

**UNIVERSIDADE DE SOROCABA  
PROGRAMA DE PÓS-GRADUAÇÃO EM PROCESSOS TECNOLÓGICOS E  
AMBIENTAIS**

**Gustavo Giancoli Martins Ferreira**

**DESENVOLVIMENTO DE APLICATIVO PARA ANÁLISE FINANCEIRA  
UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

**Sorocaba/SP  
2024**

**Gustavo Giancoli Martins Ferreira**

**DESENVOLVIMENTO DE APLICATIVO PARA ANÁLISE FINANCEIRA  
UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

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

Orientador: Prof. Dr. Norberto Aranha

**Sorocaba/SP  
2024**

### Ficha Catalográfica

F441d Ferreira, Gustavo Giancoli Martins  
Desenvolvimento de aplicativo para análise financeira utilizando inteligência artificial / Gustavo Giancoli Martins Ferreira. — 2024.  
110 f. : il.

Orientador: Prof. Dr. Norberto Aranha  
Dissertação (Mestrado em Processos Tecnológicos e Ambientais)  
– Universidade de Sorocaba, Sorocaba, SP, 2024.

1. Finanças – Inovações tecnológicas. 2. Inteligência artificial. 3. Tecnologia da informação. 4. Software - Desenvolvimento. 5. Interface de programas aplicativos (Software). I. Aranha, Norberto, orient. II. Universidade de Sorocaba. III. Título.

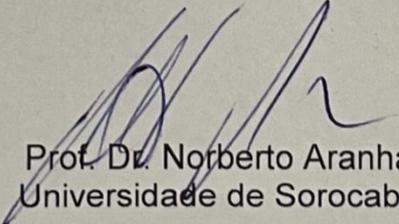
**Gustavo Giancoli Martins Ferreira**

**DESENVOLVIMENTO DE APLICATIVO PARA ANÁLISE FINANCEIRA  
UTILIZANDO INTELIGÊNCIA ARTIFICIAL**

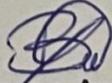
Dissertação apresentada à Banca Examinadora do Programa de Pós-graduação em Processos Tecnológicos e Ambientais - PTA da Universidade de Sorocaba, como exigência Parcial para obtenção do título de Mestre em Processos Tecnológicos e Ambientais.

Aprovado em: 05/08/2024

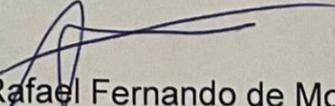
**BANCA EXAMINADORA:**



Prof. Dr. Norberto Aranha  
Universidade de Sorocaba



Prof. Dr. Bruno Aguilar da Cunha  
Universidade de Sorocaba



Prof. Dr. Rafael Fernando de Moraes Moreno  
Universidade de Sorocaba

Dedico este trabalho a minha querida esposa e aos meus pais e familiares por todo incentivo e apoio para o desenvolvimento e conclusão deste projeto.

## **AGRADECIMENTOS**

Primeiramente à Deus, que permitiu que este ciclo se completasse e por todos os acontecimentos ao longo de nossas vidas.

Agradeço a todos os professores por nos proporcionarem o conhecimento racional, a manifestação do caráter e a afetividade da educação no processo de formação profissional e pelo tanto que se dedicaram a nós. A palavra “mestre” nunca fará justiça aos professores dedicados aos quais, sem nominar, terão nossos eternos agradecimentos.

A minha querida esposa e família que, direta ou indiretamente, contribuíram na elaboração deste trabalho, em especial ao orientador Norberto, por acreditar, auxiliar e me direcionar ao longo de todo desenvolvimento deste projeto.

Muito obrigado!

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê”

(Arthur Schopenhauer)

## RESUMO

Com as grandes evoluções tecnológicas e a utilização de Inteligência artificial surgem novas ideologias e conceitos, sendo comum em muitas delas um objetivo único, melhorar a qualidade de vida do ser humano. Instituições financeiras têm utilizado a IA como forma de aplicar estratégias para conquistar e manter a fidelidade do cliente, muitas vezes se baseando como foco a satisfação do cliente. Por outro lado, levanta-se a questão se estratégias de IA utilizadas estão melhorando a qualidade de vida do cliente; se ocasionarão problemas financeiros futuros, e como a instituição pode auxiliar um cliente com problemas existentes, de modo a agir oferecendo produtos financeiros que auxiliem em seu cotidiano. Neste contexto, foi desenvolvido um aplicativo colaborativo capaz de auxiliar na tomada de decisão sobre a escolha do produto financeiro entre instituições financeiras com ênfase no financiamento imobiliária. Por meio de regras e modelos de árvore de decisão de inteligência artificial com base nos parâmetros de entrada do usuário o aplicativo é capaz de gerar *insights*, como, entrada ideal, imóvel ideal e risco de inadimplência, assim como sugestões para melhorar sua saúde financeira antes de adquirir o produto financeiro. Deste modo, o aplicativo apresentou esclarecimentos para o cliente na tomada de decisão e quais fatores podem prejudicar sua saúde financeira, os modelos de árvore de decisão apresentaram resultados favoráveis de acurácia, por outro lado as instituições podem desenvolver estratégias tecnológicas com foco em apoiar a melhor decisão para a situação financeira do cliente, assim como, evitar situações em que prejudica a qualidade de vida dos seus clientes.

**Palavras-chave:** Inteligência Artificial; Crédito Financeiro; Crédito Imobiliário; Tecnologia.

## ABSTRACT

With significant technological advancements and the rise of artificial intelligence, new ideologies and concepts have emerged, many of which share a common goal: improving the quality of human life. Financial institutions have been leveraging AI to implement strategies aimed at gaining and maintaining customer loyalty, often focusing on customer satisfaction. However, this raises questions about whether AI-driven strategies are truly enhancing the quality of life for customers, whether they might lead to future financial issues, and how institutions can assist customers facing existing problems by offering financial products that support their daily lives. In this context, a collaborative application was developed to assist in decision-making regarding the selection of financial products across institutions, with a particular emphasis on mortgage financing. Using decision tree rules and models based on user input parameters, the application can generate insights such as the ideal down payment, the most suitable property, and the risk of default, as well as offer suggestions for improving financial health before acquiring a financial product. This way, the application provides clarity for customers in their decision-making process and identifies factors that may harm their financial well-being. While the decision tree models demonstrated favorable accuracy, financial institutions can develop technological strategies focused on supporting the best decision for the customer's financial situation, while also avoiding scenarios that could negatively impact the quality of life of their clients.

**Keywords:** Artificial intelligence; Financial Credit; Real Estate financing; Technology.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Especificações de Inteligência Artificial, <i>Machine Learning</i> e <i>Deep Learning</i> .....	23
Figura 2 - Fluxo de trabalho comum no Machine Learning .....	24
Figura 3 - Fases para o desenvolvimento de algoritmos de machine learning.....	25
Figura 4 - Interação com o ChatGPT .....	26
Figura 5 - Benefícios do ChatGPT .....	27
Figura 6 - Exemplo de uma Árvore de Decisão .....	28
Figura 7 - Tipo dos Dados .....	33
Figura 8 - Api e suas rotas .....	36
Figura 9 - Uso das linguagens de programação em 2023.....	37
Figura 10 - Diagrama de Uso .....	38
Figura 11 - Fluxograma do Usuário.....	39
Figura 12 - Tela de Informações Pessoais - etapa 1 .....	45
Figura 13 - Tela de Saúde Financeira – etapa 2 .....	46
Figura 14 - Tela de Crédito Imobiliário – etapa 3 .....	47
Figura 15 - Consentimento Lei Geral de Proteção de Dados - Usar .....	48
Figura 16 - Consentimento Lei Geral de Proteção de Dados - Colaborar .....	48
Figura 17 - Rotas do software .....	50
Figura 18 - Configuração do <i>bucket</i> s3 .....	52
Figura 19 - Política de segurança do <i>bucket</i> .....	52
Figura 20 - Exemplo de Armazenamento no S3.....	53
Figura 21 - Estrutura da tabela NoSQL .....	55
Figura 22 - Política de acesso da tabela NoSQL.....	55
Figura 23 - Dado armazenado na tabela de Processos .....	56
Figura 24 - Configuração da instancia do EC2.....	58
Figura 25 - Funções AWS Lambda .....	59
Figura 26 - Configuração da AWS Lambda.....	60
Figura 27 – Estrutura da base de treinamento – Modelo de <i>Score</i> Crédito.....	62
Figura 28 - Mix de crédito por Profissão e Idade Média por Profissão .....	63
Figura 29 - Média do Salário por Mix de crédito.....	64
Figura 30 - Média da Taxa de Juros por Mix de Crédito .....	64
Figura 31 - Correlação das Variáveis - Modelo de <i>Score</i> Crédito .....	65
Figura 32 - Estrutura da base de treinamento – Modelo de Risco Inadimplência .....	69
Figura 33 - Média do valor empréstimo por target.....	70
Figura 34 - Correlação das Variáveis - Modelo de Risco de Inadimplência .....	71
Figura 35 - Contextualização do Prompt do ChatGPT .....	75
Figura 36 - Arquitetura de Solução.....	76
Figura 37 - Métricas <i>Decision Tree</i> vs XGBoost – <i>Score</i> de Crédito.....	78
Figura 38 - Curva ROC AUC <i>Decision Tree</i> - <i>Score</i> de Crédito .....	79
Figura 39 - Curva ROC AUC XGBoost - <i>Score</i> de Crédito.....	80
Figura 40 - <i>K-Fold Decision Tree</i> - <i>Score</i> de Crédito .....	81
Figura 41 - Peso das variáveis no Modelo <i>Score</i> de Crédito .....	82
Figura 42 - Métricas <i>Decision Tree</i> vs XGBoost – Risco de Inadimplência.....	82
Figura 43 - <i>K-Fold XGBoost</i> - Risco de Inadimplência .....	83
Figura 44 - Curva ROC AUC XGBoost - Risco de Inadimplência.....	84
Figura 45 - Peso das variáveis no Modelo Risco de Inadimplência .....	85
Figura 46 - Normalização Pré-Crédito - <i>Score</i> de Crédito .....	88
Figura 47 - Normalização Pós-Crédito - <i>Score</i> de Crédito .....	88

Figura 48 - Importância das Feature na Simulação - Score de Crédito.....	89
Figura 49 - Normalização - Risco de Inadimplência .....	90
Figura 50 - Importância das Feature na Simulação - Risco de Inadimplência .....	91
Figura 51 - <i>E-mail</i> de resposta do <i>software</i> .....	92
Figura 52 - <i>E-mail</i> de resposta do <i>software</i> - Gen IA .....	92
Figura 53 - Normalização Pré-Crédito - Score de Crédito .....	94
Figura 54 - Normalização Pós-Crédito - Score de Crédito .....	95
Figura 55 - Importância das Feature na Simulação - Score de Crédito.....	95
Figura 56 - Normalização - Risco de Inadimplência .....	96
Figura 57 - Importância das Feature na Simulação - Risco de Inadimplência .....	97
Figura 58 - <i>E-mail</i> de resposta do <i>software</i> .....	97
Figura 59 - <i>E-mail</i> de resposta do <i>software</i> - Gen IA .....	98

## LISTA DE TABELAS

Tabela 1 - Comparação de Bancos de Dados NoSQL.....	35
Tabela 2 - Tabela das variáveis de entrada (continua) .....	43
Tabela 3 - Tabela clusterização das variáveis de Score de Crédito(continua) .....	66
Tabela 4 - Tabela das variáveis de entrada (continuação).....	67
Tabela 5 - Tabela das variáveis de entrada (conclusão).....	68
Tabela 6 - Tabela clusterização das variáveis de Risco de Inadimplência (continua) .....	72
Tabela 7 - Variáveis de entrada para simulação (continua) .....	86
Tabela 8 - Variáveis de entrada para simulação (continua) .....	93

## LISTA DE SIGLAS E ABREVIATURAS

S3	Simple Storage Service
EC2	Elastic Cloud Compute
NIST	Instituto Nacional de Standards e Tecnologia do Departamento de Comércio dos Estados Unidos
SES	Simple E-mail Service
GSMA	Global System for Mobile Communications Association
AWS	Amazon Web Service
GCP	Google Cloud Computing
NoSQL	Not Only Sql
SQL	Structured Query Language
API	Interface de programação de aplicações
GPT	Generative Pre-Trained Transformer

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>16</b>
<b>2</b>	<b>OBJETIVO.....</b>	<b>19</b>
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>20</b>
<b>3.1</b>	<b>Inadimplência no Brasil.....</b>	<b>20</b>
<b>3.2</b>	<b>Big Data .....</b>	<b>22</b>
<b>3.3</b>	<b>Inteligência Artificial.....</b>	<b>23</b>
<b>3.4</b>	<b>Modelo de Inteligência Artificial – ChatGPT .....</b>	<b>26</b>
<b>3.5</b>	<b>Modelo de Inteligência Artificial de Classificação – Árvore de Decisão ....</b>	<b>27</b>
<b>3.6</b>	<b>Computação em Nuvem .....</b>	<b>30</b>
<b>3.7</b>	<b>Serverless.....</b>	<b>32</b>
<b>3.8</b>	<b>Banco de Dados.....</b>	<b>32</b>
3.8.1	Banco de Dados Relacional.....	34
3.8.2	Banco de Dados Não Relacional .....	34
<b>3.9</b>	<b>Application Programming Interface .....</b>	<b>35</b>
<b>3.10</b>	<b>Linguagem de Programação Python.....</b>	<b>36</b>
<b>4</b>	<b>ETAPAS DE DESENVOLVIMENTO.....</b>	<b>38</b>
<b>4.1</b>	<b>Jornada de uso .....</b>	<b>38</b>
<b>4.2</b>	<b>Levantamento de Requisitos .....</b>	<b>40</b>
4.2.1	Requisitos Funcionais.....	40
4.2.2	Requisitos não funcionais .....	41
<b>4.3</b>	<b>Entrada de Dados .....</b>	<b>42</b>
4.3.1	Página Web .....	42
4.3.2	API.....	49
<b>4.4</b>	<b>Armazenamento dos Dados.....</b>	<b>51</b>
4.4.1	Simple Storage Service – S3.....	51
4.4.2	AWS DynamoDb.....	54
<b>4.5</b>	<b>Processamento dos Dados .....</b>	<b>56</b>
4.5.1	Elastic Compute Cloud – EC2 .....	57
4.5.2	AWS Lambda.....	59
<b>4.6</b>	<b>Modelo de Inteligência artificial.....</b>	<b>61</b>
4.6.1	Modelo de Score de Crédito .....	62
4.6.2	Modelo de Risco de Inadimplência .....	68
<b>4.7</b>	<b>Modelo de Inteligência artificial - ChatGPT .....</b>	<b>74</b>
<b>4.8</b>	<b>Notificação.....</b>	<b>75</b>
4.8.1	Simple E-mail Service - SES .....	75

<b>4.9</b>	<b>Arquitetura de Solução .....</b>	<b>76</b>
<b>4.10</b>	<b>Modelo de Score de Crédito.....</b>	<b>78</b>
<b>4.11</b>	<b>Modelo de Risco de Inadimplência .....</b>	<b>82</b>
<b>5</b>	<b>RESULTADOS .....</b>	<b>86</b>
<b>5.1</b>	<b>Simulação Positiva .....</b>	<b>86</b>
<b>5.2</b>	<b>Simulação Negativa .....</b>	<b>92</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS.....</b>	<b>99</b>
	<b>REFERÊNCIAS.....</b>	<b>103</b>

## 1 INTRODUÇÃO

O crescente uso de aplicativos na atualidade é um fenômeno que pode ser explicado por uma série de fatores, incluindo o aumento da adoção de dispositivos móveis, a evolução da tecnologia e a mudança nos hábitos de consumo. Os aplicativos são *softwares* desenvolvidos para serem executados em dispositivos móveis, como smartphones e tablets. Eles são uma das formas mais populares de usar a internet e desempenham um papel importante na vida cotidiana.

O desenvolvimento de *software* no Brasil tem crescido significativamente nos últimos anos. De acordo com o relatório emitido pelo Ministério da Ciência, Tecnologia e Inovações, o crescimento foi de 6,5% em relação a 2021 e maior do que a taxa global (2022). O relatório também estima um crescimento futuro impulsionado pela economia digital, demandando investimentos em segurança de dados e na aceleração da migração para nuvem.

O crescimento da utilização de dispositivos móveis é um dos principais fatores que impulsionam a expansão do uso de aplicativos. Segundo dados do relatório “A Economia Móvel 2019”, da *Global System for Mobile Communications Association* (GSMA), a população mundial possui mais de 7 bilhões de pessoas, e mais de 5 bilhões delas possuem um smartphone, e esse número só deve crescer nos próximos anos. O crescimento dos dispositivos móveis significa que mais pessoas têm acesso a aplicativos, o que aumenta a demanda por esses produtos (GSMA apud Agência Brasil, 2019).

A evolução da tecnologia também é um fator que contribui para o crescimento do uso de aplicativos. Os smartphones estão cada vez mais poderosos e com recursos avançados. Isso permite que os aplicativos sejam cada vez mais sofisticados e ofereçam uma gama mais ampla de funcionalidades. Por exemplo, os aplicativos podem agora ser usados para fazer compras, acessar informações, se comunicar e se divertir (RM Telefonia, 2023).

A mudança nos hábitos de consumo também é um fator que impulsiona o crescimento do uso de aplicativos. Isso tem levado ao desenvolvimento de aplicativos que atendem às diversas necessidades do usuário. Por exemplo, existem aplicativos que podem ser usados para comprar produtos, pesquisar informações, fazer chamadas e enviar mensagens de texto (Wünsche, 2023).

O uso de aplicativos é uma tendência que deve continuar crescendo nos próximos anos. Os aplicativos são uma forma conveniente e eficiente de usar a internet, e eles oferecem uma gama de funcionalidades que podem tornar mais fácil e agradável a vida do usuário (Wünsche, 2023).

As instituições também têm se beneficiado da tecnologia. Metzner e Matias (2015) relatam que o sistema financeiro tem usado a seu favor as evoluções tecnológicas das últimas décadas, como a internet, computador ou dispositivo móvel. No passado, toda a interação com produtos financeiros era realizada apenas nas agências bancárias fisicamente. Desde um extrato bancário até a contratação de um produto financeiro como, por exemplo, financiamento imobiliário. O internet banking possibilitou que as operações bancárias fossem realizadas de casa, sem a necessidade de se deslocar para as agências. Essa transformação foi favorável para ambos os lados, cliente pela comodidade e instituição financeira na parte de redução de custo.

Ao longo da vida as pessoas precisam tomar decisões financeiras, desde as mais simples como parcelar a compra de um eletrodoméstico até as mais complexas como o financiamento para aquisição de um imóvel (Costa; Miranda, 2013).

Para tomar tais decisões é necessário um conhecimento básico sobre produtos financeiros e informações relevantes relacionadas ao mercado financeiro para que o indivíduo possa fazer escolhas conscientes e adequadas ao seu perfil, condições financeiras e capacidade de pagamento, evitando futuros endividamentos (Piccini; Pinzetta, 2014).

Apesar da necessidade desses conhecimentos, observou-se que na maioria das famílias brasileiras o gerenciamento das finanças pessoais ainda não é uma prática comum. Poucas delas recebem orientação sobre educação financeira ou buscam maneiras eficientes de organizar e otimizar o uso do dinheiro ou da renda disponível. No que diz respeito a investimentos, os brasileiros tendem a adotar uma postura conservadora, optando por modalidades de baixo risco, como a caderneta de poupança, conhecida por sua tradicionalidade e perfil de investimento seguro (Borges, 2014).

Em vista dessa necessidade de gerenciamento financeiro por parte das famílias, buscou-se estudar a viabilidade de um aplicativo que ajude as pessoas nas tomadas de decisões financeiras. Portanto, pensando em um aplicativo financeiro que

venha apoiar e gerar insights para as escolhas pessoais sobre o produto financeiro de crédito imobiliário, a primeira pergunta que se coloca é:

Como o aplicativo financeiro pode ser aprimorado para fornecer insights mais valiosos aos usuários, especialmente no contexto da avaliação de produtos financeiros como o crédito imobiliário, com o objetivo final de mitigar o risco de inadimplência e promover decisões financeiras mais informadas e seguras por parte dos usuários?

Observando o cenário atual, indivíduos com melhor nível de compreensão, informação e conhecimento em relação à utilização consciente do dinheiro resultam em melhor capacidade de planejar, organizar e honrar seus compromissos financeiros. Consequentemente evitando situações de endividamento e inadimplência (Cordeiro; Costa; Silva, 2018).

Neste contexto, propõe-se neste trabalho o desenvolvimento de um aplicativo que, ao coletar informações do perfil do usuário - como gastos com cartões, reservas de emergência, despesas médicas e histórico financeiro - e dados do produto financeiro - como taxa de juros, tipo de indexador e detalhes do imóvel -, seja capaz de fornecer suporte à tomada de decisão em relação ao produto de crédito imobiliário desejado. O aplicativo também tem o objetivo de destacar os riscos associados à saúde financeira do usuário, com foco especial na possibilidade de inadimplência futura.

O aplicativo foi projetado para fornecer aos usuários informações fundamentadas em dados, visando aprimorar a tomada de decisões relacionadas a crédito imobiliário. Ao oferecer esse embasamento, busca-se potencializar a assertividade das escolhas dos usuários. Por meio da entrada de dados com informações do perfil do usuário e do produto financeiro, a ferramenta é capaz de gerar um relatório que compara diferentes produtos financeiros e seus riscos.

O objetivo deste trabalho é fornecer aos usuários informações fundamentadas em dados, pois sabe-se que decisões embasadas dessa forma tendem a ser mais assertivas. O aplicativo desenvolvido está focado em auxiliar os usuários na escolha do crédito imobiliário mais adequado às suas necessidades e objetivos financeiros, proporcionando um suporte valioso durante o processo de tomada de decisão.

## 2 OBJETIVO

### Objetivo Geral

Desenvolver um aplicativo baseado em Inteligência Artificial (IA) que auxiliará as pessoas na análise do produto financeiro de crédito imobiliário. O propósito principal desse aplicativo é fornecer *insights* para apoiar a tomada de decisão na escolha do produto financeiro e as necessidades individuais do usuário, bem como sua influência na saúde financeira a curto, médio e longo prazo, incluindo o risco futuro de inadimplência.

### Objetivos Específicos

- Identificar *insights* relevantes a tomada de decisão para crédito Imobiliário, utilizando análise de dados e técnicas de inteligência artificial para extrair insights que possam identificar oportunidades no crédito imobiliário e orientar na tomada de decisões financeiras.
- Identificar situações em que modelos de inteligência artificial, por meio da identificação de padrões, possam ser aplicados para melhorar a saúde financeira do cliente, oferecendo recomendações e estratégias personalizadas para o produto de financiamento imobiliário.
- Utilizar conceitos e ferramentas de computação em nuvem para coleta e processamento de dados, aproveitando a capacidade de processamento de grandes volumes de dados (*Big Data*) para obter análises mais abrangentes e precisas.
- Estabelecer parâmetros de entrada personalizáveis para permitir a realização de análises adaptadas às necessidades e objetivos individuais de cada cliente na análise para o financiamento imobiliário.
- Definir as informações de saída disponibilizadas para o usuário, garantindo que sejam claras, relevantes e úteis para a compreensão de sua situação financeira e a tomada de decisão.
- Conduzir testes para validar a acurácia e qualidade dos modelos de inteligência artificial utilizados, garantindo que os resultados sejam confiáveis e eficazes.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 Inadimplência no Brasil

Embora o termo "inadimplência" tenha uma única definição, que é a falta de cumprimento de uma obrigação (Houaiss, 2001 apud Annibal, 2009), definir operacionalmente o conceito de inadimplência, ou *default*, para avaliação de risco de crédito, não é uma tarefa simples.

Conforme observado por Sicsú (2003 apud ANNIBAL, 2009), é difícil obter um consenso entre analistas de crédito a respeito de uma definição operacional de inadimplência. Uma visão é mais rigorosa com o propósito de obter um sistema de classificação de risco que aprove operações de crédito de forma mais parcimoniosa.

No entanto, por outro lado, existe a visão de outro grupo de analistas de crédito que está mais preocupado com a criação de um sistema que limite os possíveis negócios de uma instituição financeira, e tende a adotar uma definição menos restritiva.

Segundo o Comitê da Basileia sobre Supervisão Bancária (2006, apud Annibal, 2009), ocorre *default* em relação a um devedor específico quando um ou ambos os eventos seguintes ocorrem: o banco considera improvável que o devedor pague suas obrigações ao conglomerado financeiro sem recorrer a ações como a realização de garantias, e o devedor está atrasado em mais de 90 dias em alguma obrigação material com o conglomerado financeiro.

A inadimplência no Brasil é um problema que afeta muitas pessoas. Segundo a pesquisa realizada pela Confederação Nacional do Comércio de Bens, Serviços e Turismo (2023), a inadimplência atingiu 27% dos lares brasileiros em fevereiro de 2022.

Esse aumento da inadimplência é uma preocupação crescente, como evidenciado pelo mais recente levantamento da Serasa, com dados de janeiro de 2023, que apontou um crescimento após uma desaceleração no mês anterior. Com um aumento de mais de 600 mil pessoas, o indicador de inadimplência atingiu 70,09 milhões de brasileiros com o nome restrito (Serasa, 2024).

Os principais motivos que levam o brasileiro à inadimplência são o desemprego, a redução da renda, o descontrole financeiro e a realização de empréstimos em nome de terceiros (Serasa, 2023). Esses fatores têm impacto direto

no perfil etário dos endividados no Brasil, com idades entre 41 e 60 anos representando a maior fatia da população com nome restrito, totalizando 35%. Em seguida, estão as faixas etárias de 26 a 40 anos (34,2%), acima de 60 anos (18,8%) e os jovens entre 18 e 25 anos (12%) (Serasa, 2024).

No trabalho de mestrado de Granemann (2001), foi analisada a inadimplência na área de habitação da Caixa Econômica Federal, no âmbito do Escritório de Negócios Institucional de Florianópolis. O estudo concluiu que a inadimplência é um problema complexo e multifacetado, envolvendo questões como desemprego, falta de educação financeira e endividamento excessivo.

Essas conclusões ressaltam a importância de entender os fatores que contribuem para a inadimplência, especialmente quando se trata do financiamento habitacional, um dos sonhos mais almejados pela população (Silva, 2006).

De fato, segundo Silva (2006), o sonho da casa própria é um desejo muito comum entre as pessoas de todo o mundo. Ele está enraizado na necessidade de segurança, estabilidade e liberdade. Uma casa própria nos dá um lugar para chamar de nosso, onde pode-se criar uma vida para toda família. Também dá um senso de propriedade e orgulho, e ajuda a nos sentirmos mais seguros e estáveis.

O sonho da casa própria é um sonho que muitas pessoas nunca conseguem realizar. No entanto, isso não significa que se deva desistir dele. É importante continuar trabalhando duro, economizando dinheiro, para que um dia se possa realizar esse sonho (Silva, 2006).

Portanto, ao compreender as complexidades da inadimplência no contexto do crédito imobiliário e seus impactos socioeconômicos, é possível traçar estratégias para lidar com esse desafio.

A análise dos fatores que contribuem para a inadimplência, como desemprego, redução da renda e descontrole financeiro, oferece insights valiosos para o desenvolvimento de dispositivos de apoio a tomada de decisão para promover uma gestão mais responsável das finanças pessoais.

Em última análise, ao enfrentar a inadimplência de forma proativa e holística, é possível mitigar seus efeitos negativos e promover uma maior estabilidade econômica e social para a população.

### 3.2 Big Data

Schonberger e Cukier (2014) citaram em seus livros “*Big Data: A Revolution That Will Transform How We Live, Work, and Think*” (2014) que o termo *Big Data* marcava o início de uma grande transformação e que a verdadeira revolução não estava nas máquinas e sim nos dados e como são utilizadas. Os frutos da sociedade da informação são visíveis seja ele no celular de cada indivíduo, computador em cada mochila e nos grandes sistemas de tecnologias da informação e apenas meio século após a inclusão dos computadores, os dados começaram a se acumular até o ponto em que algo novo e especial aconteceu.

Não existe uma definição rigorosa de *Big Data*. Inicialmente, a ideia era que o volume de informações havia crescido tanto que a quantidade examinada não cabia mais na memória que os computadores usam para processamento, então os engenheiros precisavam reformular as ferramentas que usavam para analisar (Schonberger; Cukier, 2014).

Assim, o conceito de *Big Data* seguiu algumas derivativas como sugere em alguns livros em que dividem o conceito em três, quatro ou até mesmo 5 dimensões. Dentre as mais citadas está a proposta pela Meta Group (2001) que deriva o termo Big Data em três dimensões Volume, Velocidade e Variedade, sendo caracterizados respectivamente como o alto volume de dados armazenados, a velocidade em que são gerados e variedades em que são dispostos (Laney, 2001 apud Garcia, 2021, p. 5).

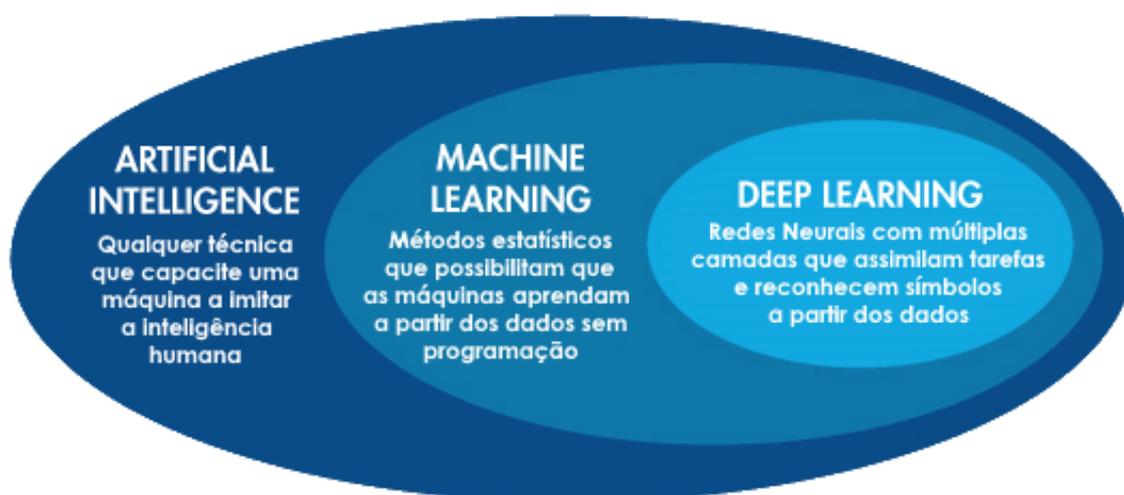
Segundo os autores McAfee e Brynjolfsson (2012 Apud Garcia, 2021, p. 5) empresas que incorporaram e maximizam o potencial da tomada de decisão tendo por base informações geradas por dados (empresas *data-driven*) obtiveram melhores resultados financeiros e operacionais.

Elia et al. (2020), apud Garcia (2021, p. 6) detalha Big Data com sendo “Big Data é o ativo de informação caracterizado por um volume, velocidade e variedade tão elevados que requer tecnologias específicas e métodos analíticos para sua transformação em valor.”

### 3.3 Inteligência Artificial

Para Brock e Von Wangenheim (2019 apud Garcia, 2021) o primeiro fator do modelo composto de sucesso para implantação da inteligência artificial consiste a respeito da informação, visto que, para que a IA resulte em informações de qualidade é requisitado uma grande quantidade de dados (*Big Data*) com alto nível de qualidade. Como exposto na figura 1 é possível compreender os diversos ramos antes de iniciar a abordagem de IA.

Figura 1 - Especificações de Inteligência Artificial, *Machine Learning* e *Deep Learning*



Fonte: Opencadd, 2021.

A inteligência artificial teve seu início em um campo de estudo a partir da década de 1950, com a intenção de simular a capacidade humana de raciocinar, perceber, aprender, resolver questões e tomar decisões (Mccarthy, 1990).

Segundo Kai-Fuu Lee (2019) a revolução completa da inteligência artificial ainda demanda um pouco mais de tempo para se concretizar e será dividida em quatro ondas: IA de Internet, Negócio, Percepção e Autônoma. Cada uma delas atuando de forma independente e de maneiras distintas nos diversos e mais profundos setores da sociedade.

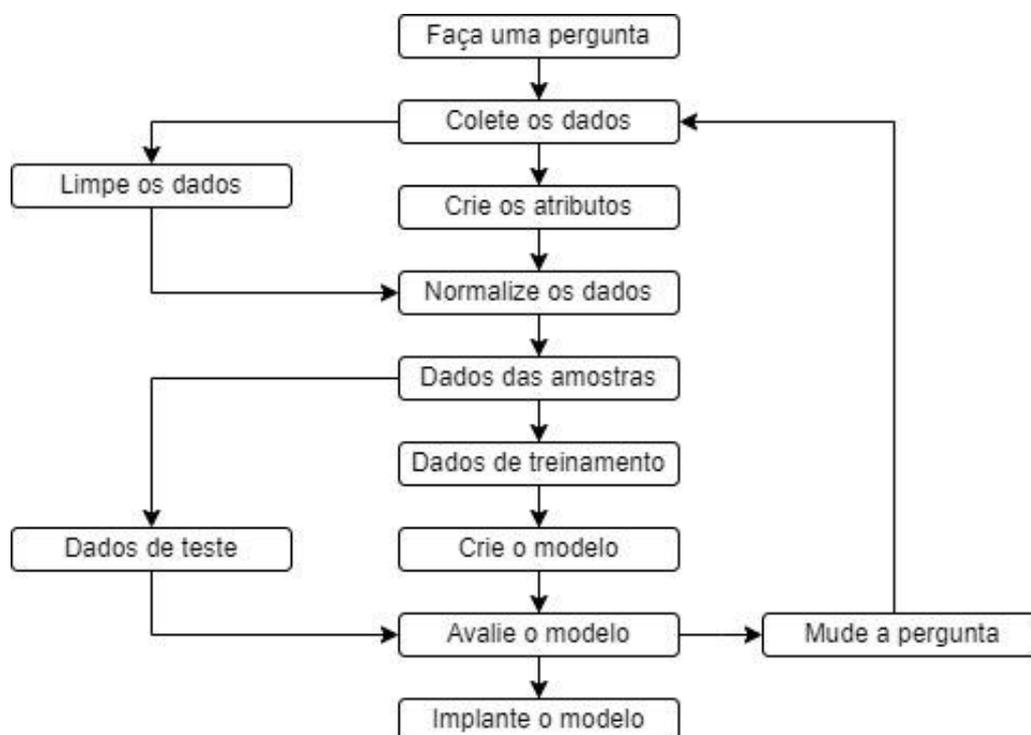
Dentre as ondas, IA de Internet e Negócio já estão em nossa sociedade modificando conceitos no mundo digital e financeiro, seguindo da onda da percepção que está se iniciando digitalizando o mundo físico e aprendendo e entendendo o mundo ao nosso redor e, com isso, pretendendo remodelar a maneira de viver e interagir no cotidiano; por último a onda Autônoma que atingirá profundamente a

sociedade à medida que carros, drones, fábricas e robôs transformarão a sociedade (Lee, 2019).

Os modelos de *machine learning* seguem um fluxo de trabalho comum conforme figura 2 e podem ser divididos em dois grupos: aprendizado supervisionado e não-supervisionado (Lecun et al, 2015 apud Garcia 2021).

De acordo com Paixão et al. (2022), a principal distinção entre os modelos de aprendizagem supervisionada e não supervisionada reside na base do algoritmo de treinamento.

**Figura 2 - Fluxo de trabalho comum no Machine Learning**



Fonte: O`Reilly *Machine Learning: Guia de Referência Rápida*, 2020.

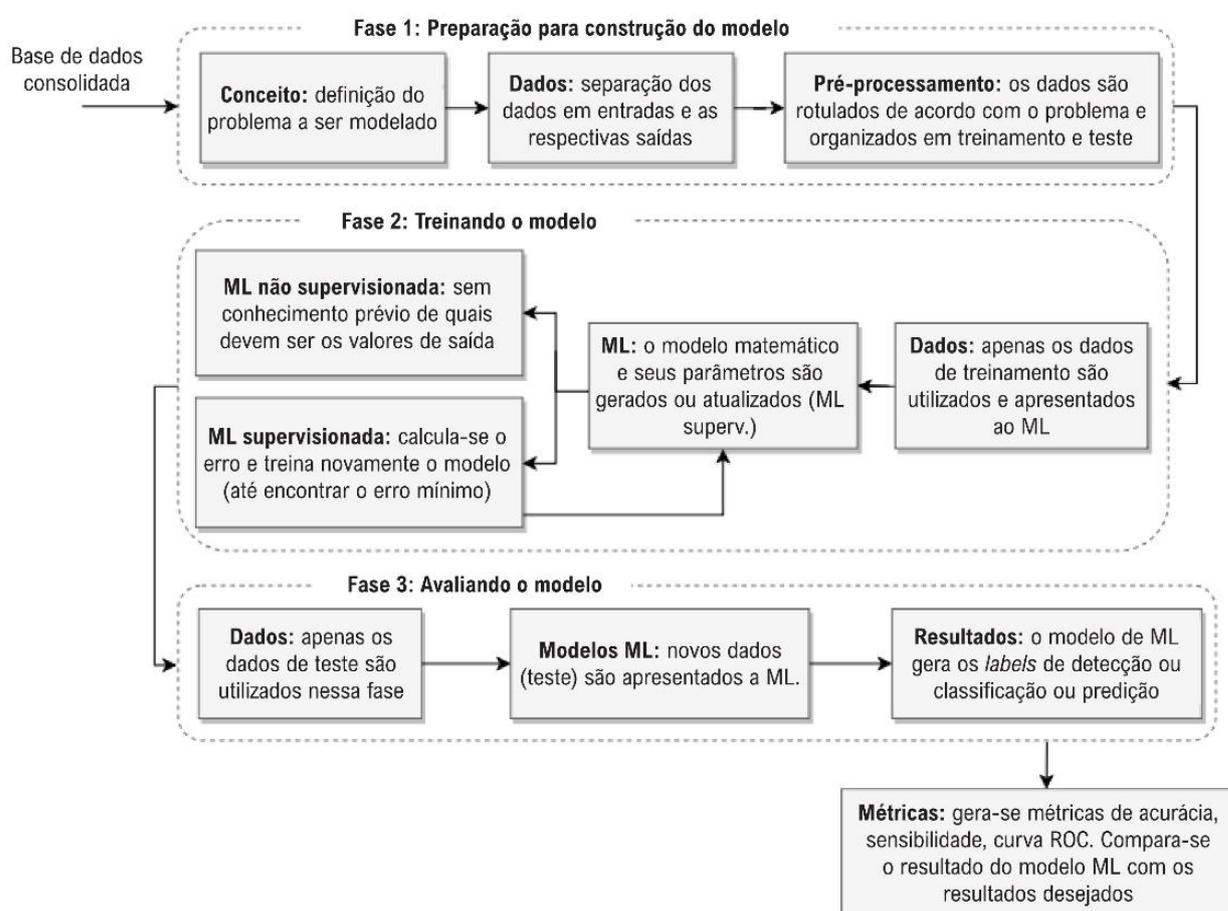
Nos modelos de aprendizagem não supervisionada, o algoritmo extrai características dos dados e constrói uma representação sem conhecimento prévio dos rótulos de cada dado. Isso implica que o algoritmo identifica padrões nas informações de classe de maneira heurística, sem a orientação de rótulos pré-determinados. Essa abordagem proporciona uma vantagem significativa, permitindo que o algoritmo analise padrões que possam não ter sido considerados anteriormente (Paixão et al., 2022).

Por outro lado, nos modelos de aprendizagem supervisionada, o algoritmo possui conhecimento dos rótulos dos dados, ou seja, as amostras estão corretamente

classificadas. Durante o treinamento, o modelo é comparado com os rótulos previamente atribuídos, e o processo é repetido até que um erro mínimo seja alcançado (Paixão et al., 2022).

Essa diferença fundamental no treinamento dos modelos influencia diretamente como eles lidam com os dados e, conseqüentemente, seus resultados, por estes motivos é preciso o estudo do modelo que melhor responde à pergunta, conforme representado na figura 3.

**Figura 3 - Fases para o desenvolvimento de algoritmos de machine learning**



Fonte: PAIXAO et al., 2022.

O *deep learning* é uma subárea do aprendizado de máquina que utiliza redes neurais artificiais para efetuar aprendizado por meio de dados. Inspiradas pelo funcionamento do cérebro humano, as redes neurais artificiais têm a capacidade de identificar padrões complexos nos conjuntos de dados. O *deep learning* tem demonstrado êxito em uma ampla gama de aplicações, abrangendo desde o

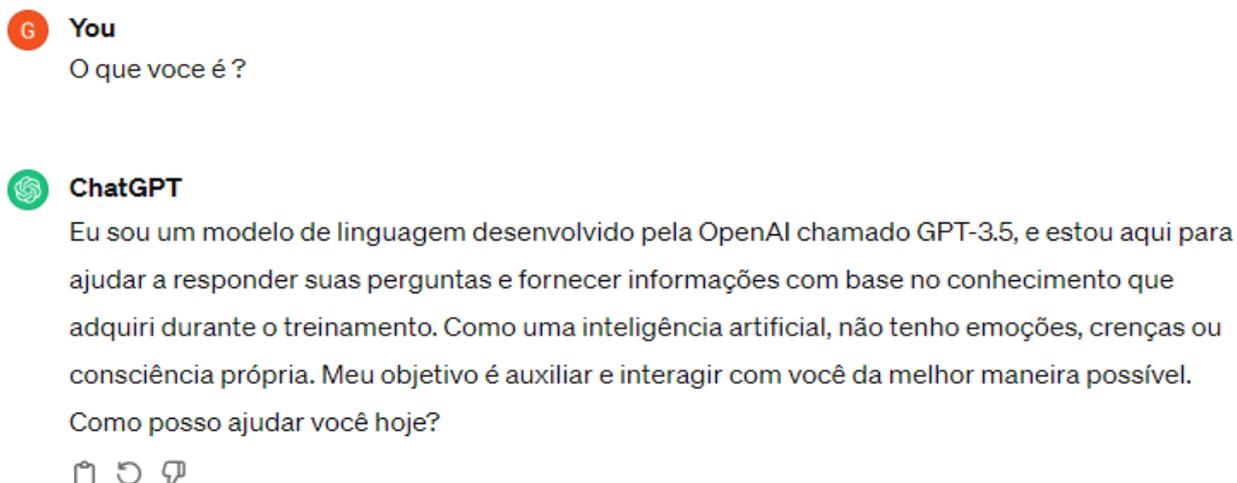
reconhecimento de imagens e fala até a tradução automática e o diagnóstico médico (Goodfellow; Bengio; Courville, 2016).

### 3.4 Modelo de Inteligência Artificial – ChatGPT

O *Chat Generative Pre-Trained Transformer*, também conhecido como ChatGPT foi desenvolvido pela empresa norte-americana OpenAI e apresentado ao público em 2022 (OpenAI, 2024).

Por meio da ferramenta ChatGPT é possível realizar interações realistas e fluentes, traduções linguísticas automáticas, resumo de textos, algoritmos matemáticos e criar conteúdo. Este processo acontece por meio de uma interface de mensageria, ou simplificando como uma conversa entre o usuário e um sistema de inteligência artificial, conforme questionado na figura 4 é possível observar uma interação simples.

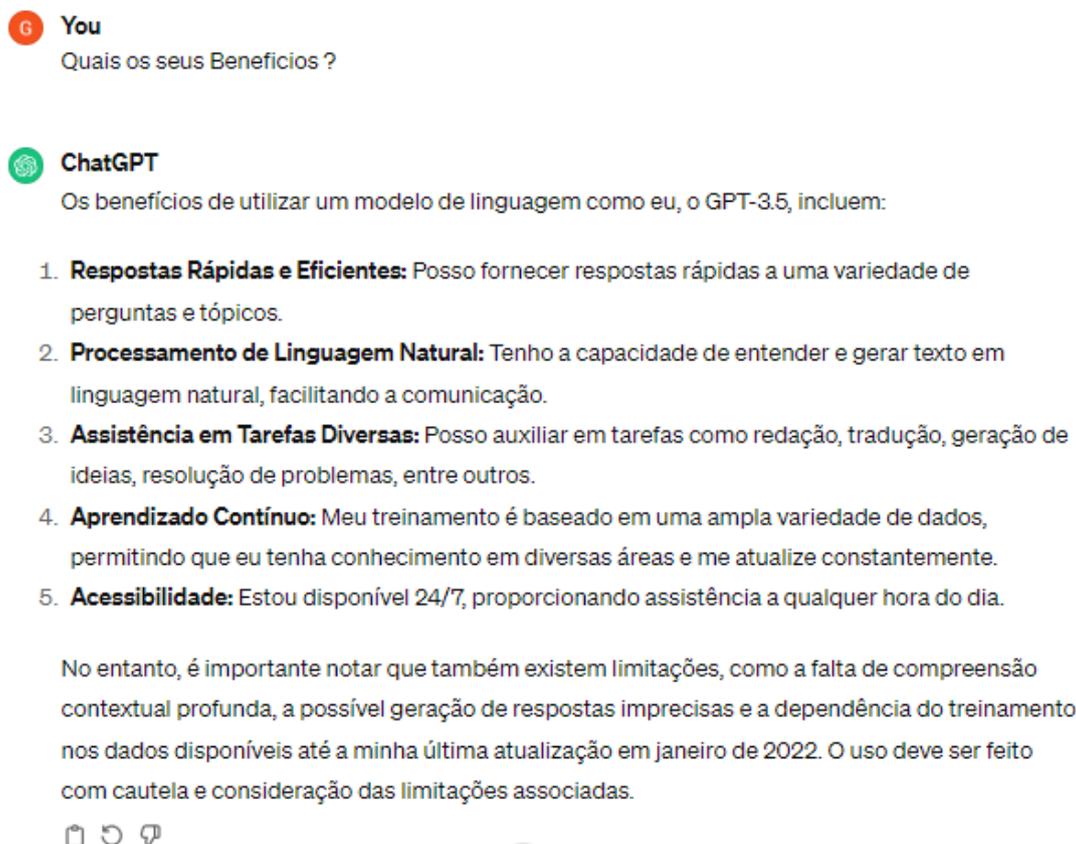
Figura 4 - Interação com o ChatGPT



Fonte: Elaboração Própria.

Quando questionado conforme figura 5 sobre seus benefícios o ChatGPT é capaz de citar alguns, mas também expor a possibilidade de não interpretar corretamente a pergunta resultando em uma possível resposta errônea.

Figura 5 - Benefícios do ChatGPT



Fonte: Elaboração Própria.

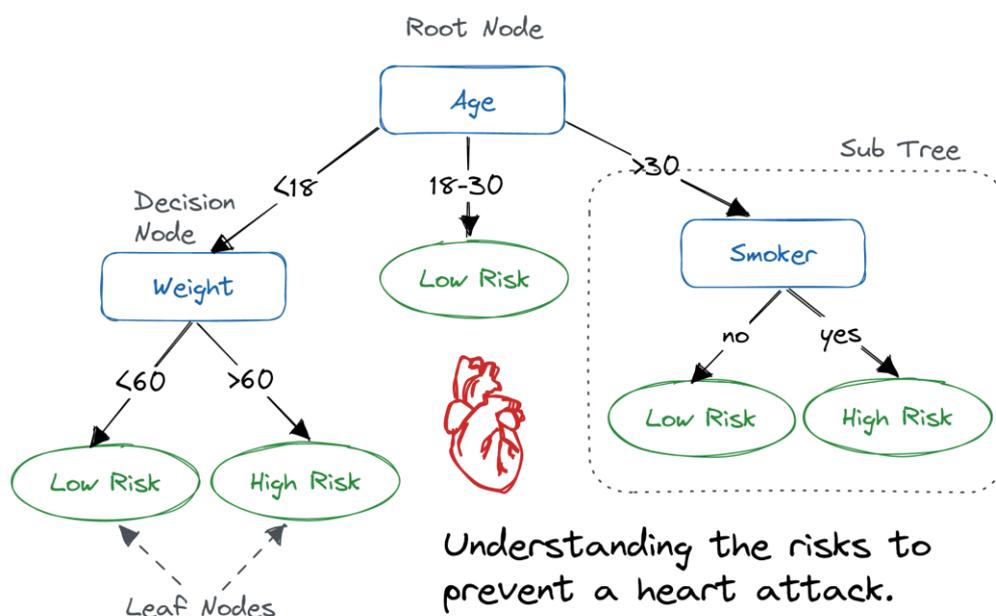
### 3.5 Modelo de Inteligência Artificial de Classificação – Árvore de Decisão

O algoritmo de aprendizado de máquina conhecido como Árvore de Decisão (*Decision Tree*) é amplamente empregado na etapa de classificação (ROKACH, 201, apud Fagundes, 2022). Uma definição proposta por Breiman et. Al (1984, apud Fagundes 2022) define que a construção de uma árvore de decisão segue a seguinte estrutura ilustrada na figura 6, sendo utilizado como exemplo uma análise do nível de risco de uma pessoa possuir problemas do coração.

Se a condição de terminação é atingida, isto é, o conjunto de treinamento é homogêneo então a árvore é definida com apenas um nó, como exemplo na figura 6 onde a idade entre 18 e 30 anos é definida pela árvore de decisão como baixo risco. Caso contrário, utilizando o conjunto de atributos de treinamento definido em azul na figura 6 (Idade, Peso e Fumante), um atributo é escolhido e a partir dele é gerado outro subconjunto (*subtree*) sucessivamente. Uma vez obtido um subconjunto

uniforme em que todos os atributos pertencem a uma mesma classe (*leaf nodes*), um nó é criado finalizando a análise na árvore de decisão.

Figura 6 - Exemplo de uma Árvore de Decisão



Fonte: DataCamp, 2024.

A definição de sucesso para uma árvore de decisão (Ferreira, 2022) depende do critério utilizado para a escolha do atributo que define o subconjunto em cada interação, buscando a menor árvore. Para Ferreira (2022) deve-se maximizar a separação de classes em cada etapa, deixando os sucessores de cada nó mais puros. Para isso métricas de impurezas, ou também chamadas de entropia (*entropy*), são definidas.

Segundo Breiman et. Al (1984, apud Fagundes 2022) ao construir uma árvore são escolhidos os nós com menores valores de entropia, por meio desta métrica o algoritmo decide qual nó deve dividir e qual é o ponto de divisão do recurso.

O uso da métrica de entropia ou impureza é comumente utilizado na teoria da informação para medir a pureza ou impureza de um determinado conjunto, com o objetivo de reduzir as impurezas variando entre 0 e 1, onde quanto mais próximo de um maior a impureza (Fagundes, 2022).

Uma métrica é definida como medida de impureza  $i(T)$  se satisfaz as seguintes propriedades (Ferreira, 2022).

- $i(T)$  é mínima somente quando  $p_i = 1$  e  $p_j = 0$  para  $j \neq i$  (é mínima quando todos os exemplos são da mesma classe)

- $i(T)$  é máxima somente quando  $p_j = \frac{1}{k}$  (é máxima quando há exatamente o mesmo número de exemplos de cada classe).
- $i(T)$  é simétrica em relação a  $p_1, p_2, \dots, p_k$  (não importa a ordem das classes, desde que as proporções sejam as mesmas)

Quando  $i(T)$  atinge seu valor mínimo,  $T$  é denominado puro ou homogêneo ou sem mistura de classe. Para qualquer valor acima do valor mínimo  $T$  é denominado impuro ou heterogêneo.

A biblioteca de programação baseada em Python denominada *scikit-learn* (2024) define como padrão a métrica de impureza denominada Índice GINI (1936), desenvolvida pelo estatístico italiano Conrado Gini no ano de 1912, o Índice GINI é definido pela equação 1:

$$i(T) = -\sum_j p_j(1 - p_j) p_j = 1 - \sum_j p_j^2 \quad (1)$$

Seja  $T$  um conjunto de exemplos e  $p_j$  as proporções de exemplos de cada classe  $C$  e  $T$  ( $j = 1, 2, \dots, k$ ) representado como  $T = [p_1, p_2, \dots, p_k]$  onde  $\sum_{j=1}^k p_j = \sum p_j = 1$ .

Onde  $i(T)$  significa o índice Gini de um conjunto  $T$ ,  $\sum_j$  representa a somatória de cada classe ou categoria dentro do conjunto  $T$ ,  $p_j$  representa a probabilidade da classe  $j$  no conjunto  $T$ , e por fim,  $(1 - p_j)$  é a proporção dos elementos que não pertencem à classe  $j$ .

Para  $k$  classes, o valor mínimo é zero e o máximo é  $\frac{k-1}{k}$ . Como pode ser observado, independentemente do número  $k$  de classes, o valor máximo é sempre inferior à unidade, tendendo à 1 quando  $k \rightarrow +\infty$ , resultando em um conjunto impuro ou heterogêneo.

Segundo Gama (2004, apud Fagundes 2022) uma árvore de decisão, contempla os seguintes pontos:

- Cada nó de decisão representa um teste para o atributo
- Cada ramo é distinto e desce a um possível resultado do atributo
- Cada folha está associada a uma classe (resultado)
- O Conjunto de raiz até a folha é o resultado de uma regra de classificação

Para avaliar o desempenho preditivo dos modelos de classificação, utiliza-se uma série de métricas que fornecem percepções sobre o quão bem o modelo está realizando suas previsões. Entre essas métricas, uma das mais amplamente utilizadas é a área sob a curva da característica operacional do receptor (ROC AUC).

Desenvolvida inicialmente por Gustav Theodor Fechner no século XIX, a curva ROC AUC foi introduzida no contexto de aprendizado de máquina por Spackman (Neto, 2020). Ao calcular a ROC AUC, são considerados todos os possíveis pares de classes no problema de classificação, e os valores ROC AUC para cada par de classes são agregados para fornecer uma métrica global do desempenho do modelo em múltiplas classes (Hand et al., 2022, apud Ferreira, 2022).

No entanto, embora a ROC AUC seja uma métrica relevante, é fundamental complementá-la com outras técnicas e métricas de avaliação, como a acurácia, que mede a proporção de previsões corretas em relação ao total de casos. Além disso, é recomendável utilizar a curva ROC em conjunto com outras métricas, como a precisão (precision), que avalia a proporção de verdadeiros positivos entre as previsões positivas, e o F1-score, que harmoniza precisão e recall, oferecendo uma visão mais equilibrada da performance do modelo. (Hand et al., 2022 apud Ferreira, 2022).

Por fim, conclui-se que a complexidade de uma árvore de decisão é geralmente avaliada através de métodos como o índice de Gini ou o índice de entropia, conforme abordado anteriormente neste capítulo. A performance do modelo, por sua vez, pode ser melhor avaliada utilizando uma combinação de métricas, incluindo a ROC AUC, precisão, F1-score e acurácia, para obter uma análise mais completa e robusta dos resultados.

### **3.6 Computação em Nuvem**

Computação em nuvem, também conhecida por *Cloud Computing*, pode ser definida segundo a publicação SP 800-145 (NIST, 2011) como um modelo que permite o acesso universal de modo proveitoso e sob demanda, através da rede de internet a um conjunto compartilhado de recursos computacionais configurável, que pode ser rapidamente provisionado, escalado e disponibilizado com o mínimo de esforço de gerenciamento ou de interações com o provedor de serviço.

A provedora de serviço em nuvem AWS distribui este serviço com definições de preço de acordo com a demanda do usuário, proporcionando a dispensabilidade de datacenter e servidores físicos por parte do usuário (Amazon Web Service, 2023).

O NIST define em sua publicação algumas características essenciais para *cloud computing*, sendo elas (Mell; Grace, 2011):

- Disponibilização de recursos e serviços sob procura: o usuário deve ser capaz de provisionar e escalar unilateralmente e de forma automática, sem depender de um recurso humano.

- *Pooling* de recursos: A *cloud* oferece um *pool* de recursos computacionais dinâmicos que pode atender múltiplas necessidades e usuários, de acordo com sua demanda e localização.

- Ubiquidade: Os recursos disponibilizados pelo provedor devem ser alcançados de forma ágil e fácil de qualquer lugar onde possua acesso a rede de internet.

- Escalabilidade: O provedor deve disponibilizar e manter o desempenho, performance e resiliência à medida que a carga sobre ela aumenta.

- Elasticidade: Os recursos computacionais devem ter flexibilidade, isto é, aumentados ou diminuídos de forma automática de acordo com a carga a que estes estão sujeitos, ou de acordo com as exigências do usuário.

- Monitorização e controle: A *cloud* deve permitir que os utilizadores tenham acesso a um sistema de controle que permita monitorar e gerir os recursos e serviços computacionais que estão ao seu dispor.

- Pagamento pelo que se utiliza: A *cloud* aplica um modelo de pagamento em que os usuários são cobrados pelos recursos que utilizam, isto é, sob demanda.

Segundo a AWS existem 3 modelos de produtos oferecidos pelas provedoras de serviço de computação em nuvem.

1. O PaaS (*Platform as a Service*), onde não precisa mais gerenciar a infraestrutura subjacente e pode manter o foco na implantação e no gerenciamento de aplicativos.

2. O SaaS (*Software as a Service*) oferece um produto completo, executado e gerenciado pelo provedor de serviços.

3. O IaaS (*Infrastructure as a Service*) oferece acesso a recursos de rede, computadores (virtuais ou em *hardware* dedicado) e espaço de armazenamento de dados.

### 3.7 Serverless

Segundo Shafiei, Khonsari e Mousavi (2022), os principais benefícios da arquitetura *serverless* estão relacionados ao desenvolvimento de aplicações sem a necessidade de trabalhar infraestrutura e escalabilidade, movendo a responsabilidade por alocar recursos para o provedor do serviço em nuvem. Outro benefício citado pelos autores é a redução de custo e a influência da cobrança por processamento, isto é, só é cobrado o que for utilizado, *On-Demand* (Adzic; Chatley, 2017).

As principais características de um serviço *serverless* são (Shafiei; Khonsari; Mousavi, 2022):

- Omissão das informações sobre o ambiente de execução para o cliente, como máquina virtual, contêiner, sistema operacional, entre outros, tornando-as transparentes para o cliente;
- Escalabilidade automática como responsabilidade do provedor, onde recursos devem estar disponíveis de acordo com a demanda;
- Somente os recursos utilizados pelo cliente devem ser cobrados pelo provedor;
- Capacidade do provedor de executar a requisição assim que ela for recebida, ou o quanto antes possível, com tempo de execução máximo limitado;
- O provedor deve estar ciente das definições e dependências das funções adicionadas pelo cliente, como também quaisquer informações relacionadas ao estado e ambiente de execução.

### 3.8 Banco de Dados

Um banco de dados é uma coleção organizada de informações e/ou dados, normalmente armazenadas eletronicamente em um sistema de computador. Além disso, os bancos de dados são projetados para lidar com grandes volumes de dados de forma eficiente e oferecer acesso rápido e confiável às informações armazenadas.

Um banco de dados é geralmente controlado por um sistema de gerenciamento de banco de dados (SGBD), que é responsável por gerenciar e organizar os dados, bem como facilitar o acesso a eles por meio de consultas e operações de manipulação (Oracle, 2023).

Juntos, os dados e o SGBD, juntamente com os aplicativos associados a eles, são chamados de sistema de banco de dados, geralmente abreviados para apenas banco de dados. Este sistema desempenha um papel fundamental em uma ampla gama de aplicações, desde sistemas bancários e sistemas de gerenciamento de estoque até sistemas de gerenciamento de conteúdo e sistemas de suporte à decisão (Oracle, 2023).

Existem muitos tipos diferentes de bancos de dados para diferentes necessidades de armazenamento e consulta dos dados (Oracle, 2023). Quando se fala em dados, estes podem ser divididos em três categorias: estruturados, semi-estruturados e não estruturados, conforme detalhado na figura 7.

Figura 7 - Tipo dos Dados



Fonte: Monteiro, 2019

### 3.8.1 Banco de Dados Relacional

Segundo Chamberlin (2012) a linguagem SQL, também conhecida como *Structured Query Language* ou “sequel”, foi desenvolvida pelo pesquisador da IBM Ted Codd em 1973 por meio do projeto R. O SQL tem como objetivo permitir que os usuários criem, acessem e manipulem bases de dados relacionais. O SQL foi projetado para permitir ao usuário criar e acessar comandos para definir, recuperar e manipular dados armazenados em uma base de dados relacional.

Atualmente a linguagem SQL é a linguagem padrão para acesso a banco de dados relacionais e é amplamente utilizada em todo o mundo. Os sistemas de banco de dados relacionais são os sistemas de banco de dados mais comuns, sendo usados por empresas de todos os tamanhos.

Os bancos de dados relacionais têm como propriedade o modelo ACID que garante que as transações são processadas com confiança (SILVA, 2011).

O acrônimo ACID significa:

1. Atomicidade: Garante que uma transação é executada completamente ou não é executada de maneira alguma. Em caso de falha, não ocorrerá nenhuma alteração.
2. Consistência: Assegura que o banco de dados permanece em um estado válido antes e depois de cada transação. Ou seja, todas as regras de integridade e restrições definidas são respeitadas durante a execução da transação.
3. Isolamento: Garante que o resultado de uma transação seja isolado das outras transações concorrentes. Cada transação deve ser tratada como se estivesse sendo executada individualmente, independentemente de outras transações que estejam ocorrendo simultaneamente.
4. Durabilidade: Garante que as alterações realizadas por uma transação sejam permanentes e persistam mesmo após uma falha do sistema, incluindo falhas de *hardware* ou *software*.

### 3.8.2 Banco de Dados Não Relacional

Para enfrentar desafios relacionados à performance e escalabilidade dos bancos de dados relacionais, surgiu a necessidade de explorar alternativas encontradas nos chamados bancos de dados não relacionais.

Esses bancos de dados são capazes de gerenciar dados semiestruturados oriundos de diversas origens e possibilitam escalabilidade mais barata e menos complexa, não necessitando de servidores muito robustos e nem muitos profissionais para administrá-los (Oliveira, 2014).

Segundo Sadalage e Fowler (2015) um banco de dados não relacional, também conhecido como NOSQL, são classificados como banco de dados chave-valor, banco de dados de documentos, armazenamento em família de colunas e banco de dados de grafos.

A seguir são descritos alguns serviços de *cloud* para bancos de dados não relacionais e seus benefícios e pontos a melhorar.

**Tabela 1 - Comparação de Bancos de Dados NoSQL**

Banco de Dados	Provedor	Benefícios	Contras
DynamoDB	AWS	Alta performance Escalabilidade Totalmente gerenciado Estritamente consistente	Custo Limitações de consistência
Google Cloud Bigtable	GCP	Alta performance Escalabilidade Preços acessíveis	Não é totalmente gerenciado Consistência
Cosmos DB	Azure Microsoft	Alta performance Escalabilidade Totalmente gerenciado	Custo Limitações de consistência

Fonte: Elaboração Própria

### 3.9 Application Programming Interface

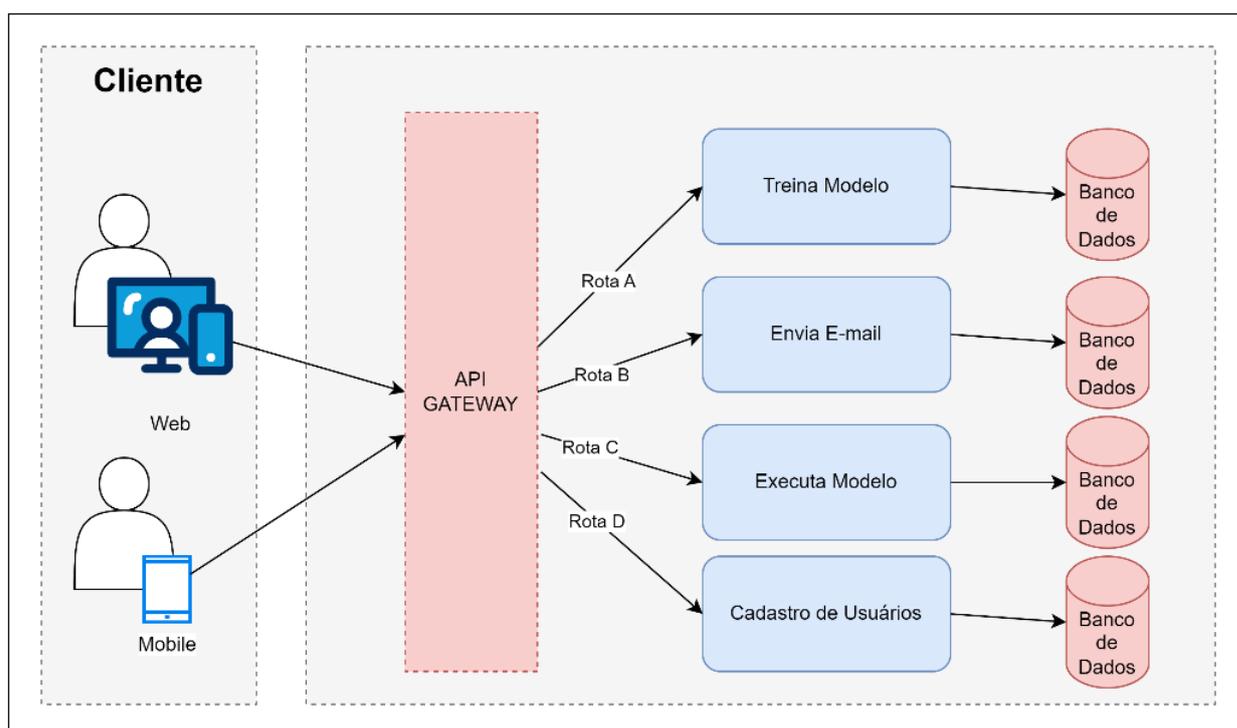
De acordo com o artigo escrito por Netto, Martins e Orlandini (2018), API é uma sigla para *Application Programming Interface*, que em português significa Interface de Programação de Aplicativos. Uma API é definida por um conjunto de rotinas e padrões de programação que permitem o acesso a um aplicativo de software ou a uma plataforma baseada na *Web*. Através da utilização de APIs, é possível obter dados e interagir com serviços oferecidos por outras empresas ou plataformas, facilitando a integração e comunicação entre diferentes sistemas (Oracle, 2023).

Integrar rotas em uma API é como definir o caminho que os usuários seguirão para acessar diferentes recursos ou funcionalidades do aplicativo. Cada rota é como uma entrada para uma parte específica do sistema, conforme ilustrado na figura 8, e

é mapeada para um ou mais métodos HTTP, como GET, POST, PUT ou DELETE. Esses métodos determinam a ação que será executada no serviço de *back-end* quando a rota for acessada (Oracle, 2023).

A função principal das APIs é permitir que desenvolvedores possam construir, interagir e integrar sistemas de maneira eficiente e padronizada. Elas são fundamentais na construção de arquiteturas de software modernas, especialmente na era da computação em nuvem e dos serviços baseados na *Web*. As APIs proporcionam um meio seguro e controlado para expor funcionalidades internas de um sistema, permitindo que outros sistemas ou desenvolvedores utilizem esses recursos sem a necessidade de conhecer ou acessar diretamente o código fonte subjacente (Oracle, 2023).

Figura 8 - Api e suas rotas



Fonte: Elaboração Própria

### 3.10 Linguagem de Programação Python

A linguagem de programação de alto nível chamada por Python (Rossum, 1991), desenvolvida por Guido Van Rossum em 1989, é conhecida por sua simplicidade, legibilidade do código e sua ampla gama de aplicações, desde

aplicações *web* até aplicações de inteligência artificial (Da Silva et al. 2020), além de ser uma linguagem de código aberto.

A linguagem de programação Python é muito interessante como primeira linguagem de programação devido à sua simplicidade e clareza. Embora simples, é também uma linguagem poderosa, podendo ser usada para administrar sistemas e desenvolver grandes projetos. É uma linguagem clara e objetiva, pois vai direto ao ponto, sem rodeios (Menezes, 2010, p.21).

Ao analisar os dados apresentados na figura 9, é evidente que o Python ocupa uma posição de destaque entre as linguagens de programação mais utilizadas em 2023. Com uma participação de 14,82%, o Python lidera o ranking, superando todas as outras linguagens listadas. Essa predominância do Python é significativa.

A linguagem de programação Python tem ganhado popularidade significativa nos últimos anos devido à sua sintaxe simples e legível, o que o torna uma escolha atrativa para desenvolvedores de todos os níveis de experiência.

Além disso, a vasta gama de bibliotecas e *frameworks open source* e uma comunidade altamente ativa torna o Python uma ferramenta versátil para uma variedade de aplicações, desde desenvolvimento *web* até análise de dados e inteligência artificial (PYTHON, 2024).

Portanto, a liderança do Python neste gráfico não apenas reflete sua popularidade entre os desenvolvedores, mas também destaca sua relevância e importância no cenário atual da programação de software.

**Figura 9 - Uso das linguagens de programação em 2023**

Oct 2023	Oct 2022	Change	Programming Language	Ratings
1	1		 Python	14.82%
2	2		 C	12.08%
3	4	▲	 C++	10.67%
4	3	▼	 Java	8.92%
5	5		 C#	7.71%
6	7	▲	 JavaScript	2.91%
7	6	▼	 Visual Basic	2.13%
8	9	▲	 PHP	1.90%
9	10	▲	 SQL	1.78%
10	8	▼	 Assembly language	1.64%

Fonte: Tiobe, 2023.

## 4 ETAPAS DE DESENVOLVIMENTO

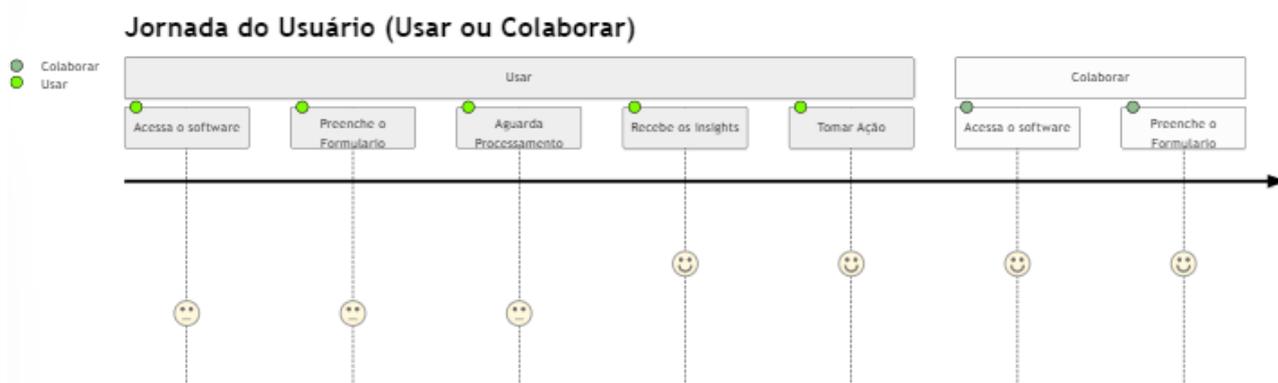
Neste capítulo, serão abordados as etapas e os produtos de tecnologia em nuvem utilizados no desenvolvimento do software. O conteúdo será estruturado de forma sequencial, seguindo a ordem de execução do processo dentro da aplicação. Dessa maneira, permitirá uma visão clara e organizada das diferentes fases e dos respectivos componentes tecnológicos empregados para viabilizar o funcionamento do *software* na nuvem.

Com essa abordagem estruturada, será possível compreender melhor a lógica e o fluxo do sistema, bem como a importância dos produtos e serviços específicos de nuvem em cada etapa do processo. Ao destacar a relação entre as etapas e os produtos utilizados, pretende-se fornecer um guia abrangente e informativo para aqueles que desejam explorar e compreender a implementação dessa aplicação baseada em nuvem.

### 4.1 Jornada de uso

Na figura 10 é definido a Jornada do Usuário para os dois tipos de usuário do sistema, o usuário com objetivo de obter apoio para tomada de decisão do crédito imobiliário e o usuário colaborativo, que tem como objetivo colaborar com enriquecimento da base de dados que será utilizada para retrainar os modelos preditivo usado pelo software.

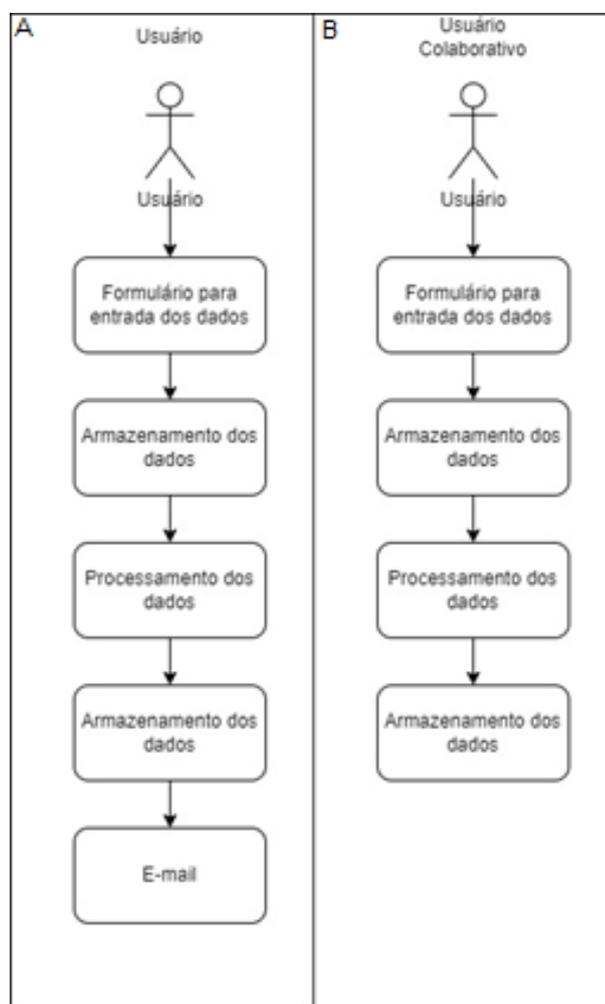
Figura 10 - Diagrama de Uso



Fonte: Elaboração própria

Dividindo o fluxo do usuário em duas personas, é representado na figura 11 o usuário com objetivo de obter apoio na tomada de decisão.

**Figura 11 - Fluxograma do Usuário**



Fonte: Elaboração própria

I- Usuário em busca de apoio na tomada de decisão, representado pela letra A na figura 11:

1. O usuário inicia o processo preenchendo um formulário de entrada com as informações relevantes.
2. Em seguida, os dados fornecidos são armazenados no sistema para posterior análise.
3. Os dados são então processados pelo software, onde são aplicados algoritmos de análise e modelagem.
4. Após o processamento, os resultados são armazenados novamente para referência futura.

5. Por fim, o usuário recebe um *e-mail* contendo os *insights* gerados pelo software, fornecendo apoio na tomada de decisão para o crédito imobiliário.

II- Usuário colaborativo em busca de enriquecimento da base de dados, representado pela letra B na figura 11:

1. O usuário preenche um formulário com os dados que deseja contribuir para a base existente.
2. Esses dados são armazenados no sistema, aumentando a quantidade de informações disponíveis.
3. Em seguida, os dados são processados pelo software para realizar o retreinamento do modelo preditivo.
4. Esse processo de retreinamento do modelo visa aprimorar a precisão e eficácia das previsões futuras com base nos novos dados fornecidos pelo usuário.

Esses dois caminhos representam diferentes formas de interação com o software, cada uma atendendo às necessidades específicas dos usuários envolvidos.

## 4.2 Levantamento de Requisitos

### 4.2.1 Requisitos Funcionais

Os requisitos funcionais descrevem as funcionalidades específicas que o sistema de inteligência artificial para análise de financiamento imobiliário busca fornecer.

Eles são as capacidades e comportamentos esperados do software. A seguir, são apresentados alguns dos principais requisitos funcionais identificados:

Cadastro de usuários:

- O sistema deve permitir o cadastro de usuários, incluindo informações pessoais e financeiras relevantes. Isso engloba a possibilidade de inserir detalhes como histórico de gastos, renda líquida, dívidas, bem como outras informações financeiras do usuário.
- O sistema deve ser capaz de receber e armazenar informações detalhadas sobre os produtos financeiros associados aos usuários,

incluindo, mas não se limitando a taxas de juros, prazos e outros dados pertinentes.

- Validar e armazenar os dados de forma segura.
- Gerenciar perfis de usuários.

Coleta e armazenamento de dados financeiros:

- Coletar dados financeiros relevantes dos clientes, como renda, despesas, patrimônio e histórico de crédito.
- Armazenar e atualizar os dados de forma segura e confidencial.

Análise de capacidade de pagamento:

- Avaliar a capacidade de pagamento do cliente com base nos dados fornecidos.
- Realizar cálculos financeiros para determinar a viabilidade do financiamento imobiliário.
- Considerar fatores como taxa de juros, prazo do empréstimo e valor do imóvel.

Geração de relatórios e *insights* financeiros:

- Gerar relatórios personalizados com informações relevantes sobre o financiamento imobiliário.
- Fornecer *insights* e recomendações com base nos dados e análises realizadas.
- Apresentar os resultados de forma clara e compreensível aos usuários.
- Considerar fatores como taxa de juros, prazo do empréstimo e valor do imóvel.

#### 4.2.2 Requisitos não funcionais

Os requisitos não funcionais buscam às características não diretamente relacionadas às funcionalidades específicas do sistema, mas que são igualmente importantes para sua eficácia, desempenho e usabilidade. A seguir, são apresentados alguns requisitos não funcionais relevantes:

Segurança e privacidade dos dados:

- Garantir a segurança dos dados financeiros e pessoais dos usuários.

- Implementar mecanismos de criptografia, autenticação e controle de acesso.
- Estar em conformidade com as regulamentações de privacidade de dados, como a Lei Geral de Proteção de Dados (LGPD).
- Análise de capacidade de pagamento

Performance e escalabilidade:

- Garantir que o sistema tenha um desempenho adequado, com tempos de resposta rápidos.
- Lidar com uma grande quantidade de dados e usuários simultaneamente.
- Ser capaz de escalar e lidar com um aumento na demanda sem comprometer a performance.

Usabilidade e interface amigável:

- Apresentar uma interface intuitiva e fácil de usar.
- Oferecer uma experiência do usuário agradável e satisfatória.

## **4.3 Entrada de Dados**

### **4.3.1 Página Web**

Para promover uma interação visual eficaz e intuitiva entre o usuário do aplicativo e as variáveis de entrada, conforme tabela 2, relacionadas ao seu perfil e ao produto de crédito imobiliário desejado, foi desenvolvida uma página web interativa no formato de um formulário. Esta abordagem foi projetada para simplificar o processo de preenchimento, assegurando simultaneamente a coerência e a relevância das informações fornecidas.

Nesta página web interativa, os usuários são conduzidos a preencher uma série de variáveis, que posteriormente serão utilizadas nos modelos de inteligência artificial, estas variáveis são fundamentais para a análise de seu perfil e para a avaliação do crédito imobiliário desejado. Essas variáveis abrangem tanto informações pessoais, como renda mensal e idade, quanto detalhes específicos do produto de crédito imobiliário, incluindo o valor desejado, o prazo de pagamento e a taxa de juros, entre outros aspectos.

Para garantir a praticidade e a eficiência do processo de preenchimento, o formulário foi projetado de forma intuitiva, com campos claramente definidos e orientações contextuais para cada variável. Adicionalmente, foram implementados recursos de validação em tempo real para auxiliar os usuários a evitarem erros e assegurar a precisão das informações fornecidas.

**Tabela 2 - Tabela das variáveis de entrada (continua)**

<b>Variável de Entrada</b>	<b>Descrição</b>	<b>Grupo</b>	<b>Tipo Variável</b>
Data Nascimento	Data de Nascimento	Cliente	Texto
Profissão	Que tipo de ocupação o cliente tem	Cliente	Texto
Renda Líquida	Salário líquido	Renda	Numérico
FGTS	Valor disponível no FGTS	Renda	Numérico
Gasto Educação Mes	Gasto mensal com educação, exemplo faculdade	Gastos	Numérico
Gasto Imobiliário Mes	Gasto mensal com imóvel, exemplo aluguel	Gastos	Numérico
Gasto Saúde Mes	Gasto mensal com saúde, exemplo plano de saúde	Gastos	Numérico
Gasto Cartão Mês 1	Gasto no cartão a 1 meses atras	Gastos	Numérico
Gasto Cartão Mês 2	Gasto no cartão a 2 meses atrás	Gastos	Numérico
Gasto Cartão Mês 3	Gasto no cartão a 3 meses atrás	Gastos	Numérico
Gasto Cartão Mês 4	Gasto no cartão a 4 meses atrás	Gastos	Numérico
Gasto Cartão Mês 5	Gasto no cartão a 5 meses atrás	Gastos	Numérico
Gasto Cartão Mês 6	Gasto no cartão a 6 meses atrás	Gastos	Numérico
Quantidade Filho	Número de filhos	Cliente	Numérico
Reserva Renda Fixa	Valor da reserva em renda fixa	Renda	Numérico
Reserva Renda Variável	Valor da reserva em renda variável	Renda	Numérico
Reserva Emergência	Possui reserva de emergência	Renda	Booleano
Valor Investido Mensalmente	Valor Investido Mensalmente	Renda	Numérico
Utilizou Reserva 12 meses	Se teve a necessidade de utilizar a reserva nos últimos 12 meses	Gastos	Booleano
Limite Cartão	Valor do limite bruto do cartão	Cliente	Numérico
Comprometimento Cartão	Valor comprometido do limite do cartão	Gastos	Numérico
Escolaridade	Nível de educação mais alto que o cliente alcançou	Cliente	Texto
Beneficiaria INSS	Se o usuário é beneficiário INSS	Cliente	Booleano
Dívida Pendente	Se sim, valor da dívida	Gastos	Numérico
Empréstimo Valor Mensal	Se sim, valor do pagamento mensal	Gastos	Numérico
Empréstimo Valor	Se sim, valor da dívida	Gastos	Numérico
Produto	Financiamento Imobiliário ou veículo	Produto	Texto
Sexo	Sexo do proponente	Cliente	Texto
Possui Carro Próprio	Se o usuário possui um carro próprio	Cliente	Booleano

**Tabela 2 - Tabela das variáveis de entrada (conclusão)**

Possui Casa Própria	Sinaliza se o cliente possui uma casa ou apartamento	Cliente	Booleano
Taxa Juros	Taxa de juros do produto	Produto	Numérico
Estado civil	Estado civil do cliente	Cliente	Texto
Mora com os pais	Se usuário reside com os pais	Cliente	Booleano
Data Nascimento	Data Nascimento do Cliente	Cliente	Texto
CEP Moradia Atual	Região do imóvel	Produto	Texto
Idade Carro Cliente	Idade do carro do cliente	Cliente	Numérico
Tipo Organização	Tipo de organização onde o cliente trabalha	Cliente	Texto
Valor Imóvel	Valor do imóvel escolhido	Produto	Numérico
m2_imovel	Metragem do imóvel escolhido	Produto	Numérico
Idade Imóvel	Idade do imóvel escolhido	Produto	Numérico
Quartos	Quantos quartos o imóvel possui	Produto	Numérico
Vagas	O imóvel escolhido possui quantas vaga	Produto	Numérico
CEP Imóvel Financiado	Região do imóvel escolhido	Produto	Texto
Area Lazer	O imóvel escolhido possui área de lazer	Produto	Booleano
Elevador	O imóvel possui elevador	Produto	Booleano
Necessita Empréstimo ITBI	Necessita realizar empréstimo para o ITBI (Imposto sobre Transmissão de Bens Imóveis)	Produto	Booleano
Indexador	Indexador de juros TR, IPCA, POUAPANÇA	Produto	Texto
Tipo Imóvel	Tipo de imóvel escolhido	Produto	Texto
Estado Documentação Imóvel	Estado da documentação do imóvel escolhido	Produto	Texto
Quantos Andares	Quantos andares possui o prédio/imóvel	Produto	Numérico
Objetivo imóvel	Objetivo do imóvel escolhido	Produto	Texto
Tempo Emprego	Tempo total em que o cliente tem empregado	Cliente	Numérico
Tipo Amortização	Tipo de amortização do financiamento (SAC ou TP)	Produto	Texto
Mix de Crédito	Pontuação junto a outras empresas exemplo Serasa	Cliente	Numérico

Fonte: Elaboração Própria

Para garantir a funcionalidade dinâmica e responsiva da página web interativa, foi escolhida uma combinação de linguagens de programação: JavaScript e HTML.

O HTML (*Hypertext Markup Language*) serve como espinha dorsal da estrutura da página, permitindo-nos criar uma interface intuitiva e bem-organizada para os usuários. Utilizando *tags* HTML, foi construída a estrutura do formulário, definindo claramente cada campo de entrada e suas respectivas etiquetas descritivas. Além

disso, aproveitou-se as capacidades do HTML para incorporar elementos visuais interativos, como botões de seleção e caixas de verificação, facilitando assim a entrada de dados por parte do usuário (Mozilla, 2023).

Na figura 12, 13 e 14, é possível observar a disposição dos campos de entrada, divididos em etapas: Informações Pessoais, Saúde Financeira e, por fim, Crédito Imobiliário, organizados de maneira ordenada e de fácil compreensão. Cada campo é acompanhado por uma etiqueta descritiva, garantindo que os usuários saibam exatamente que tipo de informação é esperada.

**Figura 12 - Tela de Informações Pessoais - etapa 1**

**VAMOS PREENCHER ALGUNS DETALHES PARA NOSSA ANÁLISE!**

Nome	Email		
dd/mm/aaaa	Possui Filho?	Escolaridade	
Genero	Estado	Cidade	
CEP	Estado Civil		
Profissão	Limite Cartão		
Tipo Organização	Tempo_emprego	Idade carro	
<input type="checkbox"/> INSS?	<input type="checkbox"/> Carro Proprio?	<input type="checkbox"/> Casa Proprio?	<input type="checkbox"/> Mora com pais?

**VOLTAR** **CONTINUAR**

Fonte: Elaboração própria

Figura 13 - Tela de Saúde Financeira – etapa 2

INFORMAÇÕES PESSOAIS → SAÚDE FINANCEIRA → CRÉDITO IMOBILIÁRIO → ENVIAR PARA ANÁLISE

### INFORMAÇÕES SOBRE SUA RENDA

Renda_Liquida	\$	FGTS na entrada	\$
Reserva_renda_	\$	Reserva_renda_	\$
		Valor_investido	\$
		Mix de Crédito	\$

### INFORMAÇÕES SOBRE SEUS GASTOS

Gasto com educação por mês	\$	Gasto com aluguel por mês	\$
Gasto com saúde por mês	\$	Comprometimento_cartao	\$
Consignado_valor	\$	Emprestimo_Valor_Mes	\$
		Emprestimo_valor	\$

Gastos com Cartão

Mês 1	\$	Mês 2	\$	Mês 3	\$
Mês 4	\$	Mês 5	\$	Mês 6	\$

por mês

Possui Plano de Saúde ?

**VOLTAR** **CONTINUAR**

Fonte: Elaboração própria

Essa metodologia de divisão em etapas aprimora a experiência do usuário, possibilitando que ele dedique sua atenção a uma seção específica por vez, o que torna o preenchimento do formulário mais claro e intuitivo. Adicionalmente, esse enfoque contribui para evitar a sobrecarga de informações, o que torna a interação mais gerenciável.

Figura 14 - Tela de Crédito Imobiliário – etapa 3

**INFORMAÇÕES SOBRE O PRODUTO FINANCEIRO**

Taxa de Juros ano	Valor do Financiamento		
Prazo para pag:	Valor Primeira I	Indexador	Tipo_Amortizacac

**Necessita empréstimo para ITBI?**

**INFORMAÇÕES SOBRE O IMÓVEL**

Valor_imovel	M² imóvel	Idadeimovel
Quartos	Vagas	
CEP_imovel_financiado	Quantos_andares	
Tipo_imovel	Objetivo_imovel	

**Possui Area de lazer ?**       **Possui Elevador?**       **Documentação está OK?**

**VOLTAR**      **CONTINUAR**

Fonte: Elaboração própria

Na figura 15, destaca-se a importância da norma que requer que o usuário concorde explicitamente com o uso de suas informações pessoais, seguindo as recomendações da Lei Geral de Proteção de Dados (Brasil, 2018). Essa medida visa garantir a privacidade e a segurança dos dados do usuário, assegurando que apenas informações estritamente necessárias sejam coletadas e utilizadas, e somente com o consentimento explícito do usuário. Esta Norma também está presente no fluxo de usuário colaborativo, conforme figura 16.

Figura 15 - Consentimento Lei Geral de Proteção de Dados - Usar



Fonte: Elaboração própria

Ao incluir essa norma na interface do aplicativo, a equipe demonstra seu compromisso com a transparência e o respeito à privacidade dos usuários. Isso cria uma relação de confiança entre o usuário e o aplicativo, além de estar em conformidade com as regulamentações de proteção de dados vigentes.

Figura 16 - Consentimento Lei Geral de Proteção de Dados - Colaborar



Fonte: Elaboração própria

O JavaScript oferece funcionalidades dinâmicas que tornam as páginas verdadeiramente interativas. Recursos como validação de dados em tempo real e *feedback* instantâneo para o usuário são implementados. Por exemplo, ao preencher um campo obrigatório incorretamente, é destacado o erro imediatamente, fornecendo orientações úteis sobre como corrigi-lo (JavaScript, 2024).

A combinação harmoniosa de HTML e JavaScript permite criar uma página web interativa que é não apenas visualmente atraente, mas também altamente funcional e fácil de usar.

Além de sua função na parte visual e interativa da página, o JavaScript desempenha um papel fundamental na comunicação entre o *front-end* e o *back-end* do aplicativo por meio de API. Ao utilizar JavaScript para se comunicar com o *back-end* via API e rota, é possível garantir uma integração fluida e eficiente entre as diferentes partes do sistema (JavaScript, 2024).

O uso de rotas no *back-end* permite que o JavaScript do *front-end* se comunique com os *endpoints* corretos do servidor, garantindo que as solicitações sejam direcionadas para as funções apropriadas.

Por meio de uma requisição *GET* via HTTP, o JavaScript encaminha os dados do formulário preenchido pelos usuários para o serviço no *back-end*, utilizando a rota correspondente. Essa comunicação permite que os dados sejam processados adequadamente, seja para validação e cadastro de novos usuários, no caso da rota "cadastrovalidacao", ou para inserção de dados na base de dados colaborativa, no caso da rota "cadastrobasededados".

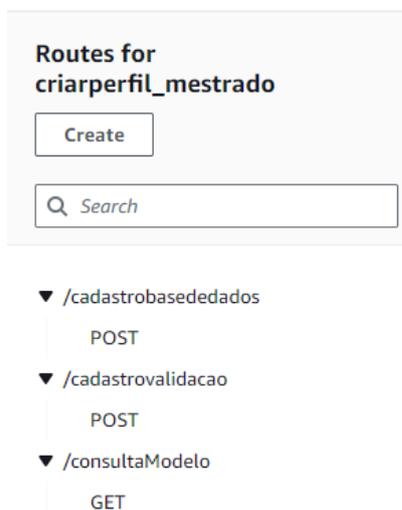
Por sua vez, o *back-end* processa essas solicitações e envia uma resposta de volta ao *front-end*, geralmente em formato JSON (*JavaScript Object Notation*).

#### 4.3.2 API

Neste projeto, para facilitar a comunicação entre a página web(*front-end*) e a serviços de processamento(*back-end*), foi adotado o AWS API GATEWAY, um serviço de infraestrutura em nuvem capaz de criar APIs baseadas em HTTPS. De acordo com a AWS (2023), o AWS API GATEWAY permite criar, publicar, manter, monitorar e proteger APIs, lidando com o tráfego de milhares de aplicativos, integrando-se de forma harmoniosa aos demais recursos da AWS.

Para atender às necessidades deste projeto, foram desenvolvidas as seguintes rotas, conforme figura 17:

**Figura 17 - Rotas do software**



Fonte: Elaboração própria

- *cadastrovalidacao*: Esta rota é utilizada pelo *front-end* e destinada aos usuários que desejam utilizar o aplicativo, e seu propósito é permitir o cadastro e a validação das informações dos usuários. A funcionalidade associada a esta rota inclui a coleta dos dados do usuário por meio de uma requisição *POST*, seguida pelo processamento desses dados para verificar sua validade e, se aprovados, armazená-los no banco de dados do aplicativo.
- *cadastrobasededados*: Esta rota é designada para os usuários colaborativos que desejam contribuir com a base de dados do aplicativo. A funcionalidade associada a esta rota permite o envio de novos dados ou a atualização de dados existentes no banco de dados. Isso pode incluir informações sobre produtos financeiros, taxas de juros, histórico de crédito dos usuários, entre outros. Os usuários colaborativos podem enviar essas informações por meio de uma requisição *POST*, e o sistema processa esses dados para integrá-los ao banco de dados principal, garantindo sua consistência e validade.
- *consultaModelo*: Esta rota tem o propósito de realizar uma requisição *GET* ao serviço responsável pela execução dos modelos preditivos. O serviço do *back-end* é acionado para executar os modelos de machine learning treinados. Esses modelos utilizam as informações fornecidas pelo usuário.

## 4.4 Armazenamento dos Dados

Neste tópico será apresentado como é feito o armazenamento dos dados do usuário, que é dividido em dois tipos de produtos de computação em nuvem: um para os dados relacionados ao treinamento do modelo, processados em modelo *batch*, e outro para os dados de entrada do usuário, processados em tempo quase real.

### 4.4.1 Simple Storage Service – S3

O *Simple Storage Service* (S3) da AWS é uma solução de armazenamento de objetos eficiente, acessível e altamente confiável, projetada para atender às demandas de empresas e usuários. Destaca-se por oferecer baixo custo, alta disponibilidade, segurança avançada e escalabilidade flexível. Com sua arquitetura distribuída e redundante, garante a disponibilidade dos dados mesmo em casos de falhas (Amazon Web Service, 2023).

Opções abrangentes de segurança permitem controlar o acesso aos dados, enquanto a escalabilidade dinâmica facilita a adaptação à crescente demanda de armazenamento. Amplamente utilizado para armazenar dados analíticos e como insumo para treinamento de modelos de *machine learning*, o S3 proporciona uma base sólida para operações de negócios e impulsiona a tomada de decisões embasadas em insights valiosos (Amazon Web Service, 2023).

A utilização desse armazenamento é de suma importância para garantir a disponibilidade e confiabilidade dos dados utilizados no processo retreinamento do modelo de *machine learning*.

Para este projeto, que inclui dados sensíveis do usuário, foi criado um *bucket* no serviço de armazenamento S3 da AWS, com o nome "enriquecimento-mestrado-giancoli". O *bucket* ilustrado na figura 18 é configurado com restrições de acesso público. Além disso, seus dados em repouso são criptografados, garantindo uma camada adicional de segurança.

Figura 18 - Configuração do *bucket* s3

**Block public access (bucket settings)** Edit

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to all your S3 buckets and objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to your buckets or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

**Block all public access**

On

► Individual Block Public Access settings for this bucket

---

**Default encryption** Info Edit

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type Info

Server-side encryption with Amazon S3 managed keys (SSE-S3)

Bucket Key

When KMS encryption is used to encrypt new objects in this bucket, the bucket key reduces encryption costs by lowering calls to AWS KMS. [Learn more](#)

Enabled

Fonte: Elaboração própria

Na figura 19, também é possível observar que o *bucket* foi configurado com políticas de restrições de acesso público, restringindo o acesso apenas para requisições originadas da rede privada da conta AWS.

Figura 19 - Política de segurança do *bucket*

Policy

```

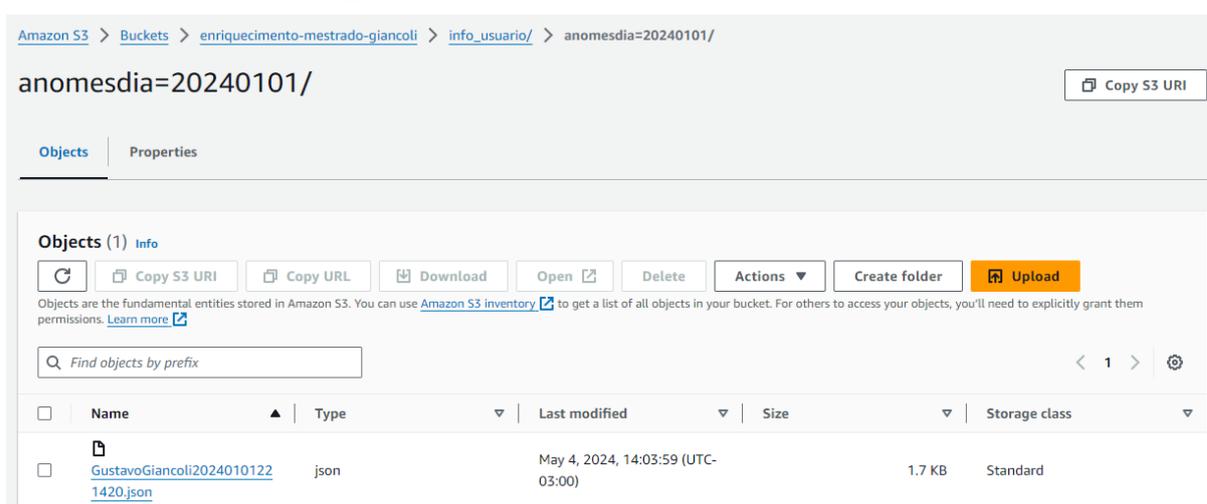
1  {
2    "Version": "2012-10-17",
3    "Id": "PolicyForVPCOnly",
4    "Statement": [
5      {
6        "Sid": "AllowAccessFromSpecificVPC",
7        "Effect": "Deny",
8        "Principal": "*",
9        "Action": "s3:*",
10       "Resource": [
11         "arn:aws:s3:::enriquecimento-mestrado-giancoli",
12         "arn:aws:s3:::enriquecimento-mestrado-giancoli/*"
13       ],
14       "Condition": {
15         "StringNotEquals": {
16           "aws:sourceVpc": "vpc-██████"
17         }
18       }
19     }
20   ]
21 }
```

Fonte: Elaboração própria

Quando um usuário utiliza o *software* para colaborar com o enriquecimento da base de dados, faz o preenchimento do formulário detalhando sua experiência passada e clica no botão de envio, uma requisição *POST* é acionada na rota "*cadastrobasededados*" da API. Isso desencadeia o seguinte fluxo de ações:

1. O *front-end* do aplicativo realiza uma requisição *POST* para a rota "*cadastrobasededados*". Esta requisição contém os dados do formulário a serem armazenados s3.
2. Quando o *back-end* recebe a requisição, ele processa os dados recebidos e prepara-os para serem armazenados no formato de arquivo json, dentro de um caminho do s3, conhecido como *Prefix*. Os dados são armazenados na estrutura ilustrada na figura 20 para organizar os dados de forma eficiente e acessível.
3. Após o armazenamento bem-sucedido dos dados, o *back-end* confirma uma resposta de confirmação para o *front-end*, indicando que a operação foi concluída com sucesso.

**Figura 20 - Exemplo de Armazenamento no S3**



Fonte: Elaboração própria

Nesse contexto, o armazenamento adequado dos dados analíticos desempenha um papel crucial na obtenção de modelos de *machine learning* mais precisos e robustos. Além disso, ao permitir o acesso rápido e fácil aos dados necessários, facilita-se o processo de desenvolvimento e teste de novos modelos, acelerando a inovação e promovendo avanços na área de inteligência artificial.

Portanto, o armazenamento de dados analíticos desempenha um papel fundamental no cenário atual, oferecendo a base necessária para impulsionar a pesquisa científica, aprimorar a tomada de decisões estratégicas e impulsionar o progresso contínuo no campo da ciência de dados e da inteligência artificial.

#### 4.4.2 AWS DynamoDb

O teorema CAP foi elaborado pelo cientista da computação Eric Brewer (2000) como uma hipótese, sendo comprovada em 2002 por Seth Gilbert e Nancy Linch, tornando assim um teorema (Cardoso, 2012).

O teorema define três características para classificar e apoiar a escolha do banco de dados ideal para seu sistema, sendo que os relacionais são definidos como consistente e disponível e os não relacionais como tolerância ao particionamento e disponibilidade (Silva, 2011).

1. Consistência de que todos os nós de um sistema distribuído necessitam que a versão dos dados seja a mesma.
2. Disponibilidade para que ainda que um dos nós do sistema distribuído falhe, outro nó garante a disponibilidade da informação.
3. Tolerância ao particionamento o dado e particionado entre os nós do sistema distribuído, mesmo que em servidores diferentes.

Portanto, o uso do AWS DynamoDB neste projeto é justificado pelo fato de que ele atende aos requisitos de consistência, disponibilidade e tolerância a falhas necessários para armazenar dados que exigem baixa latência, garantia de consistência e disponibilidade contínua para consulta pela API, essenciais para o bom funcionamento do modelo e da aplicação como um todo.

A AWS disponibiliza como seu principal produto para atender o mercado de bancos NoSQL o AWS DynamoDB, que tem uma infraestrutura totalmente gerenciada pela AWS possibilitando alta performance e escalabilidade, oferecendo serviços de criptografia, backup e expurgo dos dados, caso configurado. O AWS DynamoDB permite que você transfira os encargos administrativos de operação e escalabilidade de um banco de dados distribuído (Amazon Web Service, 2023).

A estrutura da tabela nomeada de Processos é definida com a composição da chave de partição sendo o *e-mail* do usuário e a chave de ordenação sendo a data de criação do registro, como ilustrado na figura 21.

**Figura 21 - Estrutura da tabela NoSQL**

**General information** [Info](#)

Partition key Email (String)	Sort key DataCriacao (String)
Capacity mode <u>On-demand</u>	Table status ✔ Active
Alarms ✔ No active alarms	Point-in-time recovery (PITR) <a href="#">Info</a> <u>Off</u>
Resource-based policy <a href="#">Info</a> ✔ Active	

Fonte: Elaboração própria

Além disso, os dados armazenados na tabela são criptografados para assegurar sua segurança enquanto estão em repouso. A tabela possui políticas específicas de acesso para permitir interações somente com aplicativos que estejam dentro da rede privada da conta, conhecida na AWS como VPC.

Este conjunto de medidas ilustrados na figura 21 e 22, visa garantir a confidencialidade e integridade dos dados, restringindo o acesso apenas a fontes autorizadas e confiáveis.

**Figura 22 - Política de acesso da tabela NoSQL**

**Resource-based policy for table** [Info](#) Manage policy ▼ Copy

The resource-based policy, written in JSON, helps manage access to this DynamoDB table.

```

1  {
2    "Version": "2012-10-17",
3    "Statement": [
4      {
5        "Effect": "Deny",
6        "Principal": "*",
7        "Action": "dynamodb:*",
8        "Resource": "arn:aws:dynamodb:sa-east-1:██████████:table
9          /Processos",
10       "Condition": {
11         "StringNotEquals": {
12           "aws:sourceVpc": "vpc-██████████"
13         }
14       }
15     ]
  
```

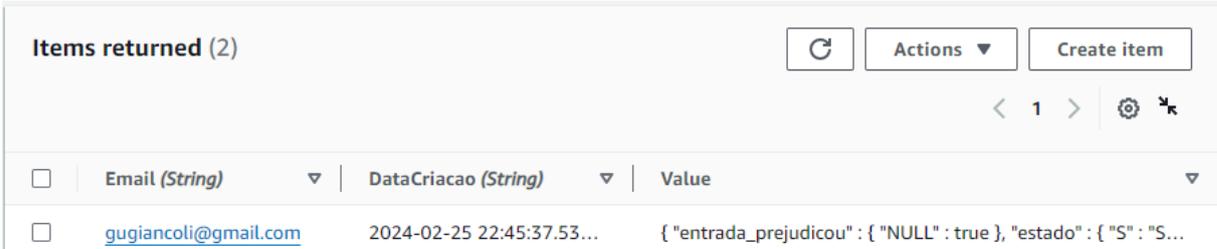
JSON Ln 1, Col 1

Fonte: Elaboração própria

Quando um usuário utiliza o *software* para solicitar a análise do seu perfil e produto de crédito imobiliário, faz o preenchimento do formulário e clica no botão de envio, uma requisição *POST* é acionada na rota "cadastrovalidacao" da API. Isso desencadeia o seguinte fluxo de ações:

1. O front-end do aplicativo envia uma requisição *POST* para a rota "cadastrobasededados", contendo os dados do formulário a serem processados.
2. Ao receber a requisição, o *back-end* prepara os dados recebidos e envia uma requisição para a rota "consultaModelo", com o objetivo de submetê-los aos modelos de inteligência artificial para análise.
3. O *back-end* armazena o dado analisado na tabela de processos do AWS DynamoDB, como ilustrado na figura 23. Em seguida, esse dado é enviado para o Amazon S3 para ser utilizado em estudos analíticos.
4. Após o armazenamento bem-sucedido dos dados, o *back-end* envia uma resposta de confirmação para o *front-end*, indicando que a operação foi concluída com sucesso.

Figura 23 - Dado armazenado na tabela de Processos



Items returned (2)			
	Email (String)	DataCriacao (String)	Value
<input type="checkbox"/>	<a href="mailto:gugiancoli@gmail.com">gugiancoli@gmail.com</a>	2024-02-25 22:45:37.53...	{ "entrada_prejudicou": { "NULL": true }, "estado": { "S": "S..."

Fonte: Elaboração própria

## 4.5 Processamento dos Dados

Neste tópico será apresentado como é feito o armazenamento dos dados do usuário, que é dividido em dois tipos de produtos de computação em nuvem: um para os dados relacionados ao treinamento do modelo, processados em modelo *batch*, e outro para os dados de entrada do usuário, processados em tempo quase real.

#### 4.5.1 Elastic Compute Cloud – EC2

No que tange às vantagens, a elasticidade do EC2 permite a escalabilidade instantânea dos recursos de computação, otimizando custos ao evitar gastos excessivos. Uma variedade de tipos de instância, combinada com facilidade de implementação e gerenciamento, proporciona uma solução flexível para diversas necessidades. *Startups* podem lançar produtos rapidamente, enquanto empresas maiores podem implementar ambientes de teste e desenvolvimento, além de cargas de trabalho de alta demanda (Amazon Web Service, 2023).

Em resumo, o Amazon EC2 permanece como um alicerce vital da infraestrutura de nuvem da AWS, oferecendo escalabilidade e flexibilidade. Seu potencial é maximizado quando os usuários compreendem profundamente suas vantagens e consideram cuidadosamente os desafios. Ao fazê-lo, as organizações podem alavancar eficazmente o EC2 para impulsionar a inovação e a eficiência operacional em um ambiente em constante mudança (Amazon Web Service, 2023).

O AWS EC2 desempenha um papel essencial neste projeto, não apenas na distribuição e processamento dos resultados via rota API ("consultaModelo"), mas também desempenha uma função crucial no retreinamento do modelo de inteligência artificial. Dentro da infraestrutura do AWS EC2, são alocados os processos que requerem maior tempo de execução, capacidades computacionais mais elevadas e tempos de resposta mais rápidos, garantindo uma disponibilidade contínua.

A configuração do EC2 representada na figura 24, inclui uma instância com alto poder computacional, otimizada para memória, conforme recomendações da AWS. Isso resulta em tempos de resposta mais rápidos para a API e reduz o tempo de execução da etapa de retreinamento do modelo (Amazon Web Services, 2023).

Essa abordagem garante que o sistema seja capaz de lidar com cargas de trabalho intensivas e exigentes, mantendo a eficiência e a confiabilidade necessárias para oferecer um serviço de alta qualidade aos usuários.

**Figura 24 - Configuração da instancia do EC2**

Attribute	r3.2xlarge
On-Demand Linux pricing	1.3990 USD per Hour
On-Demand Windows pricing	1.7670 USD per Hour
vCPUs	8 (4 core)
Memory (MiB)	62464
Storage (GB)	160
Supported root device types	ebs, instance-store
Network performance	High
Architecture	x86_64
Burstable	false
Free-tier eligible	false
Current generation	false

Fonte: Elaboração própria

Quando ocorre uma requisição na API para a rota “consultaModelo”, os dados preenchidos nos formulários são enviados para o código Python em execução na instância do AWS EC2. O processo segue os seguintes passos:

1. Ao receber a requisição, o *back-end* prepara os dados recebidos para submetê-los aos modelos de inteligência artificial para análise.
2. O *back-end* realiza a primeira análise utilizando o modelo causal do *score* de crédito do cliente. Nesta etapa, são validadas duas situações: o *score* do cliente pré-crédito imobiliário e, posteriormente, o mesmo *score* após a aquisição do crédito. Se o *score* sofrer alteração a recomendação é negativa.
3. Em seguida, o *back-end* realiza a análise utilizando o segundo modelo, responsável pelo risco de inadimplência do cliente.
4. Após o processamento, a análise é retornada para o requisitante da API.

Em nossa instancia AWS EC2 foi criado uma tarefa agendada para executar o retreinamento dos modelos diariamente. Neste caso o seguinte processo é executado:

1. O *back-end* coleta os dados disponibilizados no Amazon S3 pelos usuários que optaram por colaborar com o projeto.

2. Os dados são então consolidados, preparados e normalizados para serem usados na inferência do modelo.
3. Em seguida, o *back-end* publica os artefatos do modelo e instala o novo artefato no código de execução dos modelos.

#### 4.5.2 AWS Lambda

O AWS Lambda é o primeiro serviço de computação *serverless* para praticamente qualquer tipo de aplicação ou serviço de *back-end*. Funções podem ser configuradas a partir do envio do código em formato ZIP ou de imagem de contêiner, e o Lambda se responsabiliza por alocar os recursos necessários para a execução das funções com base na solicitação ou evento de entrada, para qualquer dimensão de tráfego (Amazon Web Service, 2023).

Considerado a primeira plataforma *serverless* disponível no mercado em alta escala, possui algumas concorrentes como:

- Google Cloud Functions
- Microsoft Azure Functions

Atualmente o AWS Lambda aceita as seguintes linguagens de programação: Java, Go, PowerShell, Node.js, C#, Python, e Ruby, bem como fornece uma API de tempo de execução que permite usar qualquer linguagem de programação adicional para criar suas funções (Amazon Web Service, 2023).

Neste projeto, a AWS Lambda é desenvolvida em linguagem de programação Python e desempenha um papel fundamental ao receber as requisições das APIs “cadastrovalidacao” e “cadastrobasededados”, cada rota possui sua AWS Lambda, respectivamente `lbdcadastrovalidacao` e `lbdcadastrobasededados` demonstrado na figura 25.

**Figura 25 - Funções AWS Lambda**

<input type="checkbox"/>	Function name	Description	Package type	Runtime
<input type="checkbox"/>	<a href="#">lbdcadastrobasededados</a>	-	Zip	Python 3.12
<input type="checkbox"/>	<a href="#">lbdcadastrovalidacao</a>	-	Zip	Python 3.12

Fonte: Elaboração própria

Ela é responsável por preparar os dados e salvá-los nos respectivos armazenamentos. Além disso, a AWS Lambda também realiza a requisição aos modelos de inteligência artificial. O processo ocorre da seguinte forma:

1. Quando uma requisição é feita pelo *front-end* às APIs “cadastrovalidacao” ou “cadastrobasededados”, a AWS Lambda é acionada para processá-la.
2. A AWS Lambda prepara os dados recebidos da requisição para que estejam prontos para serem salvos nos armazenamentos correspondentes, como o Amazon S3 ou o AWS DynamoDB.
3. Após o preparo dos dados, a Lambda os salva no armazenamento apropriado, garantindo sua integridade e segurança.
4. Além disso, a Lambda realiza a requisição aos modelos de inteligência artificial para análise dos dados, fornecendo o retorno dos insights para os usuários.

Todo esse processo é executado de forma eficiente e escalável pela AWS Lambda, garantindo uma resposta rápida e confiável às requisições com as configurações expostas na figura 26.

**Figura 26 - Configuração da AWS Lambda**

General configuration <a href="#">Info</a>			<a href="#">Edit</a>
Description	Memory	Ephemeral storage	
-	1024 MB	512 MB	
Timeout	SnapStart <a href="#">Info</a>		
0 min 10 sec	None		

Fonte: Elaboração própria

Essa abordagem permite uma integração suave entre as APIs, os armazenamentos e os modelos de inteligência artificial, facilitando a execução das tarefas e garantindo a qualidade e disponibilidade dos serviços oferecidos aos usuários.

É importante mencionar que a AWS Lambda possui alguns limites a serem considerados no contexto deste projeto, por alguns desses limites como tempo de execução os modelos são executados em outro contexto (AWS EC2) explicado no item anterior. Esses limites incluem por exemplo, a alocação de memória varia de 128

*megabytes* a 10.240 MB em incrementos de 1 *megabytes*, com potência de Unidade Central de Processamento (CPU) ajustada de acordo. O tempo limite da função é de 15 minutos, e o tamanho máximo do pacote de implantação é de 250 *megabytes* não compactados. Além disso, há limites para a quantidade de ambientes de execução, a carga de invocação e o tamanho do *payload* de solicitação e resposta (Amazon Web Service, 2023)

#### 4.6 Modelo de Inteligência artificial

Este projeto de *software* foi elaborado utilizando a linguagem de programação Python. Neste tópico será explorado detalhadamente o desenvolvimento dos modelos de árvore de decisão empregados no *software*, incluindo o modelo de cálculo de *score* de crédito e o modelo de avaliação de risco de inadimplência.

Cada modelo será abordado em uma subseção específica, seguindo o modelo delineado na figura 2: formulação de perguntas, coleta de dados, criação de atributos, normalização dos dados, limpeza dos dados, amostragem dos dados, treinamento do modelo, avaliação do modelo e implantação do modelo.

Conforme destacado por Karasan (2022), a escassez e a indisponibilidade de dados sempre foram problemas enfrentados nas finanças, e os modelos econométricos têm suas limitações. O autor ressalta que uma solução baseada em modelos com aprendizagem automática é uma alternativa para lidar com o problema de escassez de dados, sendo uma tendência na área de finanças e gestão de riscos financeiros.

Diante das dificuldades encontradas na obtenção direta dos dados, considerados sensíveis, a equipe recorreu à plataforma Kaggle, vinculada ao Google, que abriga uma comunidade ativa de cientistas de dados. Foi possível acessar a base de dados necessária para o presente estudo por meio desta plataforma (Kaggle, 2024).

Após o *download* do conjunto de dados no formato CSV (Comma Separated Values), foi realizada a transferência para o Databricks, uma poderosa ferramenta de análise e processamento de dados baseada em nuvem.

O Databricks oferece uma interface de notebook que facilita o desenvolvimento e execução de códigos em Python, vale ressaltar que para usuários universitários

dispõem de um ambiente gratuito, o que reduz o custo de desenvolvimento deste *software*.

Além disso, o Databricks destaca-se por sua capacidade de lidar eficientemente com grandes conjuntos de dados e facilitar a colaboração entre equipes de análise de dados (DataBricks, 2024).

Posteriormente, o código é migrado sem alterações para a estrutura de processamento na AWS citadas nas sessões anteriores.

#### 4.6.1 Modelo de Score de Crédito

O modelo em questão tem como objetivo responder à pergunta sobre o *score* de crédito do usuário e qual seria o impacto desse *score* após a aquisição de crédito imobiliário. Uma base de dados<sup>1</sup> disponível no Kaggle, intitulada "Credit Score Classification", fornecida pelo autor Huma Gonen, é utilizada para esse fim. Essa base de dados possui a estrutura ilustrada na figura 27 e contém 31.853 registros.

**Figura 27 – Estrutura da base de treinamento – Modelo de Score Crédito**

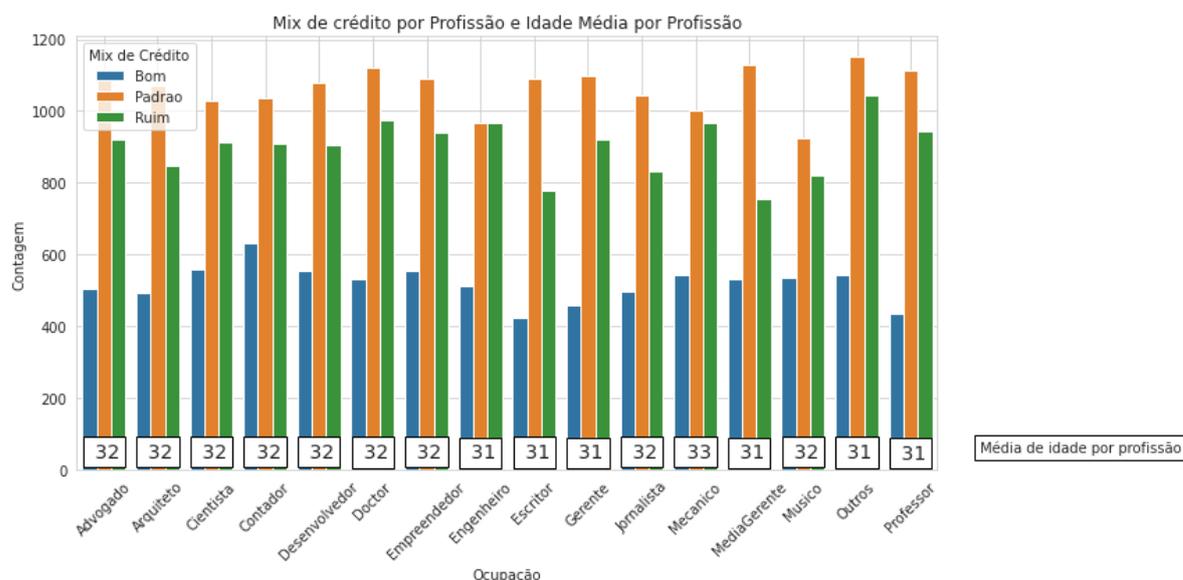
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 31853 entries, 4 to 99991
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   idade                                 31853 non-null  int32
1   ocupacao                              31853 non-null  category
2   renda_anual                           31853 non-null  int32
3   salario_mensal                        31853 non-null  int32
4   taxa_juros                             31853 non-null  int32
5   num_de_emprestimo                     31853 non-null  int32
6   tipo_de_emprestimo                    31853 non-null  category
7   Divida_pendente                       31853 non-null  int32
8   Empréstimo_Valor_Mes                  31853 non-null  int32
9   Valor_investido_mensalmente            31853 non-null  int32
10  Pagamento_Comportamento              31853 non-null  category
11  Saldo_Mensal                           31853 non-null  int32
12  Mix_de_credito_normalizado             31853 non-null  int64
dtypes: category(3), int32(9), int64(1)
```

Fonte: Elaboração própria

<sup>1</sup> Disponível em: < <https://www.kaggle.com/code/humagonen/eda-credit-score-classification-train> >

Uma análise da distribuição dos dados foi realizada, comparando a idade média, a profissão e o resultado do *score* de crédito, também denominado como mix de crédito. Essa análise foi visualizada por meio do gráfico apresentado na figura 28.

**Figura 28 - Mix de crédito por Profissão e Idade Média por Profissão**

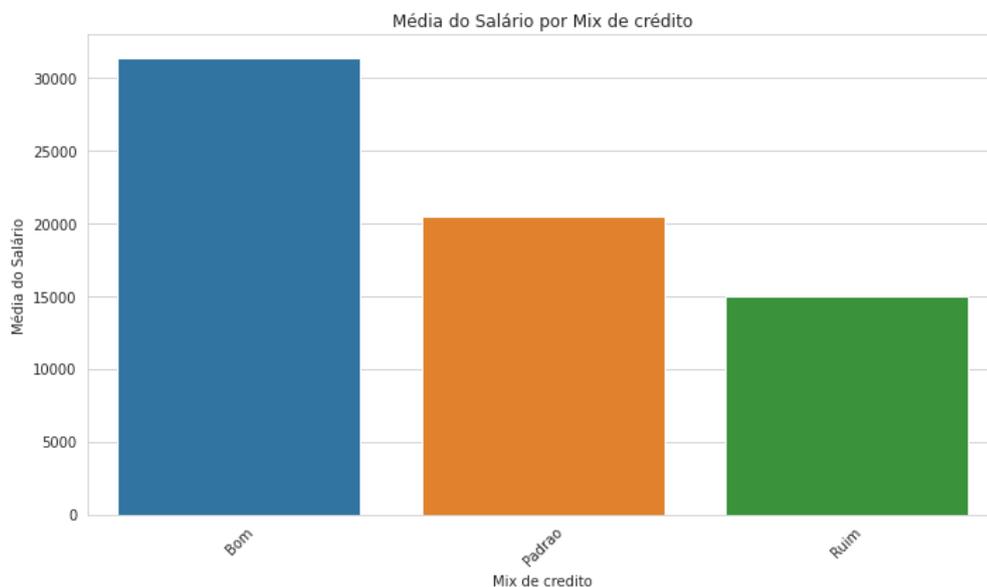


Fonte: Elaboração própria

Esse gráfico permite que se visualize a distribuição do mix de crédito por profissão, juntamente com a média de idade por profissão, indicada por meio de números sobre as barras correspondentes. Essa análise auxilia na compreensão da relação entre a ocupação dos indivíduos, o *score* de crédito e a idade média.

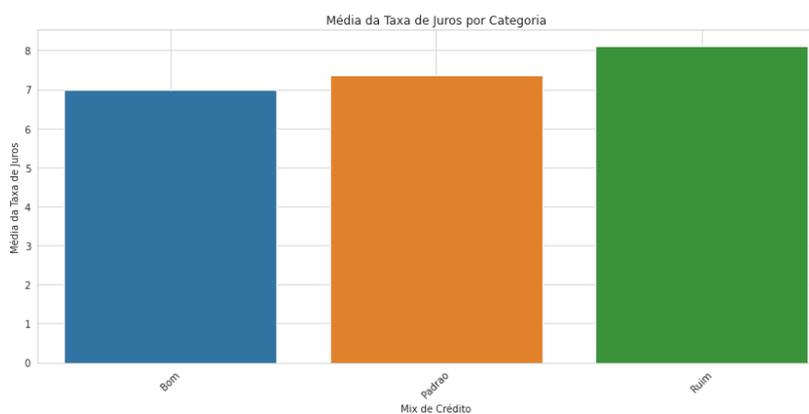
Outra análise relevante, conforme apresentado na figura 29, revela a comparação entre a média salarial e o mix de crédito. Nessa figura, observa-se claramente que o salário exerce uma influência positiva sobre o *score* de crédito dos indivíduos. Essa constatação sugere que, em linhas gerais, quanto maior o salário de uma pessoa, maior tende a ser o seu *score* de crédito. Essa relação entre salário e *score* de crédito é de suma importância, pois evidencia a relevância do aspecto financeiro individual na determinação da confiabilidade e capacidade de pagamento dos indivíduos.

Além disso, essa descoberta pode orientar estratégias financeiras e políticas de concessão de crédito, contribuindo para uma análise mais precisa e criteriosa por parte do *software*.

**Figura 29 - Média do Salário por Mix de crédito**

Fonte: Elaboração própria

Na figura 30, é apresentada a média da taxa de juros em relação ao mix de crédito. Nesse gráfico, é possível observar que a taxa de juros tem uma influência negativa no mix de crédito (*score*), ou seja, taxas de juros mais altas tendem a estar associadas a *scores* de crédito mais baixos.

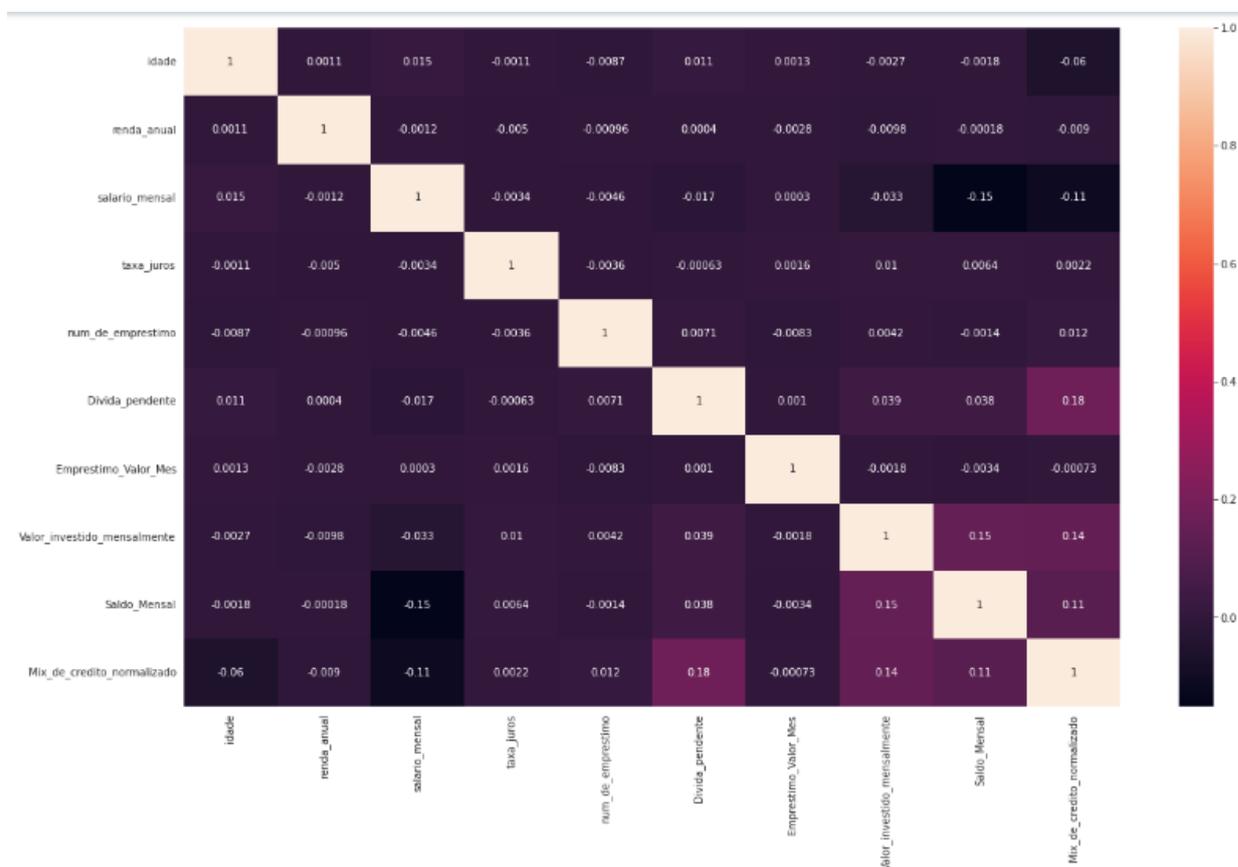
**Figura 30 - Média da Taxa de Juros por Mix de Crédito**

Fonte: Elaboração própria

Essas análises nos fornecem insights valiosos sobre a relação entre o salário, a taxa de juros e o mix de crédito, contribuindo para uma compreensão mais completa dos fatores que influenciam o *score* de crédito dos indivíduos.

Após a normalização dos dados, é realizada a correlação das variáveis. As variáveis com maior correlação com a variável "Mix de Crédito" serão selecionadas para serem utilizadas no modelo de árvore de decisão.

**Figura 31 - Correlação das Variáveis - Modelo de Score Crédito**



Fonte: Elaboração própria

A correlação é uma medida estatística fundamental que revela a associação entre duas variáveis e varia de -1 a 1, indicando o grau e direção dessa relação (Harrison, 2020).

Na figura 31, foi observada uma correlação entre o salário mensal e o mix de crédito normalizado. Foi identificada uma correlação negativa entre essas duas variáveis, indicando que à medida que o salário mensal aumenta, o mix de crédito normalizado tende a diminuir. Essa descoberta sugere uma possível relação entre salários mais elevados e menor exposição a diferentes tipos de crédito.

Outra correlação encontrada na figura 31 é entre a dívida pendente e o mix de crédito normalizado. Aqui, uma correlação positiva considerável foi identificada, indicando que um aumento na dívida pendente está associado a um mix de crédito

mais alto. Essa relação sugere uma possível ligação entre níveis mais altos de dívida e uma maior diversidade de fontes de crédito.

Além disso, uma correlação positiva foi encontrada entre o saldo mensal e o valor investido mensalmente. Isso indica que indivíduos com saldos mensais mais altos tendem a investir quantias maiores mensalmente. Essa associação entre o saldo disponível e o investimento sugere uma relação entre a estabilidade financeira e a capacidade de investir.

Com o objetivo de simplificar a análise do resultado do modelo, utilizou-se o algoritmo de clusterização *k-means*. Por meio desta abordagem, foi possível criar grupos homogêneos para as variáveis numéricas. Os resultados da clusterização para cada atributo numérico estão apresentados na Tabela 3.

**Tabela 3 - Tabela clusterização das variáveis de Score de Crédito(continua)**

Cluster	Variável	Mínimo	Máximo
Cluster 1	Idade	24.00	28.00
Cluster 2	Idade	38.00	42.00
Cluster 3	Idade	49.00	100.00
Cluster 4	Idade	14.00	18.00
Cluster 5	Idade	33.00	37.00
Cluster 6	Idade	29.00	32.00
Cluster 7	Idade	43.00	48.00
Cluster 8	Idade	19.00	23.00
Cluster 1	Renda Anual	37832.02	8254386.00
Cluster 2	Renda Anual	100255649.40	114895773.00
Cluster 3	Renda Anual	50195532.60	66935111.40
Cluster 4	Renda Anual	84694777.20	99274021.20
Cluster 5	Renda Anual	9313158.60	27549617.40
Cluster 6	Renda Anual	116161371.00	130464048.60
Cluster 7	Renda Anual	67416575.40	83070446.40
Cluster 8	Renda Anual	28978290.00	48242568.60
Cluster 1	Salário Mensal	30142.53	39781.09
Cluster 2	Salário Mensal	7020.26	12731.54
Cluster 3	Salário Mensal	65998.87	81902.77
Cluster 4	Salário Mensal	12752.30	21073.12
Cluster 5	Salário Mensal	51505.68	65563.24
Cluster 6	Salário Mensal	39883.40	51437.16
Cluster 7	Salário Mensal	1639.69	7016.53
Cluster 8	Salário Mensal	21094.56	30122.60
Cluster 1	Taxa de Juros	0.10	38.70

**Tabela 4 - Tabela das variáveis de entrada (continuação)**

Cluster 2	Taxa de Juros	257.10	332.60
Cluster 3	Taxa de Juros	503.30	579.70
Cluster 4	Taxa de Juros	107.10	176.70
Cluster 5	Taxa de Juros	418.50	499.80
Cluster 6	Taxa de Juros	179.40	255.70
Cluster 7	Taxa de Juros	335.20	417.00
Cluster 8	Taxa de Juros	40.50	106.00
Cluster 1	Número de Empréstimos	1.00	101.00
Cluster 2	Número de Empréstimos	855.00	1088.00
Cluster 3	Número de Empréstimos	-100.00	-100.00
Cluster 4	Número de Empréstimos	404.00	597.00
Cluster 5	Número de Empréstimos	143.00	352.00
Cluster 6	Número de Empréstimos	1106.00	1319.00
Cluster 7	Número de Empréstimos	1340.00	1496.00
Cluster 8	Número de Empréstimos	613.00	833.00
Cluster 1	Dívida Pendente	8652.26	11837.77
Cluster 2	Dívida Pendente	19138.79	22972.09
Cluster 3	Dívida Pendente	2760.86	5623.67
Cluster 4	Dívida Pendente	11842.79	15386.65
Cluster 5	Dívida Pendente	1.24	2756.97
Cluster 6	Dívida Pendente	22998.82	26989.58
Cluster 7	Dívida Pendente	5631.23	8642.32
Cluster 8	Dívida Pendente	15391.08	19127.56
Cluster 1	Empréstimo Valor Mês	26.55	31584.60
Cluster 2	Empréstimo Valor Mês	377184.60	444587.40
Cluster 3	Empréstimo Valor Mês	154186.20	210616.20
Cluster 4	Empréstimo Valor Mês	265096.80	317196.00
Cluster 5	Empréstimo Valor Mês	32086.80	90882.00
Cluster 6	Empréstimo Valor Mês	318394.80	376542.00
Cluster 7	Empréstimo Valor Mês	211280.40	263644.20
Cluster 8	Empréstimo Valor Mês	93754.80	152830.80
Cluster 1	Valor Investido Mensalmente	1451.09	2171.67
Cluster 2	Valor Investido Mensalmente	54000.00	54000.00
Cluster 3	Valor Investido Mensalmente	0.00	498.73
Cluster 4	Valor Investido Mensalmente	3154.42	4532.89
Cluster 5	Valor Investido Mensalmente	918.56	1450.89
Cluster 6	Valor Investido Mensalmente	2171.87	3149.32
Cluster 7	Valor Investido Mensalmente	498.75	918.40
Cluster 8	Valor Investido Mensalmente	4540.34	8679.44
Cluster 1	Saldo Mensal	923.52	1438.36
Cluster 2	Saldo Mensal	3065.50	3929.69
Cluster 3	Saldo Mensal	1856.29	2381.10

**Tabela 5 - Tabela das variáveis de entrada (conclusão)**

Cluster 4	Saldo Mensal	3933.46	5074.37
Cluster 5	Saldo Mensal	1438.45	1856.01
Cluster 6	Saldo Mensal	2381.63	3064.55
Cluster 7	Saldo Mensal	0.71	923.03
Cluster 8	Saldo Mensal	5077.36	8385.91

Fonte: Elaboração Própria

Para o treinamento do modelo, foi adotada a configuração padrão da função `train_test_split` da biblioteca `sklearn`. Essa configuração consistiu na divisão da base de dados em duas partes distintas: 70% dos registros (22.297 registros) foram designados para o conjunto de treinamento, enquanto os restantes 30% (9.556 registros) foram reservados para o conjunto de teste.

Essa abordagem de divisão equilibrada permitiu que o modelo fosse treinado em uma quantidade representativa de dados e, posteriormente, validado em uma parcela independente da base, garantindo uma avaliação de sua performance.

#### 4.6.2 Modelo de Risco de Inadimplência

O modelo desenvolvido foi concebido com o propósito de analisar o risco de inadimplência dos usuários, com uma atenção especial direcionada ao contexto da aquisição de crédito imobiliário. Para tal análise, utilizou-se como fonte de dados o conjunto a base de dados<sup>2</sup> disponível em uma competição do Kaggle denominada "HomeCreditRisk".

A figura 32 a seguir, oferece uma visão representativa da estrutura de dados desta fonte de dados, fornecendo uma referência visual útil para o entendimento da organização e da composição do conjunto de dados utilizado.

<sup>2</sup> Disponível em: < <https://www.kaggle.com/competitions/home-credit-default-risk> >

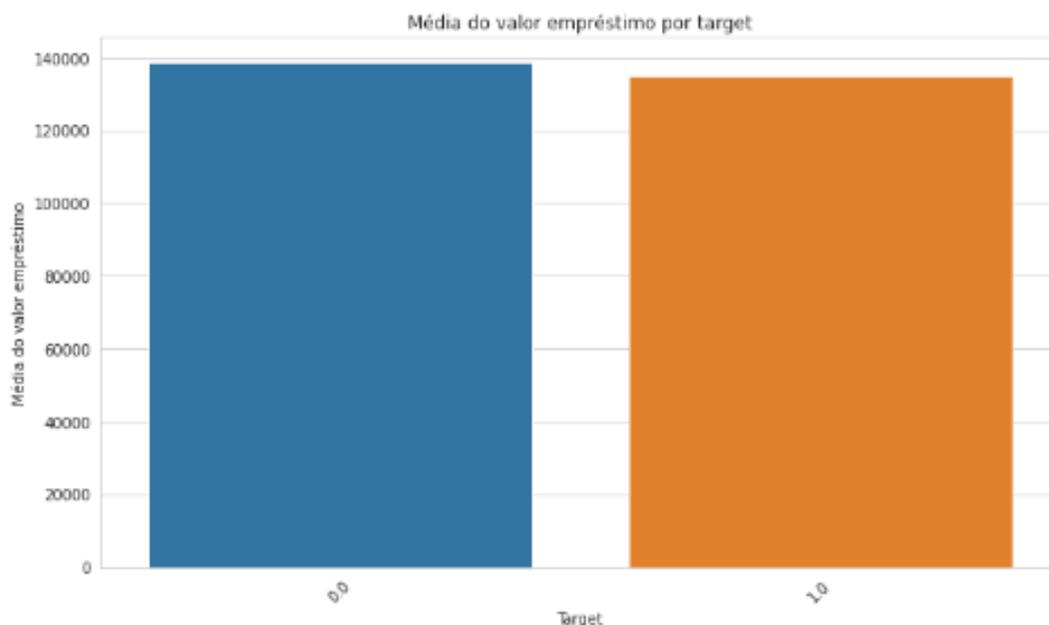
**Figura 32 - Estrutura da base de treinamento – Modelo de Risco Inadimplência**

```
pd_test.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Data columns (total 37 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   TARGET                                     307511 non-null  object
1   Tipo_de_Contrato                           307511 non-null  object
2   Genero                                       307511 non-null  object
3   Possui_Carro                               307511 non-null  object
4   Possui_Imovel_Proprio                     307511 non-null  object
5   Quantidade_de_Filhos                       307511 non-null  object
6   Renda_Total                                307511 non-null  float64
7   Prestacao_do_Emprestimo                   307499 non-null  float64
8   Preco_dos_Bens                             307233 non-null  float64
9   Tipo_de_Acompanhante                       306219 non-null  object
10  Tipo_de_Renda                               307511 non-null  object
11  Tipo_de_Educacao                           307511 non-null  object
12  Estado_Civil                               307511 non-null  object
13  Tipo_de_Habitacao                          307511 non-null  object
14  Dias_de_Idade                              307511 non-null  object
15  Dias_Empregado                             307511 non-null  object
16  Idade_do_Carro_Proprio                     104582 non-null  object
17  Possui_Email                               307511 non-null  object
18  Tipo_de_Ocupacao                           211120 non-null  object
19  Quantidade_de_Membros_da_Familia          307509 non-null  object
20  Tipo_de_Organizacao                        307511 non-null  object
21  Media_de_Apartamentos                     151450 non-null  object
22  Media_de_Ano_de_Construcao                 103023 non-null  object
23  Modo_de_Apartamentos                       151450 non-null  object
24  Modo_de_Area_do_Porao                      127568 non-null  object
25  Modo_de_Ano_de_Construcao                  103023 non-null  object
26  Modo_de_Area_Comum                         92646 non-null  object
27  Modo_de_Elevadores                         143620 non-null  object
28  Modo_de_Entradas                           152683 non-null  object
29  FLOORSMAX_MODE                             154491 non-null  object
30  FLOORSMIN_MODE                             98869 non-null  object
31  Modo_de_Area_do_Terreno                    124921 non-null  object
32  Modo_de_Apartamentos_para_Morar            97312 non-null  object
33  Modo_de_Area_de_Vida                       153161 non-null  object
34  Modo_de_Apartamentos_Nao-Residenciais      93997 non-null  object
35  Modo_de_Area_Nao-Residencial               137829 non-null  object
dtypes: float64(3), object(34)
memory usage: 86.8+ MB
```

Fonte: Elaboração própria

Este gráfico apresentado na figura 33 mostra a média do valor do empréstimo para cada categoria da variável alvo. A variável alvo indica se o cliente teve dificuldades de pagamento, ou seja, se ele teve atraso no pagamento por mais de X dias em pelo menos uma das primeiras Y parcelas do empréstimo, onde esse caso é representado como 1 e todos os outros casos como 0.

**Figura 33 - Média do valor empréstimo por target**

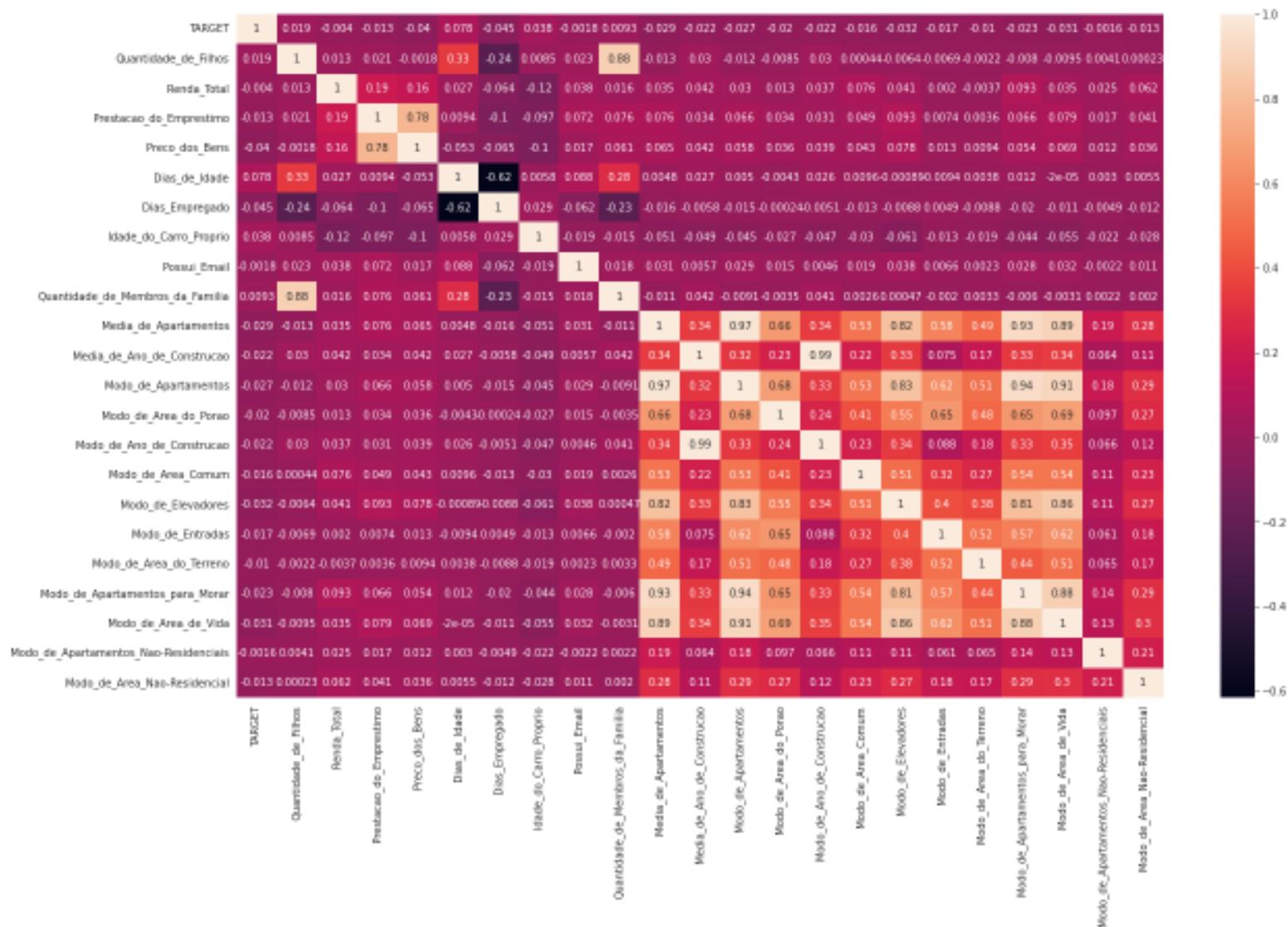
Fonte: Elaboração própria

No eixo x do gráfico, encontram-se representadas as duas categorias da variável alvo: 0 e 1. O valor 0 representa os casos em que o cliente não enfrentou dificuldades de pagamento, enquanto o valor 1 representa os casos em que houve atraso no pagamento.

No eixo y, é apresentada a média do valor do empréstimo para cada uma dessas categorias. O gráfico de barras oferece uma representação visual da variação da média do valor do empréstimo entre os clientes que enfrentaram dificuldades de pagamento (1) e aqueles que não enfrentaram (0).

A interpretação desse gráfico auxilia na compreensão de possíveis diferenças significativas no valor médio do empréstimo entre os dois grupos, fornecendo insights importantes para identificar padrões e compreender melhor a relação entre o valor do empréstimo e o risco de inadimplência.

Figura 34 - Correlação das Variáveis - Modelo de Risco de Inadimplência



Fonte: Elaboração própria

Foram examinadas as variáveis que apresentam alta correlação com a variável forte, conforme ilustrado na figura 34 e em seguida foi elaborado algumas hipóteses:

- Quantidade de Filhos: A quantidade de filhos tem uma correlação negativa moderada com a variável forte. Isso sugere que famílias com mais filhos podem ter uma tendência ligeiramente menor de cumprir com as obrigações financeiras. Famílias maiores podem enfrentar despesas adicionais com educação, saúde e alimentação, o que pode impactar sua capacidade de pagar empréstimos ou cumprir outras obrigações financeiras.
- Modo de Apartamentos: O modo de apartamentos tem uma correlação moderada com a variável forte. Isso pode indicar que o tipo de moradia pode ter uma influência significativa na capacidade de pagamento.
- Dias Empregado: A quantidade de dias empregado tem uma correlação negativa moderada com a variável forte. Isso sugere que a estabilidade no emprego pode ser um fator importante na capacidade de cumprir obrigações financeiras.

É interessante ressaltar que a variável renda total tem uma correlação negativa com a variável alvo. Comprovando uma intuição de que uma renda maior significa uma maior capacidade de cumprir obrigações financeiras.

Foi realizado o mesmo processo de clusterização utilizando a biblioteca k-means, por meio desta biblioteca foi possível criar grupos homogêneos para as variáveis numéricas. Na tabela 4, está definido a clusterização para cada atributo numérico.

**Tabela 6 - Tabela clusterização das variáveis de Risco de Inadimplência (continua)**

Cluster	Variável	Mínimo	Máximo
Cluster 1	Quantidade de Filhos	1	1
Cluster 2	Quantidade de Filhos	6	6
Cluster 3	Quantidade de Filhos	2	2
Cluster 4	Quantidade de Filhos	3	3
Cluster 5	Quantidade de Filhos	4	4
Cluster 6	Quantidade de Filhos	5	5
Cluster 1	Renda Total	0	0
Cluster 2	Renda Total	7	7
Cluster 3	Renda Total	3	3
Cluster 4	Renda Total	5	5

Tabela 4 - Tabela das variáveis de entrada (continuação)

Cluster 5	Renda Total	1	1
Cluster 6	Renda Total	4	4
Cluster 7	Renda Total	2	2
Cluster 8	Renda Total	6	6
Cluster 1	Prestação do Empréstimo	1	1
Cluster 2	Prestação do Empréstimo	4	4
Cluster 3	Prestação do Empréstimo	7	7
Cluster 4	Prestação do Empréstimo	3	3
Cluster 5	Prestação do Empréstimo	2	2
Cluster 6	Prestação do Empréstimo	0	0
Cluster 7	Prestação do Empréstimo	6	6
Cluster 8	Prestação do Empréstimo	5	5
Cluster 1	Preço dos Bens	0	0
Cluster 2	Preço dos Bens	5	5
Cluster 3	Preço dos Bens	3	3
Cluster 4	Preço dos Bens	7	7
Cluster 5	Preço dos Bens	2	2
Cluster 6	Preço dos Bens	1	1
Cluster 7	Preço dos Bens	6	6
Cluster 8	Preço dos Bens	4	4
Cluster 1	Dias de Idade	7	7
Cluster 2	Dias de Idade	2	2
Cluster 3	Dias de Idade	5	5
Cluster 4	Dias de Idade	0	0
Cluster 5	Dias de Idade	6	6
Cluster 6	Dias de Idade	1	1
Cluster 7	Dias de Idade	3	3
Cluster 8	Dias de Idade	4	4
Cluster 1	Dias Empregado	7	7
Cluster 2	Dias Empregado	2	2
Cluster 3	Dias Empregado	4	4
Cluster 4	Dias Empregado	0	0
Cluster 5	Dias Empregado	6	6
Cluster 6	Dias Empregado	1	1
Cluster 7	Dias Empregado	3	3
Cluster 8	Dias Empregado	5	5
Cluster 1	Idade do Carro Próprio	6	6
Cluster 2	Idade do Carro Próprio	2	2
Cluster 3	Idade do Carro Próprio	0	0
Cluster 4	Idade do Carro Próprio	5	5
Cluster 5	Idade do Carro Próprio	3	3
Cluster 6	Idade do Carro Próprio	4	4

**Tabela 4 - Tabela das variáveis de entrada (conclusão)**

Cluster 7	Idade do Carro Próprio	1	1
Cluster 8	Idade do Carro Próprio	7	7
Cluster 1	Média de Apartamentos	6	6
Cluster 2	Média de Apartamentos	1	1
Cluster 3	Média de Apartamentos	4	4
Cluster 4	Média de Apartamentos	0	0
Cluster 5	Média de Apartamentos	5	5
Cluster 6	Média de Apartamentos	7	7
Cluster 7	Média de Apartamentos	2	2
Cluster 8	Média de Apartamentos	3	3
Cluster 1	Quantidade de Membros da Família	2	2
Cluster 3	Quantidade de Membros da Família	1	1
Cluster 4	Quantidade de Membros da Família	3	3
Cluster 5	Quantidade de Membros da Família	5	5
Cluster 6	Quantidade de Membros da Família	4	4
Cluster 7	Quantidade de Membros da Família	6	6
Cluster 8	Quantidade de Membros da Família	0	0
Cluster 1	Ano de Construção	7	7
Cluster 2	Ano de Construção	1	1
Cluster 3	Ano de Construção	3	3
Cluster 4	Ano de Construção	0	0
Cluster 5	Ano de Construção	6	6
Cluster 6	Ano de Construção	2	2
Cluster 7	Ano de Construção	5	5
Cluster 8	Ano de Construção	4	4

Fonte: Elaboração Própria

Para o treinamento do modelo foi utilizado a configuração padrão da função `train_test_split` da biblioteca `sklearn`, isto é, a divisão da base de dados foi de 70% (18.969 registros) para treino e 30% (8130 registros) para teste.

#### 4.7 Modelo de Inteligência artificial - ChatGPT

Este *software* incorpora um modelo de inteligência artificial fundamentado no ChatGPT, que atua como uma camada adicional para explicar as decisões tomadas.

A API do ChatGPT é uma interface de programação de aplicativos que permite que desenvolvedores integrem os recursos de geração de texto do modelo de linguagem GPT (*Generative Pre-trained Transformer*) em seus próprios aplicativos, plataformas ou sistemas (OpenAI, 2024).

Após os modelos clássicos de classificação definirem um resultado, uma requisição é feita à API para análise pelo ChatGPT.

Para aumentar a assertividade das respostas, é crucial fornecer um contexto adequado no prompt de entrada e também especificar a versão específica do ChatGPT a ser utilizada. Essa configuração é representada na figura 35.

**Figura 35 - Contextualização do Prompt do ChatGPT**

```
completion = openai.ChatCompletion.create(  
    model="gpt-3.5-turbo",  
    messages=[  
        {"role": "user", "content":  
            """Como analista de política de crédito, minha responsabilidade é avaliar cuidadosamente cada  
            solicitação de financiamento imobiliário para garantir que esteja em conformidade com as  
            diretrizes da instituição financeira e que represente um investimento seguro para ambas as  
            partes envolvidas. Recentemente, fui incumbido de analisar a solicitação de um cliente em  
            potencial para um financiamento imobiliário."""},  
        {"role": "user", "content": f"{str(query)}"}  
    ]  
)
```

Fonte: Elaboração própria

## 4.8 Notificação

Por fim, neste item é abordado o produto escolhido para o envio das decisões e insights que o *software* processou para o respectivo cliente, este envio é feito via e-mail.

### 4.8.1 Simple E-mail Service - SES

Nesse contexto, o Amazon Simple Email Service (SES) da Amazon Web Services (2024) emergiu como uma solução poderosa para o envio e gerenciamento de e-mails em escala. Neste contexto, o foco recai sobre a utilização do SES como parte integrante do *software* desenvolvido, que visa gerenciar o envio de e-mails contendo respostas de análises para os usuários.

O SES é um serviço de e-mail altamente escalável que permite às organizações enviar e receber e-mails de forma confiável, eficiente e segura. Suas características abrangentes incluem a capacidade de enviar e-mails em massa, rastrear entregas, gerenciar lista de destinatários e implementar verificações de autenticidade de domínio e assinaturas digitais. Além disso, o SES oferece integração com outros

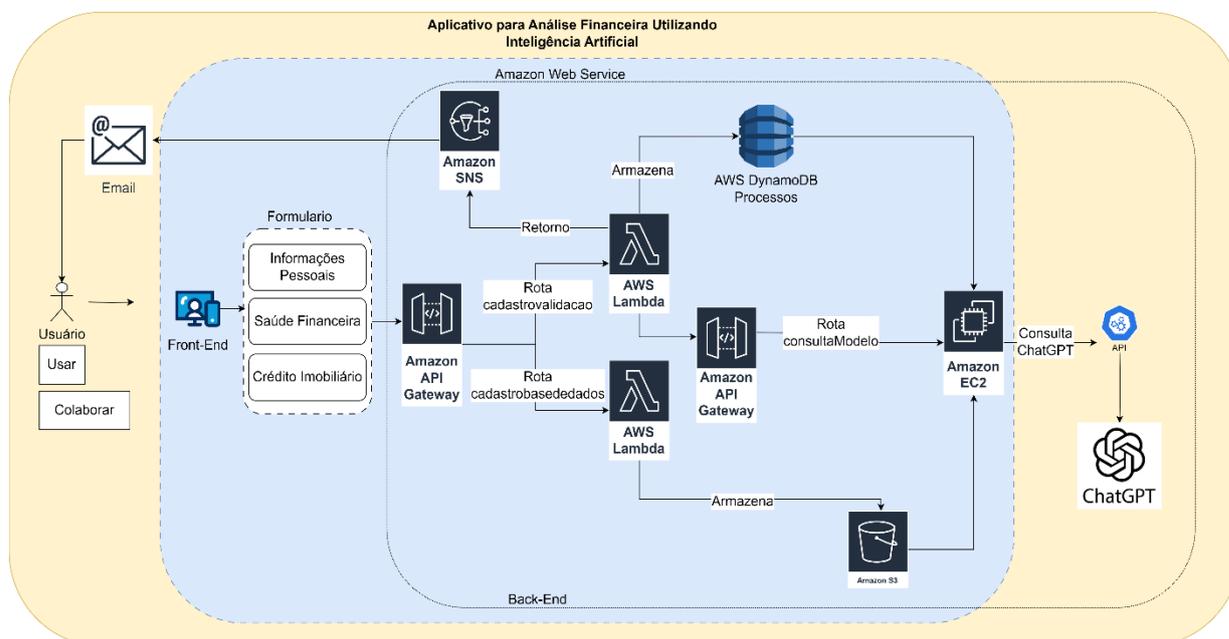
serviços AWS, como o AWS Lambda, para automatizar fluxos de trabalho complexos. (Amazon Web Service, 2023)

#### 4.9 Arquitetura de Solução

A adoção de soluções baseadas em nuvem é um dos requisitos deste *software*. Este texto aborda a arquitetura de solução desenvolvida para o *software*, utilizando serviços da AWS para garantir eficiência, escalabilidade e baixo custo operacional.

A arquitetura é projetada para atender às necessidades de dois tipos de usuários: os que solicitam análises de perfil e os que contribuem com dados para o aprimoramento dos modelos de inteligência artificial. Ao longo deste capítulo, é explorado os principais componentes da arquitetura, incluindo a integração entre os serviços AWS Lambda, AWS EC2, AWS DynamoDB e AWS S3. Cada componente desempenha um papel específico no processamento, armazenamento e análise dos dados, garantindo uma experiência fluida e confiável para os usuários.

**Figura 36 - Arquitetura de Solução**



Fonte: Elaboração própria

Por meio dessa arquitetura ilustrada na figura 36, é possível oferecer uma solução robusta e escalável, capaz de lidar com grandes volumes de dados e fornecer insights valiosos em tempo real. Combinando a flexibilidade dos serviços *serverless* da AWS com a potência dos recursos computacionais sob demanda, esta arquitetura

representa uma abordagem moderna e eficaz para o desenvolvimento de aplicativos na nuvem.

Agora, será explorado mais detalhadamente cada componente da arquitetura e como eles se integram para fornecer uma solução completa e eficiente.

A AWS Lambda emerge como a camada de processamento de eventos. Suas funções são responsáveis por receber e processar requisições da API, normalizando os dados recebidos e direcionando-os para consulta aos modelos de inteligência artificial. A grande vantagem da Lambda é sua natureza *serverless*, que proporciona escalabilidade automática e cobrança conforme o uso, otimizando custos e garantindo flexibilidade na implementação das funcionalidades necessárias.

Por sua vez, o AWS EC2 entra em cena para executar os modelos de *machine learning*. A alocação de instâncias EC2 otimizadas para memória assegura tempos de resposta rápidos e eficientes no processamento dos dados. Além disso, o EC2 é encarregado do retreinamento periódico do modelo, garantindo sua atualização e capacidade de adaptação a novos dados e cenários.

Enquanto isso, o AWS DynamoDB assume o papel de banco de dados principal do sistema. Sua arquitetura altamente escalável e gerenciada pela AWS garante alta disponibilidade e desempenho consistente, inclusive em momentos de carga elevada.

Além disso, o AWS S3 desempenha o papel de armazenamento de objetos, abrigando tanto os dados dos usuários quanto os resultados das análises. Sua escalabilidade excepcional e integração fluida com outros serviços da AWS facilitam o processamento e armazenamento eficientes dos dados, garantindo sua disponibilidade e proteção contra perdas.

Por fim, o *software* faz uso do serviço AWS SES para enviar *e-mail* aos usuários, informando as respostas das análises dos modelos. Ademais, há a chamada para o serviço ChatGPT, que permite uma interação mais personalizada e inteligente com os usuários, agregando valor à experiência geral do usuário.

#### 4.10 Modelo de Score de Crédito

Neste estudo, é explorada a aplicação de diferentes técnicas de aprendizado de máquina para prever o *score* de crédito do usuário. O modelo<sup>3</sup> foi treinado e testado utilizando duas abordagens distintas de árvores de decisão: o algoritmo XGBoost e a implementação clássica da árvore de decisão (*Decision Tree*) do scikit-learn.

As variáveis-alvo foram definidas como 0, 1 e 2, correspondendo a Ruim, Padrão e Bom, respectivamente. Ambas as técnicas foram testadas ao longo do desenvolvimento do software, como ilustrado na Figura 37.

Figura 37 - Métricas *Decision Tree* vs XGBoost – Score de Crédito

Decision Tree					XGBoost				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.94	0.95	0.95	3444	0	0.92	0.89	0.90	3455
1	0.89	0.88	0.88	4152	1	0.79	0.86	0.82	4143
2	0.83	0.83	0.83	1960	2	0.80	0.70	0.74	1958
accuracy			0.90	9556	accuracy			0.84	9556
macro avg	0.89	0.89	0.89	9556	macro avg	0.84	0.82	0.82	9556
weighted avg	0.90	0.90	0.90	9556	weighted avg	0.84	0.84	0.84	9556

Fonte: Elaboração própria

A comparação entre os resultados dos modelos *Decision Tree* e XGBoost revela diferenças significativas em suas performances de classificação. Inicialmente, ao analisar a métrica de acurácia, observa-se que o modelo *Decision Tree* alcançou uma acurácia de aproximadamente 90%, enquanto o XGBoost obteve uma acurácia ligeiramente inferior, em torno de 84%.

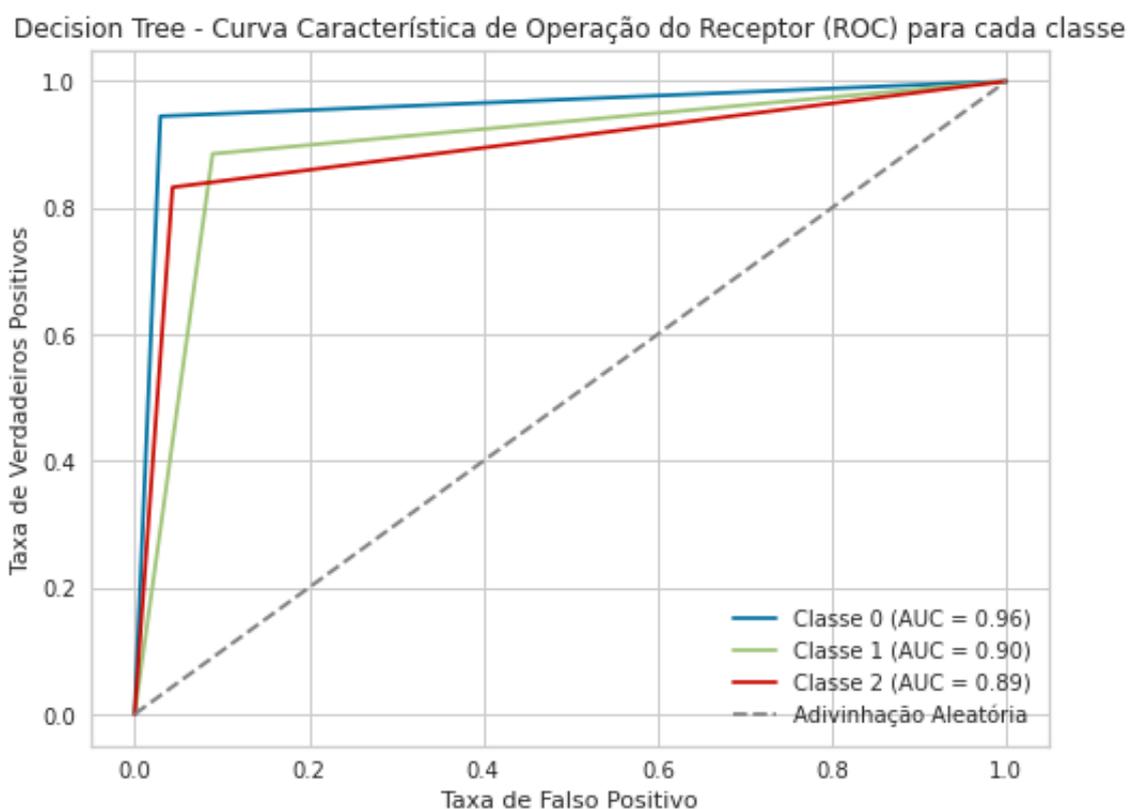
Ao examinar as métricas de *precision*, *recall* e *f1-score* para cada classe, percebe-se algumas distinções importantes. O modelo *Decision Tree* apresentou resultados ligeiramente melhores em termos de *precision*, *recall* e *f1-score* para todas as classes em comparação com o XGBoost. Especificamente, para a classe 1, o modelo *Decision Tree* alcançou uma *precision* de 89%, *recall* de 88% e *f1-score* de 88%, enquanto o XGBoost obteve uma *precision* de 79%, *recall* de 86% e *f1-score* de 82%.

<sup>3</sup> Disponível em: < <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bfcf/3329414618340995/412882276756469/4384925233469574/latest.html> >

No entanto, é crucial considerar o *trade-off* entre as métricas de *precision*, *recall* e *f1-score*, bem como a acurácia geral do modelo. Embora o *Decision Tree* tenha apresentado resultados ligeiramente melhores em algumas métricas, o XGBoost pode ser preferível em cenários onde a acurácia geral é prioritária.

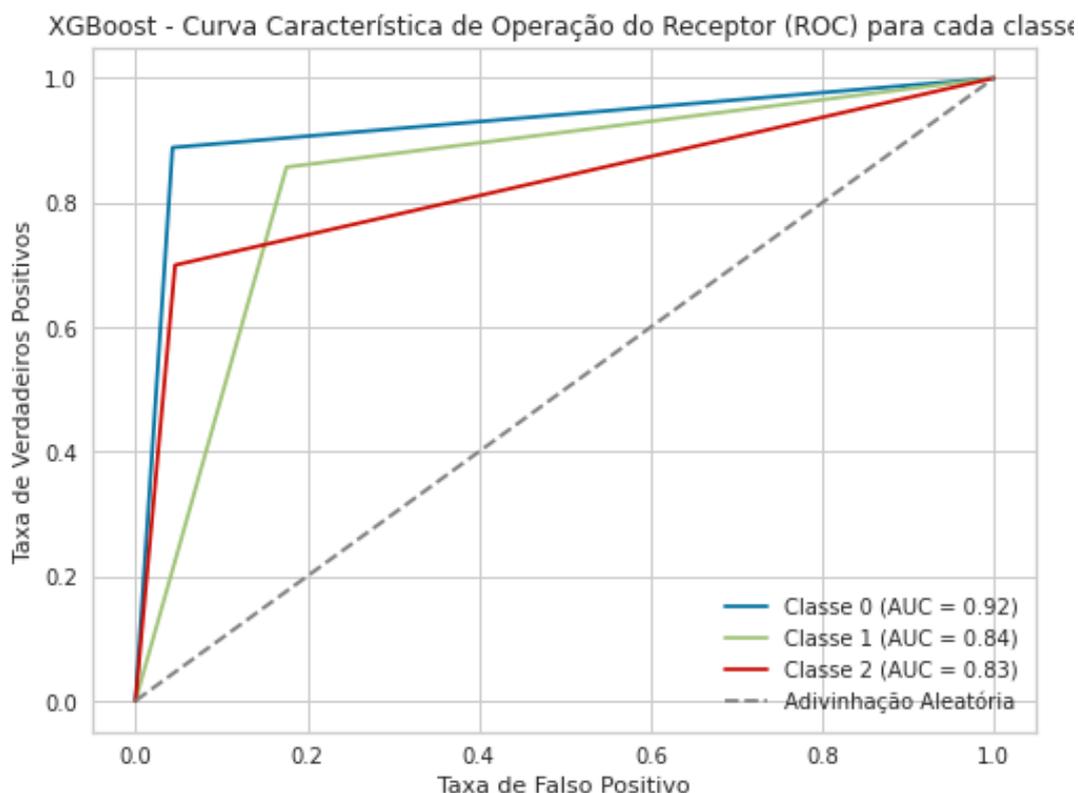
Para confirmar a escolha do modelo foi utilizada a comparação da curva ROC AUC de cada modelo, conforme figura 38 e 39.

**Figura 38 - Curva ROC AUC *Decision Tree* - Score de Crédito**



Fonte: Elaboração própria

**Figura 39 - Curva ROC AUC XGBoost - Score de Crédito**

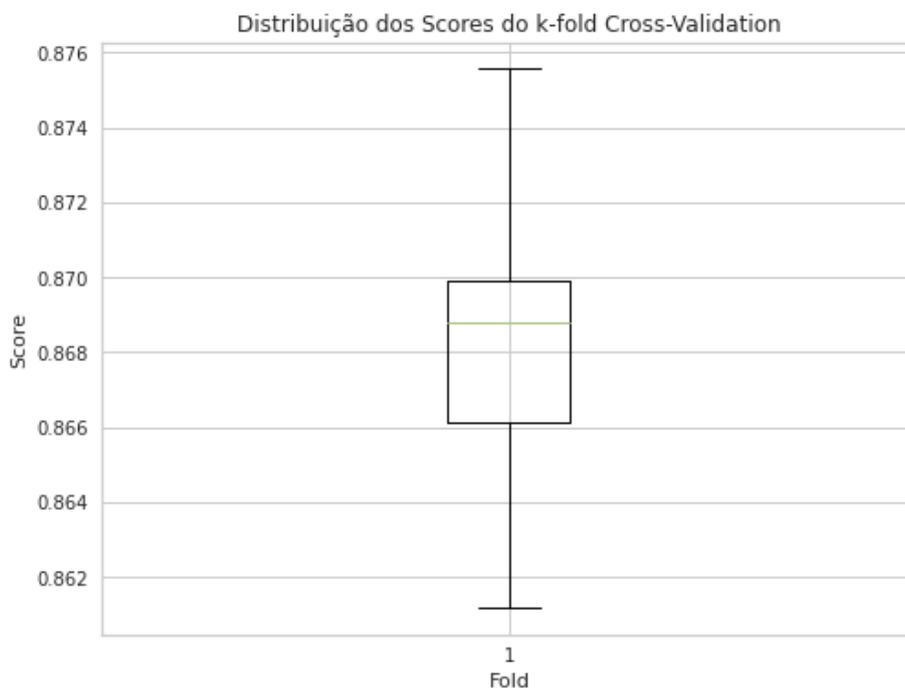


Fonte: Elaboração própria

Considerando esses resultados, para este modelo o *Decision Tree* demonstrou um desempenho geral melhor em comparação com a *Xboost*, especialmente na classe majoritária.

No cenário em questão, onde se lida com um problema de aprendizado de máquina supervisionado, é crucial evitar o *overfitting* e garantir que o modelo seja capaz de generalizar bem para dados não vistos. (Scikit-learn, 2024).

A partir da validação cruzada k-fold que é uma técnica que divide o conjunto de dados em k partes iguais e realiza k iterações de treinamento e teste. Em cada iteração, um dos k subconjuntos é retido como conjunto de teste, enquanto os outros k-1 subconjuntos são usados como conjunto de treinamento. Esse processo é repetido k vezes, garantindo que cada subconjunto seja usado como conjunto de teste uma vez (Scikit-learn, 2024).

**Figura 40 - K-Fold Decision Tree - Score de Crédito**

Fonte: Elaboração própria

Os resultados do k-fold *cross-validation* revelam uma consistência razoável no desempenho do modelo de classificação em diferentes divisões dos dados. Os *scores* de acurácia obtidos em cada *fold* individual variam entre 86.61% e 87.56%, com uma acurácia média de aproximadamente 86.83%.

Isso sugere que o modelo mantém um desempenho relativamente estável em diferentes conjuntos de dados de teste, o que é uma indicação positiva de sua capacidade de generalização.

Ao analisar o peso das variáveis na figura 41 durante o processo de tomada de decisão do modelo, torna-se evidente que cinco variáveis exercem uma influência significativa sobre a variável alvo. As variáveis, em ordem de maior impacto para menor, são: "divida\_pendente", "tipo\_de\_empréstimo", "salário\_mensal", "idade" e "ocupação".

Essa descoberta ressalta a importância desses fatores na predição ou determinação da variável de interesse e fornece *insights* para a compreensão dos mecanismos subjacentes ao modelo.

**Figura 41 - Peso das variáveis no Modelo Score de Crédito**

Weight	Feature
0.4105 ± 0.0131	Divida_pendente
0.2656 ± 0.0051	tipo_de_emprestimo
0.2000 ± 0.0073	salario_mensal
0.1628 ± 0.0042	idade
0.1449 ± 0.0060	ocupacao
0.0369 ± 0.0026	Saldo_Mensal
0.0080 ± 0.0024	Valor_investido_mensalmente
0.0036 ± 0.0018	Pagamento_Comportamento
0.0001 ± 0.0005	num_de_emprestimo
0 ± 0.0000	renda_anual
-0.0001 ± 0.0002	Emprestimo_Valor_Mes
-0.0004 ± 0.0005	taxa_juros

Fonte: Elaboração própria

#### 4.11 Modelo de Risco de Inadimplência

Neste estudo, foi explorada a aplicação de diferentes técnicas de aprendizado de máquina para prever o risco de inadimplência do usuário. O modelo<sup>4</sup> foi treinado e testado utilizando duas abordagens distintas de árvores de decisão: o algoritmo XGBoost e a implementação clássica da árvore de decisão (*Decision Tree*) do scikit-learn.

As variáveis-alvo foram definidas como 0 e 1, correspondendo a Sem Risco e Com Risco, respectivamente. Durante o desenvolvimento do *software*, ambas as técnicas foram testadas, como evidenciado na figura 42.

**Figura 42 - Métricas *Decision Tree* vs XGBoost – Risco de Inadimplência**

Decision Tree					XGBoost				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0.0	0.91	0.94	0.93	7425	0.0	0.94	1.00	0.97	7612
1.0	0.10	0.07	0.08	705	1.0	0.18	0.00	0.01	518
accuracy			0.86	8130	accuracy			0.94	8130
macro avg	0.51	0.50	0.50	8130	macro avg	0.56	0.50	0.49	8130
weighted avg	0.84	0.86	0.85	8130	weighted avg	0.89	0.94	0.91	8130

Fonte: Elaboração própria

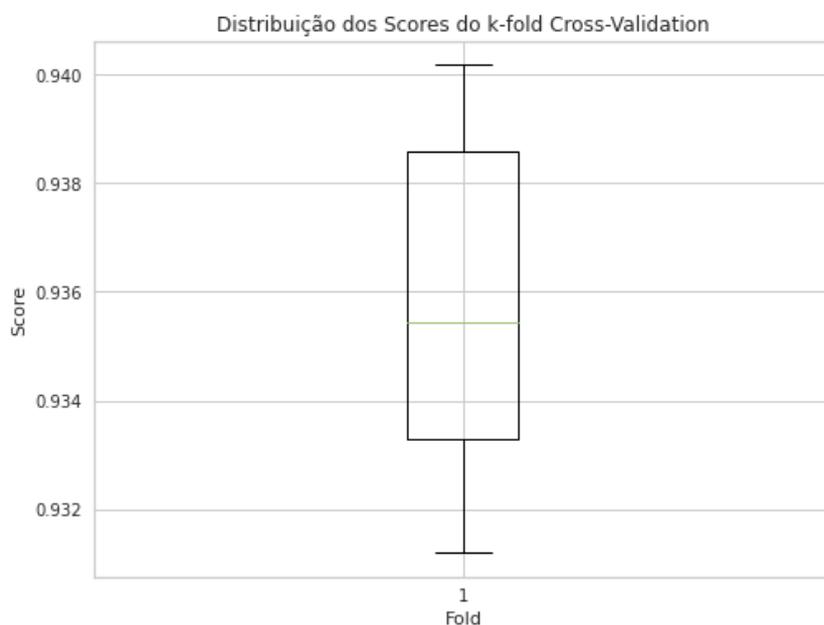
<sup>4</sup> Disponível em: < <https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bfcf/3329414618340995/412882276756472/4384925233469574/latest.html> >

Ao comparar os resultados de dois modelos de classificação, XGBoost e *Decision Tree*, é evidente que o modelo XGBoost supera a *Decision Tree* em termos de desempenho. Enquanto o XGBoost alcança uma precisão média ponderada de 0.91, a *Decision Tree* atinge 0.85.

Especificamente, o XGBoost demonstra uma precisão muito maior na previsão da classe 0.0, atingindo 0.94 em comparação com 0.91 da *Decision Tree*. Embora o *recall* e a pontuação F1 da classe 1.0 sejam baixos em ambos os modelos, o XGBoost ainda supera a *Decision Tree*, alcançando 0.18 e 0.01, respectivamente, em comparação com 0.10 e 0.08.

Assim, com base nessas métricas de desempenho, o modelo XGBoost é considerado superior devido à sua capacidade de prever com mais precisão ambas as classes, especialmente a classe minoritária. No mesmo cenário de evitar o *overfitting* do modelo, foi realizado o teste k-fold, em que os resultados são então médios para produzir uma estimativa final de desempenho do modelo.

**Figura 43 - K-Fold XGBoost - Risco de Inadimplência**

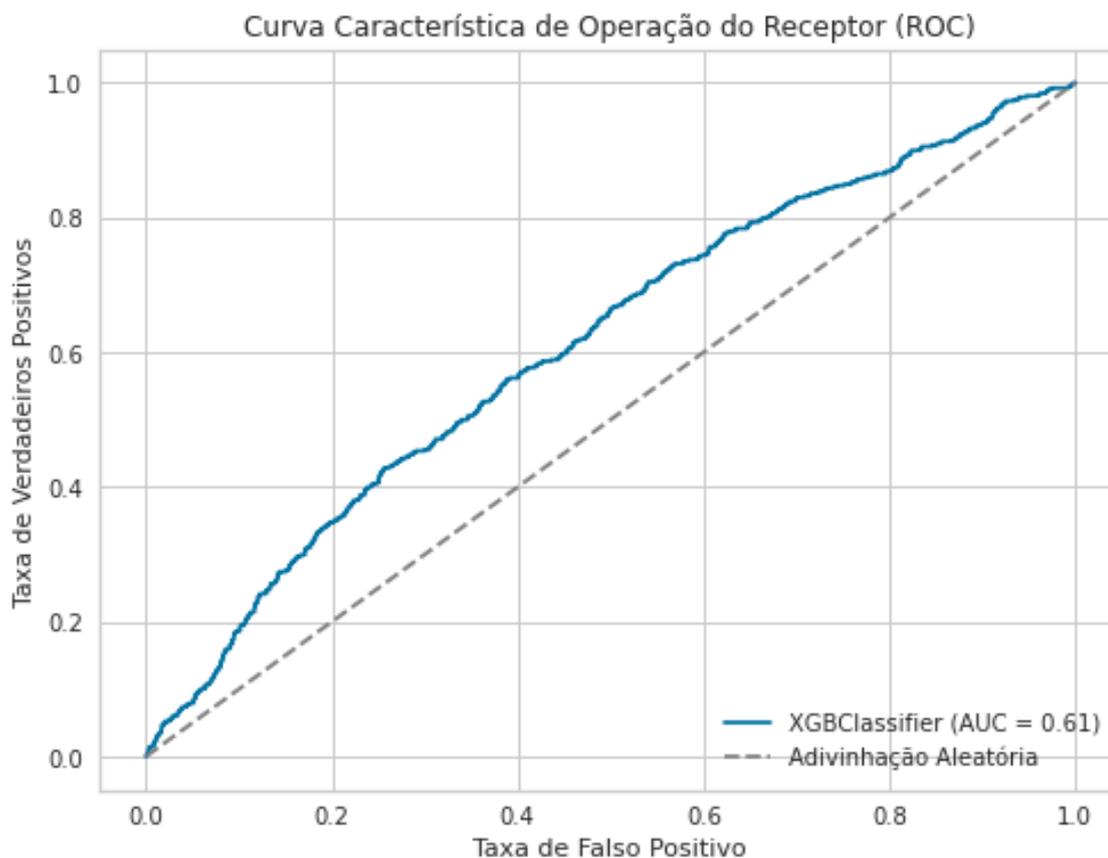


Fonte: Elaboração própria

Com base na análise de figura 43, é possível determinar que os *scores* de cada *fold* variam ligeiramente em torno da acurácia média, indicando consistência na performance do modelo em diferentes divisões dos dados. A acurácia média de cerca de 93.57% sugere que o modelo tem um bom desempenho na tarefa de classificação.

Nesse contexto ilustrado na figura 44, uma pontuação de 0.6117 sugere que o modelo possui uma habilidade moderada de classificação, mostrando uma taxa de verdadeiros positivos maior do que a taxa de falsos positivos, porém com espaço para melhorias em sua precisão e robustez na identificação das classes.

**Figura 44 - Curva ROC AUC XGBoost - Risco de Inadimplência**



Fonte: Elaboração própria

A pontuação ROC AUC de 0.6117 indica a capacidade do modelo em distinguir entre as classes positiva e negativa, onde um valor de 0.5 representa um desempenho aleatório e 1 representa um desempenho perfeito.

Ao analisar o peso das variáveis na figura 45 durante o processo de tomada de decisão do modelo, torna-se evidente que duas variáveis exercem uma influência significativa sobre a variável alvo. As variáveis, em ordem de maior impacto para menor, são: "Preco\_dos\_bens ", "Dias\_Empregado".

Essa revelação destaca a significância desses elementos na antecipação ou determinação da variável de interesse, fornecendo uma perspectiva sobre os processos fundamentais subjacentes ao modelo.

**Figura 45 - Peso das variáveis no Modelo Risco de Inadimplência**

<b>Weight</b>	<b>Feature</b>
0.0890	Preco_dos_Bens
0.0764	Dias_Empregado
0.0548	Prestacao_do_Emprestimo
0.0545	Modo_de_Elevadores
0.0543	Modo_de_Apartamentos_Nao-Residenciais
0.0471	Modo_de_Apartamentos
0.0465	Idade_do_Carro_Proprio
0.0458	Renda_Total
0.0445	Quantidade_de_Membros_da_Familia
0.0444	Modo_de_Ano_de_Construcao
0.0441	Possui_Email
0.0437	Modo_de_Area_do_Terreno
0.0391	Modo_de_Area_de_Vida
0.0390	Dias_de_Idade
0.0379	Modo_de_Apartamentos_para_Morar
0.0378	Modo_de_Entradas
0.0377	Media_de_Apartamentos
0.0372	Modo_de_Area_do_Porao
0.0347	Quantidade_de_Filhos

Fonte: Elaboração própria

## 5 RESULTADOS

Neste capítulo serão apresentados e analisados os resultados do *software* desenvolvido a fim de observar o seu comportamento frente a alguns testes de usuário.

Dois modelos foram desenvolvidos em linguagem de programação Python e baseados em algoritmos de aprendizado de máquina denominado “árvore de decisão”, sendo o primeiro modelo para análise do *score* de crédito do usuário, o segundo modelo para análise de risco de o cliente tornar-se inadimplente.

### 5.1 Simulação Positiva

Foi estabelecido um teste hipotético para demonstrar o funcionamento *end-to-end* do *software*. O teste positivo é iniciado com o preenchimento do formulário do *front-end* com os seguintes dados, conforme apresentado na tabela 5.

**Tabela 7 - Variáveis de entrada para simulação (continua)**

Variável de Entrada	Valor de Entrada
Profissão	Engenheiro de Dados
Renda Líquida	R\$ 13000
FGTS	R\$ 35000
Gasto Educação Mês	R\$ 1450
Gasto Imobiliário Mês	R\$ 1250
Gasto Saúde Mês	R\$ 600
Gasto Cartão Mês 1	R\$ 3000
Gasto Cartão Mês 2	R\$ 3100
Gasto Cartão Mês 3	R\$ 3200
Gasto Cartão Mês 4	R\$ 3000
Gasto Cartão Mês 6	R\$ 3000
Quantidade Filho	0
Reserva Renda Fixa	R\$ 50000
Reserva Renda Variável	R\$ 40000
Reserva Emergência	Sim
Utilizou Reserva 12 meses	Não
Limite Cartão	R\$ 20000
Comprometimento Cartão	R\$ 3000
Escolaridade	Nível Superior
Beneficiária INSS	Não
Consignado Valor	R\$ 0
Empréstimo Valor	R\$ 0

**Tabela 7 - Variáveis de entrada para simulação (conclusão)**

Produto	Financiamento Imobiliário
Sexo	Masculino
Possui Carro Próprio	Sim
Possui Casa Própria	Não
Estado civil	Casado
Mora com os pais	Sim
Data Nascimento	14-01-1996
CEP Moradia Atual	00000000
Idade Carro Cliente	5
Tipo Organização	Grande
Valor Imóvel	R\$ 415000
m2_imovel	59
Idade Imóvel	11
Quartos	2
Vagas	1
CEP Imóvel Financiado	0000000
Area Lazer	Sim
Elevador	Sim
Necessita Empréstimo ITBI	Sim
Indexador	IPCA
Tipo Imóvel	Apartamento
Estado Documentação Imóvel	Apto
Quantos Andares	19
Objetivo imóvel	Morar
Taxa Financiamento	7%
Tempo Emprego	5
Valor Investido Mensalmente	0
Empréstimo Valor Mês	R\$ 3000
Tipo Amortização	SAC
Mix de Crédito	850

Fonte: Elaboração própria

Para a normalização para o modelo de score de crédito, é dividido em dois processos em duas etapas: pré-crédito e pós-crédito, conforme detalhado no desenvolvimento.

A abordagem comportamental, pré e pós contratação do crédito, permite uma análise mais abrangente do histórico financeiro e do comportamento do cliente em diferentes estágios, proporcionando uma análise mais detalhada e dinâmica do perfil financeiro do cliente, permitindo uma avaliação mais precisa de sua capacidade de pagamento e de seu comportamento financeiro ao longo do tempo.

As figuras 46 e 47 representam visualmente esses dois momentos distintos de avaliação do cliente.

Na etapa de pré-crédito, são considerados os dados e informações disponíveis antes da concessão do crédito.

**Figura 46 - Normalização Pré-Crédito - Score de Crédito**

```
Out[47]: {'idade': 28,
'ocupacao': 'DESENVOLVEDOR',
'renda_anual': 169000.0,
'salario_mensal': 13000.0,
'taxa_juros': 7.0,
'num_de_emprestimo': 0,
'tipo_de_emprestimo': 'Emprestimo_Hipotecario',
'Divida_pendente': 0.0,
'Emprestimo_Valor_Mes': 0,
'Valor_investido_mensalmente': 200,
'Pagamento_Comportamento': 'Gastos_Baixos_Pagamentos_de_Grande_Valor',
'Saldo_Mensal': 6916.666666666666}
```

Fonte: Elaboração própria

Por outro lado, na etapa de pós-crédito, são consideradas as informações coletadas após a concessão do crédito e durante o período de contrato.

**Figura 47 - Normalização Pós-Crédito - Score de Crédito**

```
Out[107]: {'idade': 28,
'ocupacao': 'DESENVOLVEDOR',
'renda_anual': 169000.0,
'salario_mensal': 13000.0,
'taxa_juros': 7.0,
'num_de_emprestimo': 1,
'tipo_de_emprestimo': 'Emprestimo_Hipotecario',
'Divida_pendente': 0.0,
'Emprestimo_Valor_Mes': 3000,
'Valor_investido_mensalmente': 200,
'Pagamento_Comportamento': 'Gastos_Baixos_Pagamentos_de_Grande_Valor',
'Saldo_Mensal': 6916.666666666666}
```

Fonte: Elaboração própria

Para compreender a lógica por trás da decisão tomada neste contexto específico, é crucial analisar a figura 48 e a contribuição relativa de cada variável envolvida no processo de tomada de decisão.

**Figura 48 - Importância das *Feature* na Simulação - Score de Crédito**

Pré-Crédito

y=Ruim (probability 0.000) top features		y=Padrao (probability 0.000) top features		y=Bom (probability 1.000) top features	
Contribution?	Feature	Contribution?	Feature	Contribution?	Feature
+0.361	<BIAS>	+0.428	<BIAS>	+0.273	tipo_de_emprestimo
-0.361	Divida_pendente	+0.135	Divida_pendente	+0.225	Divida_pendente
		+0.051	idade	+0.211	<BIAS>
		-0.162	ocupacao	+0.178	salario_mensal
		-0.178	salario_mensal	+0.162	ocupacao
		-0.273	tipo_de_emprestimo	-0.051	idade

Pos-Crédito

y=Ruim (probability 0.000) top features		y=Padrao (probability 0.000) top features		y=Bom (probability 1.000) top features	
Contribution?	Feature	Contribution?	Feature	Contribution?	Feature
+0.361	<BIAS>	+0.428	<BIAS>	+0.273	tipo_de_emprestimo
-0.361	Divida_pendente	+0.135	Divida_pendente	+0.225	Divida_pendente
		+0.051	idade	+0.211	<BIAS>
		-0.162	ocupacao	+0.178	salario_mensal
		-0.178	salario_mensal	+0.162	ocupacao
		-0.273	tipo_de_emprestimo	-0.051	idade

Fonte: Elaboração própria

Ao examinar a importância atribuída a cada uma das variáveis nesta simulação, é possível identificar quais aspectos exerceram maior influência no desfecho final. Os valores listados representam a proporção de contribuição de cada variável para a decisão final.

Por exemplo, a "dívida pendente" e o "tipo de empréstimo" emergem como os fatores mais significativos. Isso sugere que esses dois elementos desempenharam um papel preponderante na determinação do resultado final da decisão, enquanto outras variáveis tiveram contribuições menos proeminentes.

Essa análise fornece uma visão mais clara sobre os principais impulsionadores por trás da decisão e ajuda a compreender melhor o raciocínio por trás das respostas obtidas.

Para o modelo de risco de inadimplência a normalização de dados se resume em uma etapa, ilustrada na figura 49.

**Figura 49 - Normalização - Risco de Inadimplência**

```
Out[108]: {'Quantidade_de_Filhos': '1',
'Renda_Total': '13000',
'Prestacao_do_Emprestimo': '317000',
'Preco_dos_Bens': 94000.0,
'Dias_de_Idade': 10340,
'Dias_Empregado': 1825,
'Idade_do_Carro_Proprio': 1460,
'Possui_Email': 1,
'Quantidade_de_Membros_da_Familia': 3,
'Media_de_Apartamentos': 59,
'Media_de_Ano_de_Construcao': 2013,
'Modo_de_Apartamentos': 1,
'Modo_de_Area_do_Porao': 1,
'Modo_de_Ano_de_Construcao': 1,
'Modo_de_Area_Comum': 1,
'Modo_de_Elevadores': 0,
'Modo_de_Entradas': 2,
'Modo_de_Area_do_Terreno': 1,
'Modo_de_Apartamentos_para_Morar': 1,
'Modo_de_Area_de_Vida': 1,
'Modo_de_Apartamentos_Nao-Residenciais': 1,
'Modo_de_Area_Nao-Residencial': 1}
```

Fonte: Elaboração própria

Ao observar a figura 50, é possível inferir que a capacidade financeira do cliente desempenha um papel crucial na decisão final. Essa observação sugere que clientes com uma renda mais alta podem ter uma probabilidade maior de obter aprovação para determinadas solicitações de crédito.

Essa explicação ilustrada na figura 50 reforça a contribuição de cada característica para a previsão feita pelo modelo. O valor positivo associado a cada característica indica que a presença dessa característica aumenta a probabilidade da classe prevista, enquanto valores negativos indicam uma diminuição na probabilidade da classe prevista.

**Figura 50 - Importância das *Feature* na Simulação - Risco de Inadimplência**

y=SemRisco (probability 0.937, score -0.957) top features

Contribution?	Feature
+0.768	<BIAS>
+0.221	Modo_de_Apartamentos_Nao-Residenciais
+0.106	Media_de_Apartamentos
+0.058	Modo_de_Area_Nao-Residencial
+0.050	Modo_de_Apartamentos_para_Morar
+0.030	Dias_Empregado
	... 5 more positive ...
	... 4 more negative ...
-0.040	Preco_dos_Bens
-0.070	Modo_de_Area_do_Terreno
-0.090	Idade_do_Carro_Proprio
-0.091	Modo_de_Area_Comum

Fonte: Elaboração própria

O viés, representado como <BIAS>, é uma constante adicionada à previsão independentemente das características específicas do exemplo. Algumas características, como "Modo\_de\_Apartamentos\_Nao-Residenciais" e "Media\_de\_Apartamentos", contribuem positivamente para a previsão, indicando que sua presença está associada a uma probabilidade maior da classe prevista.

Por outro lado, características como "Preco\_dos\_Bens" e "Idade\_do\_Carro\_Proprio" contribuem negativamente para a previsão desta simulação, sugerindo que sua presença está associada a uma probabilidade menor da classe prevista.

O *software* analisou e sugeriu conforme figura 51 que, com base nos parâmetros de entrada (tabela 5), este cliente deve seguir com o financiamento pagando a entrada mínima que a instituição prestadora do financiamento sugere e sem riscos futuros, uma vez que os modelos apresentaram os seguintes resultados:

- Análise negativa para o risco de a entrada utilizada no financiamento afetar a saúde financeira ao decorrer do contrato, isto é, o *score* de crédito não se altera.
- Análise negativa para o risco de o cliente tornar-se inadimplente.

Figura 51 - *E-mail* de resposta do *software*

**Seguindo os parâmetros informados no formulário, analisamos seu perfil e chegamos a seguinte conclusão.**

#### **Riscos**

##### **Entrada**

A entrada pode possivelmente prejudicar seu orçamento ? Não

##### **Inadimplência**

Você possivelmente tem risco de ficar inadimplente ? Não

Fonte: Elaboração própria

Após a análise dos modelos de classificação (*decision tree*) o *software* realizou a comunicação com o ChatGPT para validação da decisão tomada pelos modelos de inteligência artificial de classificação clássicos, o resultado é compatível com a decisão dos modelos. Este resultado também é enviado para o usuário do *software* no mesmo e-mail representado na figura 50.

Figura 52 - *E-mail* de resposta do *software* - Gen IA

#### **Resultando ChatGPT**

Com base nas informações fornecidas, o cliente parece ter uma situação financeira estável, com uma renda líquida sólida e reservas significativas. No entanto, é importante avaliar o comprometimento de sua renda com os gastos existentes e o novo financiamento imobiliário. Além disso, a análise de riscos associados ao tipo de indexador, taxa de financiamento e tipo de amortização é crucial. Recomendo avaliar o índice de comprometimento da renda mensal com todas as despesas, considerando o novo financiamento.

Renda Líquida e FGTS: A renda líquida do cliente é de R\$13.000, o que é considerado um bom valor para o financiamento de um imóvel. Além disso, o cliente possui um saldo de FGTS significativo, o que pode ser usado como entrada ou para reduzir o valor financiado.

Reservas Financeiras: O cliente possui reservas em renda fixa e variável, além de uma reserva de emergência. Isso demonstra responsabilidade financeira e a capacidade de enfrentar imprevistos, o que é um ponto positivo para a análise de crédito.

Gastos e Comprometimento: Os gastos mensais do cliente parecem estar dentro de um limite razoável em relação à sua renda. Embora os gastos com cartão de crédito sejam relativamente altos, eles ainda estão dentro do limite de crédito disponível e podem ser gerenciados. O comprometimento de renda com o cartão também está dentro de limites aceitáveis.

Estabilidade Profissional: O cliente tem cinco anos de tempo de emprego, o que indica estabilidade profissional. Isso é importante para garantir a capacidade de pagamento a longo prazo do financiamento.

Score de Crédito: O cliente possui um score Serasa alto (850), o que sugere bom histórico de crédito e confiabilidade financeira.

Outras Considerações: O cliente tem um carro próprio, o que pode ser considerado um ativo adicional. Além disso, ele mora com os pais, o que pode significar que não há despesas de aluguel adicionais.

Fonte: Elaboração própria

## **5.2 Simulação Negativa**

O teste negativo é iniciado com o preenchimento do formulário do front-end, utilizando os dados específicos listados na Tabela 8, esses dados foram selecionados para representar cenários em que o sistema deve recomendar o usuário a não seguir com o contrato de crédito imobiliário.

**Tabela 8 - Variáveis de entrada para simulação (continua)**

Variável de Entrada	Valor de Entrada
Profissão	Engenheiro de Dados
Renda Líquida	R\$ 5000
FGTS	R\$ 15000
Gasto Educação Mês	R\$ 1450
Gasto Imobiliário Mês	R\$ 1250
Gasto Saúde Mês	R\$ 600
Gasto Cartão Mês 1	R\$ 3000
Gasto Cartão Mês 2	R\$ 3100
Gasto Cartão Mês 3	R\$ 3200
Gasto Cartão Mês 4	R\$ 3000
Gasto Cartão Mês 6	R\$ 3000
Quantidade Filho	0
Reserva Renda Fixa	R\$ 5000
Reserva Renda Variável	R\$ 20000
Reserva Emergência	Sim
Utilizou Reserva 12 meses	Não
Limite Cartão	R\$ 20000
Comprometimento Cartão	R\$ 3000
Escolaridade	Nível Superior
Beneficiária INSS	Não
Consignado Valor	0
Empréstimo Valor	R\$ 38000
Produto	Financiamento Imobiliário
Sexo	Masculino
Possui Carro Próprio	Sim
Possui Casa Própria	Não
Estado civil	Casado
Mora com os pais	Sim
Data Nascimento	14-01-1996
CEP Moradia Atual	00000000
Idade Carro Cliente	2
Tipo Organização	Grande
Valor Imóvel	R\$ 317000
m2_imovel	59
Idade Imóvel	11
Quartos	2
Vagas	1
CEP Imóvel Financiado	00000000
Area Lazer	Sim
Elevador	Sim
Necessita Empréstimo ITBI	Sim
Indexador	IPCA
Tipo Imóvel	Apartamento
Estado Documentação Imóvel	Apto

**Tabela 8 - Variáveis de entrada para simulação (conclusão)**

Quantos Andares	19
Objetivo imóvel	Morar
Taxa Financiamento	12%
Tempo Emprego	5
Valor Investido Mensalmente	0
Empréstimo Valor Mês	R\$ 3000
Tipo Amortização	SAC
Mix de Crédito	850

Fonte: Elaboração própria

Na etapa de pré-crédito, são considerados os dados e informações disponíveis antes da concessão do crédito.

**Figura 53 - Normalização Pré-Crédito - Score de Crédito**

```
Out[63]: {'idade': 28,
'ocupacao': 'DESENVOLVEDOR',
'renda_anual': 65000.0,
'salario_mensal': 5000.0,
'taxa_juros': 12.0,
'num_de_emprestimo': 1,
'tipo_de_emprestimo': 'Emprestimo_Hipotecario',
'Divida_pendente': 38000.0,
'Emprestimo_Valor_Mes': 0,
'Valor_investido_mensalmente': 200,
'Pagamento_Comportamento': 'Gastos_Baixos_Pagamentos_de_Pequeno_Valor',
'Saldo_Mensal': -1083.333333333334}
```

Fonte: Elaboração própria

Por outro lado, na etapa de pós-crédito, são consideradas as informações coletadas após a concessão do crédito e durante o período do contrato.

**Figura 54 - Normalização Pós-Crédito - Score de Crédito**

```
Out[70]: {'idade': 28,
'ocupacao': 'DESENVOLVEDOR',
'renda_anual': 65000.0,
'salario_mensal': 5000.0,
'taxa_juros': 7.0,
'num_de_emprestimo': 0,
'tipo_de_emprestimo': 'Emprestimo_Hipotecario',
'Divida_pendente': 38000,
'Emprestimo_Valor_Mes': 0,
'Valor_investido_mensalmente': 200,
'Pagamento_Comportamento': 'Gastos_Baixos_Pagamentos_de_Pequeno_Valor',
'Saldo_Mensal': -1083.3333333333334}
```

Fonte: Elaboração própria

Para compreender a lógica por trás da decisão tomada neste contexto específico, é crucial analisar a figura 55 e a contribuição relativa de cada variável envolvida no processo de tomada de decisão.

**Figura 55 - Importância das *Feature* na Simulação - Score de Crédito**

Pré-Crédito

y=Ruim (probability 1.000) top features		y=Padrao (probability 0.000) top features		y=Bom (probability 0.000) top features	
Contribution?	Feature	Contribution?	Feature	Contribution?	Feature
+0.539	Divida_pendente	+0.428	<BIAS>	+0.211	<BIAS>
+0.361	<BIAS>	+0.008	salario_mensal	-0.103	Divida_pendente
+0.101	salario_mensal	-0.436	Divida_pendente	-0.108	salario_mensal

Pos-Crédito

y=Ruim (probability 1.000) top features		y=Padrao (probability 0.000) top features		y=Bom (probability 0.000) top features	
Contribution?	Feature	Contribution?	Feature	Contribution?	Feature
+0.539	Divida_pendente	+0.428	<BIAS>	+0.211	<BIAS>
+0.361	<BIAS>	+0.008	salario_mensal	-0.103	Divida_pendente
+0.101	salario_mensal	-0.436	Divida_pendente	-0.108	salario_mensal

Fonte: Elaboração própria

Ao examinar a importância atribuída a cada uma das variáveis nesta simulação, é possível identificar que a "dívida pendente" e o "salário mensal" emergem como os fatores mais significativos. Isso sugere que esses dois elementos desempenharam um papel preponderante na determinação do resultado final da decisão, enquanto outras variáveis tiveram contribuições menos proeminentes.

Para o modelo de risco de inadimplência a normalização de dados se resume em uma etapa, ilustrada na figura 56.

**Figura 56 - Normalização - Risco de Inadimplência**

```
Out[62]: {'Quantidade_de_Filhos': '1',
'Renda_Total': '5000',
'Prestacao_do_Emprestimo': '317000',
'Preco_dos_Bens': 25000.0,
'Dias_de_Idade': 10437,
'Dias_Empregado': 730,
'Idade_do_Carro_Proprio': 730,
'Possui_Email': 1,
'Quantidade_de_Membros_da_Familia': 3,
'Media_de_Apartamentos': 59,
'Media_de_Ano_de_Construcao': 2013,
'Modo_de_Apartamentos': 1,
'Modo_de_Area_do_Porao': 1,
'Modo_de_Ano_de_Construcao': 1,
'Modo_de_Area_Comum': 1,
'Modo_de_Elevadores': 0,
'Modo_de_Entradas': 2,
'Modo_de_Area_do_Terreno': 1,
'Modo_de_Apartamentos_para_Morar': 1,
'Modo_de_Area_de_Vida': 1,
'Modo_de_Apartamentos_Nao-Residenciais': 1,
'Modo_de_Area_Nao-Residencial': 1}
```

Fonte: Elaboração própria

Ao observar a figura 57, é possível inferir que o tempo de emprego do cliente desempenha um papel crucial na decisão final. Essa observação sugere que clientes com um tempo empregado pequeno podem ter uma probabilidade correr riscos ao realizar o contrato de crédito imobiliário.

**Figura 57 - Importância das *Feature* na Simulação - Risco de Inadimplência**  
 y=ComRisco (probability 0.911, score -0.579) top features

Contribution?	Feature
+0.768	<BIAS>
+0.331	Dias_Empregado
+0.230	Dias_de_Idade
+0.093	Idade_do_Carro_Proprio
	... 5 more positive ...
	... 6 more negative ...
-0.045	Modo_de_Apartamentos_para_Morar
-0.072	Preco_dos_Bens
-0.079	Prestacao_do_Emprestimo
-0.132	Modo_de_Area_do_Porao
-0.151	Renda_Total
-0.330	Modo_de_Area_Comum

Fonte: Elaboração própria

O *software* analisou e sugeriu conforme figura 58 que, com base nos parâmetros de entrada (tabela 8), este cliente não deve seguir com o financiamento, uma vez que os modelos apresentaram os seguintes resultados:

- Análise positiva para o risco de a entrada utilizada no financiamento afetar a saúde financeira ao decorrer do contrato, isto é, o *score* de crédito é ruim.
- Análise positiva para o risco de o cliente tornar-se inadimplente.

**Figura 58 - E-mail de resposta do *software***

Seguindo os parâmetros informados no formulário, analisamos seu perfil e chegamos a seguinte conclusão.

#### Riscos

##### Entrada

A entrada pode possivelmente prejudicar seu orçamento ? Sim

##### Inadimplência

Você possivelmente tem risco de ficar inadimplente ? Sim

Fonte: Elaboração própria

Após a análise dos modelos de classificação (*decision tree*) o *software* realizou a comunicação com o ChatGPT para validação da decisão tomada pelos modelos de

inteligência artificial de classificação clássicos, o resultado é compatível com a decisão dos modelos. Este resultado também é enviado para o usuário do *software* no mesmo *e-mail* representado na figura 59.

### Figura 59 - *E-mail* de resposta do *software* - Gen IA

#### Resultado ChatGPT

E Com base nas informações fornecidas, o cliente apresenta uma situação financeira estável, com uma renda líquida mensal de R\$ 5.000,00 e reservas financeiras que totalizam R\$ 25.000,00. No entanto, ao analisar o perfil de crédito, é importante considerar o nível de comprometimento da renda com as despesas existentes, o financiamento pretendido, e as dívidas pendentes. Atualmente, o cliente possui um empréstimo ativo de R\$ 38.000,00, o que adiciona uma obrigação financeira significativa ao seu orçamento mensal. Somando-se aos gastos fixos de R\$ 3.150,00 (educação, moradia e saúde), a dívida pendente aumenta o comprometimento da renda, deixando pouco espaço para o pagamento do novo financiamento imobiliário. Com a primeira parcela do financiamento imobiliário estipulada em R\$ 6.000,00, o comprometimento total da renda pode ultrapassar níveis seguros. Isso representa um risco elevado de inadimplência, especialmente considerando que o cliente já possui obrigações financeiras relevantes. Além disso, a taxa de juros de 12% e o indexador IPCA introduzem variáveis que podem aumentar os custos ao longo do tempo, especialmente em períodos de inflação alta. Dado o comprometimento existente com a dívida e as despesas mensais, o cliente pode enfrentar dificuldades para manter os pagamentos do financiamento imobiliário. Recomendo uma reavaliação cuidadosa da capacidade de pagamento do cliente, considerando todos os compromissos financeiros. É aconselhável revisar o valor do financiamento solicitado ou explorar alternativas, como ajustar o prazo de pagamento ou buscar melhores condições de juros.

Fonte: Elaboração própria

## 6 CONSIDERAÇÕES FINAIS

O desenvolvimento deste projeto proporcionou uma oportunidade única de reexaminar a análise de crédito para financiamento imobiliário. Diferentemente das abordagens convencionais das instituições financeiras, que se concentram na capacidade do cliente em cumprir as cláusulas contratuais, este projeto visa preparar o cliente para ingressar em um financiamento imobiliário, minimizando os riscos de inadimplência a longo prazo.

Com referenciais teóricos robustos e processos tecnológicos de análise de dados, o *software* desenvolvido demonstrou ser uma ferramenta capaz de enriquecer a tomada de decisões dos usuários no contexto de financiamento imobiliário.

O objetivo geral de desenvolver um aplicativo baseado em Inteligência Artificial (IA) para auxiliar na análise do produto financeiro de crédito imobiliário foi plenamente alcançado, resultando em um *software* robusto e com base nas estatísticas dos modelos eficaz.

Os objetivos específicos delineados no início do projeto foram cuidadosamente abordados e alcançados com sucesso.

O primeiro objetivo específico, de identificar *insights* relevantes para a tomada de decisão no crédito imobiliário, utilizando análise de dados e técnicas de inteligência artificial, foi alcançado por meio da criação deste *software* que fornece ferramentas e recursos para auxiliar na tomada de decisão de crédito imobiliário.

O segundo objetivo de identificar situações em que eles podem ser utilizados na tomada de decisão para crédito imobiliário. Ambos os objetivos foram alcançados e validados por meio da aplicação de técnicas de inteligência artificial, incluindo árvores de decisão, XGBoost, clusterização com k-means e inteligência artificial generativa. Esses métodos foram essenciais para interpretar variáveis explicativas, identificar oportunidades, validar a qualidade preditiva dos modelos e enriquecer a tomada de decisões financeiras. Além disso, os modelos desenvolvidos foram capazes de prever o risco de inadimplência do cliente e fornece recomendações personalizadas para a tomada de decisão no crédito imobiliário, cumprindo assim os objetivos estabelecidos com sucesso.

É importante justificar a ausência do diagrama das árvores de decisão, que se deve ao extenso tamanho final de cada árvore, tornando inviável a sua representação neste trabalho.

O quarto objetivo, de utilizar conceitos e ferramentas de computação em nuvem para coleta e processamento de dados, também foi plenamente alcançado, permitindo uma análise abrangente e precisa dos dados financeiros dos usuários.

O quinto objetivo, de estabelecer parâmetros de entrada personalizáveis e definir informações de saída claras e relevantes, foi atendido por meio da implementação de interfaces intuitivas e da geração de relatórios para os usuários do *software*.

O sexto objetivo, de definir as informações de saída disponibilizadas para o usuário foi abordado no desenvolvimento do projeto, com foco em fornecer dados claros, para a compreensão da situação financeira do usuário e para facilitar a tomada de decisão. Para atingir esse objetivo, o *software* foi projetado para fornecer informações detalhadas sobre o risco de redução do *score* de crédito e o risco de inadimplência, fundamentais para que o usuário compreenda sua posição financeira e os potenciais impactos de suas decisões ao iniciar um crédito imobiliário.

Além disso, o uso de inteligência artificial generativa foi incorporado para explicar detalhadamente as decisões dos modelos por meio das variáveis explicativas. Isso significa que o *software* é capaz de fornecer *insights* sobre como as variáveis específicas influenciam as decisões dos modelos de inteligência artificial, permitindo uma compreensão mais profunda dos fatores que afetam a situação financeira do usuário. Essa abordagem transparente e informativa ajuda o usuário a tomar decisões orientado a dados e a gerenciar melhor seu crédito imobiliário.

O sétimo objetivo, de conduzir testes para validar a acurácia e qualidade dos modelos de inteligência artificial utilizados, garantindo que os resultados sejam confiáveis e eficazes, foi realizado com sucesso, assegurando que o *software* desenvolvido apresente resultados confiáveis e eficientes em sua aplicação.

Em suma, os objetivos delineados neste projeto foram completamente atendidos, resultando em um *software* que oferece suporte à tomada de decisões financeiras e contribui para a saúde financeira dos usuários.

A revisão bibliográfica ressaltou pontos relevantes, como o elevado e crescente índice de famílias brasileiras endividadas. Revelado pela Confederação Nacional do

Comércio de Bens, Serviços e Turismo em 2023, de que 27% das famílias brasileiras estão inadimplentes, segundo pesquisa do Serasa Levantando o ponto de que a população brasileira tem um alto índice de analfabetismo financeiro.

Este *software* é relevante para que os usuários em busca do sonho da casa própria atinjam esse objetivo de maneira financeiramente saudável, onde acredita-se que devido ser uma “nova” maneira de realizar análises de crédito, não somente pensando se o cliente vai pagar, mas também se não irá prejudicar a saúde financeira do cliente, pensando e centralizando a decisão dos modelos no cliente e não no produto financeiro, empoderando o cliente a incluir em suas decisões o uso dos dados.

Além disso, o *software* visa combater a escassez de fontes confiáveis para cursos de qualidade sobre educação financeira. Espera-se que, ao contribuir para a redução do analfabetismo financeiro, tem-se a expectativa que o *software* influencie as instituições financeiras a não apenas vender produtos, mas também considerar o bem-estar do cliente. Todos os aspectos abordados são documentados para servir como base para pesquisas futuras.

Embora o foco atual do projeto seja exclusivamente em financiamento imobiliário, considera-se a possibilidade de expansão para crédito de veículos ou uso de cartão de crédito. A educação financeira, um campo em ascensão, promete crescimento nos próximos anos.

Ao longo do processo de análise de dados, foi notável o impacto positivo e desafiador das facilidades proporcionadas pelo ambiente de computação em nuvem. Esse ambiente permitiu não apenas aumentar o poder computacional conforme necessário, mas também viabilizou um *software* de baixo custo devido à sua escalabilidade.

Embora não se tenha informações suficientes para explicar a relevância da inclusão de novos dados aos modelos, sugere-se como uma área de pesquisa futura o quão relevante para outras visões os dados coletados podem contribuir, por exemplo, a possibilidade de arrependimento na aquisição do imóvel. Contudo, com base nas referências bibliográficas e técnicas utilizadas, é possível afirmar que o *software* tem potencial para evoluir de maneira positiva.

O desafio inicial durante o desenvolvimento e treinamento dos modelos clássicos concentrou-se na dificuldade de adquirir bases de dados, considerando a sensibilidade das informações e a necessidade de conformidade com a Lei Geral de

Proteção de Dados. Entretanto, questões relativas ao armazenamento e manipulação de dados sensíveis do usuário, bem como a monetização, exigirão abordagens mais aprofundadas em trabalhos futuros, considerando a complexa jurisdição atual. Sobre o uso da ChatGPT ele veio com o propósito de complementar e explicar a informação gerada pelos modelos clássicos, há uma grande margem de estudo de evolução no uso da ferramenta inclusive fica como sugestão uma evolução da autonomia e eliminar o uso dos modelos clássicos.

Em virtude da relevância e atualidade do tema, sugerem-se estudos futuros:

1. Identificação dos motivos que levam a população brasileira a enfrentar dificuldades financeiras,
2. Avaliação da possibilidade de utilizar Inteligência Artificial Generativa para promover a saúde financeira individual.
3. Avaliação do impacto da clusterização das variáveis de entrada nos resultados dos modelos de árvore de decisão.

## REFERÊNCIAS

ADZIC, G.; CHATLEY, R. Serverless computing: Economic and architectural impact. Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. New York, NY, USA: Association for Computing Machinery, 2017. (ESEC/FSE 2017), p. 884–889. ISBN 9781450351058. Disponível em: <https://dl.acm.org/doi/10.1145/3106237.3117767>. Acesso em: 19 de março de 2023.

AGENCIA BRASIL. Mais de 5 bilhões de pessoas já contam com celulares em todo mundo. 2019. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-09/mais-de-5-bilhoes-de-pessoas-usam-aparelho-celular-revela-pesquisa/>. Acesso em: 25 de abril de 2023.

ARAUJO, Luis Cesar G. de. Teoria Geral da Administração; aplicação e resultados nas empresas brasileiras. 2a. ed. São Paulo: ATLAS, 2014. v. 1. 291 p. Empresas Brasileiras (2a. Ed.). São Paulo: Atlas, 2004.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/what-is-cloud-computing/>. Acesso em: 12 de março de 2023.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/ec2/>. Acesso em: 12 março de 2023.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/lambda/>. Acesso em: 12 de março de 2023.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/api-gateway/>. Acesso em: 11 de março de 2023.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/ses/>. Acesso em: 11 de março de 2023.

AMAZON WEB SERVICES. Disponível em: <https://aws.amazon.com/pt/s3/>. Acesso em: 11 de março de 2023.

ANNIBAL, Clodoaldo Aparecido. "Inadimplência do Setor Bancário Brasileiro: uma avaliação de suas medidas." Working Papers Series 192, Central Bank of Brazil, Research Department, 2009. Disponível em:

<https://www.bcb.gov.br/content/publicacoes/WorkingPaperSeries/wps192.pdf>.

Acesso em: 01 de maio de 2024.

AZURE MICROSOFT. Disponível em: <https://azure.microsoft.com/pt-br/products/cosmos-db>. Acesso em: 08 de abril de 2023.

BANCO CENTRAL DO BRASIL. Disponível em:

<https://www.bcb.gov.br/detalhenoticia/590/noticia>. Acesso em: 11 maio de 2022.

BORGES, Paulo Roberto Santana. Educação financeira: o novo perfil das famílias na administração das finanças pessoais. IX Encontro de Produção Científica e Tecnológica, Gabriela Martins dos Santos, Marlette Cassia Oliveira Ferreira, Flávio Santino Bizarrias, Jussara da Silva Teixeira Cucato, Jussara Goulart da Silva, 2014.  
CARDOSO, R., Bases de Dados NoSQL, Instituto Superior de Engenharia do Porto, 2012.

BRASIL. Lei n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais. Diário Oficial da União: seção 1, Brasília, n. 157, p. 59, 15 ago. 2018.

CHAMBERLIN, Donald D., Early History of SQL. IEEE Annals of the History of Computing, vol. 34, no. 4, pp. 78-82. Disponível em:

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6359709>. Acesso em: 06 abril de 2023.

Confederação Nacional do Comércio de Bens, Serviços e Turismo. Endividamento atinge 78,3% das famílias brasileiras. Disponível em:

<https://agenciabrasil.ebc.com.br/economia/noticia/2023-05/endividamento-atinge-783-das-familias-brasileiras-diz-cnc#:~:text=A%20previs%C3%A3o%20%C3%A9%20que%20o,6%25%20de%20abril%20de%202022..>

Acesso em agosto de 2023.

CORDEIRO, Nilton José Neves; COSTA, Manoel Guto Vasconcelos; SILVA, Marcio Nascimento da. Educação financeira no Brasil: uma perspectiva panorâmica. *Ensino da Matemática em Debate*, v. 5, n. 1, p.69-84, 2018.

COSTA, Cristiano Machado; MIRANDA, Cléber José de. Educação financeira e taxa de poupança no Brasil. *Revista de Gestão, Finanças e Contabilidade*, v. 3, n. 3, p. 57-74, 2013.

DA SILVA, G. L., Diniz, P. S., Ferreira, J. L., Franca, J. V., Silva, A. C., de Paiva, A. C., and de Cavalcanti, E. A. (2020). Superpixel-based 137 deep convolutional neural networks and active contour model for automatic prostate segmentation on 3d mri scans. *Medical & Biological Engineering & Computing*, 58(9):1947–1964.

DATA CAMP. Disponível em: <https://www.datacamp.com/tutorial/decision-tree-classification-python>. Acesso em: 11 de janeiro de 2024.

DATABRICKS. Disponível em: <https://community.cloud.databricks.com/>. Acesso em: 11 de janeiro de 2024.

DE ANDRADE, Elias Sebastião. Desenvolvimento Sustentável E Sociedade 5.0: Rumo À Felicidade E Ao Bem-Estar. *Revista Humanitaris-B3*, v. 2, n. 2, p. p. 6-25, 2021.

ELMASRI, Ramez; NAVATHE, Shamkant B. *Sistemas de Banco de Dados*. 6. ed. São Paulo: Addison Wesley, 2011. eSocial. Disponível em: <<http://portal.esocial.gov.br/institucional/conheca-o>>. Acesso em: 06 de abril de 2023.

FAGUNDES, Julio Cesar Franke. Abordagem baseada em Árvores de Decisão para detecção e identificação de intrusões em ambientes da Internet das Coisas baseados em Computação Nevoeiro. Universidade Federal de Santa Catarina. 2022. Disponível em: [https://repositorio.ufsc.br/bitstream/handle/123456789/232654/TCC\\_JulioFranke.pdf?sequence=1&isAllowed=y](https://repositorio.ufsc.br/bitstream/handle/123456789/232654/TCC_JulioFranke.pdf?sequence=1&isAllowed=y). Acesso em: 6 de junho de 2022.

FERREIRA, Caio. Indução de Árvore de Decisão utilizando Meta-Aprendizado. Universidade de São Paulo. 2022. Disponível em: <https://www.teses.usp.br/teses/disponiveis/59/59143/tde-05122022->

145141/publico/Dissertacao\_Corrigida\_Caique\_Ferreira.pdf. Acesso em: 6 junho de 2022.

GARCIA, Miguel Alexandre Da Cruz. Impacto da Inteligência Artificial no setor financeiro. Disponível em: [https://repositorio.iscte-iul.pt/bitstream/10071/23975/1/master\\_miguel\\_cruz\\_garcia.pdf](https://repositorio.iscte-iul.pt/bitstream/10071/23975/1/master_miguel_cruz_garcia.pdf). Acesso em: 6 junho de 2022.

GINI, Conrado. On the measure of concentration with special reference to income and statistics. Colorado College Publication, General Series, 1936.

GRANEMANNR Oliveira Roney. Inadimplência na área de habitação da Caixa Econômica Federal, no âmbito do escritório de negócios institucional de Florianópolis: um estudo de caso. Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/82167>. Acesso em: 6 junho de 2022.

GOOGLE CLOUD PLATAFORM (GCP). Disponível em: <https://cloud.google.com/bigtable/docs?hl=pt-br>. Acesso em: 08 de abril de 2023.

GOODFELLOW J. Ian; BENGIO Yoshua; COURVILLE Aaron: Deep learning. Genet Program Evolvable, 2016. Disponível em: <https://doi.org/10.1007/s10710-017-9314-z>. Acesso em 07 julho de 2022.

HARRISON, Matt. Machine Learning: Guia de Referência Rápida (2020). São Paulo: Novatec O'Reilly.

JAVASCRIPT. Disponível em: <https://www.javascript.com/>. Acesso em 1 Maio de 2024.

KAGGLE. Disponível em: <https://www.kaggle.com/>. Acesso em 1 Maio de 2024.

KPMG. Estratégia de negócios e novos modelos operacionais – Open Banking, 2021. Disponível em: <https://home.kpmg/br/pt/home/industrias/servicos-financeiros/open-banking/estrategias-negocios.html>. Acesso em 07 julho de 2022.

KARASAN, Abdullah. Machine Learning for Financial Risk Management with Python: Algorithms for Modeling Risk (2021). O'Reilly Media. Disponível em: <https://www.studocu.com/vn/document/hanoi-university-of-science/simple-machine/machine-learning-for-financial-risk-management-with-python-algorithms-for-modeling-risk-1nbsped-1492085251-9781492085256/41172131>. Acesso em 07 julho de 2022.

LABS NEWS: 34 milhões de brasileiros não têm acesso a serviços bancários. Disponível em: <https://ilocomotiva.com.br/clipping/labs-news-34-milhoes-de-brasileiros-nao-tem-acesso-a-servicos-bancarios/>. Acesso em 07 julho de 2022.

LAVAGNOLI, S. (ED.). Especificações de Inteligência Artificial, Machine Learning e Deep Learning. Disponível em: <https://opencadd.com.br/machine-learning-ou-deep-learning/>. Acesso em 07 julho de 2022.

LEE, Kai-Fu. Inteligência artificial (2019). Rio de Janeiro: Globo Livros.

MAYER-SCHÖNBERGER, Viktor; CUKIER, K. Big Data: A revolution that will transform how we live, work and think. London: John Murray, 2017.

MCCARTHY, J Artificial intelligence, logic and formalizing common sense. In: Philosophical logic and artificial intelligence. Springer, Dordrecht, 1989. p. 161-190.

MELL, Peter; GRANCE, Tim. The NIST definition of Cloud Computing. NIST, 2011. Disponível em: <https://csrc.nist.gov/publications/detail/sp/800-145/final>. Acesso em: 12 março de 2023.

MENEZES, N. N. C. Introdução a programação com python. São Paulo: Novatec, 2010.

METZNER, Talita Dayane; MATIAS, Alberto Borges. O setor bancário brasileiro de 1990 a 2000. São Paulo: Manole, 2015.

Ministério de Ciência, Tecnologia e Inovações. Relatório do MCTI aponta que indústria de Software e Serviços de TIC cresceu 6,5% no Brasil em 2021. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2022/07/relatorio-do->

mcti-aponta-que-industria-de-software-e-servicos-de-tic-cresceu-6-5-no-brasil-em-2021. Acesso em: 08 de abril de 2023.

MONTEIRO, Leandro Pinho. "Dados Estruturados e Não Estruturados". Blog da Universidade de Tecnologia. Disponível em: <https://universidadedatecnologia.com.br/dados-estruturados-e-nao-estruturados/>. Acesso em: 01 de Maio de 2024.

MOZILLA. "HTML: HyperText Markup Language". Disponível em: <https://developer.mozilla.org/en-US/docs/Web/HTML>. Acesso em: 01 de Maio de 2024.

NETTO, V. Antônio; MARTINS E. Luiz; ORLANDINI, Guilherme. Development of api for autonomous navigation of robotic platform. Unifacs, 2018. Disponível em: <https://revistas.unifacs.br/index.php/rsc/article/view/5131>. Acesso em: 11 maio. 2022.

NETO, Zeferino G. D. S. Curva ROC para comparação de modelos de predição para variáveis dicotômicas. Universidade Federal de Viçosa, 2020. Disponível em: <https://www.locus.ufv.br/bitstream/123456789/28623/1/texto%20completo.pdf>. Acesso em: 12 março. 2023

OpenAI. Disponível em: [openai.com](https://openai.com). Acesso em: 17 de fevereiro de 2024.

OLIVEIRA, S. Bancos de dados não-relacionais: Um novo paradigma para armazenamento de dados em sistemas de ensino colaborativo. Disponível em: <https://www2.unifap.br/oliveira/files/2016/02/35-124-1-PB.pdf>. Acesso em: 12 março de 2023.

ORACLE. Disponível em: <https://docs.oracle.com/pt-br/iaas/Content/APIGateway/Concepts/apigatewayconcepts.htm>. Acesso em: 12 março. 2023

PAIXÃO, Gabriela Miana de Mattos; SANTOS, Bruno Campos; ARAUJO, Rodrigo Martins de; RIBEIRO, Manoel Horta; MORAES, Jermana Lopes de; RIBEIRO, Antonio L.. Machine Learning na Medicina: Revisão e Aplicabilidade. Arq. Bras. Cardiol., v. 118, n. 1, p. 95-102, jan. 2022.

PICCINI, Ruberlan Alex Bilha; PINZETTA, Gilberto. Planejamento financeiro pessoal e familiar. Unoesc & Ciência – ACSA, v. 5, n. 1, p.95-102, 2014.

PYTHON. Python Programming Language. Disponível em: [https://  
https://www.python.org/](https://https://www.python.org/). Acesso em: 12 março. 2023

RM TELEFONIA. O futuro dos smartphones e a tecnologia móvel emergente. Disponível em: <https://rmtelefonica.com.br/o-futuro-dos-smartphones-e-a-tecnologia-movel-emergente/>. Acesso em agosto de 2023.

ROSSUM, G. V. Python. Versão 3.11. 1991. Disponível em: <https://docs.python.org/pt-br/dev/whatsnew/3.11.html>. Acesso em: 06 de janeiro de 2023.

SADALAGE, P. J.; FOWLER, M. NoSQL Essencial: Um guia conciso para o mundo emergente da persistência poliglota. 1. ed. São Paulo: Novatec, 2015.

SCKIKIT-LEARN. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>. Acesso em: 12 março. 2023

SCKIKIT-LEARN. Disponível em: [https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html). Acesso em: 5 maio. 2024

SERASA. Mapa da Inadimplência e Negociação de Dívidas no Brasil. Disponível em: <https://www.serasa.com.br/limpa-nome-online/blog/mapa-da-inadimplencia-e-renogociacao-de-dividas-no-brasil/>. Acesso em: 01 de maio de 2024.

SERASA. Conheça as 7 principais causas de inadimplência no Brasil hoje. Disponível em: <https://www.serasaexperian.com.br/conteudos/estudos-e-pesquisas/conheca-as-7-principais-causas-de-inadimplencia-no-brasil-hoje/>. Acesso em: 08 de abril de 2023.

SHAFIEI, H.; KHONSARI, A.; MOUSAVI, P. Serverless Computing: A Survey of Opportunities, Challenges and Applications. ACM Computing Surveys, 2022. Disponível em: <https://arxiv.org/abs/1911.01296#>. Acesso em: 19 mar. 2023.

SILVA, C. A. R. F. O. Da., Data modeling with NoSQL: How, when and why, Faculdade de Engenharia da Universidade do Porto, 2011.

SILVA, M. J. O sonho da casa própria: um estudo sobre as motivações das pessoas que compram uma casa. São Paulo: Editora Atlas, 2006.

TIOBE. Uso das linguagens de programação em 2023, 2023. Disponível em: <https://www.tiobe.com/tiobe-index/>. Acesso em: 28 novembro. 2023.

VARGAS, Isadora Formenton; SANTOS, Nicolas Peixoto dos. A Inclusão Financeira por meio de Plataformas Digitais: experiências de países emergentes e análise do estado da arte no Brasil. Revista da PGBC, v. 14, n. 2, p. 49-62, dez. 2020. Disponível em: <https://revistapgbc.bcb.gov.br/revista/article/view/1078/56>. Acesso em: 11 maio. 2022.

WÜNSCHE, Gabriela Melissa. Mudanças de hábitos de consumo dos usuários de aplicativos de economia compartilhada em decorrência da pandemia da covid 19. Universidade Federal de Santa Catarina 2023. Disponível em: [https://repositorio.ufsc.br/bitstream/handle/123456789/237211/Melissa\\_TCC.pdf?sequence=1&isAllowed=y](https://repositorio.ufsc.br/bitstream/handle/123456789/237211/Melissa_TCC.pdf?sequence=1&isAllowed=y). Acesso em: 19 mar. 2023