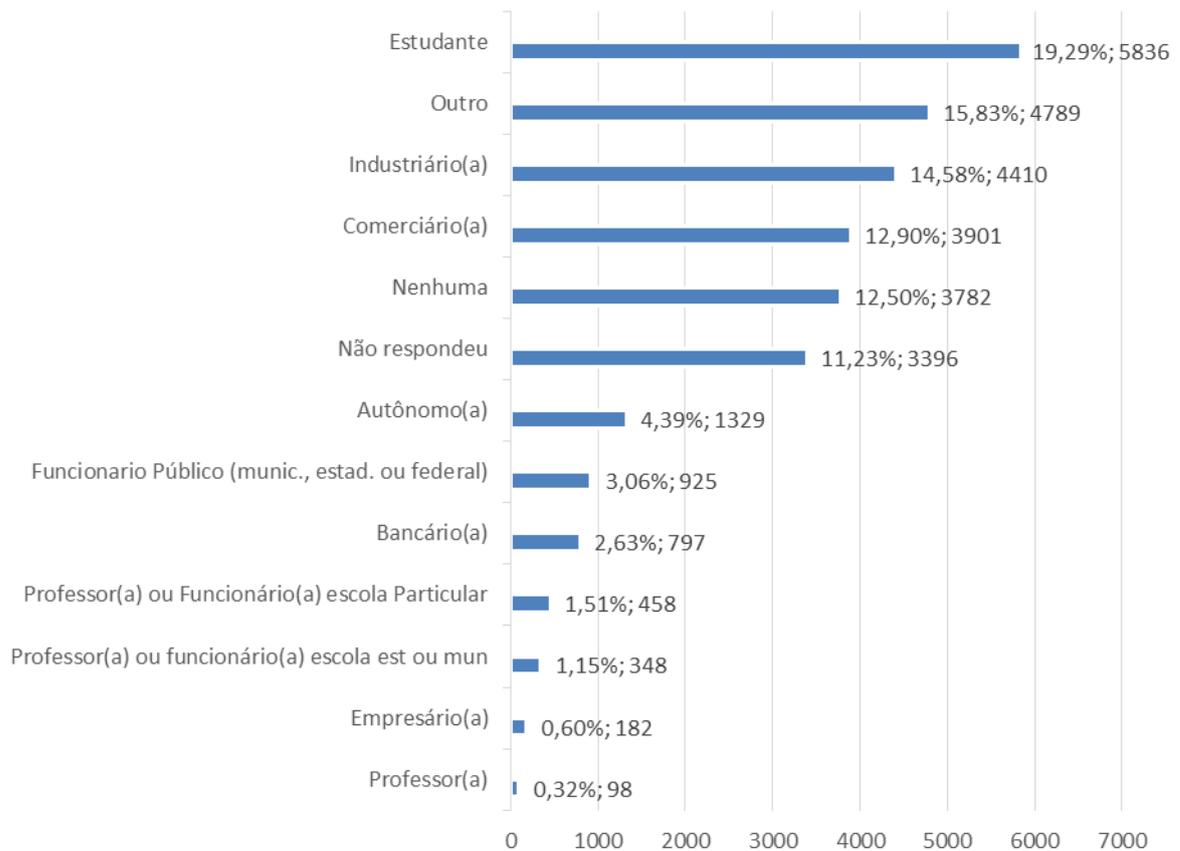
Resposta Original	Resposta Padronizada
Nenhuma	Nenhuma
Nenhum	
Comerciário	Comerciário(a)
Comerciário(a)	
Não respondeu	Não respondeu
Não Informada	
Bancário	Bancário(a)
Bancário(a)	
Autônomo	Autônomo(a)
Autônomo(a)	
Funcionário(a) Público(a)(exceto professor(a))	Funcionário Público (munic., estad. ou federal)
Funcionário(a) Público(a) (mun, est, oe Fed)	
Funcionario Público (munic., estad. ou federal)	
Funcionário(a) público(a) (mun, est, fed)	
Industriário(a)	Industriário(a)
Industriário	
Professor ou Funcionário de Escola Est. ou Munic.	Professor(a) ou

Professor(a) ou Func(a) de escola Est ou Mun	funcionário(a) escola est ou mun
Professor(a) ou funcionário(a) escola est ou mun	
Professor ou Funcionário de Escola Particular	Professor(a) ou Funcionário(a) escola Particular
Professor(a) ou Funcionário(a) escola Particular	
Professor(a) ou funcionário(a) escola particular	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 17 - Atividade que desempenha



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Meio que utiliza para manter-se informado

Quadro 5 - Meio que utiliza para manter-se informado (Original/Padronizada)

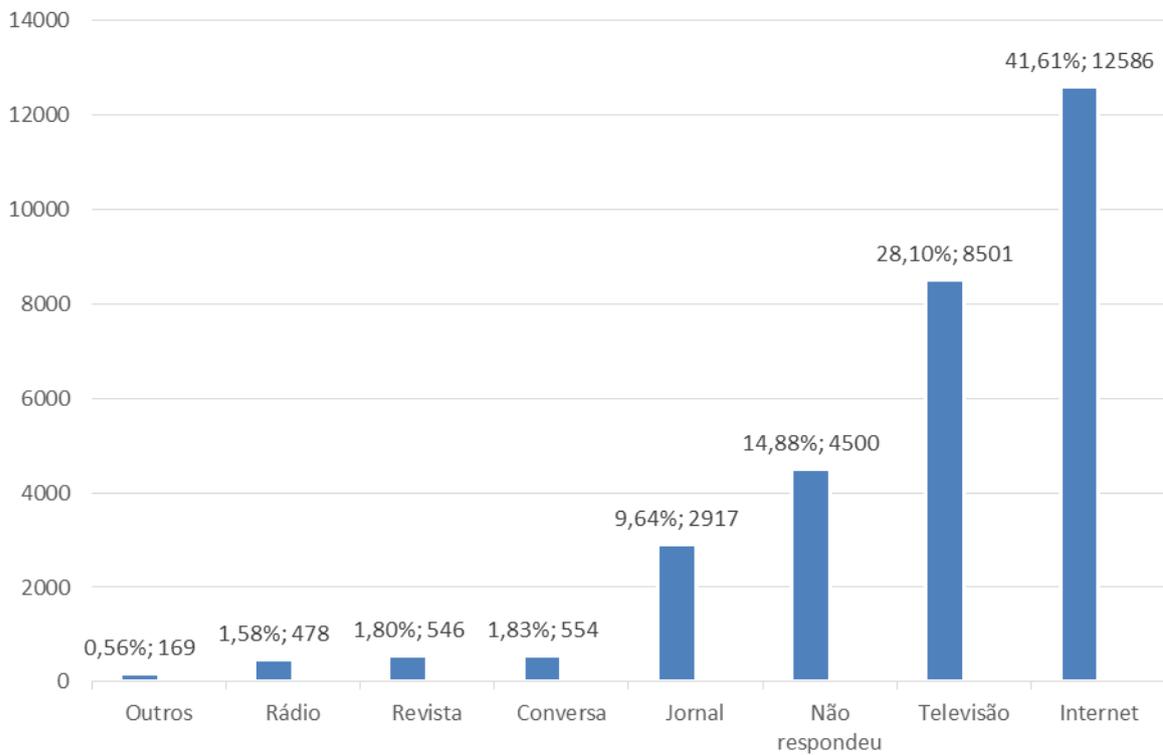
Resposta Original	Resposta Padronizada
Internet	Internet
Televisão	Televisão
Jornal	Jornal

Conversa	Conversa
Revista	Revista
Outros	Outros
Rádio	Rádio
Não respondeu	Não respondeu
Não Informada	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 18 - Meio que utiliza para manter-se informado



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Candidato possui acesso a internet

Quadro 6 - Candidato possui acesso a internet (Original/Padronizada)

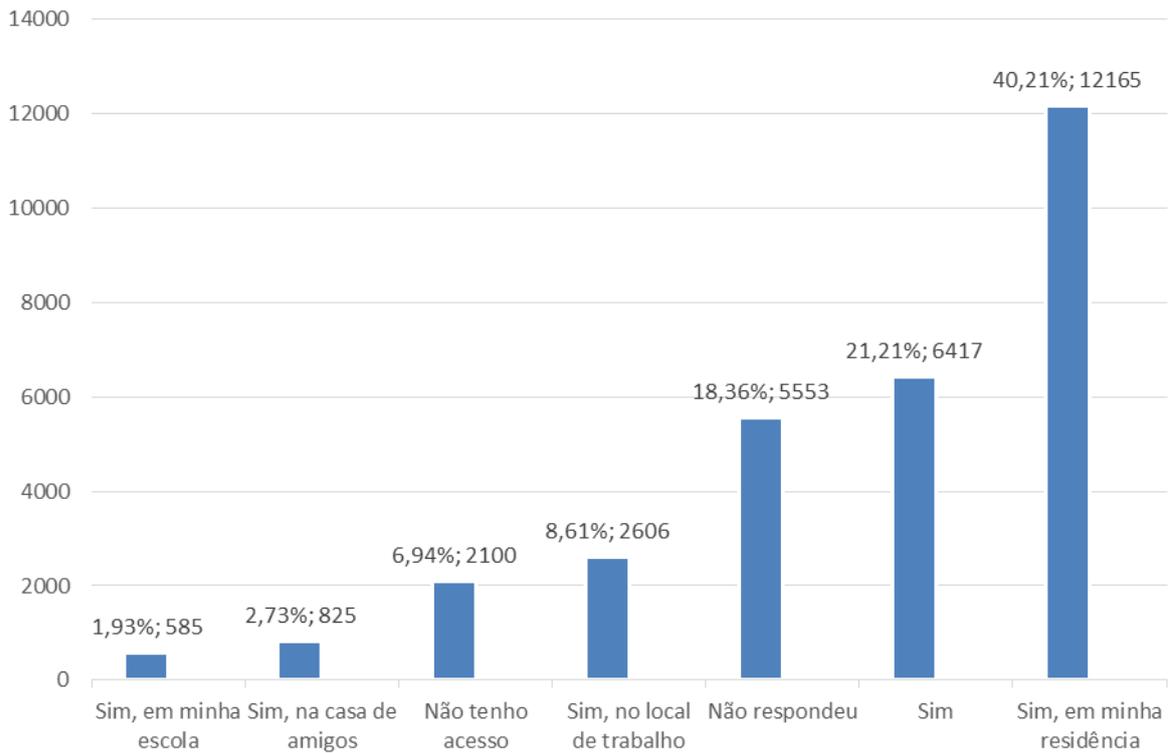
Resposta Original	Resposta Padronizada
Sim, na escola	Sim, em minha escola
Sim, em minha escola	
Sim, em casa	Sim, em minha residência
Sim, em minha residência	

Sim, com os amigos	Sim, na casa de amigos
Sim, com amigos	
Sim, na casa de amigos	
Não respondeu	Não respondeu
Não Informada	
Sim, no local de trabalho	Sim, no local de trabalho
Sim, no trabalho	
Não tenho acesso	Não tenho acesso
Não	
Sim, em outros locais	Sim
Sim	
Sim, em vários lugares	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 19- Candidato possui acesso a internet



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Motivo da escolha do curso

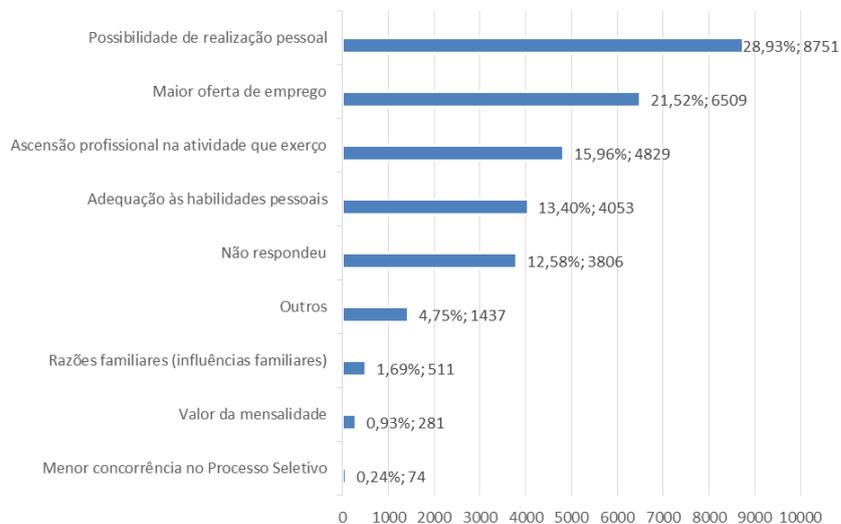
Quadro 7 - Motivo da escolha do curso (Original/Padronizada)

Resposta Original	Resposta Padronizada
Ascensão profissional	Ascensão profissional na atividade que exerço
Ascensão profissional na atividade que exerço	
Ascensão profissional na ativ que exerço	
Realização pessoal	Possibilidade de realização pessoal
Possibilidade de realização pessoal	
Razões familiares	Razões familiares (influências familiares)
Razões familiares (influências familiares)	
Razões familiares / Influências familiares	
Por ser o curso reconhecido no mercado	Maior oferta de emprego
Por ser o curso reconhecido pelo mercado	
Ampla mercado de trabalho	
Maior oferta de emprego	
Menor concorrência no processo seletivo	Menor concorrência no Processo Seletivo
Menor concorrência no Processo Seletivo	
Não respondeu	Não respondeu
Não Informada	
Adequação às habilidades pessoais	Adequação às habilidades pessoais
Adequação às habilidades pessoais	

Fonte: Banco de dados da UNISO

Nota: Elaborado pelo autor.

Gráfico 20 - Motivo da escolha do curso



Fonte: Banco de dados da UNISO

Nota: Elaborado pelo autor.

- Conheceu a UNISO através de qual meio

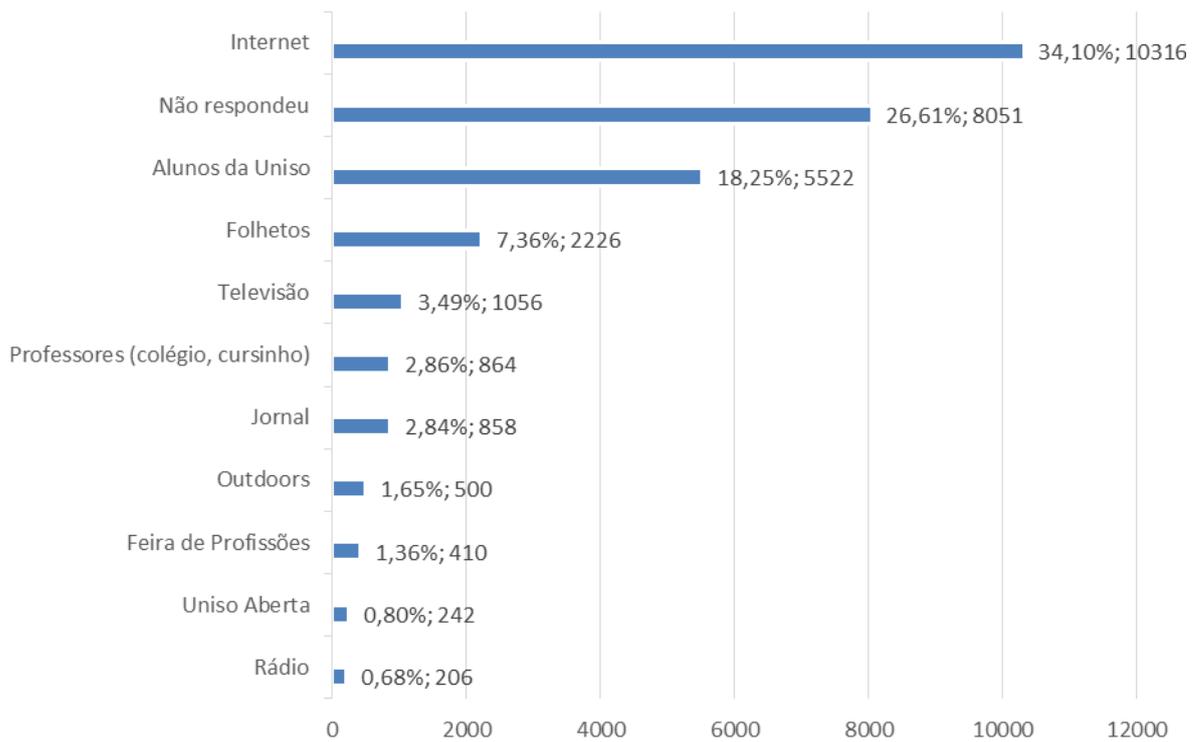
Quadro 8 - Conheceu a UNISO através de qual meio (Original/Padronizada)

Resposta Original	Resposta Padronizada
Não respondeu	Não respondeu
Não Informada	
Professores (colégio, cursinho)	Professores (colégio, cursinho)
Professores	
Feira de Profissões	Feira de Profissões
Feiras de Profissões	
Feira de Profissões	

Fonte: Banco de dados da UNISO

Nota: Elaborado pelo autor.

Gráfico 21 - Conheceu a UNISO através de qual meio



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Quantidade de livros que lê em média por ano

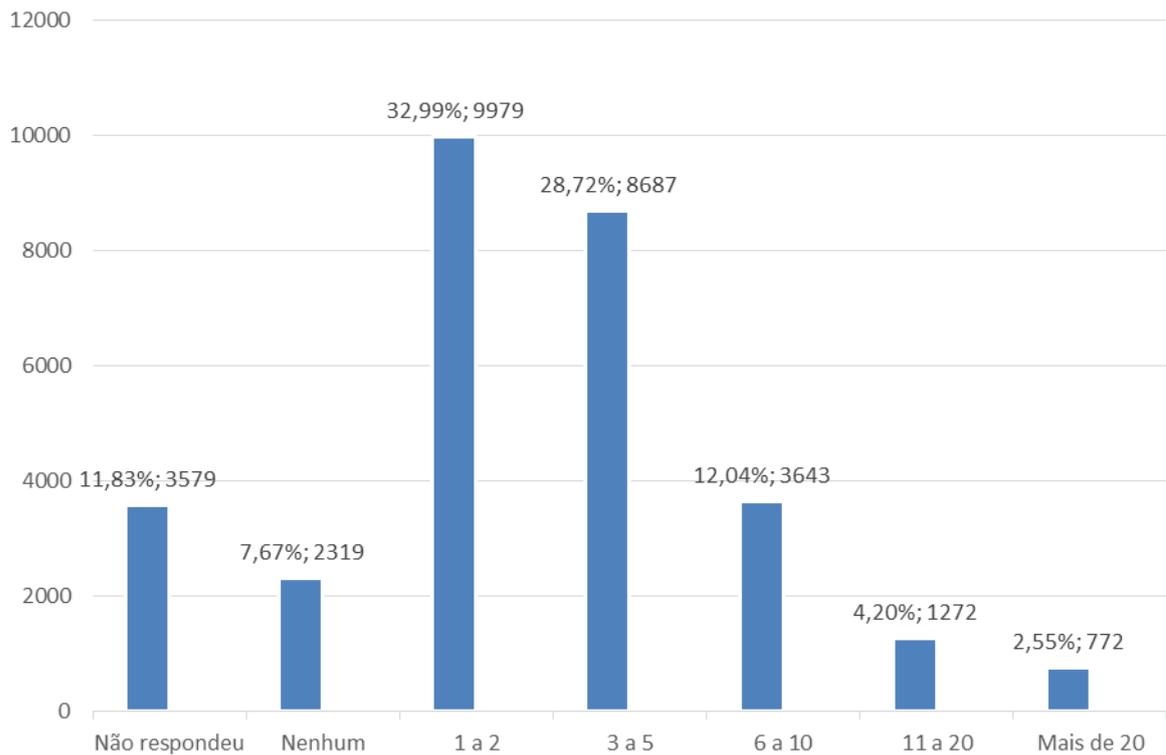
Quadro 9 - Quantidade de livros que lê (Original/Padronizada)

Resposta Original	Resposta Padronizada
1 a 2	1 a 2
3 a 5	3 a 5
6 a 10	6 a 10
11 a 20	11 a 20
Mais de 20	Mais de 20
Nenhum	Nenhum
Não Informada	Não respondeu
Não respondeu	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 22 - Quantidade de livros que lê



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Exerce alguma atividade remunerada

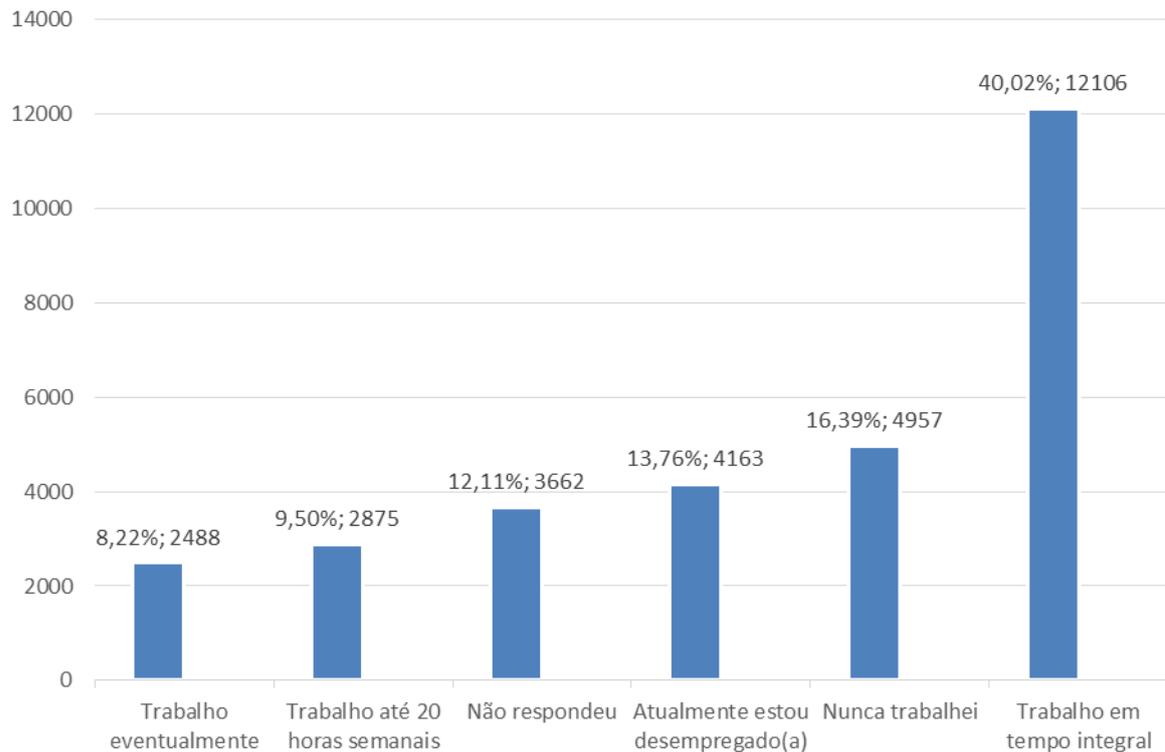
Quadro 10 - Exerce alguma atividade remunerada (Original/Padronizada)

Resposta Original	Resposta Padronizada
Trabalho eventualmente	Trabalho eventualmente
Trabalho em tempo integral	Trabalho em tempo integral
Trabalho até 20 horas semanais	Trabalho até 20 horas semanais
Nunca trabalhei	Nunca trabalhei
Atualmente estou desempregado	Atualmente estou desempregado(a)
Atualmente estou desempregado(a)	
Atualmente estou desempregado(a)	
Não respondeu	Não respondeu
Não Informada	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 23 - Exerce alguma atividade remunerada



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

- Turno de conclusão do ensino médio

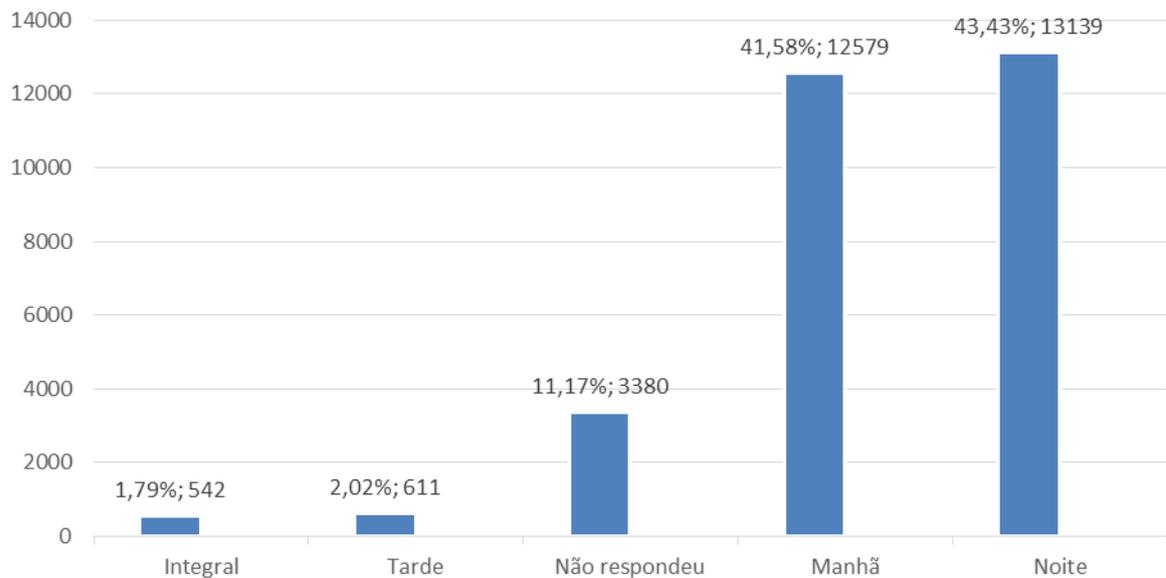
Quadro 11 - Turno de conclusão ensino médio (Original/Padronizada)

Resposta Original	Resposta Padronizada
Manhã	Manhã
Noite	Noite
Tarde	Tarde
Integral	Integral
Não Informada	Não respondeu
Não respondeu	

Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Gráfico 24 - Turno de conclusão do ensino médio



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Ao analisar os dados brutos extraídos do questionário informativo, podemos obter a quantidade de respostas que tivemos por questão. Nosso objetivo é garimpar esses dados a fim de classificá-los mediante a escolha do curso.

A primeira medida tomada foi cruzar cada uma das questões individualmente com a variável curso, para assim verificar se existe uma possível relação da resposta sobre um curso. Após essa primeira análise, começaremos a trabalhar com múltiplas questões aplicadas sobre um determinado curso, visando encontrar uma classificação das questões do questionário informativo sobre determinado curso.

5.2.1 Renda do candidato e curso escolhido

O algoritmo não encontrou nenhuma classificação ao minerar a variável renda com o curso, ou seja, dentro de um determinado curso temos várias pessoas com os mais distintos tipos de renda, não há uma concentração suficiente de uma determinada renda, fazendo com que o algoritmo não conseguisse criar a árvore.

5.2.2 Número de dependentes que vivem com a renda e curso escolhido

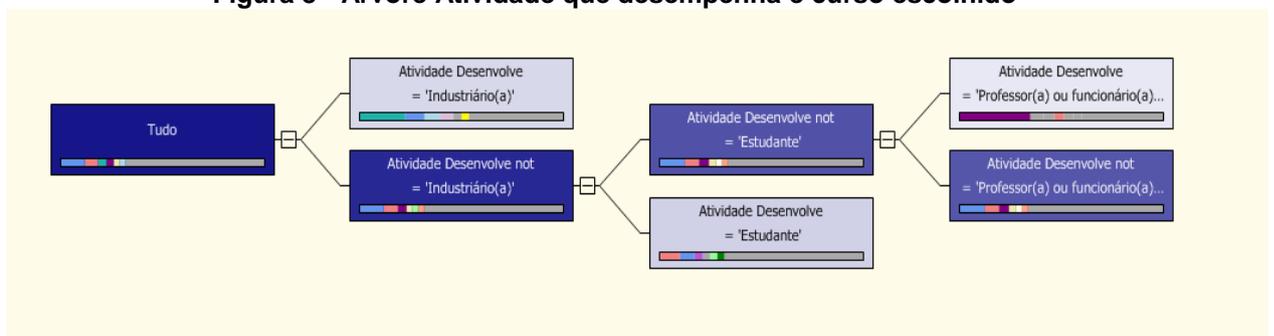
Algoritmo não encontrou nenhuma classificação ao minerar a variável “número de dependentes que vivem com a renda”. Desta forma, essa variável não se distingue com uma maior concentração sobre um determinado curso.

5.2.3 Candidato já iniciou algum curso superior e curso escolhido

Algoritmo não encontrou nenhuma classificação, os candidatos que já possuem ou não um curso superior estão bem distribuídos entre os cursos.

5.2.4 Atividade que desempenha e curso escolhido

Figura 8 - Árvore Atividade que desempenha e curso escolhido



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Algoritmo de árvore de decisão encontrou as seguintes classificações:

1. Se a atividade que desempenha for “Industriário” então curso que tem maior probabilidade é o de “Curso Superior de Tecnologia em Gestão da Produção Industrial”.
2. Se atividade for “Estudante”, o curso que tem maior probabilidade são os de: Direito, Administração e Farmácia. Levando em consideração

que Administração e Direito são os que têm maior procura e portanto maior volume de dados, o que se destaca é o curso de farmácia.

3. Se atividade for “Professor(a) ou funcionário(a) de escola estadual ou municipal,” o curso que tem maior procura é o de “Pedagogia”.

5.2.5 Meio que utiliza para manter-se informado e curso escolhido

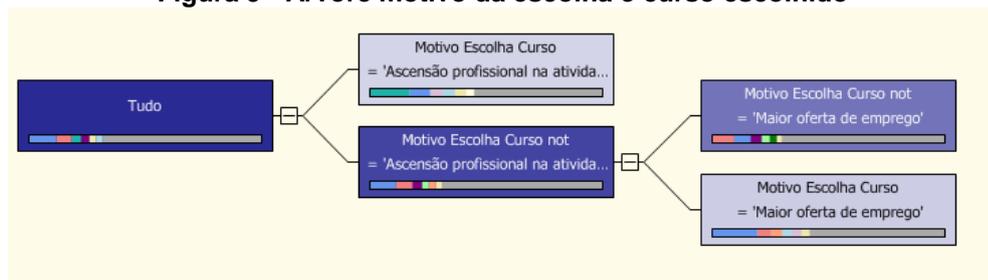
Algoritmo não encontrou nenhuma classificação.

5.2.6 Acesso à internet e curso escolhido

Algoritmo não encontrou nenhuma classificação.

5.2.7 Motivo da escolha do curso e curso escolhido

Figura 9 - Árvore motivo da escolha e curso escolhido



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Algoritmo de árvore de decisão encontrou as seguintes classificações:

1. Se motivo da escolha do curso for “Ascensão profissional na atividade que exerço” o curso que tem maior evidência é o: Curso Superior de Tecnologia em Gestão da Produção Industrial, acompanhado dos cursos: Administração, Curso Superior de Tecnologia em Logística, Engenharia de Produção, Ciências Contábeis e Curso Superior de Tecnologia em Gestão de Recursos Humanos. Administração aparece em segundo lugar mesmo possuindo a maior quantidade de dados. Porém os que têm muita evidência são os cursos tecnológicos.
2. Se motivo da escolha for “Maior oferta de emprego” os cursos que têm maior evidência são os: Administração, Direito, Administração de Negócios e Engenharia de Produção.

5.2.8 Conheceu a UNISO através de qual meio e curso escolhido

Algoritmo não encontrou nenhuma classificação.

5.2.9 Quantidade de livros que lê em média por ano e curso escolhido

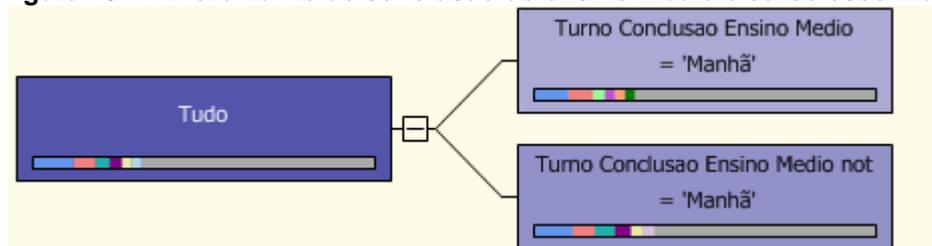
Algoritmo não encontrou nenhuma classificação.

5.2.10 Exerce alguma atividade remunerada e curso escolhido

Algoritmo não encontrou nenhuma classificação.

5.2.11 Turno de conclusão do ensino médio e curso escolhido

Figura 10 – Árvore Turno de conclusão do ensino médio e curso escolhido



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

1. Se o turno de conclusão do ensino médio for “Manhã”, cursos com maior evidência foram os de: Administração, Direito, Comunicação Social - Habilitação em Publicidade e Propaganda, Farmácia.
2. Se o turno de conclusão do ensino médio não for “Manhã”, então os cursos são: Administração, Direito, Curso Superior de Tecnologia em Gestão da Produção Industrial e Pedagogia.

Se levarmos em consideração que Administração e Direito apresentam o maior volume de dados e, por tanto, concentram um maior número de pessoas, distorcendo as classificações encontradas, podemos interpretar que quem se forma no ensino médio de manhã tem uma grande probabilidade de procurar o curso de Publicidade e Propaganda e Farmácia, e se conclui em outro turno destacam-se os cursos de Gestão de Produção Industrial e Pedagogia.

Nesta primeira análise individual das questões sobre o curso, conseguimos extrair algumas regras, que foram analisadas e comentadas acima, porém vale ressaltar as seguintes:

- Se atividade que desempenha for “Industriário” então curso que tem maior probabilidade é o de Curso Superior de Tecnologia em Gestão da Produção Industrial.
- Se atividade for “Estudante”, o curso que tem maior probabilidade são os de: Direito, Administração e Farmácia.
- Se atividade for “Professor(a) ou funcionário(a) escola estadual ou municipal,” o curso que tem maior procura é o de Pedagogia.

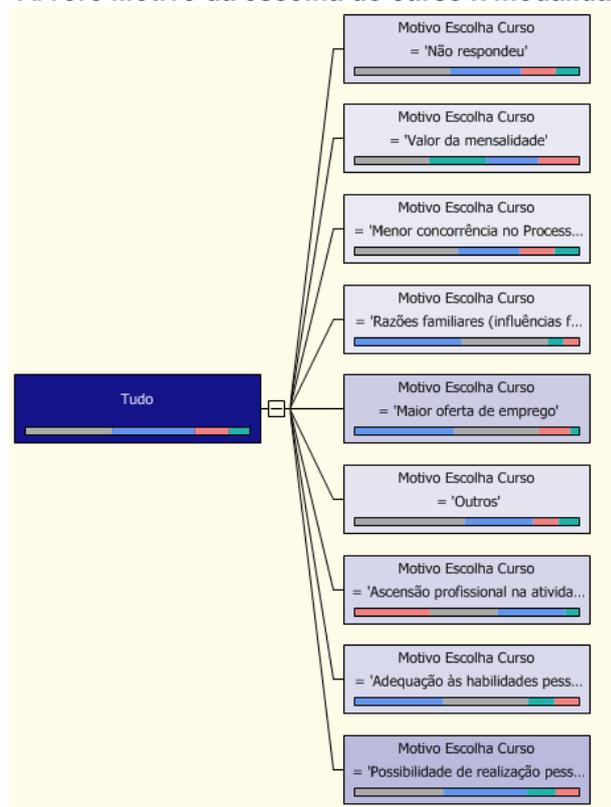
Observa-se que algumas profissões se relacionam muito com um determinado curso, como industrial e professor; ou seja, a pessoa já exerce uma profissão e vem à Universidade para se aprimorar. Observa-se também que o curso de farmácia com a condição estudante tem grande relação; isso pode se dar devido ao fato de ser um curso oferecido apenas no período da manhã, e que provavelmente os alunos são pessoas que não trabalham.

- A modalidade tecnológica está fortemente relacionada com o motivo da escolha do curso “Ascensão profissional na atividade que exerço”.

Provavelmente essa relação se dá pelo fato de serem pessoas que já trabalham na área e buscam uma formação para crescer na empresa e assim melhorar sua posição.

Para melhor compreender a relação da variável “motivo da escolha do curso” com a modalidade tecnológica, foi utilizada a variável modalidade como previsão ao invés do curso e foi obtido o seguinte resultado:

Figura 11 – Árvore Motivo da escolha do curso x modalidade do curso



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Interpretando essa árvore, temos alguns “nós” que chamaram a atenção como:

- Se o motivo da escolha do curso for “Valor da Mensalidade”, temos uma forte relação com a modalidade licenciatura, sendo a modalidade licenciatura a primeira, antes mesmo da graduação que tem a maior concentração dos dados.
- Se motivo da escolha do curso for “Razões familiares (influências familiares)”, temos forte relação com modalidade bacharel.
- Se motivo da escolha do curso for “Maior oferta de emprego”, temos forte relação com modalidade bacharel.
- Se motivo da escolha do curso for “Ascensão profissional na atividade que exerço” há forte relação com modalidade tecnológica. Ou seja, comprovando nosso questionamento, pessoas que escolhem o curso por ascensão profissional na atividade que exercem escolhem cursos tecnológicos.

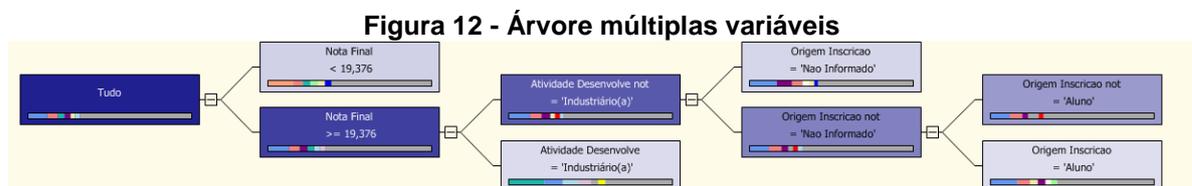
- Se motivo da escolha do curso for “Adequação às habilidades pessoais”, há forte relação com modalidade bacharel.
- Se motivo da escolha do curso for “Possibilidade de realização pessoal”, há forte relação com modalidade bacharel.

Até esse ponto, trabalhamos com uma única variável aplicada sobre o curso. Nosso próximo passo será explorar múltiplas questões aplicadas sobre a variável curso.

5.2.12 Múltiplas variáveis

A primeira mineração realizada foi a aplicação de todas as questões do processo seletivo com relação ao curso. Para tanto, trabalhamos com as variáveis: acesso a internet, nota final, atividade que desenvolve, atividade remunerada, conhecimento da UNISO através de, iniciou algum curso superior, quantidade de livros que lê, meio de informação, motivo da escolha do curso, número de dependentes da renda, origem da inscrição e turno de conclusão do ensino médio.

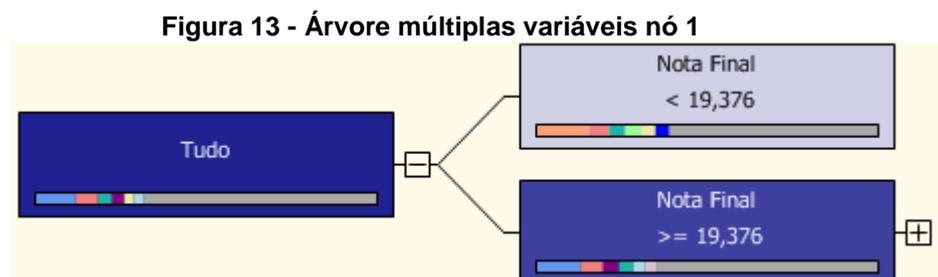
Nesta primeira garimpagem foi obtida a seguinte árvore:



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

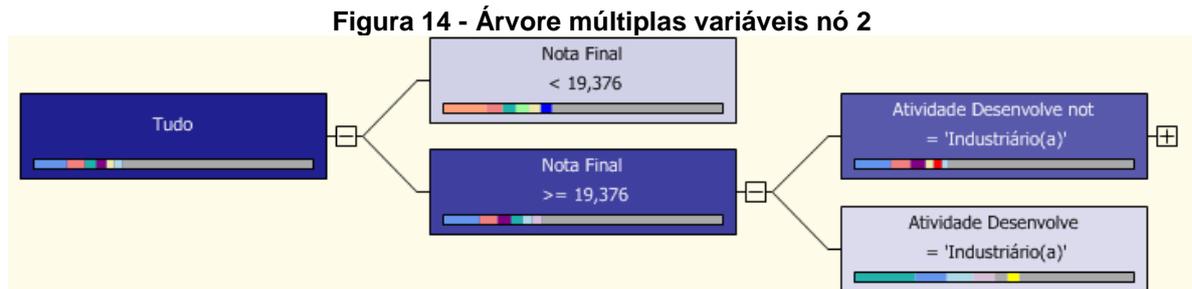
Para melhor visualização apresentaremos cada nó da árvore isoladamente:



Fonte: Banco de dados da UNISO.

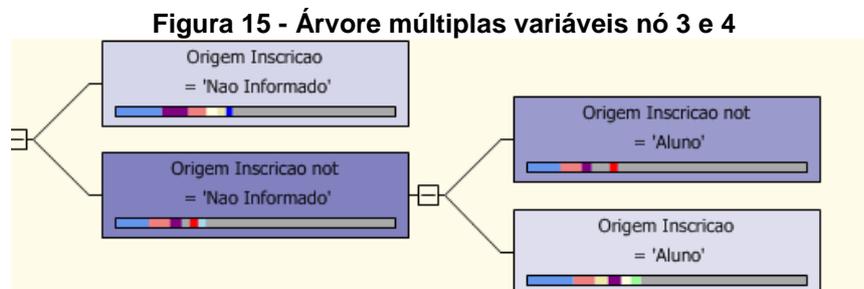
Nota: Elaborado pelo autor.

Neste primeiro nível podemos encontrar uma ramificação com a nota que o candidato obteve. Se a nota for menor que 19,376, então os cursos que têm maior probabilidade são os: Administração de Negócios, Direito, Curso Superior de Tecnologia em Gestão da Produção Industrial e Comunicação Social - Habilitação em Publicidade e Propaganda.



Fonte: Banco de dados da UNISO.
Nota: Elaborado pelo autor.

Caso a nota seja maior ou igual a 19,376 e a atividade for industriário, os cursos com maior probabilidade são: Curso Superior de Tecnologia em Gestão da Produção Industrial, Administração e Engenharia de Produção, conforme a imagem acima.

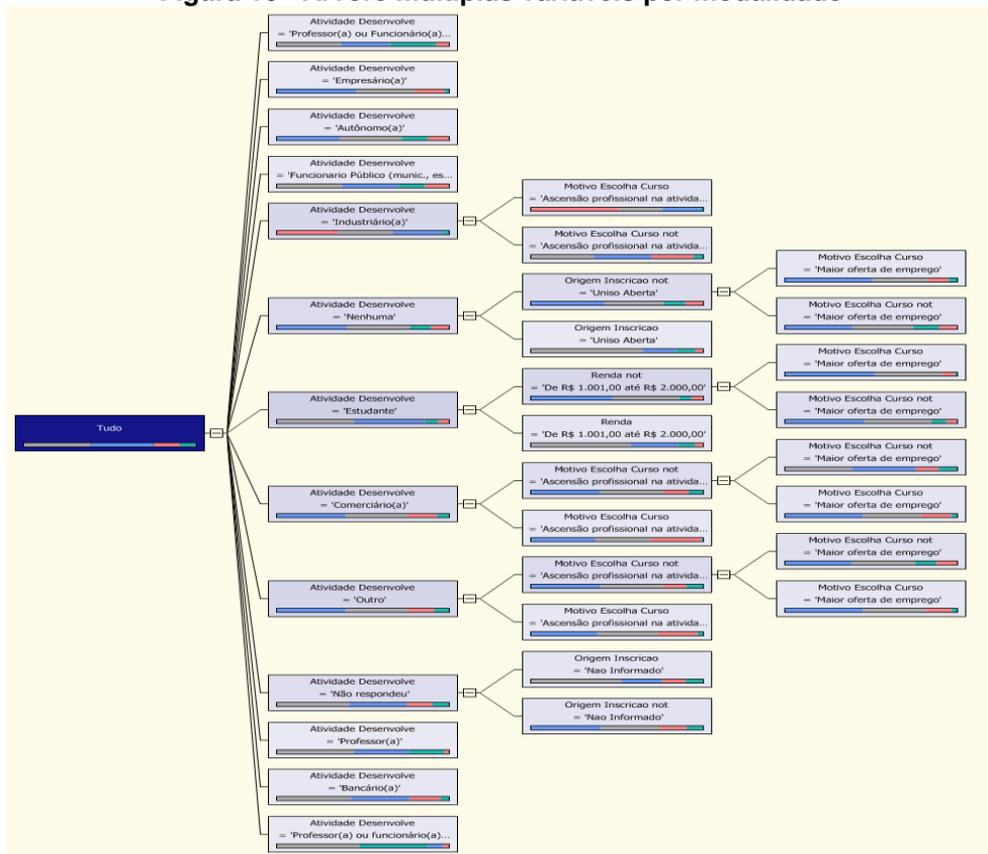


Fonte: Banco de dados da UNISO.
Nota: Elaborado pelo autor.

Caso a nota seja maior ou igual a 19,376 e a atividade não for industriário e origem da inscrição for igual “Aluno”, ou seja, indicação de um aluno da UNISO, os cursos são: Administração, Direito, Ciências Contábeis e Pedagogia.

Como a UNISO possui diversos cursos dentro de uma mesma modalidade, fazendo com que nosso dado fique disperso e assim não encontrando uma possível classificação, utilizamos as mesmas variáveis sobre a modalidade do curso, ao invés do curso. Obtivemos a seguinte árvore:

Figura 16 - Árvore múltiplas variáveis por modalidade



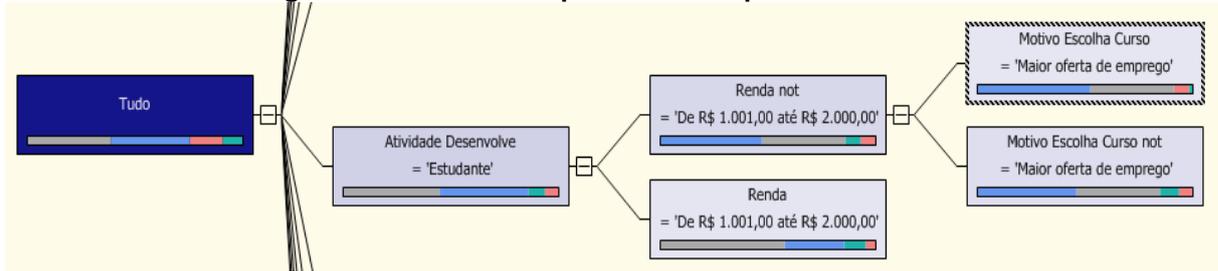
Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Vale destacar nesta árvore as seguintes tendências:

1. Se atividade que desenvolve for “professor(a) ou funcionário(a) escola estadual ou municipal” então a modalidade licenciatura é a que predomina.
2. Se atividade que desenvolve for “professor(a) ou funcionário(a) escola particular”, então modalidade é bacharel, acompanhada bem próximo de licenciatura.
3. Se atividade que desenvolve for “estudante” e renda de R\$ 1.001,00 até R\$ 2.000,00 então modalidade de curso é bacharel, acompanhado de licenciatura. Se renda for maior que R\$ 2.000,00 e o motivo da escolha for “maior oferta de emprego”, a modalidade é bacharel acompanhado da modalidade tecnológico. Se renda maior que R\$ 2.000,00 e motivo não for “maior oferta de emprego”, a modalidade é bacharel acompanhada de licenciatura, conforme mostra a seguinte ramificação da árvore:

Figura 17 - Árvore múltiplas variáveis por modalidade 1



Fonte: Banco de dados da UNISO.

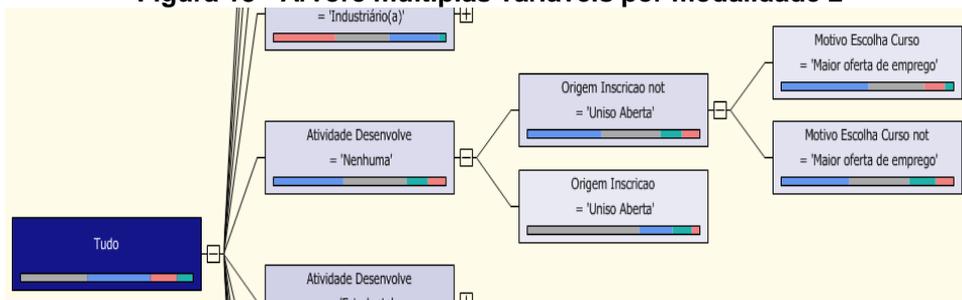
Nota: Elaborado pelo autor.

Vale observar nesta ramificação que a pessoa que é apenas estudante e tem renda entre R\$ 1.001,00 a R\$ 2.000,00 tem uma grande probabilidade de escolher a modalidade licenciatura. Também, se a renda for maior que R\$ 2.000,00 e o motivo da escolha do curso não for “maior oferta de emprego” observa-se novamente a presença da licenciatura.

Já se a renda é maior que R\$ 2.000,00 e motivo da escolha do curso for “maior oferta de emprego”, notamos a presença da modalidade tecnológica, logo após bacharel. Vale ressaltar que, como a maioria dos alunos e cursos é da modalidade bacharel, essa provavelmente sempre apontará em primeiro lugar, mas o que mais chama atenção nessa garimpagem é essa distinção do segundo lugar entre as modalidades licenciatura e tecnológica.

4. Se a pessoa não desenvolve atividade nenhuma o motivo da escolha do curso não for “maior oferta de emprego”, a modalidade com maior probabilidade é bacharel acompanhada em segundo lugar da modalidade licenciatura. Caso não desenvolva atividade e o motivo da escolha do curso seja “maior oferta de emprego”, a modalidade com maior probabilidade é a de bacharel acompanhada da tecnológica.

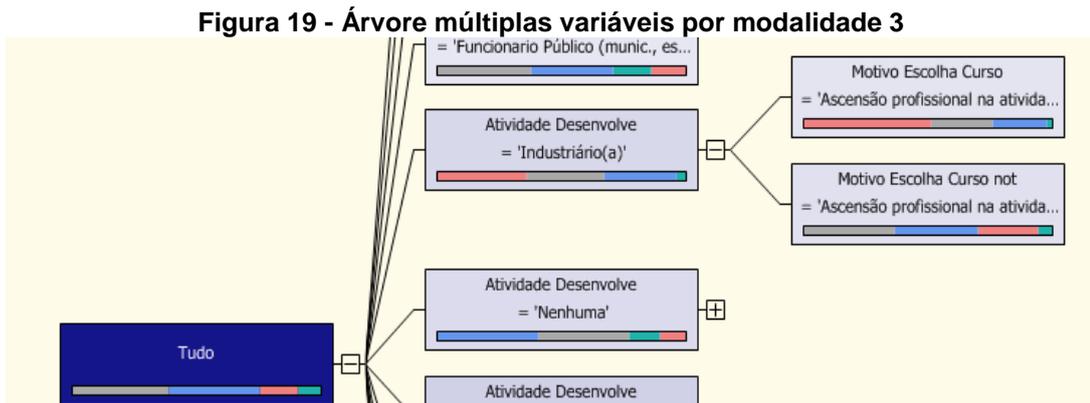
Figura 18 - Árvore múltiplas variáveis por modalidade 2



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

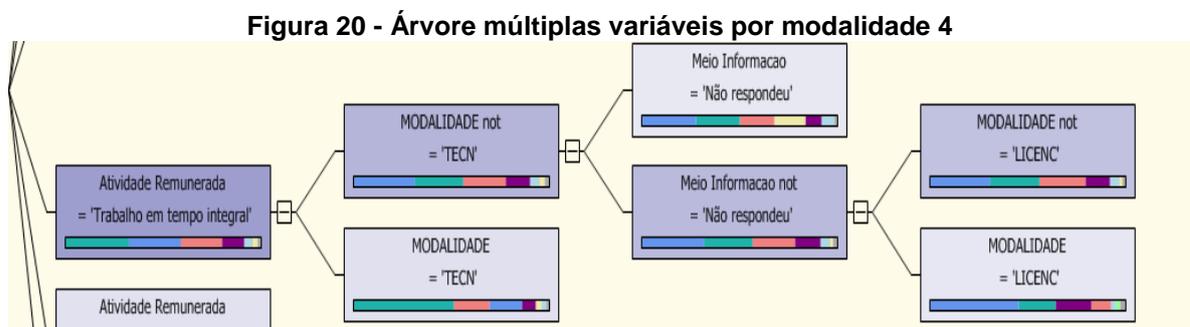
5. Se atividade que realiza for “industrialário” e o motivo da escolha do curso for “ascensão profissional na atividade que exerce”, então a modalidade é a tecnológica. Caso a atividade seja industrialário e o motivo da escolha do curso não for “ascensão profissional na atividade que exerço”, então a modalidade é bacharel acompanhada de tecnológica.



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Uma variável que foi de grande impacto para essa análise foi o motivo da escolha do curso. Por tanto, configuramos a ferramenta de mineração de dados para que ao invés de prever o campo curso, utilizássemos a variável “motivo da escolha do curso” como previsão e aplicássemos as demais variáveis sobre essa para, assim, verificarmos a possível existência de uma classificação com o motivo da escolha do curso. Obtivemos a seguinte ramificação da árvore que chamou atenção:



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

1. Se atividade remunerada for “trabalho tempo integral” e a modalidade do curso escolhido for “tecnológica”, então o motivo da escolha do curso é “Ascensão Profissional na Atividade que Exerço”.

2. Se atividade remunerada for “Trabalho em tempo integral” e a modalidade do curso escolhido for “licenciatura”, então o motivo da escolha do curso é “Possibilidade de realização pessoal”.

Como o algoritmo de árvore de decisão é utilizado para criarmos classificações e assim definirmos estereótipos, podemos dizer que a pessoa que escolhe um curso da modalidade tecnológica e trabalha em tempo integral escolhe esse curso para obter ascensão profissional na atividade que exerce.

Também podemos classificar que a pessoa que escolhe um curso na modalidade licenciatura e trabalha em tempo integral o motivo que a levou a optar por essa modalidade é a possibilidade de realização pessoal.

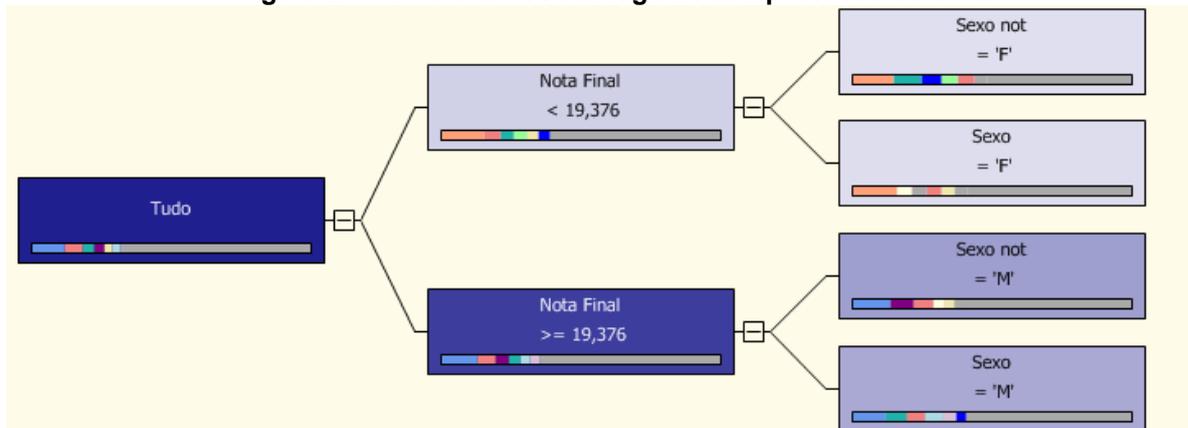
O objetivo do nosso trabalho não é explorar a campo essas probabilidades, mas uma questão que ficou em evidência e inquietável sobre a modalidade licenciatura é: será que com a mudança das leis e a exigência do diploma para lecionar, os professores que já estavam em sala de aula se viram obrigados a estudar, ou isso já era um desejo pessoal? Ou seja, é por isso que a grande probabilidade da modalidade licenciatura é o motivo de realização pessoal? Ou ainda, as pessoas que escolhem a modalidade licenciatura o fazem por um desejo pessoal enquanto as que escolhem a modalidade tecnológica buscam ascensão profissional em uma atividade que já exerce? O que essas pessoas têm a dizer?

Como podemos observar, essas minerações abrem um leque para muitos questionamentos que podem ser trabalhados, respondidos ou compreendidos com trabalhos futuros. No que diz respeito à presente pesquisa é que tentaremos classificar e encontrar novas regras cruzando os dados socioeconômicos/demográficos com o do processo seletivo. Para isso, o capítulo seguinte detalhará essa nova garimpagem.

5.3 Dados demográficos e processo seletivo

Após termos trabalhado separadamente com dados socioeconômicos do aluno e do processo seletivo, o próximo passo será a mineração cruzando essas duas vertentes. Primeiramente, trabalhamos com as variáveis: nota final, sexo e faixa etária, para verificarmos a existência de uma classificação. Após a aplicação da ferramenta, obtivemos a seguinte árvore:

Figura 21 - Árvore dados demográficos x processo seletivo



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

Analisando a árvore gerada obtivemos as seguintes classificações:

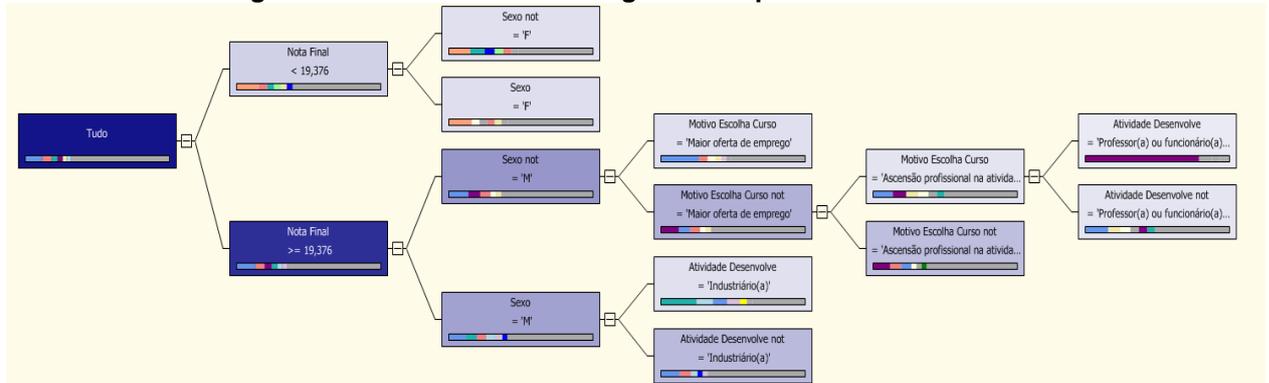
1. Se nota final for menor que 19,376 e sexo for “feminino”, então os cursos que têm maior probabilidade de escolha são: Administração de Negócios, Curso Superior de Tecnologia em Gestão de Recursos Humanos e Pedagogia - Docência nas Séries Iniciais do Ensino Fundamental.
2. Se nota final for menor que 19,376 e sexo for “masculino”, então os cursos são: Administração de Negócios, Sistemas de Informação e Curso Superior de Tecnologia em Gestão da Produção Industrial.
3. Se nota final for maior ou igual a 19,376 e sexo for “feminino”, então os cursos com maior probabilidades são: Direito, Administração e Pedagogia.
4. Se nota final for maior ou igual a 19,376 e sexo for “masculino” então os cursos são: Direito, Curso Superior de Tecnologia em Gestão da Produção Industrial e Administração.

Na segunda análise utilizamos todas as variáveis do questionário informativo (acesso a internet, nota final, atividade que desenvolve, atividade remunerada, conhecimento da UNISO através de, iniciou algum curso superior, quantidade de livros que lê, meio de informação, motivo da escolha do curso, número de

dependentes da renda, origem da inscrição e turno de conclusão do ensino médio), juntamente com as variáveis sexo e faixa etária.

A ferramenta de mineração de dados gerou as seguintes classificações:

Figura 22 - Árvore dados demográficos e processo seletivo 2



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

1. Se nota final for menor que 19,376 e sexo for “feminino”, os cursos com maior probabilidade são: Administração de Negócios, Curso Superior de Tecnologia em Gestão de Recursos Humanos e Pedagogia - Docência nas Séries Iniciais do Ensino Fundamental.
2. Se nota final for menor que 19,376 e sexo for “masculino”, os cursos com maior probabilidade são: Administração de Negócios, Sistemas de Informação e Curso Superior de Tecnologia em Gestão da Produção Industrial.
3. Se nota final for maior que 19,376, sexo “masculino” e atividade que desenvolve for “Industriário”, os cursos com maior probabilidade são: Curso Superior de Tecnologia em Gestão da Produção Industrial, Administração e Engenharia de Produção.
4. Se nota final for maior que 19,376, sexo “masculino” e atividade que desenvolve não for “Industriário”, os cursos com maior probabilidade são: Administração, Direito e Engenharia da Produção.
5. Se nota final for maior que 19,376, sexo “feminino” e motivo da escolha do curso for “maior oferta de emprego”, os cursos com maior probabilidade são: Administração, Direito e Curso Superior de Tecnologia em Gestão de Recursos Humanos.

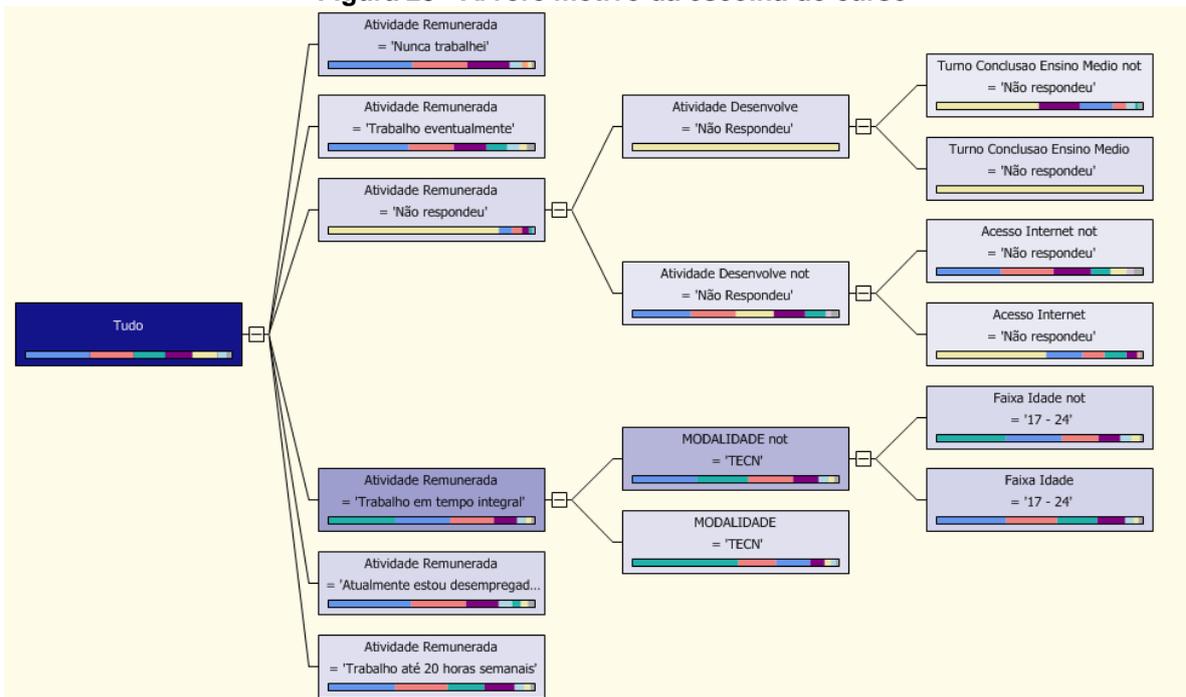
6. Se nota final maior que 19,376; sexo for “feminino” e motivo da escolha do curso não for “maior oferta de emprego” e não for “ascensão profissional na atividade que exerço”, os cursos com maior probabilidade de escolha são: Pedagogia, Direito e Administração.
7. Se nota final maior que 19,376, sexo for “feminino” e motivo da escolha do curso for ascensão profissional na atividade que exerce e a atividade que desenvolve for “Professor(a) ou funcionário(a) escola estadual ou municipal”, os cursos com maior probabilidade é: Pedagogia.
8. Se nota final maior que 19,376; sexo for “feminino” e motivo da escolha do curso for “ascensão profissional na atividade que exerço” e a atividade que desenvolve não for “Professor(a) ou funcionário(a) escola estadual ou municipal”, os cursos com maior probabilidade são: Administração, Ciências Contábeis e Curso Superior de Tecnologia em Gestão de Recursos Humanos.

Podemos observar na regra número 6 que mais uma vez o curso de pedagogia sobressaiu dos demais e reforçou a ideia de que as pessoas que optam pela modalidade de licenciatura o fazem por motivos de realização pessoal.

Observamos na regra 7 que aqueles que exercem atividade “professor(a) ou funcionário(a) escola estadual ou municipal” e tem o motivo da escolha do curso como “ascensão profissional na atividade que exerço” têm uma probabilidade de 77,73% de cursarem o curso de pedagogia.

Para estudar melhor o motivo da escolha do curso, configuramos a ferramenta para provisioná-lo de acordo com outras variáveis, sendo obtidos os seguintes resultados:

Figura 23 - Árvore motivo da escolha do curso



Fonte: Banco de dados da UNISO.

Nota: Elaborado pelo autor.

O que chamou a atenção e vale a pena destacar é a classificação que o algoritmo gerou quando a atividade for “Trabalho em tempo integral”:

1. Se a modalidade for “tecnológica” 49,19%, há probabilidade que o motivo da escolha do curso seja ascensão profissional.
2. Se a modalidade não for “tecnológica” e a faixa etária estiver entre os 17 e 24 anos, o motivo da escolha do curso é “possibilidade de realização pessoal”.
3. Se a modalidade não for “tecnológica” e a faixa etária for superior a 24 anos, o motivo da escolha do curso é “Ascensão profissional na atividade que exerço”.

Podemos constatar que novamente a modalidade tecnológica está ligada ao motivo ascensão profissional, mas uma coisa interessante é que a faixa etária do estudante também está diretamente associada ao motivo da escolha do curso. Como o algoritmo identificou, os jovens até 24 anos têm uma maior probabilidade de procurar um curso pelo motivo de realização pessoal, enquanto os com idade superiores a 24 anos o fazem pelo motivo de ascensão profissional. Conseguimos compreender que os jovens, com menos de 24 anos o que provavelmente não trabalham e, quando esses optam por um curso na Universidade, escolhem aquele

que eles sonham em fazer, enquanto os mais velhos, acima de 24 anos, na maioria das vezes escolhem um curso na área em que já trabalham.

6 CONSIDERAÇÕES FINAIS

No decorrer da pesquisa foram aplicadas várias combinações nas variáveis do banco de dados da Universidade de Sorocaba para classificá-las e assim obter uma probabilidade estatística que classifique um determinado grupo de estudante sobre um curso.

Essas classificações não retratam a realidade absoluta; são apenas dados estatísticos que podem ajudar a entender o cenário analisado e, no caso desta pesquisa, classificar um determinado grupo de estudantes com relação ao curso.

Com a utilização desta ferramenta aplicada sobre os dados socioeconômicos/demográficos dos alunos obtivemos alguns resultados:

- Se o sexo for feminino e a faixa etária maior que 24 anos, o curso que tem maior procura é o de Pedagogia, com 26,90% das matrículas.
- Se o sexo for masculino e a faixa etária maior que 24 anos o curso que tem maior procura é o de Tecnologia em Gestão de Produção Industrial, com 13,96% de probabilidade das matrículas.

Como descrito acima, o curso de Pedagogia tem uma grande probabilidade de ser escolhido por mulheres com mais de 24 anos. Já os homens com mais de 24 anos têm como primeira probabilidade de escolha o curso de Gestão de Produção Industrial.

Outra análise realizada nos dados socioeconômicos/demográficos demonstrou que a cor dos estudantes não está ligada diretamente a um curso, ou seja, não há uma concentração de determinada cor em um curso.

Quando cruzado com a modalidade (bacharel, licenciatura e tecnológico) ao invés do curso, foi constatado que o sexo feminino está fortemente correlacionado com a faixa etária de “51 a 60 anos” e com a modalidade licenciatura. Isso reforça ainda mais a descoberta anterior: mulheres com mais de 24 anos têm grande probabilidade de optar pelo curso de Pedagogia.

Após um trabalho de “limpeza” dos dados do processo seletivo e padronização dos mesmos, foi possível extrair as questões e respostas dos questionários informativos. Assim, Foi possível identificar inicialmente que algumas atividades profissionais se relacionam muito com um determinado curso, como o caso da atividade industriário com o Curso Superior de Tecnologia em Gestão da

Produção Industrial e professor de escola pública ou municipal com o curso de Pedagogia; ou seja, a pessoa já exerce uma profissão e vem à Universidade para se aprimorar ou progredir na carreira. Observou-se também que o curso de Farmácia tem uma grande relação com a ocupação “estudante”. Isso pode se dar devido ao fato de ser um curso oferecido apenas no período da manhã e provavelmente os estudantes não trabalhem.

Após identificar que a pergunta do questionário informativo “motivo da escolha do curso” está fortemente correlacionada ao curso, foi configurada a ferramenta de mineração para prever o motivo da escolha em função do curso. Assim, foram obtidos os seguintes resultados:

- Se o motivo da escolha do curso for “Valor da Mensalidade”, temos uma forte relação com a modalidade licenciatura.
- Se motivo da escolha do curso for “Ascensão profissional na atividade que exerço” há forte relação com modalidade tecnológica.

Pôde ser constatado que as pessoas que buscam ascensão profissional optam por cursos da modalidade tecnológica, por ser um curso de menor duração e muitas vezes com o valor da mensalidade mais baixo do que os de bacharelado. Já os cursos de licenciatura estão fortemente relacionados ao motivo de escolha “Valor da mensalidade”. Isso se dá pelo fato da Universidade oferecer incentivo acadêmico com desconto de 35% para os cursos de licenciatura.

Após aprimorar o modelo de mineração de dados e trabalhar com outras combinações das variáveis, foi possível identificar que a pessoa que é apenas estudante e tem renda entre R\$ 1.001,00 a R\$ 2.000,00 tem uma grande probabilidade de escolher a modalidade licenciatura. Também, se a renda for maior que R\$ 2.000,00 e o motivo da escolha do curso não for “maior oferta de emprego” observa-se novamente a presença da licenciatura. Já se a renda é maior que R\$ 2.000,00 e motivo da escolha do curso for “Maior oferta de emprego”, notamos a presença da modalidade tecnológica.

Podemos entender que os estudantes que têm uma renda de até R\$ 2.000,00 optam pelos cursos da modalidade licenciatura, fato que pode estar ligado ao motivo da escolha do curso “Valor da mensalidade”.

Outra classificação identificada é que a pessoa que escolhe um curso na modalidade licenciatura e trabalha em tempo integral tende a optar por essa

modalidade pela possibilidade de realização pessoal. Isso explica a descoberta descrita acima, na qual estudantes com renda maior que R\$ 2.000,00 não buscam o curso por maior oferta de emprego escolhendo a modalidade licenciatura.

Com essas classificações encontradas, fica evidente que os estudantes que possuem renda abaixo de R\$ 2.000,00 optam preferencialmente pelos cursos de licenciatura pela questão financeira, e os demais, que ganham acima de R\$ 2.000,00 e não buscam maior oferta de emprego, apresentam interesse pelos cursos de licenciatura pela questão da realização pessoal. Isto possivelmente se deve ao fato de já trabalhar na área.

Foi identificado também que a procura pelo curso por maior oferta de emprego está ligada à modalidade tecnológica e bacharel. Já os cursos da modalidade tecnológica são mais procurados por pessoas que buscam uma ascensão profissional em uma atividade que já desempenham.

Outra importante descoberta no cruzamento dos dados do processo seletivo com os socioeconômicos/demográficos é que as faixas etárias dos estudantes estão relacionadas com o motivo da escolha do curso. Os estudantes com até 24 anos, têm uma maior probabilidade de procurar um curso pelo motivo de realização pessoal, enquanto os demais o fazem por motivo de ascensão profissional.

Com essa classificação identificada, foi possível compreender que os jovens, com menos de 24 anos provavelmente não trabalham e, quando esses optam por um curso da Universidade, escolhem aquele que eles sonham em fazer, enquanto os mais velhos, acima de 24 anos, na maioria das vezes escolhem um curso na área em que já trabalham.

Uma tendência que despertou curiosidade é o fato dos cursos de licenciatura atraírem pessoas que o fazem objetivando realização pessoal. Já os cursos da modalidade tecnológica atraem pessoas que o buscam visando ascensão profissional ou maior oferta de emprego.

Foi compreendido, então, que os cursos tecnológicos da Universidade possuem um público que vem se qualificar profissionalmente para o mercado, enquanto nos cursos de licenciatura são pessoas que o fazem em busca da realização pessoal ou por questões financeiras.

A modalidade bacharel por ser muito ampla e possuir muitos cursos não destacou nenhuma classificação. Podemos concluir que os estudantes ali inseridos possuem motivos diversificados e distintos.

As três hipóteses levantadas apresentaram classificações distintas, porém, quando analisadas e interpretadas, foi possível identificar alguns estereótipos e assim conhecer melhor o estudante inserido na Instituição, apontando o público que frequenta determinado curso. Os resultados mostram tendências que podem não se verificar em termos absolutos. Com as devidas restrições e limitações de toda pesquisa a respeito de fenômenos humanos, valem para este caso que tomou por objeto a Universidade de Sorocaba nesse período de seis anos. Embora muitas de suas conclusões possam ser observadas em outras instituições similares, não há por parte do autor nenhuma pretensão a universalizá-las. Para conclusões mais amplas e seguras seriam necessários muitos outros estudos e certamente a técnica de mineração de dados é uma ferramenta de grande utilidade.

Para trabalhos vindouros nesta mesma linha, indica-se que a modalidade bacharel seja tratada separadamente das demais, pois esta apresenta um grande volume de cursos, fazendo com que a ferramenta de mineração de dados não encontre uma classificação pela heterogeneidade que existe nos dados. Assim, uma análise curso a curso desta modalidade seria o mais recomendado, pois novas classificações do perfil discente poderiam ser esboçadas.

Essa foi a primeira pesquisa empírica utilizando uma ferramenta de mineração nos dados da Universidade de Sorocaba. Uma importante contribuição para a gestão da Universidade e para o meio acadêmico, abrindo novos horizontes para pesquisas futuras.

REFERÊNCIAS

Reserch in agricultural & applied economics. **Aspectos teóricos do Data Mining e aplicação das redes neurais em previsões de preços agropecuários**. Disponível em: <<http://ageconsearch.umn.edu/bitstream/148202/2/681.pdf>>. Acesso em: 10 set. 2013.

BARATA-MOURA, José. Educação Superior: Direito ou Mercadoria?. **Avaliação**, Campinas, v. 8, n. 4, 31-36, dez. 2003.

BERNHEIM, Carlos Tünnermann. Desarrollo histórico de la Universidad. **Cincuentenario de la Revista Paraguaya de Sociología**, Asunción, jun. 2010. Serie Educación.

CÂMARA DE SOROCABA. **História da Cidade de Sorocaba**. Disponível em: <<http://www.camarasorocaba.sp.gov.br/sitecamara/historiasorocaba.html>> Acesso em: 16 de set. 2013

CARVALHO, Luís Alfredo Vidal de. **Datamining**: a mineração de dados no marketing, medicina, economia, engenharia e administração. São Paulo: Érica, 2001. 234 p.

CUNHA, Luiz Antônio. Desenvolvimento desigual e combinado no ensino superior – estado e mercado. **Educação & Sociedade**, Campinas, v. 25, n. 88, p. 795-817, out. 2004.

DEVMEDIA. **Aspectos teóricos da mineração de dados e aplicação das regras de classificação**. Disponível em: <<http://www.devmedia.com.br/aspectos-teoricos-da-mineracao-de-dados-e-aplicacao-das-regras-de-classificacao-para-apoiar-o-comercio/25429>>. Acesso em: 14 de out. 2012.

DIAS SOBRINHO, José. Educação Superior: Bem público, Equidade e Democratização. **Avaliação**: Revista da Avaliação da Educação Superior, Campinas, Sorocaba, v. 18, n. 01, 107-128, mar. 2013.

DIAS SOBRINHO, José. **Dilemas da educação superior no mundo globalizado: sociedade do conhecimento ou economia do conhecimento?**. São Paulo: Casa do Psicólogo, 2005. 258 p.

DIAS SOBRINHO, José. Calidad, pertinencia y responsabilidad social de la Universidad Latinoamericana y Caribeña, In: GAZZOLA, Ana Lúcia; DIDRIKSSON, Axel (Orgs.), **Tendencias de la Educación Superior en América Latina y el Caribe**. Caracas: IESALC-UNESCO, 2008a, p. 87-112.

DIAS SOBRINHO, José. **Dilemas da educação superior no mundo globalizado: sociedade do conhecimento ou economia do conhecimento?**. São Paulo: Casa do Psicólogo, 2005. 258 p

DIAS SOBRINHO, José . Cambios y reformas en la educación superior. In: Carlos Tünnermann Bernheim. (Org.). **La educación superior en América Latina y el**

Caribe: diez años después de la Conferencia Mundial de 1998. Cali; Caracas: Pontificia Universidad Javeriana / IESALC-UNESCO, 2008b, v. 01, p. 95-140.

DIDRIKSSON, Axel. Contexto Global y Regional de la Educación Superior en América Latina y el Caribe, in GAZZOLA, A. L.; DIDRIKSSON, A.: **Tendencias da la Educación Superior en América Latina y el Caribe.** Caracas: IESALC/UNESCO.

DOURADO, Luiz Fernandes. Reforma do Estado e as políticas para a educação Superior no Brasil nos anos 90. **Educação & Sociedade**, Campinas, v. 23, n. 80, p. 234-252, set. 2002.

FANTIN, Monica. Diante do abismo digital: mídia-educação e mediações culturais. **Perspectiva**, Florianópolis, v. 27, n. 1, 69-96, jan/jun. 2009.

FAYYAD, Usama; PIATETSKI-SHAPIRO, Gregory; SMYTH, Padhraic. The KDD Process for Extracting Useful Knowledge from Volumes of Data. In: **Communications of the ACM**, v.39, n.11, p.27-34, nov.1996

GOERGEN, Pedro. Educação superior e responsabilidade Social: do Discurso à Ação. In: Ana Lúcia Gazzola; Axel Didriksson. (Org.). **Tendencias de la Educación Superior en América Latina y el Caribe.** Bogotá: UNESCO, 2008, v.1, p. 1-25.

GOERGEN, Pedro. Universidade e compromisso social. In: UNIVERSIDADE e compromisso social. Brasília-DF: Instituto Nacional de Estudos e Pesquisas Anísio Teixeira, 2005.

GOERGEN, Pedro. Da Formação ao Ensino: Um ponto cego nas políticas de Pós-Graduação. **Avaliação:** Revista da Avaliação da Educação Superior, Campinas, Sorocaba, v. 18, n. 01, 45-68, mar. 2013.

GOMES, Luis Fernando. EAD no Brasil: Perspectivas e Desafios. **Avaliação:** Revista da Avaliação da Educação Superior, Campinas, Sorocaba, v. 18, n. 01, 13-22, mar. 2013.

GROTH, Robert. **Data Mining:** a hands-on approach for business professionals. New Jersey: Prentice Hall PTR, 1997. 264 p.

HAN, Jiawei; KAMBER, Micheline. **Data mining:** concepts and techniques. 2th ed. San Francisco: Morgan Kaufmann, 2006

HIRJI, Karim. Exploring data mining implementation. **Communications of the ACM**, v. 44, n. 7. Jul. 2001.

Instituto Brasileiro de Geografia e Estatística. **Censo Demográfico 2010.** Disponível em: <<http://www.ibge.gov.br/cidadesat/painel/painel.php?codmun=355220>>. Acesso em 16 set 2013.

Instituto Nacional de Estudos e Pesquisas. **Censo da educação superior 2010.** Disponível em: <<http://portal.inep.gov.br/web/censo-da-educacao-superior/resumos-tecnicos>>. Acesso em 01 mai 2012.

LANDINELLI, Jorge. Escenarios de Diversificación, Diferenciación y Segmentación de la Educación Superior en América Latina y el Caribe. In GAZZOLA, A. L.; DIDRIKSSON, A. (2008). **Tendencias da la Educación Superior en América Latina y el Caribe**. Caracas: IESALC/UNESCO.

LÓPEZ SEGRERA, Francisco. América Latina y el Caribe: principales tendencias de la educación superior. **Avaliação** : Revista da Avaliação da Educação Superior, Campinas, SP , v. 11, n. 3,p. 9-36, set. 2006.

MACLENNAN, Jamie; CRIVAT, Bogdan; TANG, ZhaoHui. **Data Mining with Microsoft SQL Server 2008**. Indianapolis: Wiley Publishing, 2009

MARQUES, Waldemar. Expansão e Oligopolização da Educação Superior no Brasil. **Avaliação**: Revista da Avaliação da Educação Superior, Campinas, Sorocaba, v. 18, n. 01, 69-84, mar. 2013.

MARTINHAGO, Sergio. **Descoberta de conhecimento sobre o processo seletivo da UFPR**. 2005. 114 p. Dissertação (Mestrado) – Universidade Federal do Paraná, Curitiba, 2005.

MELLO, Hivy Damasio Araújo. O Banco Mundial e a Educação no Brasil: Convergências em torno de uma agenda global. 17/12/2012, p. 379. Tese Doutorado – UNICAMP.

MICROSOFT. **Conteúdo do modelo de mineração para modelos de árvore de decisão**: Analysis Services – Mineração de Dados. Disponível em: <<http://msdn.microsoft.com/pt-br/library/cc645758.aspx>>. Acesso em 13 set 2013.

MICROSOFT. **Algoritmo Microsoft Clustering**: Analysis Services – Mineração de Dados: Disponível em: <<http://msdn.microsoft.com/pt-br/library/ms174879.aspx>>. Acesso em 13 set 2013.

MINAYO, Maria Cecília de Souza; DESLANDES, Suely Ferreira. **Pesquisa social**: teoria, método e criatividade. 30. ed. Petrópolis, RJ: Vozes, 2011. 108 p.

MORROW, Raymond A.; TORRES, Carlos Alberto. Estado, Globalização e políticas Educacionais. In: BURBULES, Nicholas, C.; TORRES, Carlos Alberto. **Globalização e Educação**: perspectivas críticas. Porto Alegre: Artmed Editora, 2004.

MOROZ, Melania; GIANFALDONI, Mônica Helena Tieppo Alves. **O processo de pesquisa**: iniciação. Brasília: Plano, 2002. 108 p.

NAMEN, Anderson A; BORGES, Sonia; SADALA, Maria. Indicadores de qualidade do ensino fundamental: o uso das tecnologias de mineração de dados e de visões multidimensionais para apoio à análise e definição de políticas públicas. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 94, n. 238, 677-700, set/dez. 2013.

NEVES, José Carlos de Araújo. **Fundação Dom Aguirre**: 35 anos de história. Sorocaba , SP: FDA, 1998. 215 p.

PALESTINO, Carlos Barbieri. **BI-business intelligence** : modelagem e tecnologia. Rio de Janeiro: Axcel Books do Brasil, 2001. 418 p.

SILVA, José Luís Conceição. **Proposta para reforma do setor primário da economia brasileira**. Brasília : Thesaurus, 2002. 126p.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistema de banco de dados**. Rio de Janeiro: Elsevier, 2006

TAN, Pang-ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao data mining: mineração de dados**. Rio de Janeiro: Ciência Moderna, 2009.

TAVARES, Claudio; BOZZA, Daniel; KONO, Frank. Descoberta de conhecimento aplicado a dados eleitorais. **Gestão & Conhecimento**, v. 5, n. 01, 54-94, jan./jun. 2007.

TURBAN, Efraim.; RAINER, Kelly; POTTER, Richard. **Administração de tecnologia da informação: tecnologia e prática**. Rio de Janeiro: Elsevier, 2005.

VANNUCCHI, Aldo. **A caminho da Uniso: história: casos e causos**. Sorocaba, SP: EDUniso, 2012. 157 p.

APÊNDICE A - CURSOS DA GRADUAÇÃO – UNIVERSIDADE DE SOROCABA**Ciências Agrárias**

Agronomia

Medicina Veterinária

Ciências Biológicas e da Saúde

Biotecnologia

Educação Física (Bacharelado)

Educação Física (Licenciatura)

Enfermagem

Farmácia

Fisioterapia

Nutrição

Terapia Ocupacional

Ciências Exatas e da Terra

Ciência da Computação

Física

Matemática

Química

Química Industrial

Sistemas de Informação

Ciências Humanas

Filosofia

Geografia

História

Pedagogia

Psicologia

Relações Internacionais

Ciências Sociais Aplicadas

Administração

Arquitetura e Urbanismo

Ciências Contábeis

Ciências Econômicas

Comércio Exterior

Comunicação Social: Jornalismo

Comunicação Social: Publicidade e Propaganda

Comunicação Social: Relações Públicas

Design

Direito

Hotelaria

Engenharias

Engenharia Ambiental

Engenharia Civil

Engenharia da Computação

Engenharia de Alimentos

Engenharia de Bioprocessos e Biotecnologia
Engenharia de Controle e Automação (Engenharia Mecatrônica)
Engenharia de Materiais
Engenharia de Produção
Engenharia Elétrica
Engenharia Química

Linguística, Letras e Artes

Artes Visuais
Dança
Letras: Português/Espanhol
Letras: Português/Inglês
Música
Teatro: Arte-Educação

Cursos Tecnológicos

Design de Interiores
Design de Produto
Design Gráfico
Design de Moda
Estética e Cosmética
Eventos
Gastronomia
Gestão Ambiental
Gestão Comercial
Gestão da Qualidade
Gestão da Produção Industrial
Gestão de Equinocultura
Gestão de Recursos Humanos
Gestão Financeira
Jogos Digitais
Logística
Marketing
Processos Gerenciais
Segurança no Trabalho

APÊNDICE B - CURSOS DA PÓS-GRADUAÇÃO LATO SENSU – UNIVERSIDADE DE SOROCABA

Controle de Qualidade em Indústrias de Alimentos
Engenharia Ambiental Industrial
Engenharia de Avaliação e Controle de Riscos Ambientais
Engenharia de Produção
Estética e Cosmetologia
Gestão Escolar
MBA em Comércio Exterior e Logística Internacional
MBA em Controladoria e Finanças
MBA em Gestão Contábil e Tributária
MBA em Gestão Estratégica
MBA em Gestão Estratégica de Pessoas
MBA em Operações e Logística
Psicopedagogia: Abordagem Clínica e Institucional

**APÊNDICE C – QUANTIDADE DE INGRESSANTES 2006-2011 POR CIDADE DO
ESTADO DE SÃO PAULO**

CIDADE	QUANTIDADE
SOROCABA	15151
VOTORANTIM	1973
TATUI	651
SALTO DE PIRAPORA	480
PIEDADE	476
ARACOIABA DA SERRA	427
SAO ROQUE	404
ITAPETININGA	373
PORTO FELIZ	370
MAIRINQUE	369
ITU	337
IBIUNA	319
BOITUVA	310
PILAR DO SUL	302
ALUMINIO	265
CERQUILHO	237
SALTO	204
TIETE	141
IPERO	132
CAPELA DO ALTO	132
SAO MIGUEL ARCANJO	84
SAO PAULO	83
ARACARIGUAMA	62
ITAPEVA	60
INDAIATUBA	57
CESARIO LANGE	44
CAPAO BONITO	42
VARGEM GRANDE PAULISTA	41
TAPIRAI	40
SARAPUI	33
ANGATUBA	26
LARANJAL PAULISTA	22
PEREIRAS	21
CAMPINAS	18
CONCHAS	15
PIRACICABA	12
ITAPORANGA	12
OSASCO	12
GUAREI	11

BURI	11
REGISTRO	10
ALAMBARI	9
COTIA	8
SANTO ANDRE	7
ITARARE	7
JUNDIAI	7
PARANAPANEMA	7
PORANGABA	6
BAURU	6
SAO BERNARDO DO CAMPO	6
BRAGANCA PAULISTA	6
CORONEL MACEDO	6
ITAI	6
SANTOS	6
PIRAJU	6
APIAI	5
TAQUARITUBA	5
AVARE	5
ITAPEVI	5
GUARULHOS	5
QUADRA	5
GUAPIARA	4
SAO JOSE DOS CAMPOS	4
CABREUVA	4
MARILIA	4
BARUERI	4
BOTUCATU	4
ATIBAIA	3
SOCORRO	3
CATANDUVA	3
ITATIBA	3
VINHEDO	3
RIBEIRAO BRANCO	3
ITABERA	3
RIBEIRAO PRETO	3
ILHA COMPRIDA	3
RIO CLARO	3
CARAPICUIBA	3
PEDREIRA	2
NOVA ODESSA	2
CONCHAL	2
ARARAS	2
TORRE DE PEDRA	2

OURINHOS	2
SETE BARRAS	2
LIMEIRA	2
BIRIGUI	2
SANTA BARBARA D'OESTE	2
IBITINGA	2
JANDIRA	2
ITAOCA	2
JOSE BONIFACIO	2
PRAIA GRANDE	2
JUMIRIM	2
MOGI DAS CRUZES	2
SAO CAETANO DO SUL	2
TABOAO DA SERRA	2
SAO CARLOS	2
TAQUARITINGA	2
SAO JOAO DA BOA VISTA	2
TAQUARIVAI	2
PIRASSUNUNGA	2
LEME	2
SAO MANUEL	2
CANDIDO MOTA	2
MINEIROS DO TIETE	2
CERQUEIRA CESAR	2
SALTO GRANDE	2
ADAMANTINA	2
SAO JOSE DO RIO PRETO	1
CAMPOS DO JORDAO	1
MANDURI	1
CAPIVARI	1
BOFETE	1
BORBOREMA	1
FRANCO DA ROCHA	1
PINDAMONHANGABA	1
SODRELIA	1
IPORANGA	1
HOLAMBRA II	1
IRACEMAPOLIS	1
PENAPOLIS	1
BARRA BONITA	1
ESTIVA GERBI	1
AGUAI	1
GUAIRA	1
Brigadeiro Tobias	1

BERNARDINO DE CAMPOS	1
BATATAIS	1
SUD MENUCCI	1
PRESIDENTE VENCESLAU	1
MOGI-GUACU	1
AMERICANA	1
ARARAQUARA	1
CAJAMAR	1
PAULINIA	1
ITAPIRA	1
JUNQUEIROPOLIS	1
RIBEIRAO GRANDE	1
ELIAS FAUSTO	1
CAJURU	1
LENCOIS PAULISTA	1
CAMPINA DO MONTE ALEGRE	1
SAO PEDRO	1
RIVERSUL	1
SAO ROQUE DA FARTURA	1
CRAVINHOS	1
SERRA NEGRA	1
CRUZEIRO	1
MAUA	1
JACAREI	1
CAMPO LIMPO PAULISTA	1
JALES	1
SUMARE	1
SANTA BRANCA	1
TAGUAI	1
SANTA CRUZ DAS PALMEIRAS	1
MURUTINGA DO SUL	1
SANTANA DE PARNAIBA	1
IARAS	1
CUBATAO	1
TAUBATE	1
VOTUPORANGA	1
PARDINHO	1
DIADEMA	1
BARIRI	1
DIVINOLANDIA	1
ILHABELA	1
ELDORADO	1
JAU	1

APÊNDICE D – QUANTIDADE DE INGRESSANTES POR CURSO

Curso	2006	2007	2008	2009	2010	2011	Total Geral
Administração	11	637	336	321	242	430	1977
Pedagogia	84	571	389	337	163	204	1748
Direito	203	171	144	244	242	413	1417
Curso Superior de Tecnologia em Gestão da Produção Industrial	245	329	244	181	100	104	1203
Engenharia de Produção			135	248	225	351	959
Ciências Contábeis	273	151	100	115	83	188	910
Curso Superior de Tecnologia em Logística		202	138	154	121	179	794
Administração de Negócios	706	46	22	7	1		782
Curso Superior de Tecnologia em Gestão de Recursos Humanos	92	229	142	105	86	86	740
Comunicação Social - Habil. em Publicidade e Propaganda	173	112	87	76	82	135	665
Engenharia Ambiental		2	156	158	110	210	636
Comunicação Social - Habil. em Jornalismo	94	77	90	80	64	128	533
Curso Superior de Tecnologia em Gestão Ambiental	7	132	110	71	101	106	527
Sistemas de Informação	154	133	51	61	55	38	492
Engenharia Civil					147	330	477
Arquitetura e Urbanismo					164	298	462
Curso Superior de Tecnologia em Gestão Financeira	98	105	68	80	50	47	448
Comércio Exterior	2	87	76	80	87	103	435
História	89	74	56	51	53	84	407
Farmácia	98	67	35	69	68	64	401
Ciências Econômicas	94	73	54	46	49	65	381
Hotelaria	99	59	25	63	55	43	344
Curso Superior de Tecnologia em Gestão da Qualidade				121	92	111	324
Nutrição	57	30	30	34	55	116	322
Química	80	61	62	61	37	19	320
Ciência da Computação	49	56	43	54	12	100	314
Engenharia Química					119	186	305
Biotecnologia	55	53	48	54	37	42	289
Curso Superior de Tecnologia em Gastronomia	19	42	48	60	62	53	284
Terapia Ocupacional	61	44	25	43	51	47	271
Filosofia	77	74	14	35	31	39	270
Curso Superior de Tecnologia em Design Gráfico	53	46	30	61	32	42	264
Comunicação Social - Habil. em Relações Públicas	78	54	33	45	43	10	263
Curso Superior de Tecnologia em Gestão de Equinocultura			35	61	87	74	257
Matemática	76	53	16	32	34	31	242
Teatro - Habilitação: Arte Educação	85	59	24	22	25	21	236
Design				79	80	77	236
Letras - Habilitação em Português e Inglês	2	3	2	66	79	72	224
Curso Superior de Tecnologia em Gestão Ambiental Industrial	216				1		217
Pedagogia- Docência nas Séries Iniciais do Ensino Fundamental	185	10	3		2		200
Curso Superior de Tecnologia em Marketing		61	43	37	30	25	196

Letras - Habilitação em Português	74	55	42	14	1		186
Curso Superior de Tecnologia em Estética e Cosmética					66	110	176
Fisioterapia					53	121	174
Engenharia de Controle e Automação					64	110	174
Letras - Habilitação em Inglês	65	47	33	24	3		172
Música					114	55	169
Física	56	35	15	15	13	10	144
Engenharia da Computação					50	76	126
Enfermagem					51	70	121
Turismo	49	22	22	15	1		109
Administração em Comércio Exterior	101	7					108
Educação Física					46	55	101
Curso Superior de Tecnologia em Gestão de Marketing de Varejo	85						85
Engenharia Elétrica					30	52	82
Curso Superior de Tecnologia em Processos Gerenciais					25	35	60
Curso Superior de Tecnologia em Design de Interiores				28	14	18	60
Curso Superior de Tecnologia em Gestão Comercial					29	31	60
Curso Superior de Tecnologia em Gestão de Logística	48	6	2	2			58
Pedagogia- Docência nas Séries Iniciais do Ensino Fundamental	26	4	4	1	2	2	39
Dança					19	12	31
Curso Superior de Tecnologia em Design de Produto				11	11	2	24
Artes Visuais					10	12	22
Curso Superior de Tecnologia em Fotografia					11	6	17
Relações Internacionais					11	6	17
Curso Superior de Tecnologia em Produção Multimídia					10		10
Secretariado Executivo					8	2	10
Ciências Biológicas (Licenciatura)					10		10
Curso Superior de Tecnologia em Gestão Hospitalar					8	1	9
Curso Superior de Tecnologia em Eventos					9		9
Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas					8		8
Curso Superior de Tecnologia em Segurança no Trabalho					2	6	8
Geografia				3	4		7
Curso Superior de Tecnologia em Redes de Computadores					7		7
Curso Superior de Tecnologia em Alimentos					2	3	5
Administração de Empresas	1	1	2		1		5
Letras - Habilitação em Português e Espanhol		1		3	1		5
Ciências Biológicas (Bacharelado)					5		5
Curso Superior de Tecnologia em Design de Moda			5				5
Curso Superior de Tecnologia em Gestão da Tecnologia da Informação					4		4
Curso Superior de Tecnologia em Gestão do Turismo					4		4
Pedagogia- Hab. Doc. na Educação Infantil e Adm. Esc.		2	1				3
Curso Superior de Tecnologia em Gestão de Segurança Pública					2		2

Curso Superior de Tecnologia em Sistemas para Internet					2		2
Curso Superior de Tecnologia em Negócios Imobiliários					1		1
Curso Superior de Tecnologia em Gestão Imobiliária						1	1
Curso Superior de Tecnologia em Banco de Dados					1		1
Agronomia (Bacharelado)						1	1
Componentes Curriculares para Alunos não Regulares	1						1
Total Geral	4121	4083	3040	3528	3970	5468	24210

ANEXO A – POPULAÇÃO CENSO DEMOGRÁFICO 2010



Tabela 1.1.1 - População residente, por situação do domicílio e sexo, segundo os grupos de idade - Brasil - 2010

Grupos de Idade	População residente								
	Total	Homens	Mulheres	Situação do domicílio e sexo					
				Urbana			Rural		
				Total	Homens	Mulheres	Total	Homens	Mulheres
Total	190 755 799	93 406 990	97 348 809	160 925 792	77 710 174	83 215 618	29 830 007	15 696 816	14 133 191
0 a 4 anos	13 796 158	7 016 987	6 779 171	11 301 146	5 745 123	5 556 023	2 495 012	1 271 864	1 223 148
Menos de 1 ano	2 713 244	1 378 532	1 334 712	2 246 033	1 141 784	1 104 249	467 211	236 748	230 463
1 ano	2 694 909	1 369 425	1 325 484	2 217 676	1 126 489	1 091 187	477 233	242 936	234 297
2 anos	2 726 957	1 387 143	1 339 814	2 230 049	1 133 749	1 096 300	496 908	253 394	243 514
3 anos	2 790 782	1 419 295	1 371 487	2 274 676	1 156 129	1 118 547	516 106	263 166	252 940
4 anos	2 870 266	1 462 592	1 407 674	2 332 712	1 186 972	1 145 740	537 554	275 620	261 934
5 a 9 anos	14 969 375	7 624 144	7 345 231	12 135 294	6 169 531	5 965 753	2 834 091	1 454 613	1 379 478
5 anos	2 931 988	1 493 253	1 438 735	2 381 802	1 211 695	1 170 107	550 186	281 558	268 628
6 anos	2 894 419	1 472 635	1 421 784	2 349 036	1 193 318	1 155 718	545 383	279 317	266 066
7 anos	2 959 192	1 506 828	1 452 364	2 394 890	1 217 869	1 177 011	564 312	288 959	275 353
8 anos	2 995 612	1 528 003	1 467 609	2 422 432	1 232 734	1 189 698	573 180	295 289	277 911
9 anos	3 188 164	1 623 425	1 564 739	2 587 134	1 313 915	1 273 219	601 030	309 510	291 520
10 a 14 anos	17 166 761	8 725 413	8 441 348	13 956 994	7 062 055	6 894 929	3 209 777	1 663 358	1 546 419
10 anos	3 505 216	1 793 146	1 712 070	2 844 120	1 450 908	1 393 212	661 096	342 238	318 858
11 anos	3 352 844	1 709 291	1 643 553	2 731 861	1 386 682	1 345 179	620 983	322 609	298 374
12 anos	3 402 242	1 727 220	1 675 022	2 765 478	1 397 964	1 367 514	636 764	329 256	307 508
13 anos	3 412 748	1 728 588	1 684 160	2 774 748	1 399 281	1 375 467	639 000	329 307	308 693
14 anos	3 493 711	1 767 168	1 726 543	2 840 777	1 427 220	1 413 557	652 934	339 948	312 986
15 a 19 anos	16 990 872	8 558 868	8 432 004	14 038 999	6 998 102	7 040 897	2 951 873	1 560 786	1 391 107
15 anos	3 574 929	1 802 172	1 772 757	2 913 134	1 456 230	1 456 904	661 795	345 942	315 853
16 anos	3 410 704	1 719 415	1 691 289	2 800 800	1 398 165	1 402 635	609 904	321 250	288 654
17 anos	3 372 241	1 701 889	1 670 352	2 779 182	1 386 897	1 392 285	593 059	314 992	278 067
18 anos	3 367 172	1 699 061	1 668 111	2 801 550	1 397 096	1 404 454	565 622	301 965	263 657
19 anos	3 265 826	1 636 331	1 629 495	2 744 333	1 359 714	1 384 619	521 493	276 617	244 876
20 a 24 anos	17 245 192	8 630 229	8 614 963	14 706 068	7 276 963	7 429 105	2 539 124	1 353 266	1 185 858
20 anos	3 389 729	1 698 081	1 691 648	2 863 687	1 418 305	1 445 382	526 042	279 776	246 266
21 anos	3 421 736	1 713 268	1 708 468	2 916 244	1 444 621	1 471 623	505 492	268 647	236 845
22 anos	3 510 497	1 760 477	1 750 020	2 992 003	1 482 841	1 509 162	518 494	277 636	240 858
23 anos	3 420 254	1 709 675	1 710 579	2 927 520	1 446 578	1 480 942	492 734	263 097	229 637
24 anos	3 502 976	1 748 728	1 754 248	3 006 614	1 484 618	1 521 996	496 362	264 110	232 252
25 a 29 anos	17 104 414	8 460 995	8 643 419	14 772 956	7 225 732	7 547 224	2 331 458	1 235 263	1 096 195
30 a 34 anos	15 744 512	7 717 658	8 026 854	13 611 920	6 596 876	7 025 044	2 132 592	1 130 782	1 001 810
35 a 39 anos	13 888 579	6 766 664	7 121 915	11 975 407	5 750 498	6 224 909	1 913 172	1 016 166	897 006
40 a 44 anos	13 009 364	6 320 568	6 688 796	11 187 427	5 344 981	5 842 446	1 821 937	975 587	846 350
45 a 49 anos	11 833 352	5 692 014	6 141 338	10 181 394	4 806 322	5 375 072	1 651 958	885 692	766 266
50 a 54 anos	10 140 402	4 834 995	5 305 407	8 708 339	4 074 679	4 633 660	1 432 063	760 316	671 747
55 a 59 anos	8 276 221	3 902 344	4 373 877	7 025 474	3 238 531	3 786 943	1 250 747	663 813	586 934
60 a 64 anos	6 509 120	3 041 035	3 468 085	5 474 944	2 479 882	2 995 062	1 034 176	561 153	473 023
65 a 69 anos	4 840 810	2 224 065	2 616 745	4 040 016	1 792 798	2 247 218	800 794	431 267	369 527
70 a 74 anos	3 741 636	1 667 372	2 074 264	3 142 173	1 349 329	1 792 844	599 463	318 043	281 420
75 a 79 anos	2 563 447	1 090 517	1 472 930	2 174 038	889 908	1 284 130	389 409	200 609	188 800
80 a 84 anos	1 666 972	668 623	998 349	1 423 603	546 865	876 738	243 369	121 758	121 611
85 a 89 anos	819 483	310 759	508 724	695 385	251 112	444 273	124 098	59 647	64 451
90 a 94 anos	326 558	114 964	211 594	273 348	90 960	182 388	53 210	24 004	29 206
95 a 99 anos	98 335	31 529	66 806	81 121	24 365	56 756	17 214	7 164	10 050
100 anos ou mais	24 236	7 247	16 989	19 766	5 562	14 204	4 470	1 685	2 785

Fonte: IBGE, Censo Demográfico 2010.

ANEXO B – EVOLUÇÃO DO NÚMERO DE INSTITUIÇÕES DE EDUCAÇÃO SUPERIOR

Tabela 1 – Evolução do Número de Instituições de Educação Superior por Categoria Administrativa – Brasil – 2001-2010

Ano	Total	Pública								Privada	%
		Total	%	Federal	%	Estadual	%	Municipal	%		
2001	1.391	183	13,2	67	4,8	63	4,5	53	3,8	1.208	86,8
2002	1.637	195	11,9	73	4,5	65	4,0	57	3,5	1.442	88,1
2003	1.859	207	11,1	83	4,5	65	3,5	59	3,2	1.652	88,9
2004	2.013	224	11,1	87	4,3	75	3,7	62	3,1	1.789	88,9
2005	2.165	231	10,7	97	4,5	75	3,5	59	2,7	1.934	89,3
2006	2.270	248	10,9	105	4,6	83	3,7	60	2,6	2.022	89,1
2007	2.281	249	10,9	106	4,6	82	3,6	61	2,7	2.032	89,1
2008	2.252	236	10,5	93	4,1	82	3,6	61	2,7	2.016	89,5
2009	2.314	245	10,6	94	4,1	84	3,6	67	2,9	2.069	89,4
2010	2.378	278	11,7	99	4,2	108	4,5	71	3,0	2.100	88,3

Fonte: MEC/Inep

ANEXO C – EVOLUÇÃO DO NÚMERO DE MATRÍCULAS

**Tabela 10 – Evolução do Número de Matrículas por Categoria Administrativa –
Brasil – 2001-2010**

Ano	Total	Pública								Privada	
		Total	%	Federal	%	Estadual	%	Municipal	%	Privada	%
2001	3.036.113	944.584	31,1	504.797	16,6	360.537	11,9	79.250	2,6	2.091.529	68,9
2002	3.520.627	1.085.977	30,8	543.598	15,4	437.927	12,4	104.452	3,0	2.434.650	69,2
2003	3.936.933	1.176.174	29,9	583.633	14,8	465.978	11,8	126.563	3,2	2.760.759	70,1
2004	4.223.344	1.214.317	28,8	592.705	14,0	489.529	11,6	132.083	3,1	3.009.027	71,2
2005	4.567.798	1.246.704	27,3	595.327	13,0	514.726	11,3	136.651	3,0	3.321.094	72,7
2006	4.883.852	1.251.365	25,6	607.180	12,4	502.826	10,3	141.359	2,9	3.632.487	74,4
2007	5.250.147	1.335.177	25,4	641.094	12,2	550.089	10,5	143.994	2,7	3.914.970	74,6
2008	5.808.017	1.552.953	26,7	698.319	12,0	710.175	12,2	144.459	2,5	4.255.064	73,3
2009	5.954.021	1.523.864	25,6	839.397	14,1	566.204	9,5	118.263	2,0	4.430.157	74,4
2010	6.379.299	1.643.298	25,8	938.656	14,7	601.112	9,4	103.530	1,6	4.736.001	74,2

Fonte: MEC/Inep

ANEXO D – EVOLUÇÃO DAS TAXAS DE ESCOLARIZAÇÃO BRUTA E LÍQUIDA NA EDUCAÇÃO SUPERIOR

Tabela 8 – Evolução das Taxas de Escolarização Bruta e Líquida na Educação Superior – Brasil e Regiões – 2001-2009

Brasil/ Regiões	Educação Superior (18 a 24 anos)																	
	Escolarização Bruta									Escolarização Líquida								
	2001	2002	2003	2004	2005	2006	2007	2008	2009	2001	2002	2003	2004	2005	2006	2007	2008	2009
Brasil	15,1	16,6	18,6	18,6	19,9	22,6	24,3	25,5	26,7	8,9	9,8	10,6	10,5	11,2	12,6	13,1	13,7	14,4
Centro-Oeste	18,2	21,8	23,6	23,2	25,9	27,0	28,8	31,1	32,6	9,7	11,9	12,3	12,2	14,0	14,8	15,6	16,3	17,9
Nordeste	9,1	9,5	10,9	11,1	11,9	14,3	15,9	16,9	18,4	5,1	5,1	5,8	5,9	6,1	7,1	7,7	8,3	9,4
Norte	11,3	15,3	14,7	12,2	14,6	17,0	19,3	21,7	23,7	5,2	6,7	6,1	5,7	7,0	7,6	9,0	9,9	11,0
Sudeste	17,3	19,0	21,2	22,0	23,3	26,7	28,6	29,4	30,1	10,9	12,0	12,8	13,0	13,8	15,7	16,4	16,6	16,8
Sul	21,3	22,8	26,7	26,5	27,8	29,7	30,8	32,9	33,5	12,7	13,7	15,9	15,3	16,2	17,1	16,8	18,7	19,2

Fonte: IBGE/Pnad; elaborado por MEC/Inep

Nota: Para os anos 2001, 2002 e 2003, exclusive a população rural de RO, AC, AM, RR, PA e AP.