

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
ESCOLA POLITÉCNICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**RECOMENDAÇÃO
PERSONALIZADA DE
CARREIRAS EM TECNOLOGIA
DA INFORMAÇÃO: UMA
ABORDAGEM BASEADA EM
MACHINE LEARNING E
ANÁLISE DE PERFIS.**

IGOR PEREIRA DOURADO

Trabalho de Conclusão II apresentado
como requisito parcial à obtenção
do grau de Bacharel em Ciência da
Computação na Pontifícia Universidade
Católica do Rio Grande do Sul.

Orientador: Prof. Silvia Maria Wanderley Moraes

**Porto Alegre
2024**

RECOMENDAÇÃO PERSONALIZADA DE CARREIRAS EM TECNOLOGIA DA INFORMAÇÃO: UMA ABORDAGEM BASEADA EM MACHINE LEARNING E ANÁLISE DE PERFIS.

RESUMO

O avanço tecnológico nas últimas décadas transformou de maneira intensa o mercado de trabalho, especialmente no setor de Tecnologia da Informação (TI). A rápida evolução desta área tem gerado uma diversidade crescente em carreiras e áreas de trabalho, criando novas oportunidades, mas juntamente com isso, desafios significativos para profissionais que buscam se posicionar de acordo com suas habilidades e características ou requalificar-se dentro deste mercado dinâmico. Um problema relevante no contexto da orientação de carreiras em TI é a dificuldade de alinhar de maneira eficaz as habilidades, interesses e aspirações do indivíduo com as carreiras disponíveis. A complexidade do setor e a falta de conhecimento, torna a escolha profissional um processo muitas vezes marcado por incertezas e indecisões. Diante deste cenário, foi criado e desenvolvido o aplicativo para iOS denominado "Carreira pra TI: Uma Abordagem Baseada em Machine Learning e Análise de Perfis" ou apenas "Carreira pra TI". Empregando técnicas de Machine Learning para, a partir de respostas do usuário sobre o seu nível de conhecimento, interesse e afinidade em relação a áreas de estudo da TI e sobre suas habilidades sociais e comportamentais, o objetivo do aplicativo é sugerir carreiras da TI que melhor se alinhem com o perfil do usuário. Feito com um modelo de Machine Learning treinado com uma base de dados de mais de nove mil respostas e baseada também em dados reais obtidos ao decorrer de trabalho, o aplicativo provê recomendações personalizadas e informadas, buscando contribuir para a realização e eficiência profissional dos usuários e ajudar a fazer escolhas de carreira mais alinhadas com suas habilidades e interesses.

Palavras-Chave: Recomendação de Carreira, Tecnologia da Informação, Machine Learning, Sistema de Recomendação, Análise de Perfis, Aplicação.

PERSONALIZED CAREER RECOMMENDATIONS IN INFORMATION TECHNOLOGY: A MACHINE LEARNING AND PROFILE ANALYSIS APPROACH.

ABSTRACT

Technological advancements in recent decades have profoundly transformed the labor market, particularly in the Information Technology (IT) sector. The rapid evolution of this field has led to a growing diversity of careers and work opportunities, creating new possibilities but also significant challenges for professionals seeking to align their skills and characteristics with career paths or requalify themselves within this dynamic market. A key issue in IT career guidance is the difficulty of effectively aligning an individual's skills, interests, and aspirations with the available career options. The complexity of the sector and the lack of knowledge often make career choice a process marked by uncertainty and indecision. In response to this scenario, the iOS application "Carreira pra TI: A Machine Learning and Profile Analysis-Based Approach", or simply "Carreira pra TI", was developed. By leveraging Machine Learning techniques, the app analyzes users' responses regarding their level of knowledge, interest, and affinity for various IT study areas, as well as their social and behavioral skills, to recommend IT careers that best align with their profiles. Built using a Machine Learning model trained on a dataset of over nine thousand responses and further enriched with real-world data collected throughout the project, the app delivers personalized and informed career recommendations. This initiative aims to contribute to users' professional fulfillment and efficiency by assisting them in making career choices that align with their skills and interests, ultimately supporting their success in the dynamic IT job market.

Keywords: Career Recommendation, Information Technology, Machine Learning, Recommendation System, Profile Analysis, Application.

LISTA DE FIGURAS

Figura 3.1 – Relação da carreira de Administrador de Banco de Dados em com as áreas de estudo e as habilidades sociais	38
Figura 3.2 – Relação da carreira de Engenheiro de Hardware com as áreas de estudo e as habilidades sociais	38
Figura 3.3 – Relação da carreira de Testador de Software com as áreas de estudo e as habilidades sociais	39
Figura 3.4 – Relação da carreira de Administrador de Banco de Dados em com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada	46
Figura 3.5 – Relação da carreira de Engenheiro de Hardware com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada	46
Figura 3.6 – Relação da carreira de Testador de Software com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada	47
Figura 4.1 – Tela 1 do OnBoarding	55
Figura 4.2 – Tela 2 do OnBoarding	55
Figura 4.3 – Tela do questionário sobre áreas de estudo e habilidades sociais . . .	56
Figura 4.4 – Tela para tirar a dúvida sobre a área ou habilidade	56
Figura 4.5 – Tela do resultado	58
Figura 4.6 – Tela de detalhes da carreira carreira	60
Figura 5.1 – Gráfico da primeira pergunta sobre qual a satisfação geral com o aplicativo como uma ferramenta de orientação de carreira e entender melhor suas opções na TI	66
Figura 5.2 – Gráfico da primeira pergunta sobre o quanto você considera o aplicativo intuitivo, agradável e com uma boa experiência de usuário	67
Figura 5.3 – Gráfico da pergunta 3	68
Figura 5.4 – Gráfico da pergunta 4	69
Figura 5.5 – Gráfico da pergunta 5	70
Figura 5.6 – Gráfico da pergunta 6	71
Figura 5.7 – Gráfico da pergunta 7	72
Figura 5.8 – Gráfico da pergunta 8	73
Figura 5.9 – Gráfico da pergunta 9	74
Figura 5.10 – Gráfico da pergunta 10	75

LISTA DE SIGLAS

TI: Tecnologia da Informação

ML: Machine Learning

IA: Inteligência Artificial

iOS: iPhone Operating System

GBC: Gradient Boosting Classifier

HMM: Hidden Markov Model

SVM: Support Vector Machine

MIT: Massachusetts Institute of Technology

coreML / Create ML: Core / Create Machine Learning

US: United States

IEEE: Institute of Electrical and Electronic Engineers

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONTEXTUALIZAÇÃO	11
1.2	OBJETIVOS	12
2	DESCRIÇÃO DO PROBLEMA E TRABALHOS RELACIONADOS	13
2.1	DESCRIÇÃO DO PROBLEMA	13
2.1.1	IMPACTO DO PROBLEMA	13
2.1.2	ESPECIFICIDADE NO CONTEXTO DE TI	15
2.1.3	A ROTATIVIDADE COMO REFLEXO DO DESALINHAMENTO	16
2.2	FUNDAMENTAÇÃO TEÓRICA	16
2.2.1	A EVOLUÇÃO DO MERCADO DE TRABALHO EM TECNOLOGIA DA INFORMAÇÃO	17
2.2.2	DESAFIOS NA ORIENTAÇÃO DE CARREIRAS EM TECNOLOGIA DA INFORMAÇÃO	18
2.2.3	MACHINE LEARNING E SEU USO NO SISTEMA DE RECOMENDAÇÕES DE CARREIRA	19
2.2.4	ALGORITMOS PARA TREINAR O MODELO DE CLASSIFICAÇÃO DA MACHINE LEARNING	21
2.2.5	TÉCNICAS DE TREINAMENTO DE MACHINE LEARNING USADAS NO ALGORITMO	23
2.2.6	RELATÓRIO DE CLASSIFICAÇÃO E MÉTRICAS DE AVALIAÇÃO	26
2.2.7	SISTEMA DE RECOMENDAÇÃO	27
2.3	TRABALHOS RELACIONADOS	29
2.3.1	PREDICTING PERSONALIZED ACADEMIC AND CAREER ROADS: FIRST STEPS TOWARD A MULTI-USES RECOMMENDER SYSTEM	29
2.3.2	GUIA-ME: UM SISTEMA DE RECOMENDAÇÃO PARA AJUDAR A INICIAR OU MIGRAR DE CARREIRA NA ÁREA DE TI	30
2.3.3	NEMO: NEXT CAREER MOVE PREDICTION WITH CONTEXTUAL EMBEDDING	31
2.3.4	UNLOCKING FUTURES: A NATURAL LANGUAGE DRIVEN CAREER PREDICTION SYSTEM FOR COMPUTER SCIENCE AND SOFTWARE ENGINEERING STUDENTS	32

2.3.5	PERSONALIZED CAREER-PATH RECOMMENDATION MODEL FOR INFORMATION TECHNOLOGY STUDENTS IN INDONESIA	33
3	DESENVOLVIMENTO DO MODELO	35
3.1	CONJUNTO DE DADOS	35
3.1.1	PESQUISA SOBRE O CONJUNTO DE DADOS	35
3.1.2	VOLUMETRIA	36
3.1.3	ANÁLISE DOS DADOS E TREINAMENTO INICIAL DO CONJUNTO DE DADOS BASE	36
3.1.4	PESQUISA PARA EQUILIBRAR O CONJUNTO DE DADOS	39
3.1.5	USANDO O RESULTADO DA PESQUISA PARA MELHORAR O CONJUNTO DE DADOS	40
3.1.6	MÉTRICAS DO CONJUNTO DE DADOS	47
3.2	CONSTRUÇÃO DO MODELO	48
3.2.1	CONFIGURAÇÃO INICIAL E PRÉ-PROCESSAMENTO	48
3.2.2	ESCOLHA DO ALGORITMO	48
3.2.3	TESTES DO MODELO DE MACHINE LEARNING	52
3.2.4	ASPECTOS TÉCNICOS DO DESENVOLVIMENTO DO MODELO	53
4	PROJETO DO APLICATIVO CARREIRA PRA TI	54
4.1	OBJETIVO	54
4.2	DESENVOLVIMENTO DO APLICATIVO	54
4.3	INTERFACE DO USUÁRIO E FLUXO DO APLICATIVO	54
4.3.1	TELA DE BOAS-VINDAS E INTRODUÇÃO	55
4.3.2	FORMULÁRIO DE PERFIL DO USUÁRIO	56
4.3.3	RECOMENDAÇÃO DA CARREIRA	58
4.3.4	TELA DE DETALHES SOBRE A CARREIRA	60
4.4	DESENVOLVIMENTO	61
4.4.1	USO DO MODELO NO APLICATIVO	61
4.4.2	TESTES DE USABILIDADE DO APLICATIVO	62
4.4.3	ASPECTOS TÉCNICOS DO DESENVOLVIMENTO DO APLICATIVO	62
4.4.4	PUBLICAÇÃO E DISPONIBILIDADE	63
5	RESULTADOS GERAIS	64
5.1	O CONJUNTO DE DADOS E O MODELO DE APRENDIZADO DE MÁQUINA ..	64
5.2	VALIDAÇÃO E APLICABILIDADE DO APLICATIVO	64

5.3	CONTRIBUIÇÃO ACADÊMICA	65
5.4	IMPACTO GERAL	65
5.5	SATISFAÇÃO DO USUÁRIO	65
5.6	AVALIAÇÃO	66
5.6.1	PESQUISA SOBRE A EXPERIÊNCIA AO USAR O APLICATIVO E O MODELO	66
5.6.2	MÉTRICAS DOS NÚMEROS DE DOWNLOADS, VISUALIZAÇÕES DA PÁGINA E FALHAS	75
6	CONCLUSÃO E TRABALHOS FUTUROS	77
6.1	CONCLUSÃO	77
6.2	TRABALHOS FUTUROS	78
6.2.1	EXPANSÃO DO CONJUNTO DE DADOS REAIS PARA APRIMORAMENTO DO MODELO	78
6.2.2	PUBLICAÇÃO E DIVULGAÇÃO DO PROJETO PARA A COMUNIDADE ACADÊMICA	79
6.2.3	ESTUDO DE IMPACTO A MÉDIO E LONGO PRAZO PARA ESTUDANTES E PROFISSIONAIS DE TI	80
6.2.4	ACESSIBILIDADE NO APLICATIVO	80
	REFERÊNCIAS	82

1. INTRODUÇÃO

1.1 Contextualização

O avanço tecnológico das últimas décadas transformou profundamente o mercado de trabalho, especialmente no setor de Tecnologia da Informação (TI). A rápida evolução desse campo tem gerado uma diversidade crescente na gama de carreiras, áreas de trabalho e especializações disponíveis, refletindo não apenas em novas oportunidades, mas também em desafios significativos para profissionais que buscam se posicionar de acordo com suas habilidades e características ou requalificar-se dentro deste mercado dinâmico, visto que diversas são as áreas. Conforme discutido por Frey e Osborne no artigo "The Future of Employment: How Susceptible Are Jobs to Computerization?"[FO13] as inovações tecnológicas têm provocado transformações profundas no mercado de trabalho, redefinindo segmentos já estabelecidos e impulsionando o surgimento e estabelecimento de áreas e carreiras no setor da Tecnologia da Informação. Essas mudanças criam oportunidades, mas sobretudo criam também desafios para trabalhadores e estudantes, que precisam alinhar suas habilidades às demandas emergentes e entender quais se encaixam melhor com o seu perfil técnico e emocional. Esse cenário ressalta a importância de compreender melhor as diversas áreas disponíveis na TI a fim identificar a que melhor combina consigo e extrair o máximo da capacidade numa carreira adequada às habilidades.

Além disso, em várias dessas áreas de atuação o estudante e/ou profissional de Tecnologia da Informação não sabe ao certo do que se trata e não sabe que suas características se encaixam como um profissional desse ramo da TI, diminuindo o leque de opções de qual carreira seguir ao ingressar no mercado de trabalho por conta dessa falta de informação. Ademais, alinhar suas competências técnicas e interpessoais com a área que se está ingressando vem se mostrando cada vez mais importante, como discutido por David Deming no artigo da Oxford Academic "The Growing Importance of Social Skills in the Labor Market"(2017) [Dem17], que destaca a importância crescente das habilidades técnicas e sociais alinhadas à vaga no mercado de trabalho atual, sobretudo as habilidades sociais, e como a combinação dessas habilidades é valorizada pelos empregadores, além de melhorar bastante a satisfação pessoal no trabalho.

Dito isso, um problema relevante no contexto da orientação de carreiras em TI é perceptível para quem está ingressando no mercado de trabalho ou mudando de área: a dificuldade de alinhar de maneira eficaz as habilidades, interesses e aspirações do indivíduo com as carreiras disponíveis. A complexidade do setor, somada à escassez de recursos que ofereçam orientação personalizada baseada em dados concretos e que explicam o que realmente faz um profissional dessa área do TI, torna a escolha profissional um processo

muitas vezes marcado por incertezas e indecisões. Essa lacuna na orientação eficaz pode levar a escolhas profissionais menos satisfatórias e, em última análise, a uma menor eficiência e produtividade tanto para os indivíduos quanto para o mercado de TI como um todo, como destaca o Mignone Center for Career Success da Harvard University no artigo "The Benefits of Aligning Your Career With Your Strengths"(2022)[Bat22], alinhar uma carreira com as forças individuais não só aumenta a confiança e o desempenho, mas também reduz o estresse e melhora o engajamento no trabalho, evidenciando a importância do profissional estar se dedicando a uma área que combine mais consigo e com suas forças, habilidades e interesses individuais.

1.2 Objetivos

Diante do cenário apresentado, de crescente diversidade e complexidade no mercado de Tecnologia da Informação (TI), este trabalho tem como objetivo entregar uma solução prática e acadêmica que auxilie estudantes e profissionais a tomarem decisões mais informadas sobre suas carreiras. Para isso, o projeto abrange duas frentes principais: o desenvolvimento de um modelo de aprendizado de máquina e a criação de um aplicativo que utilize desse modelo de classificação de aprendizado de máquina.

Na primeira frente, o objetivo foi ter desenvolvido um modelo de classificação baseado em técnicas avançadas de aprendizado de máquina, utilizando um conjunto de dados enriquecido e validado com dados reais e testes com usuários. Este modelo foi projetado para processar informações sobre o grau de conhecimento, afinidade e interesse do usuário em 17 áreas de estudo da TI, além de avaliar 10 habilidades sociais e comportamentais, ligados a uma área de atuação na TI. A meta é garantir que as recomendações geradas sejam precisas, personalizadas e fundamentadas em análises robustas, promovendo um processo de orientação que contribua para escolhas de carreira mais seguras e bem embasadas, para então ser usado no aplicativo.

Na segunda frente, o objetivo foi desenvolver o aplicativo Carreira pra TI, desenvolvido para a plataforma iOS, que se utiliza desse modelo. O aplicativo visa traduzir as recomendações do modelo em uma experiência prática, acessível e intuitiva para o usuário. Ele permite que os usuários explorem áreas recomendadas, obtenham gráficos detalhados sobre suas compatibilidades e acessem descrições práticas sobre cada área de atuação. Além disso, busca fomentar o autoconhecimento ao incentivar a reflexão sobre competências técnicas e comportamentais.

Com essas duas frentes, o objetivo central deste trabalho é oferecer uma ferramenta que promova o alinhamento entre habilidades individuais e as demandas do mercado de TI, contribuindo para o desenvolvimento acadêmico da área, a realização profissional dos usuários e a escolha mais informada de carreiras.

2. DESCRIÇÃO DO PROBLEMA E TRABALHOS RELACIONADOS

Esta seção apresenta a descrição do problema, a fundamentação teórica, ancorada pela literatura, para apoiar o desenvolvimento do projeto destinado à recomendação personalizada de carreiras em TI e justificando a sua importância, e os trabalhos relacionados com a temática do presente trabalho.

2.1 Descrição do Problema

O mercado de Tecnologia da Informação (TI) apresenta uma crescente diversidade e complexidade de carreiras, reflexo da rápida evolução tecnológica. Essa diversidade, embora crie novas oportunidades, também impõe desafios significativos para estudantes e profissionais que buscam alinhar suas habilidades e interesses às diversas possibilidades de atuação.

Devido a cada vez mais a área de TI se encontrar ramificada e específica, muitos estudantes enfrentam incertezas ao tentar identificar a área de TI mais adequada para suas aspirações. Problemas como a falta de conhecimento sobre as diferentes áreas, ausência de ferramentas específicas para orientação e a dificuldade em integrar habilidades técnicas e comportamentais ao processo decisório dificultam a escolha de carreira.

Além disso, o mercado de TI exige cada vez mais profissionais que não apenas possuam conhecimentos técnicos, mas também habilidades interpessoais e comportamentais alinhadas às exigências das diferentes áreas.

Esse cenário motivou o desenvolvimento deste trabalho, cujo objetivo é abordar o problema por meio de um sistema de recomendação integrado a um aplicativo. Essa solução busca oferecer recomendações precisas e personalizadas, ajudando estudantes e profissionais a tomarem decisões mais informadas sobre suas carreiras em TI.

2.1.1 Impacto do Problema

A falta de orientação clara na escolha de carreiras tem causado impactos significativos entre estudantes e profissionais da área de Tecnologia da Informação (TI). Esse problema é amplificado pela diversidade e complexidade crescentes no setor, que dificultam o alinhamento entre habilidades individuais e as oportunidades de carreira disponíveis, com os aspectos citados a seguir:

Insatisfação profissional

insatisfação com escolhas de carreira é um dos principais efeitos de uma orientação inadequada. Segundo o relatório "State of the Global Workplace 2023" da Gallup, disponibilizado pela Terra [Ter23], apenas 23% dos funcionários em todo o mundo se consideram engajados em seus trabalhos. No Brasil, esse percentual é ligeiramente maior, com 31% de engajamento na força de trabalho, mas ainda indica que a maioria dos profissionais não está plenamente satisfeita com suas ocupações. Esses dados sugerem que escolhas de carreira desalinhadas com as expectativas, afinidades, habilidades e interesses dos profissionais podem contribuir para baixos níveis de engajamento e satisfação no trabalho.

Baixa produtividade

A falta de alinhamento entre as habilidades e interesses dos profissionais e suas funções pode resultar em baixa produtividade. Um estudo feito pela Springer [AA23] indicam que trabalhadores que não se identificam com seus papéis têm menor probabilidade de atingir altos níveis de desempenho e satisfação, principalmente a médio e a longo prazo. No setor de TI, onde a demanda por alta performance é constante, essa desconexão pode levar a atrasos em projetos e qualidade inferior nos entregáveis, comparado ao potencial que o profissional poderia alcançar se estivesse em uma área mais adequada.

Necessidade de requalificação

A necessidade de requalificação profissional é um desafio significativo no setor de Tecnologia da Informação (TI). O relatório "The Future of Jobs Report 2020" do World Economic Forum [wef20] prevê que, até 2025, aproximadamente 50% de todos os trabalhadores precisarão de requalificação devido às mudanças nas demandas do mercado e à rápida evolução tecnológica. Essa projeção reflete não apenas a velocidade das inovações tecnológicas, mas também a importância de um alinhamento inicial adequado entre os profissionais e suas áreas de atuação. Estar em um segmento da TI que corresponda às aptidões e interesses individuais é crucial para a evolução contínua e para enfrentar os desafios impostos pelas transformações do mercado.

Consequências financeiras

Consequências financeiras da falta de orientação profissional são significativas. Estudos do National Bureau of Economic Research (NBER) [DvW11] revelam que profissionais que enfrentam demissões durante recessões podem sofrer perdas salariais de até 20% ao longo de 15 a 20 anos subsequentes. Essas perdas são atribuídas à dificuldade de realocação em posições equivalentes e à necessidade de aceitar empregos com remunerações inferiores.

neração inferior, quando não se identifica mais com um área ou com um segmento em que se atua. Esses dados enfatizam a importância de uma orientação profissional eficaz para mitigar impactos financeiros adversos.

2.1.2 Especificidade no Contexto de TI

O setor de Tecnologia da Informação (TI) apresenta características únicas que tornam a escolha de carreiras desafiadora para estudantes e profissionais. A rápida evolução tecnológica, aliada à diversidade de áreas e à demanda por habilidades técnicas e comportamentais específicas, intensifica as dificuldades no alinhamento entre aptidões individuais e as exigências do mercado. Essa variedade de opções, muitas delas recentes, pode levar à confusão sobre qual caminho seguir para aplicar melhor as habilidades técnicas e características pessoais. Conforme destacado nesse artigo do SENAC [Sen23], a constante inovação no setor de TI exige que os profissionais estejam sempre atualizados e preparados para se adaptar às mudanças, o que pode ser desafiador para aqueles que buscam uma direção clara em suas carreiras.

Falta de conhecimento sobre áreas disponíveis

A ampla gama de possibilidades de carreira no setor de Tecnologia da Informação (TI), que inclui áreas como desenvolvimento de software, cibersegurança, ciência de dados e gestão de projetos, é pouco compreendida por muitos estudantes e profissionais. A pesquisa feita pelo G1 [G121] indica que apenas 14,87% dos recém-formados em 2019 e 2020 conseguiram vagas em suas áreas de formação até três meses após a formatura. Esse desconhecimento pode levar à escolha de carreiras desalinhadas com as aspirações e habilidades individuais, resultando em insatisfação profissional.

Dificuldade em alinhar habilidades técnicas e comportamentais

A TI exige uma combinação equilibrada de competências técnicas e comportamentais, como habilidades interpessoais, adaptabilidade e criatividade. No entanto, muitos profissionais subestimam a importância dessas competências ou não conseguem identificar como suas habilidades se aplicam às diversas áreas do setor. De acordo com a International Labour Organization (ILO), 67% dos profissionais de TI sentem dificuldade em identificar como suas habilidades comportamentais podem ser aplicadas no mercado, limitando suas oportunidades de crescimento [Wes23].

Impactos na saúde mental e na produtividade

Escolhas de carreira inadequadas podem resultar em impactos negativos na saúde mental e na produtividade. Um estudo conduzido por Graziotin et al. (2017) [GFWA17] identificou 49 consequências da insatisfação entre desenvolvedores de software, incluindo efeitos prejudiciais no bem-estar mental, no processo de desenvolvimento e nos produtos finais, afetando tanto o indivíduo e a sua saúde mental, quanto a empresa e o mercado como um todo. Essas descobertas destacam a importância de um direcionamento assertivo na escolha de carreiras e áreas de atuação em TI.

2.1.3 A rotatividade como reflexo do desalinhamento

A alta rotatividade no setor de Tecnologia da Informação (TI) é um reflexo da falta de orientação adequada. Dados mostrados no artigo [Hub23] indicam que a taxa de rotatividade no setor de tecnologia é de aproximadamente 13,2%, superior à média de outras indústrias, que é de 10,5%. Esse índice reflete o desalinhamento inicial entre as expectativas dos profissionais e a realidade das áreas escolhidas, indicando a necessidade de soluções que promovam decisões mais informadas e alinhadas antes mesmo do ingresso no mercado de trabalho.

Como mostrado, a diversidade e complexidade do setor de TI demandam ferramentas mais personalizadas para orientação de carreira no início da carreira e no mercado de trabalho do profissional de TI, como o sistema proposto neste trabalho, para auxiliar a mitigar esses problemas supracitados.

2.2 Fundamentação Teórica

Nesta seção, são apresentadas as bases teóricas que fundamentam o desenvolvimento do projeto "Carreira pra TI". Foram realizadas revisões na literatura com o objetivo de identificar conceitos, estudos e metodologias relevantes que ofereçam suporte sólido à proposta do trabalho. Essa abordagem busca garantir embasamento acadêmico e técnico, fornecendo uma base estruturada para o desenvolvimento do modelo de aprendizado de máquina, bem como para a implementação do aplicativo, alinhando teoria e prática no contexto da orientação profissional em Tecnologia da Informação.

2.2.1 A Evolução do Mercado de Trabalho em Tecnologia da Informação

O setor de TI tem experimentado uma transformação significativa ao longo das últimas décadas, impulsionada por avanços tecnológicos rápidos e pela digitalização de diversos setores da economia. Este panorama é discutido por Castells em "The Rise of the Network Society"(2009)[Cas09], onde o autor explora a emergência da sociedade em rede e seu impacto na estrutura do trabalho, mesmo sendo um livro de anos atrás, ainda se mostra atual e como a tecnologia afeta todo o mercado de trabalho, principalmente o de Tecnologia da informação. A crescente demanda por profissionais qualificados em TI é evidenciada pela diversificação de carreiras e pela necessidade de habilidades especializadas, conforme destacado por David Autor, do MIT, em "The Polarization of Job Opportunities in the U.S. Labor Market"(2010)[Aut10], o aumento da demanda por habilidades cognitivas e interpessoais avançadas tem exacerbado as disparidades salariais e a importância de uma orientação profissional eficaz, evidenciando a importância de ingressar-se numa área e carreira que consiga explorar e valorizar os pontos fortes e que combine mais com as características do indivíduo, de maneira que tornará mais possível se destacar e se sentir mais realizado com a escolha de carreira que foi tomada. Além disso, conforme observado também por David Autor em "Skills, education, and the rise of earnings inequality among the 'other 99 percent'"(Science, 2014)[Aut14], que analisa a disparidade de renda como uma consequência da diferença das oportunidades para se adquirir novas habilidades técnicas gerais e específicas, se qualificar é algo que custa caro e não é de fácil acesso para as pessoas em geral, isso evidencia que escolher uma área adequada de acordo com suas habilidades e preferências se mostra muito importante, para valorizar o tempo, dinheiro e satisfação pessoal do indivíduo.

Adicionalmente, a crescente integração de tecnologias emergentes, como inteligência artificial, aprendizado de máquina, computação em nuvem, segurança cibernética, softwares e aplicativos para diferentes dispositivos, inteligência das coisas, entre outras, tem moldado o panorama do mercado de trabalho em TI e criado novas áreas que já se encontram bem estabelecidas, mas ainda desconhecida por muitas pessoas da própria área da TI. Estudos recentes, como os apresentados por Brynjolfsson e McAfee em "The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies"(2014) [DM14], destacam como essas tecnologias estão criando novas profissões enquanto modificam radicalmente as já existentes. Esses avanços, embora empolgantes, também trazem o desafio de manter os trabalhadores atualizados em um ritmo que acompanha a evolução tecnológica e como trazer profissionais de qualidade, que tem o perfil da área, para uma área emergente e uma incógnita para as próprias pessoas que possuem as características da área.

Por fim, como dito, a evolução tecnológica também trouxe maior segmentação dentro da própria área de TI, criando nichos de especialização como desenvolvimento de software, análise de dados, engenharia de redes, cibersegurança e design de interfaces. A escolha de uma carreira dentro desse ecossistema, que mais combine com as aptidões e interesses do profissional, não apenas potencializa suas chances de sucesso, mas também contribui para um mercado mais eficiente e produtivo.

2.2.2 Desafios na Orientação de Carreiras em Tecnologia da Informação

A orientação profissional em TI enfrenta desafios únicos devido à natureza volátil e altamente especializada do campo. Savickas, em *Career Construction Theory and Practice* (2005)[Sav05], ressalta a importância da adaptabilidade de carreira e de estratégias de construção personalizadas, destacando a necessidade de profissionais estarem atentos e adaptados ao inconstante mercado de trabalho, explorando mais profundamente áreas onde se sentem mais confortáveis e confiantes para atuar. No contexto da Tecnologia da Informação, onde novas especializações surgem com frequência e mudanças tecnológicas tornam habilidades obsoletas rapidamente, essa adaptabilidade é ainda mais crucial. Mesmo profissionais experientes frequentemente enfrentam a necessidade de mudar de área, tornando fundamental saber onde seus pontos fortes podem ser melhor aplicados e quais carreiras e áreas são mais alinhadas às suas competências e aspirações.

Além disso, a ampla gama de opções de carreiras na TI, que inclui desde desenvolvimento de software, análise de dados e engenharia de redes até áreas emergentes como inteligência artificial e segurança cibernética, muitas vezes dificulta para os indivíduos identificar qual caminho seguir. A falta de conhecimento sobre as especificidades de cada área, bem como sobre as habilidades técnicas e comportamentais necessárias para se destacar, é um dos principais obstáculos enfrentados por aqueles que buscam ingressar no setor ou mudar de carreira dentro dele. Segundo Brown e Lent em *Career Development and Counseling: Putting Theory and Research to Work* (2005)[BL05], o processo de tomada de decisão em carreiras envolve a integração de informações sobre interesses pessoais, competências e oportunidades disponíveis, destacando a importância de ferramentas e recursos que ofereçam informações claras e personalizadas para auxiliar nessa escolha.

Outro desafio significativo é a disparidade de informações disponíveis sobre as diferentes áreas da TI. Embora haja abundância de recursos técnicos e materiais disponíveis online, poucos desses materiais oferecem uma visão abrangente e acessível das possibilidades de carreira. Muitos profissionais e estudantes desconhecem as oportunidades oferecidas por áreas emergentes ou mesmo por segmentos tradicionais, simplesmente pela falta da exposição ou compreensão das informações do que se trata cada área e seg-

mentos na TI. Essa lacuna informacional pode fortemente levar a decisões de carreira mal fundamentadas, resultando em insatisfação e subutilização do potencial individual.

Adicionalmente, as exigências crescentes por habilidades híbridas – que combinam competências técnicas e sociais – tornam a orientação de carreira ainda mais complexa. Conforme discutido por Deming no artigo já citado, *The Growing Importance of Social Skills in the Labor Market* (2017)[Dem17], habilidades interpessoais são tão importantes quanto habilidades técnicas no mercado de trabalho moderno, especialmente em setores como TI, onde a colaboração e a comunicação eficaz desempenham papéis críticos no sucesso de projetos. Assim, é essencial que a orientação de carreiras considere tanto os aspectos técnicos quanto as habilidades sociais ao sugerir áreas e profissões.

2.2.3 Machine Learning e Seu Uso No Sistema De Recomendações de Carreira

O uso de Machine Learning tem se consolidado como uma abordagem essencial para resolver problemas complexos que demandam análise de grandes volumes de dados e personalização de soluções, ao identificar padrões complexos entre os atributos de dados reais e apresentar insights que dificilmente seria percebido pela pura percepção humana. No contexto deste projeto, a escolha pelo modelo de classificação como o modelo aprendido de máquina, não apenas enriquece as recomendações de carreiras na área de Tecnologia da Informação (TI), mas também responde a proposta do aplicativo e do presente trabalho. A aplicação de Machine Learning neste trabalho é guiada pela necessidade de oferecer recomendações personalizadas e fundamentadas baseadas em dados reais, ajudando profissionais e estudantes a alinhar suas habilidades e interesses com oportunidades de carreira relevantes, e entender, como as suas características técnicas e sociais podem definir de maneira direta a escolha da sua carreira e área de atuação.

Conforme discutido por Murphy em *Machine Learning: A Probabilistic Perspective* (2013)[Mur13], algoritmos de aprendizado de máquina são particularmente eficazes na identificação de padrões complexos em dados, permitindo uma análise precisa e detalhada. Essa capacidade é complementada pelo trabalho de Bishop em *Pattern Recognition and Machine Learning* (2006)[Bis06], que explora como técnicas de classificação podem ser aplicadas para mapear entradas de usuários – como habilidades técnicas, afinidades e interesses – para saídas predefinidas, como áreas de carreira e atuação. Assim, o modelo de aprendizado de máquina escolhido para este projeto é de classificação, uma abordagem que se mostrou ideal para correlacionar os dados de entrada dos usuários com as opções de carreira mais adequadas para as suas aptidões.

O modelo foi treinado com uma base de dados robusta, composta por mais de nove mil respostas, onde foi mesclado os dados do conjunto de dados existentes juntamente com as respostas coletadas durante o desenvolvimento deste trabalho, se tornando um

modelo robusta e realista para o fim. Essa base de dados abrange informações sobre 17 áreas de estudo da TI e 10 habilidades sociais, permitindo uma análise abrangente e multidimensional dos perfis dos usuários. O treinamento do modelo foi realizado utilizando o algoritmo de classificação Gradient Boosting Classifier, modelo de aprendizado supervisionado detalhado mais na próxima sessão, mas ideal para classificação de modelos grandes, com muitos atributos e relações complexas, onde foi identificar padrões consistentes, garantindo que as recomendações geradas sejam não apenas precisas, mas também personalizadas.

A escolha do modelo de classificação para este projeto foi motivada por três fatores principais. Primeiro, a precisão: algoritmos de classificação são capazes de identificar correlações complexas entre variáveis, proporcionando resultados altamente alinhados com os perfis dos usuários. Segundo, a escalabilidade: o modelo foi projetado para processar grandes volumes de dados e atender a um número crescente de chamadas sem perda de desempenho. Terceiro, a adaptabilidade: o modelo pode ser continuamente atualizado com novos dados, assegurando que as recomendações se mantenham relevantes à medida que o mercado de TI evolui.

Outro ponto fundamental abordado neste projeto é a personalização de acordo com a necessidade da ideia do aplicativo. Pedro Domingos, em *The Few Useful Things to Know About Machine Learning* (2012)[Dom12], destaca que os algoritmos de aprendizado de máquina de classificação se tornam mais eficazes quando conseguem adaptar suas saídas a partir dos dados de entrada, aprendendo com as interações e identificando os padrões complexos entre os atributos. Aplicando essa perspectiva, o uso de aprendizado de máquina se vê como parte fundamental do projeto, pois o modelo de classificação desse projeto utiliza as informações fornecidas pelos usuários para gerar recomendações específicas, alinhadas às suas habilidades, afinidades e objetivos profissionais, separadas pelas que mais combinam com o usuário.

A aplicação prática de Machine Learning neste projeto busca não apenas oferecer recomendações de carreira, mas também que o usuário perceba coisas no decorrer do aplicativo através do uso e da obtenção das respostas sobre quais as carreiras mais recomendadas ao final. Relações essas como a correlação existente entre a habilidade social da criatividade e a habilidade técnica de design gráfico tem e como elas podem influenciar você ser um bom desenvolvedor de jogos.

Por fim, este projeto demonstra como o aprendizado de máquina pode ser uma ferramenta poderosa para resolver desafios na orientação de carreiras em um setor tão dinâmico quanto a Tecnologia da Informação. A combinação de precisão, escalabilidade e adaptabilidade torna o modelo de classificação a opção ideal para atender às necessidades deste trabalho, oferecendo um impacto significativo tanto para os usuários quanto para o mercado como um todo.

2.2.4 Algoritmos Para Treinar O Modelo de Classificação Da Machine Learning

Neste trabalho, o algoritmo escolhido para treinar o modelo de classificação foi o Gradient Boosting Classifier (GBC), devido à sua capacidade de combinar precisão, flexibilidade e eficiência em cenários de alta dimensionalidade, como os apresentados no conjunto de dados utilizado. O GBC foi selecionado após uma análise comparativa com outros algoritmos populares, incluindo Random Forest, Support Vector Machines (SVM), MLP Classifier, Logistic Regression e Naive Bayes. Cada um desses algoritmos foi avaliado com base nas métricas apresentados na hora do treinamento (sessão de resultados), características, vantagens e limitações no contexto de um sistema de recomendação de carreiras na área de Tecnologia da Informação (TI).

Fundamentação teórica da preferência pelo algoritmo GBC

O Gradient Boosting Classifier se destaca por ser um algoritmo de aprendizado de máquina baseado em árvores de decisão que combina múltiplos modelos de baixa profundidade (árvores com baixa profundidade) em uma abordagem iterativa para criar um modelo robusto. Ele se mostrou superior por sua capacidade de lidar com dados complexos, detectar padrões não lineares e ajustar hiperparâmetros para minimizar erros de classificação, como foi possível observar através de suas métricas de precisão, recall, f1-score, Cohen's Kappa e Mean Squared Error (MSE), mostrado com mais detalhes na seção de Resultados. Conforme destacado por Hastie et al. em *The Elements of Statistical Learning* (2009)[HTF09], o Gradient Boosting é especialmente eficaz em cenários que exigem precisão e capacidade de generalização, tornando-o ideal para sistemas de recomendação baseados em perfis de usuários.

Friedman, J. H. em seu artigo "Greedy Function Approximation: A Gradient Boosting Machine"(2001) [Fri01a], destaca que o Gradient Boosting é altamente flexível devido à sua capacidade de ajustar parâmetros como taxa de aprendizado, número de estimadores e profundidade das árvores. Esses ajustes permitem que o modelo encontre um equilíbrio ideal para evitar tanto o overfitting, que é quando o modelo se adapta muito aos dados do treino e não generaliza bem, quanto o underfitting, que é quando o modelo não consegue pegar os padrões do conjunto de dados de treino, refinando as previsões iterativamente e garantindo um desempenho robusto, mesmo em problemas de alta dimensionalidade.

Além disso, a escolha do Gradient Boosting Classifier (GBC) foi motivada por sua habilidade em lidar eficientemente com alguns dados outliers e variáveis altamente correlacionadas, pelas técnicas de subsampling e regularização, características presentes no conjunto de dados utilizado neste projeto. O que contribui para aumentar a robustez do modelo e prevenir problemas de overfitting, uma característica crucial num modelo para

sistemas que requerem alta precisão e confiabilidade, como a recomendação de carreiras. Essa capacidade de classificar e gerenciar de forma eficaz os padrões de dados reais e complexos, somada a sua capacidade de evitar overfitting, posiciona o GBC como uma solução ideal para o objetivo deste trabalho.

Comparação Teórica com Outros Algoritmos Testados

Random Forest

O Random Forest, assim como o GBC, utiliza árvores de decisão para classificação. Sua principal vantagem está na robustez contra overfitting devido à agregação de múltiplas árvores independentes. Contudo, diferentemente do GBC, o Random Forest não ajusta as árvores de forma iterativa, o que limita sua capacidade de aprendizado progressivo. Breiman, em *Random Forests (2001)*[Bre01], aponta que o Random Forest é ideal para situações em que a interpretabilidade das árvores individuais é essencial, mas para o contexto deste trabalho, o refinamento iterativo do GBC mostrou-se mais eficaz para melhorar a precisão e as recomendações personalizadas, além de mostrar melhores métricas de precisão, recall, f1-score, assim como mostrou em relação a todos os outros em comparação.

Support Vector Machines (SVM)

Os SVMs são algoritmos poderosos para problemas de classificação em alta dimensionalidade, destacando-se por seu uso de hiperplanos para separar as classes. Conforme Vapnik descreve em *The Nature of Statistical Learning Theory (2000)*[Vap00], o SVM é particularmente eficaz em classificações binárias. No entanto, sua aplicabilidade em problemas multiclasse, como neste projeto, requer técnicas adicionais que aumentam a complexidade computacional. Além disso, os SVMs não possuem a capacidade de aprendizado iterativo do GBC, o que os tornou menos adequados para capturar padrões complexos nos dados.

MLP Classifier (Redes Neurais)

O MLP Classifier (Multi-Layer Perceptron) oferece grande flexibilidade para capturar relações não lineares nos dados, sendo amplamente utilizado em problemas de classificação. Porém, sua necessidade de um ajuste detalhado de hiperparâmetros e sua sensibilidade a dados desbalanceados foram desafios no contexto deste trabalho. Conforme discutido por Goodfellow et al. em *Deep Learning (2017)*[Hea17], redes neurais são poderosas, mas frequentemente requerem grandes volumes de dados interligados e balanceados, e alto poder computacional para alcançar bons resultados, tornando o GBC uma alternativa mais eficiente.

Logistic Regression

A Regressão Logística é uma abordagem simples e eficaz para problemas de classificação linear. No entanto, como destacado por Hosmer et al. em *Applied Logistic Regres-*

sion (2013)[GHL13], ela tem limitações significativas ao lidar com dados não lineares ou de alta dimensionalidade. Para o contexto de recomendações personalizadas em TI, a Regressão Logística mostrou-se inadequada para capturar a complexidade e as interdependências das variáveis envolvidas.

Naive Bayes

O algoritmo Naive Bayes é conhecido por sua simplicidade e rapidez, sendo especialmente eficaz em classificações textuais. No entanto, sua suposição de independência condicional entre as variáveis limita sua aplicabilidade em cenários onde as variáveis possuem fortes correlações, como neste projeto. McCallum e Nigam, em *A Comparison of Event Models for Naive Bayes Text Classification* (1998)[MN01], destacam que essa suposição frequentemente reduz a precisão em problemas de alta complexidade, deixando-o pra trás.

Dito isso, dentre os algoritmos avaliados, o Gradient Boosting Classifier foi selecionado por sua capacidade de:

- **Lidar com dados complexos e não lineares:** O modelo é eficaz em capturar padrões complexos que outros algoritmos, como Regressão Logística ou Naive Bayes, não conseguem identificar.
- **Refinamento iterativo:** Sua abordagem de aprendizado progressivo garante que cada árvore contribua para corrigir os erros das árvores anteriores, aumentando a precisão.
- **Controle de overfitting:** Parâmetros como taxa de aprendizado e número de estimadores podem ser ajustados para evitar overfitting, garantindo generalização para novos dados.

Os resultados numéricos que comprovam a superioridade do GBC para o presente projeto em termos de precisão, recall e outras métricas serão apresentados na sessão de Resultados, enquanto as técnicas utilizadas para otimizar o modelo, como GridSearchCV e validação cruzada, serão detalhadas na próxima sessão, que unidos formaram um modelo com ótimos números e resultados práticos.

2.2.5 Técnicas de Treinamento de Machine Learning Usadas no Algoritmo

Neste trabalho, foram aplicadas diversas técnicas de machine learning para otimizar o treinamento do modelo de classificação, com o objetivo de aumentar sua precisão e robustez. A implementação combinou etapas de pré-processamento, otimização de hiperparâmetros e validação para assegurar a criação de um modelo eficiente e confiável,

chegando a até 10% de melhora nas métricas com a combinação correta do conjunto de técnicas. A seguir, detalhamos as principais técnicas utilizadas no desenvolvimento do algoritmo do modelo.

Pré-processamento dos Dados

Normalização com StandardScaler:

A normalização é uma etapa essencial em problemas de aprendizado de máquina e de classificação, especialmente quando os dados apresentam variáveis com valores e padrões muito diferentes. Conforme explicado por Bishop em *Pattern Recognition and Machine Learning* (2006)[Bis06], a normalização melhora a convergência de algoritmos baseados em gradiente e evita que variáveis com magnitudes maiores dominem o processo de aprendizado. No conjunto de dados utilizado neste trabalho, cada feature representa diferentes áreas de estudo da TI e habilidades sociais, que podem ter valores em diferentes faixas, dependendo das respostas dos usuários. A disparidade nos padrões pode causar problemas durante o treinamento do modelo, já que algoritmos baseados em gradiente, como o Gradient Boosting Classifier (GBC), são sensíveis às diferenças de magnitude e padrão entre variáveis.

Ao aplicar o StandardScaler, os dados são ajustados para possuir uma média igual a zero e uma variância unitária. Essa transformação garante que todas as features tenham a mesma importância durante o treinamento, evitando que variáveis com valores maiores dominem o aprendizado e que as com valores menores sejam negligenciadas. Esse equilíbrio é especialmente importante em conjuntos de dados de alta dimensionalidade, como o utilizado neste projeto.

Codificação de Rótulos com LabelEncoder

Para transformar as variáveis categóricas do conjunto de dados (Label) em valores numéricos, foi utilizado o LabelEncoder, que mapeia cada categoria a um valor inteiro único. Essa transformação é fundamental para algoritmos de machine learning que não conseguem lidar diretamente com dados categóricos. Hastie et al., em *The Elements of Statistical Learning* (2009)[HTF09], destacam a importância da preparação dos dados para garantir que o modelo compreenda adequadamente as relações subjacentes.

Seleção e Otimização do Modelo

Validação Cruzada com StratifiedKFold:

A validação cruzada com StratifiedKFold foi uma etapa essencial na seleção e otimização do modelo deste projeto, garantindo uma avaliação consistente e confiável da performance. Essa técnica é particularmente importante em cenários de classificação com valores desbalanceados e outliers, como o presente cenário, pois assegura que cada sub-

conjunto gerado para o treinamento e teste preserve as proporções originais das classes no conjunto de dados. Isso evita que o modelo seja enviesado para classes mais importantes pro modelo que está sendo treinado, promovendo uma avaliação mais justa e representativa. Conforme destacado por Kohavi em *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection (2001)*[Koh01], a validação cruzada é uma estratégia robusta para estimar a capacidade de generalização de um modelo, sendo amplamente utilizada em tarefas de aprendizado supervisionado, como é o caso do algoritmo que está sendo usado nesse projeto.

A abordagem *StratifiedKFold* divide os dados em k subconjuntos, no presente caso foi escolhido $k = 5$, alternando os subconjuntos utilizados para treinamento e teste em cada iteração. Essa técnica não apenas maximiza o uso eficiente dos dados disponíveis, mas também garante uma análise mais abrangente, considerando diferentes combinações de amostras para validação. No contexto deste projeto, a validação cruzada permitiu avaliar o desempenho do *Gradient Boosting Classifier (GBC)* em múltiplos cenários, assegurando que o modelo fosse capaz de generalizar adequadamente para novos usuários. Além disso, ao integrar essa estratégia com *GridSearchCV* para otimização de hiperparâmetros, foi possível refinar o modelo de forma iterativa, selecionando combinações de parâmetros que maximizassem a acurácia sem comprometer a capacidade de generalização, reduzindo assim os riscos de *overfitting* ou *underfitting*.

Otimização de Hiperparâmetros com *GridSearchCV*:

A otimização de hiperparâmetros é um passo crucial no desenvolvimento de modelos de machine learning, especialmente quando o objetivo é alcançar um equilíbrio adequado entre precisão e capacidade de generalização. Para este projeto, utilizou-se o *GridSearchCV*, uma ferramenta que realiza uma busca sistemática pelos melhores hiperparâmetros do modelo *Gradient Boosting Classifier (GBC)* a partir de possibilidades diferentes para cada tipo de hiperparâmetro, para alcançar a melhor combinação de hiperparâmetros pro treinamento do modelo. Essa abordagem, como dito, consiste em testar combinações predefinidas de valores para cada hiperparâmetro, como número de estimadores, taxa de aprendizado (*learning rate*), profundidade máxima das árvores e número mínimo de amostras por nó, entre outros, e selecionar a combinação que proporciona o melhor desempenho com base em métricas como acurácia.

O *GridSearchCV* é amplamente reconhecido por sua eficácia em encontrar configurações ideais de hiperparâmetros em espaços de busca bem definidos, conforme discutido por Bergstra e Bengio em *Random Search for Hyper-Parameter Optimization (2012)*[BB12]. Embora o artigo enfatize os méritos da busca aleatória (*random search*), os autores também destacam a relevância de métodos sistemáticos como o *Grid Search* em cenários onde o espaço de busca é limitado, mas as combinações testadas precisam ser exaustivamente avaliadas. No contexto deste projeto, o uso de *GridSearchCV* em conjunto com validação

cruzada permitiu não apenas refinar os parâmetros do modelo, mas também garantir que as melhorias fossem consistentes em diferentes divisões do conjunto de dados.

A aplicação do GridSearchCV foi essencial para explorar o impacto de diferentes configurações no desempenho do modelo, assegurando que a versão final do GBC estivesse otimizada para lidar com a alta dimensionalidade dos dados e a complexidade das inter-relações entre as variáveis. A escolha final dos hiperparâmetros será detalhada numericamente na seção de resultados, mas foi escolhida a combinação de hiperparâmetros que apresentou as melhores métricas e acurácias, evidenciando como essa estratégia contribuiu para alcançar um modelo mais robusto e eficiente.

2.2.6 Relatório de Classificação e Métricas de Avaliação

A avaliação do modelo foi conduzida de forma abrangente, utilizando um conjunto diversificado de métricas que permitiram analisar seu desempenho em diferentes perspectivas. O classification report foi gerado para fornecer detalhes sobre as métricas de precisão (precision), revocação (recall), f1-score e suporte (support) para cada classe. Essas métricas são cruciais para avaliar como o modelo equilibra previsões corretas e erros, especialmente em um cenário multi-classe como o deste projeto, onde a diversidade de classes exige uma análise detalhada de cada categoria. A precisão mede a proporção de previsões corretas em relação ao total de previsões feitas para uma classe, enquanto o recall avalia a capacidade do modelo de capturar corretamente os exemplos reais de uma classe específica. O f1-score, uma média harmônica entre precisão e recall, fornece uma métrica única para balancear esses dois aspectos, sendo especialmente útil em conjuntos de dados desequilibrados e como valores bem diferentes entre si no conjunto.

Além disso, métricas complementares foram empregadas para obter uma visão mais ampla do desempenho do modelo. O Mean Squared Error (MSE), que calcula o erro médio quadrático entre os valores previstos e os reais, foi usado para avaliar a capacidade do modelo de minimizar erros em relação ao conjunto de dados de validação. O Cohen's Kappa Score foi incluído como uma métrica robusta para medir a consistência entre as previsões do modelo e os rótulos reais, levando em consideração o acaso. Essa métrica é particularmente relevante para problemas de classificação multi-classe, pois fornece uma visão mais ajustada da qualidade do modelo além da acurácia bruta. O modelo final apresentou bons números e resultados, mostrando a importância de cada etapa e processo desse algoritmo para o treinamento do modelo de classificação, juntamente com as técnicas de aprendizado de máquina para aperfeiçoá-lo e tornar mais adequado.

Conforme discutido por Caruana e Niculescu-Mizil em "An Empirical Comparison of Supervised Learning Algorithms"(2006)[CNM06], o uso de múltiplas métricas de avaliação é essencial para assegurar que o modelo atenda aos requisitos específicos do problema

e seja capaz de capturar as nuances dos dados. A integração dessas métricas no processo de avaliação garantiu uma análise detalhada do desempenho do modelo, fornecendo insights que orientaram as iterações e ajustes no treinamento. Os resultados quantitativos dessas avaliações serão detalhados na seção de resultados, destacando o impacto de cada métrica na escolha do modelo final.

Apanhado das Técnicas:

A escolha dessas técnicas foi guiada pela necessidade de lidar com um conjunto de dados complexo e garantir que o modelo resultante fosse capaz de gerar recomendações precisas e personalizadas. O pipeline desenvolvido foi projetado para integrar as etapas de pré-processamento, treinamento e validação de forma coesa, permitindo ao Gradient Boosting Classifier operar de forma otimizada. Como apontado por Chen e Guestrin em "XG-Boost: A Scalable Tree Boosting System"(2016)[CG16], a combinação de técnicas como normalização, redução de dimensionalidade e otimização de hiperparâmetros é essencial para maximizar o desempenho de modelos baseados em aprendizado supervisionado, em especial os de Gradient.

Com essas técnicas, o modelo demonstrou capacidade de processar informações de entrada complexas e correlacionar habilidades e preferências dos usuários às carreiras mais adequadas. Os resultados quantitativos obtidos estão apresentados e discutidos na seção de Resultados, destacando numericamente a respeito do impacto das técnicas aplicadas na precisão e robustez do modelo.

2.2.7 Sistema De Recomendação

No contexto do projeto de recomendação de carreiras em Tecnologia da Informação (TI), um sistema de recomendação baseado em classificação desempenha um papel fundamental ao oferecer sugestões personalizadas de carreiras alinhadas às características únicas de cada usuário, que se apresenta como parte mais importante do trabalho. Como dito nas sessões acima, utilizando técnicas avançadas de Machine Learning e análise de perfis, o sistema analisa dados fornecidos pelos usuários, incluindo suas habilidades técnicas, sociais e interesses, para identificar as carreiras mais adequadas e fazer as recomendações. Essa abordagem não apenas simplifica a escolha de carreira em um setor complexo e em constante evolução, mas também promove decisões mais informadas e alinhadas às aspirações dos usuários.

Sistemas de recomendação são ferramentas essenciais para gerenciar a sobrecarga de informações e personalizar experiências. Como definido por Ricci, Rokach e Shapira, no livro *Introduction to Recommender Systems Handbook* (2011)[RRS10], esses sistemas filtram informações para prever preferências, fornecendo sugestões que atendem às necessidades e expectativas individuais de cada usuário. Empresas como Amazon, Net-

flix e Spotify têm demonstrado o poder dos sistemas de recomendação personalizados em melhorar a experiência do usuário e aumentar a retenção ao longo do tempo, justamente por oferecerem uma experiência personalizada. Inspirado por esses casos de sucesso, o sistema de recomendação desenvolvido neste projeto busca trazer essa mesma ideia para o campo da orientação de carreiras na TI, o de oferecer uma experiência personalizada em um campo cheio de informações.

O modelo de recomendação utilizado neste trabalho foi treinado com um conjunto de dados robusto contendo mais de 8600 amostras, representando uma ampla gama de perfis e preferências. Essa base de dados permitiu ao modelo identificar padrões e correlações que são fundamentais para oferecer recomendações precisas e personalizadas. Diferentemente de abordagens tradicionais, o sistema combina elementos de filtragem colaborativa e baseada em conteúdo. A filtragem colaborativa identifica padrões ao analisar preferências de múltiplos usuários, enquanto a abordagem baseada em conteúdo utiliza diretamente as características fornecidas pelo usuário, como interesses e níveis de conhecimento, para criar recomendações altamente contextuais e alinhadas ao perfil individual.

Além disso, a escolha de um algoritmo de classificação de machine learning para o desenvolvimento do sistema é justificada por sua capacidade de lidar com dados de alta dimensionalidade e detectar padrões complexos que seriam difíceis de identificar manualmente, sendo um modelo ideal para um problema de sistema de recomendação. Conforme destacado por Aggarwal em *Recommender Systems: The Textbook* (2016)[Agg16], sistemas de recomendação baseados em aprendizado de máquina são superiores em cenários onde a personalização e a escalabilidade são críticas. A aplicação de algoritmos de machine learning, como o Gradient Boosting Classifier, permite que o sistema não apenas aprenda com os dados históricos, mas também aprenda com a relação entre atributos muito diferentes, como por exemplo a habilidade técnica em computação forense e a habilidade social de extroversão.

Por fim, o sistema desenvolvido neste projeto vai além das abordagens tradicionais ao incorporar variáveis que avaliam tanto habilidades técnicas quanto sociais do indivíduo para oferecer recomendações, afinal um bom profissional não se faz apenas de boas habilidades técnicas, mas sim de todo o conjunto. Essa integração de múltiplas dimensões no processo de recomendação reflete uma visão holística do usuário, promovendo não apenas escolhas de carreira mais alinhadas, mas também um maior autoconhecimento e valorização de suas competências.

2.3 TRABALHOS RELACIONADOS

2.3.1 Predicting Personalized Academic and Career Roads: First Steps Toward a Multi-Uses Recommender System

O artigo *Predicting Personalized Academic and Career Roads: First Steps Toward a Multi-Uses Recommender System*, de Nadjem et al. (2020) [NTMEB+18], explora um sistema de recomendação focado em prever trajetórias acadêmicas e profissionais personalizadas. Este sistema utiliza dados provenientes de currículos e redes sociais profissionais, como Viadeo, para analisar e categorizar passos acadêmicos e de carreira em conceitos mais amplos, como "ciência da computação" ou "energia e meio ambiente". O objetivo central do trabalho é sugerir próximos passos ao usuário, baseando-se em informações anteriores, como diplomas obtidos, experiências profissionais e intenções futuras.

Um aspecto relevante destacado pelos autores é a capacidade do sistema de lidar com reorientações de carreira, ou seja, situações em que o usuário opta por mudar significativamente sua trajetória. Para isso, o sistema utiliza categorias amplas (conceitos) para simular uma visão imprecisa do futuro do usuário, garantindo que as recomendações sejam flexíveis e personalizadas. O estudo enfatiza a importância de interações iterativas com o usuário para refinar as sugestões e aumentar a precisão das previsões. Os resultados apresentados mostram que conceitos extraídos do primeiro emprego após a obtenção de um diploma têm um impacto significativo na escolha da próxima etapa da carreira, destacando a importância de capturar e modelar as intenções do usuário.

No contexto do projeto *Carreira pra TI*, o trabalho de Nadjem et al. é particularmente relevante, pois compartilha objetivos similares de personalização e análise de trajetórias. Enquanto o sistema de Nadjem et al. foca em categorizar dados de usuários para prever passos futuros, o *Carreira pra TI* utiliza técnicas de aprendizado de máquina, como o Gradient Boosting Classifier, para analisar habilidades técnicas e sociais, correlacionando-as a áreas de TI específicas. Ambos os trabalhos demonstram a importância de abordar a complexidade das trajetórias profissionais com sistemas flexíveis e interativos, capazes de incorporar intenções e dados passados do usuário, a fim de recomendar carreiras e passos pro futuro.

Além disso, o estudo de Nadjem et al. reforça a necessidade de normalizar dados biográficos e padronizar nomenclaturas para melhorar a precisão do sistema de recomendação. Essa abordagem é refletida no *Carreira pra TI*, que normaliza as entradas do usuário para garantir uma análise mais eficiente e comparável entre diferentes perfis. A ênfase em interações iterativas para refinar as previsões também está alinhada com os objetivos do *Carreira pra TI*, que usa de um algoritmo de treinamento supervisionado para treinar o

modelo, destacando a importância de sistemas que não apenas sugerem carreiras, mas também ajudam os usuários a explorar opções de forma informada e proativa, e aprenda com seus dados.

2.3.2 Guia-me: Um Sistema de Recomendação para Ajudar a Iniciar ou Migrar de Carreira na Área de TI

O artigo Guia-me: Um Sistema de Recomendação para Ajudar a Iniciar ou Migrar de Carreira na Área de TI [ELL24] propõe uma solução para os desafios enfrentados por iniciantes e profissionais em transição de carreira no competitivo mercado de trabalho de Tecnologia da Informação (TI). O sistema, denominado "Guia-me", busca orientar usuários por meio de recomendações personalizadas que abrangem cursos, certificações, vagas e empresas relevantes, com base no perfil do usuário. Ele se destaca por fornecer um roteiro estruturado e acessível, visando capacitar e qualificar indivíduos para enfrentar as exigências técnicas e sociais do setor da TI.

O sistema utiliza um algoritmo de recomendação baseado em filtragem de conteúdo com similaridade de cossenos, que analisa as características do usuário e os itens disponíveis para oferecer sugestões alinhadas às suas preferências e objetivos. Além disso, o Guia-me também explora técnicas de filtragem colaborativa para identificar padrões de comportamento semelhantes entre usuários, ampliando as possibilidades de recomendação. A plataforma considera informações como habilidades técnicas, preferências de linguagens de programação, localização geográfica e nível de experiência, permitindo que as recomendações sejam ajustadas de forma precisa às necessidades específicas de cada usuário.

Outro ponto relevante é a ênfase do sistema em expandir o horizonte dos usuários, apresentando-lhes opções de carreira que talvez não tenham considerado anteriormente, assim como acontece no Carreira pra TI. Essa abordagem visa não apenas facilitar a entrada no mercado de TI, mas também incentivar os profissionais a explorarem novas áreas de atuação. Por meio de uma interface acessível e intuitiva, o Guia-me promove um ambiente de aprendizado contínuo, ajudando os usuários a identificar lacunas em suas competências e planejar sua qualificação de maneira eficiente. A abordagem estruturada do sistema busca mitigar os desafios impostos pela rápida evolução tecnológica e pela competitividade do mercado, garantindo que os usuários tenham acesso a recursos atualizados e relevantes, e de carreiras que eles podem seguir.

No contexto do Carreira pra TI, o Guia-me oferece um exemplo significativo de como sistemas de recomendação podem ajudar a superar as barreiras enfrentadas por profissionais em transição de carreira. Enquanto o Guia-me foca em sugerir vagas e cursos, o Carreira pra TI utiliza aprendizado de máquina avançado, como o Gradient Boosting

Classifier, para correlacionar habilidades técnicas e sociais a carreiras específicas da TI. Ambos compartilham a premissa de personalizar as recomendações sobre TI com base em dados detalhados e de promover um aprendizado contínuo, permitindo que os usuários alcancem sucesso em um setor dinâmico.

A integração de técnicas como a filtragem de conteúdo e a abordagem iterativa para refinar as recomendações destacam a relevância de sistemas como o Guia-me para enfrentar os desafios do mercado de trabalho. O Carreira pra TI compartilha essa visão, utilizando dados robustos e algoritmos supervisionados para oferecer sugestões que não apenas ajudam os usuários a planejarem sua trajetória profissional de forma informada e proativa, mas também atendem às demandas atuais. Ambos os sistemas demonstram a importância de combinar tecnologia avançada com estratégias centradas no usuário para maximizar o impacto das recomendações.

2.3.3 NEMO: Next Career Move Prediction with Contextual Embedding

O artigo NEMO: Next Career Move Prediction with Contextual Embedding, de Li et al. (2017)[LJT⁺17], aborda o desafio de prever os próximos passos na trajetória profissional de indivíduos, considerando a crescente mobilidade e globalização do mercado de trabalho. O estudo destaca a lacuna existente em soluções preditivas para identificar o próximo cargo e a empresa em que o indivíduo deve atuar, um problema muitas vezes negligenciado nas análises tradicionais de mobilidade profissional. O sistema proposto, denominado NEMO, é fundamentado em um modelo contextual baseado em Long Short-Term Memory (LSTM), que combina duas principais fontes de sinais preditivos: o contexto do perfil profissional e a mineração de trajetórias de carreira.

A metodologia do NEMO utiliza embeddings contextuais para representar, de forma compacta, informações sobre usuários, competências e empresas, permitindo capturar as relações latentes entre essas entidades. A partir dessas representações, o sistema explora dependências sequenciais nas trajetórias profissionais por meio de LSTMs, modelando padrões de continuidade e reorientação nas carreiras. Essa abordagem é particularmente eficaz para lidar com dados de alta dimensionalidade e heterogeneidade, como os encontrados em um extenso conjunto de dados do LinkedIn utilizado no estudo, que inclui informações sobre diplomas, cargos e empresas. Além disso, o artigo apresenta um critério adaptado de Mean Reciprocal Rank (MRR) para avaliar a eficácia das recomendações, mostrando que o NEMO superou significativamente métodos tradicionais em termos de precisão e relevância das previsões.

Outro ponto notável do artigo é a capacidade do NEMO de integrar intenções e preferências dos usuários no processo de recomendação. A partir de interações simuladas, o sistema refina suas sugestões com base nos feedbacks dos usuários, garantindo maior

aderência às suas expectativas e objetivos. Além disso, o estudo destaca a importância de utilizar informações de múltiplos níveis, como habilidades adquiridas, experiência passada e intenção futura, para construir um sistema de recomendação mais robusto e confiável. A análise revela que a inclusão de conceitos amplos, como categorias profissionais ou áreas de estudo, é fundamental para lidar com a incerteza associada às trajetórias profissionais.

No contexto do Carreira pra TI, o NEMO oferece insights valiosos sobre como integrar informações contextuais e sequenciais para personalizar recomendações de carreira. Enquanto o NEMO utiliza LSTMs para modelar trajetórias temporais, o Carreira pra TI se baseia no Gradient Boosting Classifier para identificar padrões complexos em conjuntos de dados que combinam habilidades técnicas e sociais. Ambos os projetos compartilham a premissa de que recomendações eficazes precisam considerar tanto o histórico quanto as intenções futuras dos usuários, oferecendo sugestões alinhadas às suas características únicas.

Além disso, o enfoque do NEMO em refinar as previsões por meio de interações iterativas com o usuário é um aspecto que se alinha diretamente com os objetivos do Carreira pra TI. Embora os métodos específicos de aprendizado de máquina sejam distintos, ambos os sistemas reconhecem a importância de capturar a complexidade das trajetórias profissionais para oferecer recomendações mais precisas e personalizadas. O estudo de Li et al. reforça a relevância de combinar dados contextuais e históricos para enfrentar os desafios de um mercado de trabalho em constante transformação, fornecendo uma base teórica sólida para o desenvolvimento contínuo do Carreira pra TI.

2.3.4 Unlocking Futures: A Natural Language Driven Career Prediction System for Computer Science and Software Engineering Students

O artigo *Unlocking Futures: A Natural Language Driven Career Prediction System for Computer Science and Software Engineering Students*, de Faruque et al. (2024) [FKA24], propõe um sistema de predição de carreiras baseado em técnicas de Processamento de Linguagem Natural (PLN) e aprendizado de máquina (ML). O objetivo central do estudo é auxiliar estudantes de Ciência da Computação (CS) e Engenharia de Software (SWE) na escolha de carreiras que melhor se alinhem às suas habilidades, interesses e atividades acadêmicas. Utilizando dados coletados diretamente de estudantes por meio de formulários, os autores desenvolveram um modelo de predição que considera o desempenho acadêmico, as competências adquiridas e os interesses pessoais para gerar recomendações de carreira.

O sistema utiliza técnicas de PLN para pré-processamento de dados e extração de características importantes dos perfis dos estudantes. Posteriormente, algoritmos de classificação de aprendizado de máquina e redes neurais são empregados para prever tra-

jetórias profissionais com base nas informações analisadas. O artigo destaca a relevância de considerar o desempenho em cursos específicos e as preferências individuais como fatores determinantes para as decisões de carreira. Além disso, os autores apontam que o sistema também pode ser útil para conselheiros acadêmicos, orientadores de carreira e especialistas da indústria, ampliando o impacto do modelo para além do contexto estudantil.

A relevância do trabalho reside na sua abordagem integrada, que combina técnicas avançadas de PLN e aprendizado de máquina para lidar com os desafios inerentes à tomada de decisão profissional. O estudo reconhece a complexidade de fatores que influenciam as escolhas de carreira e busca fornecer insights práticos para que os estudantes naveguem com confiança pelas diversas opções disponíveis no mercado de trabalho de TI. Além disso, os resultados indicam que o modelo é escalável e pode ser aplicado em contextos mais amplos, como o aconselhamento profissional e a gestão de talentos.

No contexto do projeto Carreira pra TI, o trabalho de Faruque et al. é relevante por utilizar técnicas de pré-processamento para o melhor tratamento dos dados na hora de treinar o modelo, no caso foi usado de processamento linguagem natural, normalização e redimensionalidade, para assim usar do aprendizado de máquina para prever trajetórias de carreira, alinhando-se ao objetivo de fornecer recomendações personalizadas aos usuários. Ambos os trabalhos destacam a importância de sistemas flexíveis e interativos que incorporam dados reais passados e intenções dos usuários para recomendar carreiras e passos futuros de forma informada e proativa.

2.3.5 Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia

O artigo "Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia," de Siswipraptini et al. (2024) [SSWRB24], propõe um modelo de recomendação de carreiras personalizado (CPRM) para ajudar estudantes de ciência da computação a selecionar carreiras alinhadas às suas habilidades, interesses e perfis de personalidade. Este estudo aborda um desafio significativo enfrentado por estudantes de TI: a dificuldade de identificar carreiras compatíveis com suas características individuais, devido à falta de orientação adequada e ao excesso de informações disponíveis. A proposta visa otimizar o processo de escolha de carreira utilizando um modelo baseado em dados educacionais e teoria fundamentada (EDM-GT), implementado por meio de um algoritmo Naïve Bayes personalizado (p-NB).

O CPRM combina dados sobre perfis de trabalho, tipos de personalidade e disciplinas acadêmicas para gerar recomendações. A coleta de dados foi realizada com 104 estudantes universitários indonésios de ciência da computação, que forneceram informações sobre seus interesses, preferências e trajetórias acadêmicas. Os resultados foram

validados por profissionais da área de TI e psicólogos, que avaliaram a relevância e eficácia do modelo em grupos focais. O estudo também analisou a aceitação do modelo entre os usuários, com mais de 83% dos participantes relatando satisfação com as recomendações fornecidas, destacando a robustez e a aplicabilidade prática do CPRM.

A metodologia do CPRM incorpora técnicas de mineração de dados educacionais para identificar padrões em grandes conjuntos de dados e propor recomendações personalizadas. Segundo os autores, o modelo também serve como um orientador acadêmico automatizado, ajudando os estudantes a selecionar disciplinas e carreiras com base em suas aspirações profissionais e necessidades do mercado. Essa abordagem inovadora não apenas considera habilidades técnicas, mas também enfatiza a importância de alinhar interesses pessoais e características de personalidade para melhorar a satisfação e o desempenho no ambiente de trabalho.

No contexto do Carreira pra TI, o trabalho de Siswipraptini et al. oferece insights valiosos sobre a integração de perfis acadêmicos e profissionais na geração de recomendações de carreira. Assim como o Carreira pra TI, que utiliza técnicas avançadas de aprendizado de máquina, como o Gradient Boosting Classifier, para analisar habilidades técnicas e sociais, o CPRM também foca na combinação de personalidade e competências acadêmicas para sugerir carreiras e disciplinas. Ambos os trabalhos compartilham o objetivo de personalizar recomendações para aumentar a precisão e a relevância das opções apresentadas aos usuários, contribuindo para a orientação profissional de forma eficaz.

Além disso, o CPRM destaca a importância da validação prática das recomendações, um aspecto também abordado pelo "Carreira pra TI." A utilização de especialistas em TI e psicologia para avaliar a eficácia do modelo ressalta a necessidade de combinar abordagens qualitativas e quantitativas para melhorar a experiência do usuário. Assim, ambos os sistemas reforçam a ideia de que a personalização e a integração de múltiplas dimensões dos dados são fundamentais para criar soluções mais abrangentes e impactantes no campo da recomendação de carreiras.

3. DESENVOLVIMENTO DO MODELO

3.1 Conjunto de Dados

3.1.1 Pesquisa sobre o conjunto de dados

Com o tema definido, iniciou-se a busca pelo conjunto de dados mais adequado para o treinamento do modelo de classificação. Essa etapa foi essencial para garantir que as features refletissem de forma significativa e atualizada as características relevantes para carreiras em Tecnologia da Informação (TI). Para isso, explorou-se extensivamente fontes especializadas, como Kaggle e Google Dataset Search, entre outros repositórios de dados confiáveis.

Após várias análises e comparações, foi encontrado um conjunto de dados promissor no Google Dataset Search, hospedado na plataforma Mendeley Data (<https://data.mendeley.com/datasets/5z68cvxssn/1>). Esse conjunto destacava-se por suas 27 features e 9179 linhas, incluindo informações detalhadas que correlacionavam habilidades técnicas e emocionais com áreas específicas de atuação em TI. Essa característica era um diferencial valioso, pois atendia ao objetivo de incluir dimensões técnicas e comportamentais no treinamento do modelo. Após uma análise comparativa com outros candidatos, concluiu-se que esse conjunto era o mais apropriado para o projeto, por possuir boas features, com áreas de estudo atuais e habilidades sociais importantes, como features ou entradas o conjunto conta com 17 áreas técnicas, que, de forma traduzida, são elas:

Conceitos de Banco de Dados, Arquitetura de Computadores, Sistemas Distribuídos, Segurança de Sistemas, Redes de Computadores, Desenvolvimento de Software, Habilidade de Programação, Gestão de Projeto, Computação Forense, Comunicação Técnica, IA e Machine Learning, Arquitetura de Software, Análise de Negócios, Habilidades de Comunicação, Ciência de Dados, Aptidão em Resolver Problemas, Design Gráfico.

Ou seja, áreas técnicas bastante atuais e de fato estudadas durante os cursos de TI, direta ou indiretamente. Além das 17 áreas técnicas, o conjunto também tem outras 10 entradas que correspondem a habilidades sociais, sendo elas:

Flexibilidade, Consciência, Agradabilidade, Inteligência Emocional, Conversação, Abertura à mudanças, Liderança, Adaptabilidade, Criatividade, Extroversão.

Habilidade sociais importantes na área da TI, algumas mais, e outras menos, para cada respectiva área de atuação na TI. Além disso, como saída do conjunto, o conjunto conta com 15 profissões diferentes no decorrer do conjunto, sendo elas, traduzidas para sua equivalente em português e a sua versão mais atual:

Especialista em IA e ML, Especialista em APIs, Desenvolvedor Mobile, Analista de Negócios, Executivo de Suporte ao Usuário, Especialista em Segurança Cibernética, Administrador de Banco de Dados, Desenvolvedor De Jogos, Engenheiro de Hardware, Engenheiro de Helpdesk, Engenheiro DevOps, Engenheiro de Redes, Gerente de Projetos, Engenheiro de Software, Testador de Software, Documentador Técnico

Os valores das áreas de estudos no conjunto iam de 0 (nenhum conhecimento na área de estudo e nenhum interesse) até 6 (dominante da área de estudo / profissional) e as habilidades sociais representava uma porcentagem, ou seja, iam de 0.0 (total desprovisamento da habilidade) até 1.0 (domínio da habilidade).

Devido a todas essas características, como a robustez, a temática, o número de linhas e a diversidade das saídas (áreas de atuação), o conjunto de dados foi o escolhido como base para o modelo de machine learning.

3.1.2 Volumetria

Número de amostras

O conjunto de dados dispõe de 9127 linhas, e como foi utilizado o método de validação cruzada nos algoritmos, 80% foram usadas como treino e 20% para teste e validação.

Distribuição das classes

Para cada uma das 15 saídas possíveis, o conjunto de dados disponibiliza entre 514 a 542 linhas, mostrando-se um conjunto balanceado entre as saídas.

Quantidade de features

O modelo dispõe de 27 features e 1 saída, no caso das 27, 17 são features relacionadas a áreas de estudo e 10 sobre habilidades sociais, além da saída, relativo às áreas de atuações, onde cada uma foi citada na sessão anterior.

3.1.3 Análise dos dados e treinamento inicial do conjunto de dados base

Após a escolha pelo conjunto de dados, iniciou-se o processo de análise mais profunda acerca do conjunto para obter insights e começar o treinamento. Ao começar a treinar com diferentes algoritmos básicos de classificação, como o Random Forest e o Support Vector Machine (SVM), notou-se que o conjunto tinha índices de precisão e Cohen's

Kappa sempre igual ou extremamente próximo a 100% no treinamento, mesmo nos primeiros treinos e testes, sem o ajuste dos hiperparâmetros do algoritmo e usar técnicas de aprendizado de máquina, algo raro de se ver, o que indicou que o conjunto de dados levava os modelos ao overfitting, não conseguindo generalizar, e ao desequilíbrio. Com essas suspeitas então foi realizado o processo de investigar visualmente, através de um gráfico, a relação de cada label (área de atuação) com as entradas (áreas de estudo e habilidades sociais) para identificar se estava ocorrendo desequilíbrio entre as diferentes entradas para cada label, o resultado encontrado para três labels diferentes se encontram abaixo, nesse gráfico que mostra a relação das features com a respectiva label:

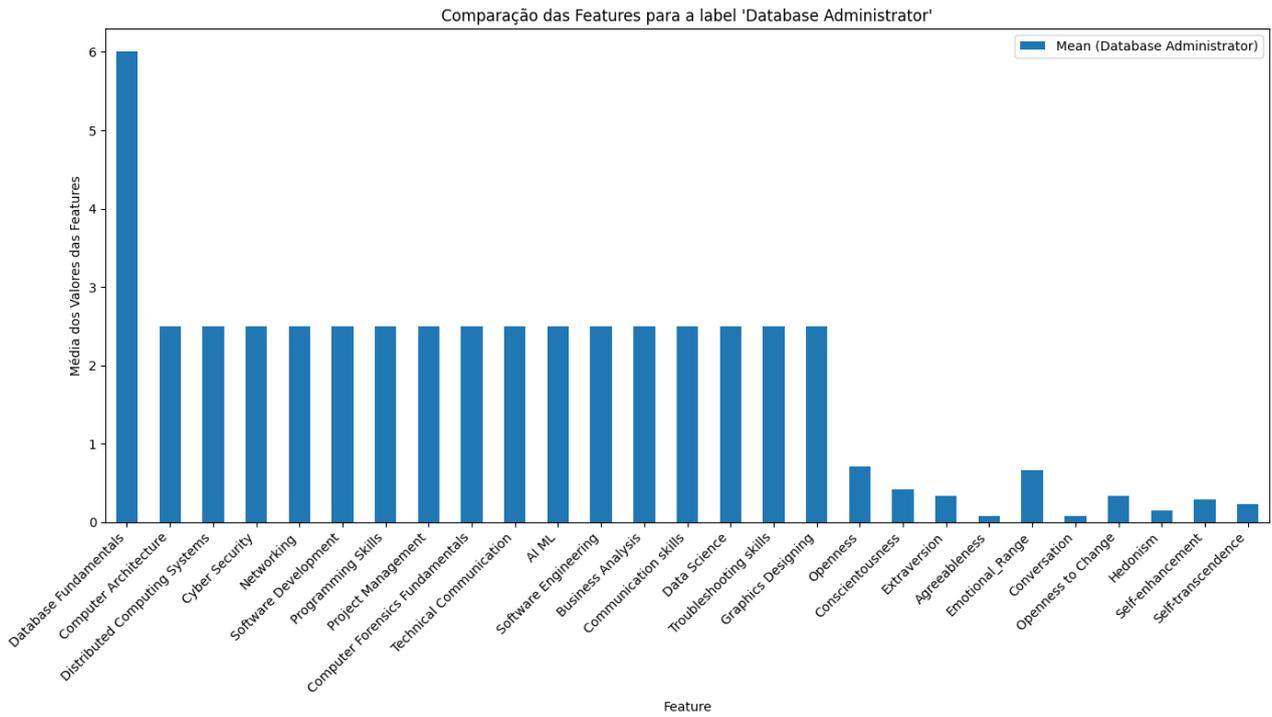


Figura 3.1 – Relação da carreira de Administrador de Banco de Dados em com as áreas de estudo e as habilidades sociais

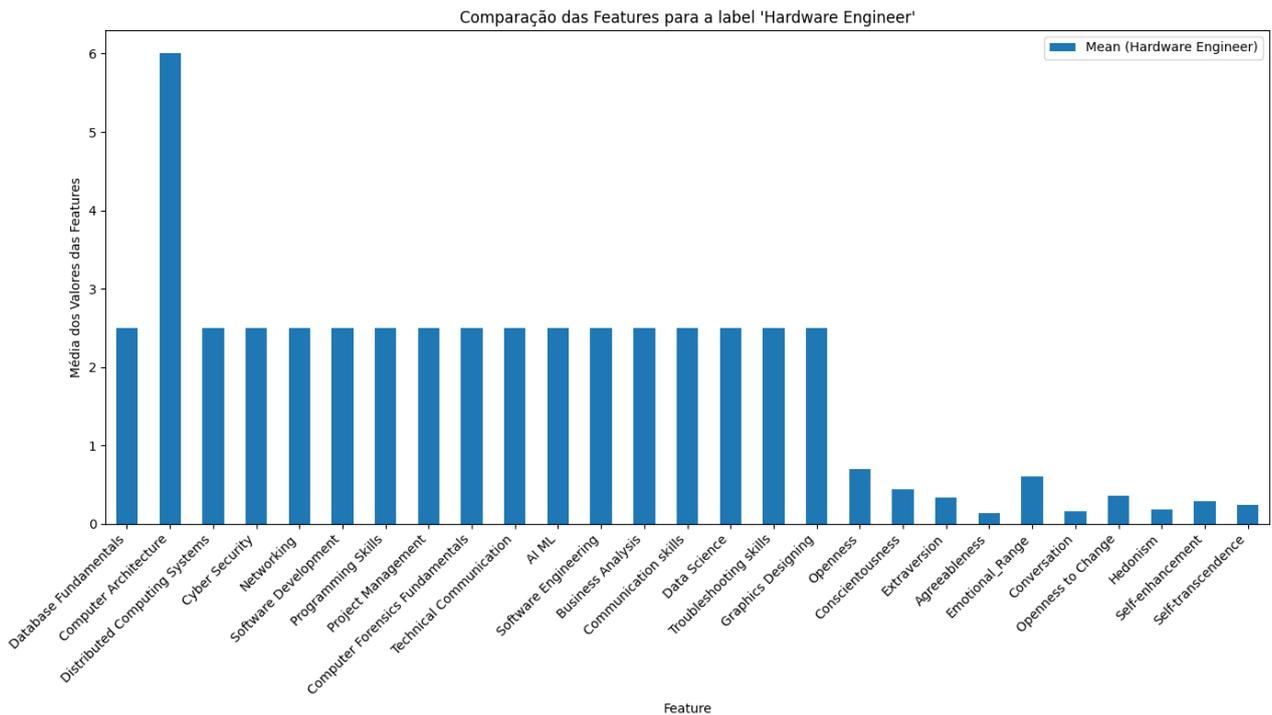


Figura 3.2 – Relação da carreira de Engenheiro de Hardware com as áreas de estudo e as habilidades sociais

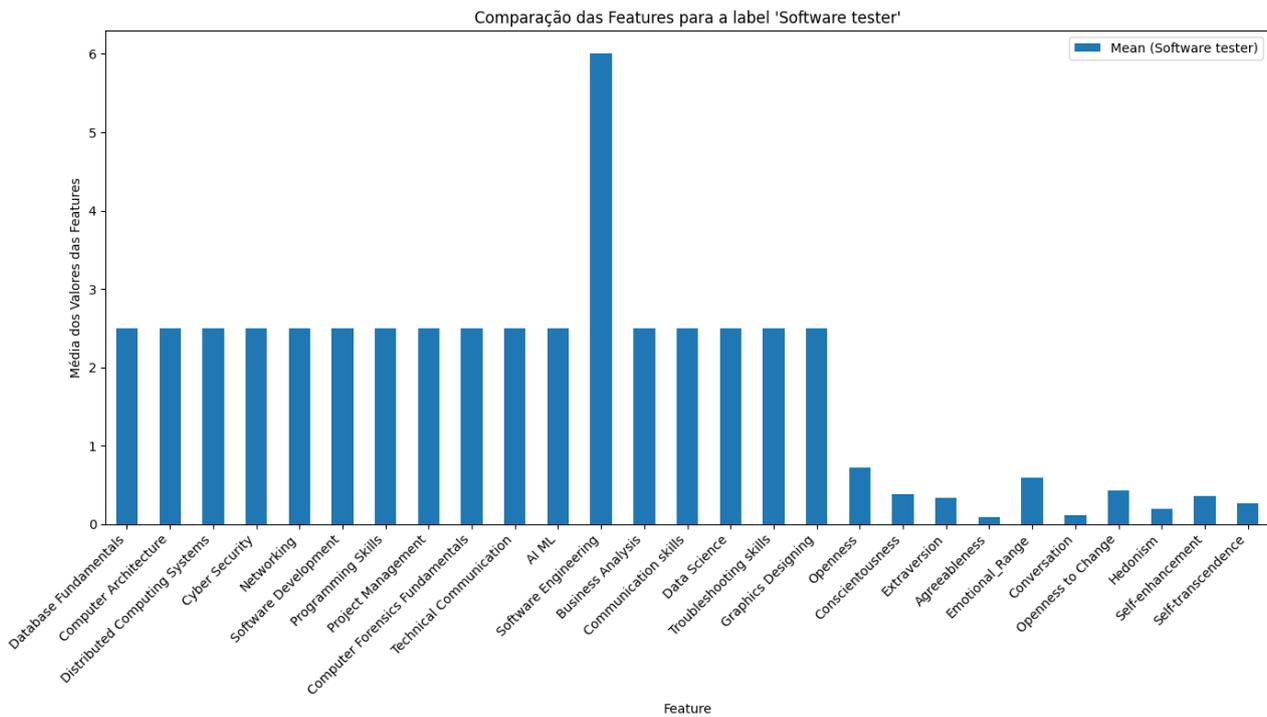


Figura 3.3 – Relação da carreira de Testador de Software com as áreas de estudo e as habilidades sociais

Como foi possível observar nesses três gráficos, uma das features se correlacionava diretamente com a label, e as outras tendiam a seguir um padrão de valores medianos ou muito abaixo. Então chegou-se a conclusão que o modelo majoritariamente tinha a característica da relação óbvia entre features e labels, que ocorre quando os dados contêm uma ou mais features (nesse caso, uma) que estão altamente correlacionadas ou que determinam diretamente a classe (como nesse caso, em que por vezes uma feature agia como um identificador da classe), isso mostrou que o modelo provavelmente estava se baseando majoritariamente nessa respectiva feature, praticamente ignorando as outras.

3.1.4 Pesquisa para equilibrar o conjunto de dados

Após identificar os problemas de overfitting e viés no conjunto de dados original, tornou-se evidente a necessidade de equilibrar as entradas para cada profissão. O objetivo então passou a ser mitigar a influência desproporcional de uma única feature nas previsões e tornar o modelo mais robusto, permitindo que ele generalizasse melhor para novas entradas.

Para alcançar esse objetivo, foi conduzida uma pesquisa prática, na forma de um questionário online e entrevista presencial, que consistiu na coleta de respostas de estudantes em final de curso e profissionais da área de Tecnologia da Informação (TI). Essa pesquisa teve como foco obter informações por meio de um questionário sobre os níveis de

conhecimento, afinidade e interesse em áreas de estudo e habilidades sociais do respondente, para isso a pesquisa foi conduzida com as mesmas questões do conjunto de dados original, ou seja, o respondente respondia o seu interesse na área de estudo entre 1 a 7, de acordo com a escala original (Nada interessado / Fraco / Iniciante / Médio / Intermediário avançado / Excelente / Profissional) e entre 0 a 100% o quanto possui da determinada habilidade social, assim então garantindo consistência com as features do conjunto original.

Essa pesquisa, por meios digitais, foi divulgada através da rede social LinkedIn, onde também foi feito contato com 8 à 12 profissionais de cada área de atuação do conjunto (16 áreas de atuação ao total), ou uma área muito similar em relação a buscada, e também pelo WhatsApp, para profissionais de TI conhecidos de cada determinada área, onde foi enviado o link da pesquisa para ser respondido. E a entrevista presencialmente ocorreu de forma similar, onde, quando possível, eu cheguei até o profissional de determinada área e fiz as mesmas perguntas do questionário online, nos mesmos moldes. Ao final era perguntado também a área de atuação do respondente / entrevistado, qual daquelas 16 áreas de atuação do conjunto era a sua ou estava diretamente ligado ao que exerce.

Com isso, foram obtidas 102 respostas, no qual 87 dessas foram obtidas pelo formulário online e 15 dessas através da entrevista presencial, que foram colocadas juntas. Em que no mínimo 5 profissionais de cada área de atuação responderam, ou seja, uma grande amostra e dados bastante representativos.

3.1.5 Usando o resultado da pesquisa para melhorar o conjunto de dados

Após a conclusão da pesquisa com 102 respondentes, foi necessário processar os dados coletados e integrá-los ao conjunto de dados original de forma que o modelo final equilibrasse adequadamente as entradas para todas as áreas de atuação. O objetivo foi mitigar os problemas de overfitting e garantir que o modelo de classificação pudesse generalizar melhor e também ficasse mais realista.

Processamento Inicial dos Dados

O primeiro passo foi baixar os dados da pesquisa em formato CSV e os organizar em uma planilha, e como segundo passo, foi realizado a conversão de todas as respostas das entradas em textos para valores numéricos compatíveis com a escala utilizada no conjunto original. Para isso, as respostas categóricas referentes às áreas de estudo foram convertidas para a escala do modelo original, de 0 a 6 (onde 0 correspondia a "Nada interessado", 1 a "Muito Básico", 2 a "Iniciante/Básico", 3 a "Intermediário", 4 a "Alto", 5 a "Avançado", 6 a "Dominante"). Já as respostas relacionadas às habilidades sociais, originalmente em uma escala de 0 a 100, foram normalizadas para o intervalo de 0.0 a 6.0.

Esse processo garantiu consistência e compatibilidade entre os dados coletados e o conjunto original, além da padronização deixando todas as entradas com um intervalo padrão, entre 0.0 e 6.0, de forma a toda ajudar o algoritmo de classificação na hora do treinamento.

Cálculo das Médias por Área de Atuação

Com os dados padronizados, os resultados foram organizados por área de atuação (16 no total). Para cada área, foram calculadas as médias das respostas para as áreas de estudo e habilidades sociais. Esses cálculos resultaram em valores médios entre a escala original entre 0.0 a 6.0, com isso essas médias foram convertidas em percentuais relativos, para cada característica em relação a área de atuação. Por exemplo, uma média de 4.2 para uma habilidade social relativa a uma área de atuação resultou em um percentual relativo de 70%, e assim foi sendo gerado para cada área de estudo / habilidade social em relação a área de estudo, isso ocorreu para melhor visualização das médias, porém o uso do conjunto permaneceu com os dados normalizados entre 0.0 e 6.0, assegurando uniformidade com os dados originais. Essas médias proporcionaram um conjunto estatístico realista sobre profissionais da TI e representativo das suas preferências e competências reais.

Como resultado final da pesquisa, foram obtidas os seguintes resultados de médias percentuais de relação ao quanto o respondente sente afinidade, interesse, conhecimento e confortável em relação às áreas técnicas e habilidades sociais:

Observação: para melhor visualização e para caber no texto, as 28 colunas foram divididas entre as 17 sobre as áreas de estudo + label e 10 sobre as habilidades sociais + label na tabela

TABELA COM MÉDIA DAS HABILIDADES SOCIAIS

Para referência, use os valores a seguir como nome original da label e as entradas, em ordem:

Label: Administrador de Banco de Dados, Engenheiro de Hardware, Desenvolvedor Mobile, Especialista em Segurança Cibernética, Engenheiro de Redes, Testador de Software, Especialista em APIs, Gerente de Projetos, Engenheiro DevOps, Redator Técnico, Especialista em IA e Aprendizado de Máquina, Desenvolvedor de Software, Analista de Negócios, Executivo de Atendimento ao Cliente, Engenheiro de Suporte Técnico, Desenvolvedor de Jogos

Valores da coluna para referência, em ordem: Flexibilidade, Consciência, Extroversão, Amabilidade, Inteligência Emocional, Conversação, Abertura para Mudanças, Liderança, Adaptabilidade, Criatividade

Label	Flex	Cons	Extra	Agree	Emoti	Conve	Open	Lead	Adapt	Creat
Database Admin	58.68	69.22	61.04	74.61	60.03	50.32	56.57	47.04	64.06	42.87
Hardware Engin	51.02	65.63	31.27	58.46	73.46	30.13	37.60	62.79	69.15	47.38
Mobile Develop	52.68	65.71	64.09	72.41	62.09	76.39	62.48	49.75	72.84	72.21
Cyber Security	23.20	85.40	45.34	43.42	76.09	39.22	39.47	72.95	42.53	37.52
Networking Eng	45.09	63.50	33.91	53.00	62.45	55.34	51.80	28.97	61.79	51.54
Software Teste	43.30	73.33	48.20	47.72	69.84	47.25	68.70	42.08	60.23	37.82
API Specialist	57.50	56.97	41.25	64.05	65.83	41.07	65.83	37.26	50.70	63.47
Project Manage	65.09	83.56	81.39	35.53	77.57	80.62	34.07	86.17	32.14	68.74
DevOps Enginee	51.65	58.45	70.03	72.43	74.90	75.10	71.86	65.24	65.45	69.78
Technical Writ	68.54	77.73	41.55	32.93	79.98	38.38	28.05	62.60	40.14	59.90
AI ML Speciali	59.01	74.18	51.24	65.20	68.26	48.62	47.71	54.91	69.10	77.67
Software Devel	67.54	77.40	71.97	67.59	55.24	72.15	59.35	39.07	63.62	77.70
Business Analy	69.75	44.47	71.15	65.76	57.87	68.20	48.30	63.59	65.20	59.40
Customer Servi	71.20	60.22	74.42	77.40	61.58	75.12	47.98	47.56	51.80	60.54
Helpdesk Engin	64.96	52.19	68.34	67.94	73.15	77.54	42.09	56.67	72.13	71.05
Game Developpe	65.46	74.75	33.98	47.92	57.31	37.54	37.52	53.92	49.98	78.12

Tabela 3.1 – Tabela com os valores das habilidades sociais do conjunto de dados obtidos através da pesquisa.

TABELA COM OS VALORES DAS ÁREAS DE ESTUDO PARTE 1:

Observação: A tabela precisou ser dividida em duas partes pois o texto não cabe uma tabela com 18 colunas

Label: Administrador de Banco de Dados, Engenheiro de Hardware, Desenvolvedor Mobile, Especialista em Segurança Cibernética, Engenheiro de Redes, Testador de Software, Especialista em APIs, Gerente de Projetos, Engenheiro DevOps, Redator Técnico, Especialista em IA e Aprendizado de Máquina, Desenvolvedor de Software, Analista de Negócios, Executivo de Atendimento ao Cliente, Engenheiro de Suporte Técnico, Desenvolvedor de Jogos

Valores da coluna para referência, em ordem: Fundamentos de Banco de Dados, Arquitetura de Computadores, Sistemas de Computação Distribuída, Segurança Cibernética, Redes, Desenvolvimento de Software, Habilidades de Programação, Gerenciamento de Projetos

Label	Data	Comp	Dist	Cyber	Netw	Soft	Prog	Proj
Adm de Banco de D	93.81	46.38	71.49	62.88	65.41	37.69	58.97	37.12
Eng de Hardware	45.71	93.23	61.19	66.42	61.6	54.37	60.56	42.66
Desenvolvedor M	59.0	42.77	35.33	41.31	57.73	82.57	70.63	27.65
Esp em Seg	57.3	54.26	52.89	91.88	68.91	31.55	62.97	23.32
Enge de Redes	66.13	80.24	59.97	72.58	86.65	54.62	62.13	35.33
Testador de Soft	58.18	39.68	55.39	63.7	33.25	75.99	70.34	54.02
Esp em APIs	72.41	49.8	72.57	58.17	85.7	75.19	84.5	65.95
Gerente de Proj	47.9	31.94	63.98	50.18	59.31	58.4	63.15	91.87
Eng DevOps	62.83	61.48	59.12	83.74	60.51	49.77	65.93	63.08
Redator Técnico	40.77	46.18	54.23	66.36	66.86	59.89	52.29	56.6
Espeem IA e ML	75.69	49.52	66.83	42.36	62.51	74.75	82.47	49.14
Dev de Soft	72.81	53.68	70.58	67.15	39.48	88.55	89.9	67.57
Analista de Neg	80.59	27.39	48.46	35.26	58.02	71.96	65.99	78.01
Exec de Atendimento	42.34	51.15	42.62	57.57	51.51	47.56	68.64	64.0
Eng de Suporte	41.62	52.7	48.54	55.33	77.93	42.06	46.54	50.02
Dev de Jogos	50.52	56.43	58.15	47.32	32.61	66.72	65.46	46.06

Tabela 3.2 – Tabela com a parte 2 dos valores das áreas técnicas do conjunto de dados obtidos através da pesquisa.

TABELA COM OS VALORES DAS ÁREAS DE ESTUDO PARTE 2:

Label: Administrador de Banco de Dados, Engenheiro de Hardware, Desenvolvedor Mobile, Especialista em Segurança Cibernética, Engenheiro de Redes, Testador de Software, Especialista em APIs, Gerente de Projetos, Engenheiro DevOps, Redator Técnico, Especialista em IA e Aprendizado de Máquina, Desenvolvedor de Software, Analista de Negócios, Executivo de Atendimento ao Cliente, Engenheiro de Suporte Técnico, Desenvolvedor de Jogos

Valores da coluna para referência, em ordem: Fundamentos de Banco de Dados, Arquitetura de Computadores, Sistemas de Computação Distribuída, Segurança Cibernética, Redes, Desenvolvimento de Software, Habilidades de Programação, Gerenciamento de Projetos

Label	Comp	Tec	AI ML	Soft	Busi	Comu	Data	Trou	Graphic
Adm de Banco de D	27.29	36.97	72.31	53.3	71.28	38.55	79.19	58.65	47.47
Eng de Hardware	69.77	79.07	44.19	53.6	37.91	57.93	43.59	74.58	62.38
Dev Mobile	42.34	57.03	31.08	71.58	36.35	67.51	39.63	82.87	54.39
Esp em Seg	75.31	52.32	32.66	48.19	39.83	57.92	44.79	81.29	50.52
Eng de Redes	58.77	36.14	49.01	56.84	36.69	59.76	28.92	73.85	40.66
Testador de Soft	38.9	70.8	63.02	77.56	38.3	71.22	62.83	87.29	43.95
Esp em APIs	38.63	76.3	23.07	59.49	71.99	80.59	29.39	59.51	52.14
Gerente de Proj	60.5	82.46	62.31	54.08	92.27	92.96	42.6	59.17	47.61
Eng DevOps	85.43	47.85	41.35	44.76	35.63	70.15	60.17	77.11	52.65
Redator Téc	46.79	90.88	61.5	53.69	50.84	94.29	45.59	64.13	60.19
Espem IA e ML	32.38	50.17	95.82	78.56	72.24	57.83	97.96	65.43	35.83
Dev de Software	37.72	54.75	72.52	97.49	63.78	52.01	58.0	74.05	51.88
Analista de Neg	40.14	81.96	69.8	78.83	91.07	80.09	83.52	45.79	30.88
Exec de Atendi	38.41	85.97	24.6	48.68	72.34	91.29	37.65	71.95	56.71
Eng de Suporte	43.08	71.37	37.16	38.8	39.3	86.45	31.11	92.73	49.39
Dev de Jogos	39.97	57.75	64.17	43.81	41.63	67.32	35.58	60.42	93.53

Tabela 3.3 – Tabela com a parte 2 dos valores das áreas técnicas do conjunto de dados obtidos através da pesquisa.

Com essas valores então foi feito a média foi feito a mescla entre 80% o valor da média da pesquisa e 20% do valor original do conjunto, isso foi feito para cada célula presente no conjunto de dados original, e assim foi feita o conjunto de dados final para ser treinado pelo algoritmo de classificação no modelo, descrito melhor na sessão posterior.

Integração com o Conjunto de Dados Original

Com as médias estatísticas da pesquisa em mãos, foi desenvolvido um método de integração que combinou esses valores com os dados originais do conjunto de treinamento. O processo, alinhado a práticas reconhecidas na literatura científica, foi realizado em três etapas principais:

Análise e Normalização dos Dados Originais: Primeiramente, os valores do conjunto original foram padronizados para a escala de 0.0 a 6.0, assegurando uniformidade com os dados da pesquisa. Essa normalização foi essencial para que os dois conjuntos pudessem ser combinados de forma harmoniosa. Estudos como "Combining datasets to increase the number of samples and improve model fitting" destacam que a padronização é um passo crítico para assegurar compatibilidade entre diferentes fontes de dados, permitindo a integração de forma robusta e consistente. **Combinação:** Para cada feature (área de

estudo ou habilidade social), foi atribuída uma ponderação relativa aos valores originais e às médias calculadas da pesquisa. Essa estratégia priorizou as médias da pesquisa (80%), considerando-as como uma base mais equilibrada e representativa, enquanto os valores originais contribuíram com 20% para preservar características do conjunto inicial. No artigo "Imputation procedures in surveys using nonparametric and machine learning methods: an empirical comparison", é discutido a eficácia de combinar múltiplas fontes de dados para melhorar a qualidade dos conjuntos utilizados em modelos de aprendizado de máquina.

Como fórmula de combinação foi aplicada em cada célula do conjunto, seguindo o cálculo abaixo:

$$\text{Média} = (0.8 \times \text{Pesquisa}) + (0.2 \times \text{ValorOriginal})$$

Essa técnica de mescla de dados é também uma forma comum nas técnicas do segmento de data augmentation, amplamente utilizada para enriquecer conjuntos de dados e aumentar a robustez de modelos, especialmente em cenários onde o conjunto original apresenta alguns problemas ou limitações.

Geração do Novo Conjunto de Dados: A partir dessa combinação ponderada, foi gerada uma nova planilha contendo as mesmas 9127 entradas e 27 features do conjunto original, agora ajustadas com os valores finais. O novo conjunto equilibrou as médias da pesquisa como uma base sólida, enquanto manteve elementos do conjunto original para preservar variações úteis para o treinamento do modelo. Trabalhos como "Combining datasets to increase the number of samples and improve model fitting" reforçam a validade de integrar múltiplas fontes de dados para aprimorar modelos, especialmente em contextos de aprendizado supervisionado.

Ao empregar essa abordagem, o trabalho buscou assegurar que o conjunto de dados resultante fosse representativo, robusto e adequado para treinamento, reduzindo vieses do conjunto original e garantindo maior generalização do modelo.

Resultados Encontrados pro Conjunto Final

O conjunto resultante refletiu uma abordagem balanceada, na qual os dados coletados da pesquisa corrigiram os desequilíbrios identificados no conjunto original, enquanto preservaram aspectos de sua estrutura inicial. Esse processo permitiu a criação de um conjunto de dados robusto, representativo e adequado para treinar o modelo de classificação com maior precisão e capacidade de generalização. Após esse processo, foi gerado novamente o gráfico para as três label que havia sido gerado anteriormente:

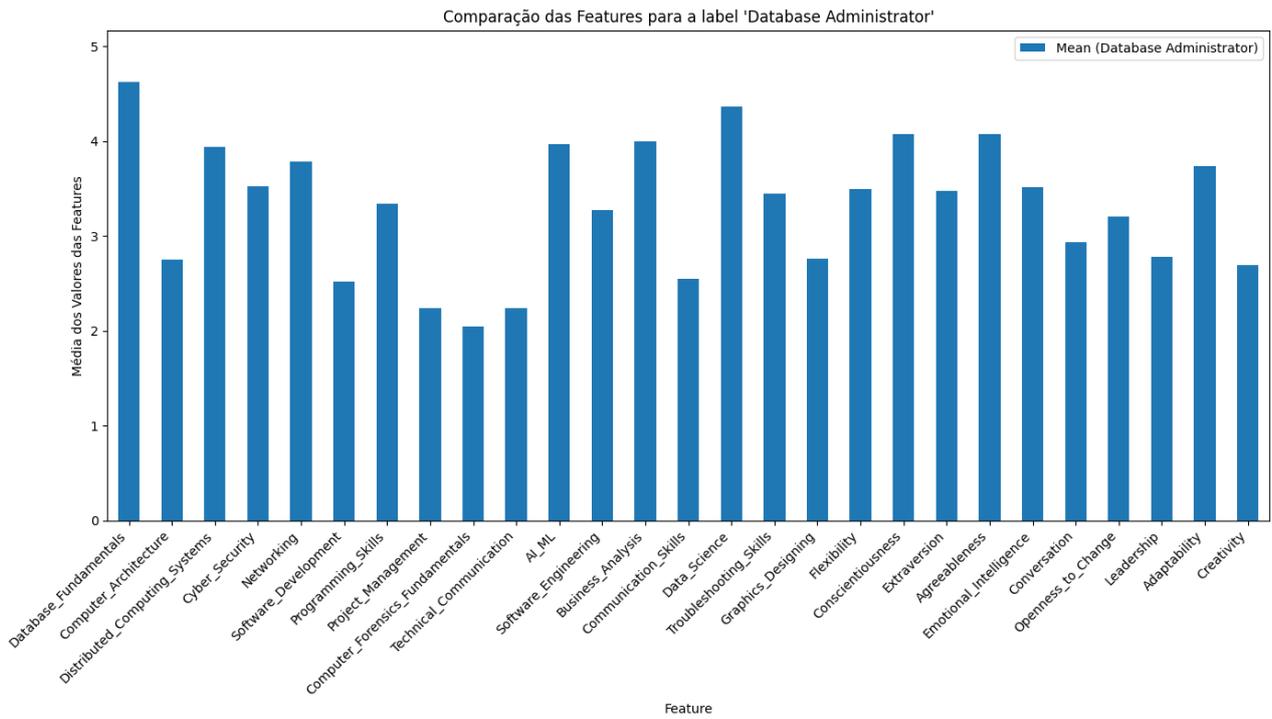


Figura 3.4 – Relação da carreira de Administrador de Banco de Dados em com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada

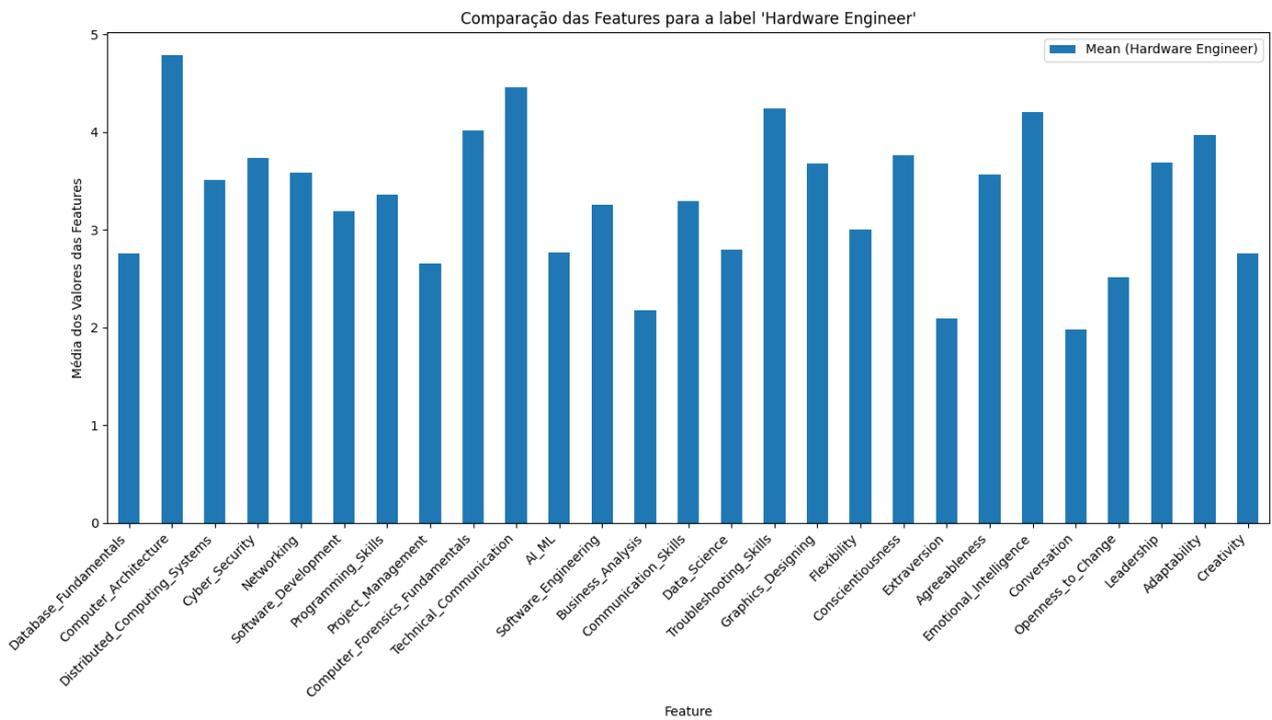


Figura 3.5 – Relação da carreira de Engenheiro de Hardware com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada

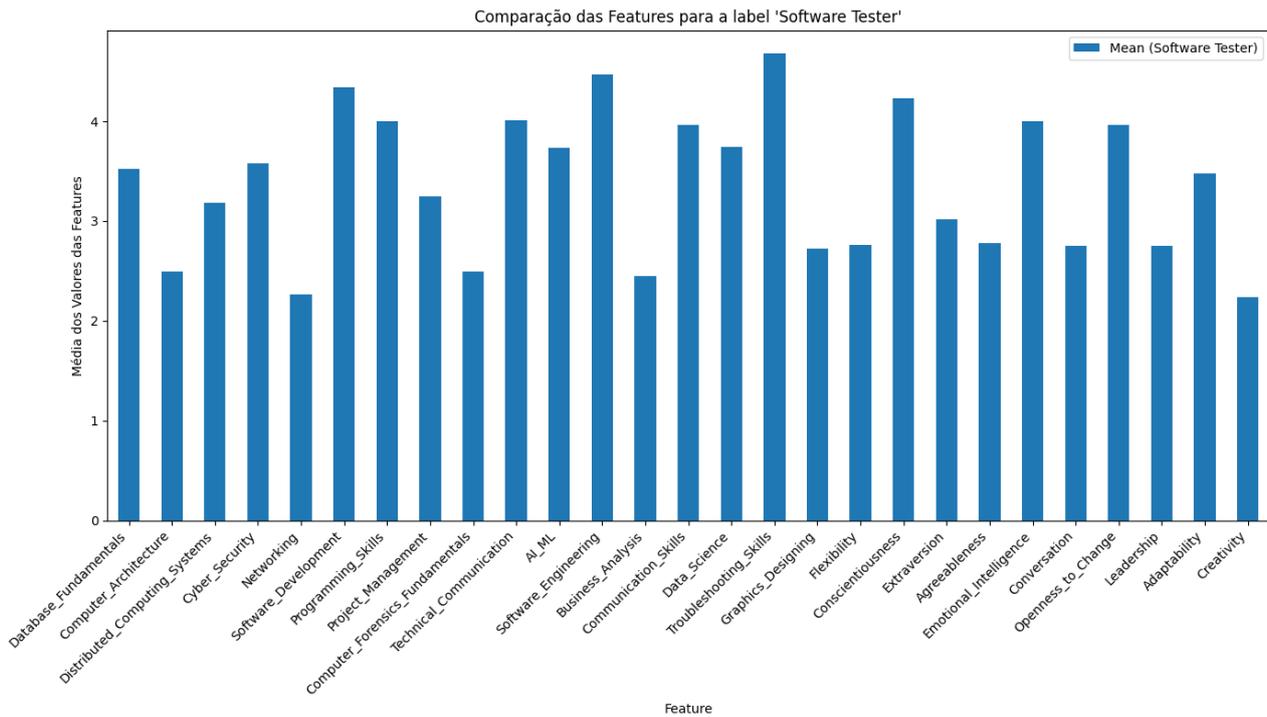


Figura 3.6 – Relação da carreira de Testador de Software com as áreas de estudo e as habilidades sociais após o equilíbrio do conjunto com a pesquisa realizada

Como é possível observar após a aplicação do resultado da pesquisa, em relação aos gráficos anteriores, as features para as áreas de entrada ficaram com um nível bem maior de equilíbrio e de ajuste, garantindo um modelo com bem mais generalização e de maneira a evitar o overfitting.

3.1.6 Métricas do Conjunto de Dados

Após as pesquisas com o questionário online e as entrevistas presenciais com estudantes de TI em final de curso e profissionais da área e todo o processo citado acima, chegou-se como resultado às seguintes médias para as features (áreas de estudo e habilidades sociais) em relação as labels (áreas de estudo):

Tabela 3.4 – Estatísticas Gerais do Conjunto de Dados

Métrica	Valor
Média Geral	3.4126
Desvio Padrão Geral	1.1646
Variância Geral	1.3659

Como o intervalo entre os valores do conjunto é pequeno (entre 0.0 a 6.0), o valores encontrado como Variância e Desvio Padrão são interpretados como bons e equilibrados.

3.2 Construção do modelo

O treinamento do modelo de classificação foi uma etapa essencial para o desenvolvimento deste projeto, envolvendo uma combinação de técnicas avançadas de aprendizado de máquina e estratégias de otimização. O objetivo central era garantir que o modelo fosse capaz de generalizar adequadamente, fornecendo recomendações precisas e personalizadas para os usuários, com base em dados consistentes e equilibrados.

3.2.1 Configuração Inicial e Pré-processamento

Após o processo de preparação e melhoria do conjunto de dados, descrito na seção anterior, o primeiro passo para o treinamento do modelo foi realizar o pré-processamento dos dados. Para isso, foram aplicadas as seguintes técnicas:

Normalização dos Dados

Para garantir uniformidade entre as variáveis e melhorar o desempenho do modelo, foi utilizado o `StandardScaler`, que ajusta os valores das features para uma distribuição com média próxima a 0 e desvio padrão próximo a 1. Essa abordagem é especialmente eficaz em cenários com alta dimensionalidade, como o presente conjunto de dados.

Codificação dos Rótulos

As classes de saída (labels), representando as áreas de atuação, foram transformadas em valores numéricos utilizando o `LabelEncoder`. Essa transformação era necessária para que o algoritmo de classificação pudesse processar adequadamente os dados.

3.2.2 Escolha do Algoritmo

O processo de escolha do algoritmo para o modelo de classificação foi uma etapa fundamental para garantir que o sistema atendesse aos objetivos do projeto: fornecer recomendações personalizadas, robustas e com alta capacidade de generalização. Essa deci-

são envolveu uma análise criteriosa baseada em fundamentação teórica, testes empíricos e, principalmente, métricas de avaliação. Para isso, diferentes algoritmos de aprendizado supervisionado foram submetidos a ciclos de treinamento intensivos, chegando alguns a demorar até aproximadamente 40 horas de treinamento para ser concluído o treinamento com as diferentes combinações de hiperparâmetros e diferentes técnicas de pré-processamento. Os algoritmos selecionados e testados para o treinamento do modelo de classificação foram Gradient Boosting Classifier (GBC), Random Forest, Support Vector Machines (SVM), MLP Classifier, Logistic Regression e Naive Bayes. A escolha desses algoritmos baseou-se em sua ampla aceitação na literatura científica e na disponibilidade de materiais que detalham seus funcionamentos, facilitando a compreensão e aplicação no contexto de aprendizado de máquina. Esses métodos são frequentemente citados como os mais eficazes para lidar com grandes volumes de dados, apresentando robustez e eficiência na manipulação de conjuntos com um elevado número de amostras e variáveis. Além disso, a diversidade dos algoritmos testados permitiu uma análise comparativa abrangente, garantindo a seleção de uma abordagem que melhor atendesse aos objetivos do projeto, tanto em precisão quanto em desempenho computacional.

Processo de Seleção, Treinamento e validação

Os algoritmos avaliados – Gradient Boosting Classifier (GBC), Random Forest, Support Vector Machines (SVM), MLP Classifier, Logistic Regression e Naive Bayes – foram submetidos a um processo estruturado de treinamento e validação cruzada, utilizando diferentes combinações de hiperparâmetros e técnicas de pré-processamento, a fim de achar a melhor combinação para cada algoritmo. O objetivo foi identificar o modelo mais adequado para capturar as complexas relações entre as 27 features e as labels do conjunto de dados, minimizando problemas como overfitting e underfitting, enquanto apresentava as melhores métricas de desempenho, como acurácia, precisão, recall, f1-score, Cohen's Kappa e Mean Squared Error (MSE).

Esse processo de avaliação criteriosa buscou não apenas analisar as métricas individuais, mas também a capacidade de cada algoritmo em equilibrar precisão, adaptabilidade e escalabilidade frente aos desafios impostos pela alta dimensionalidade e pelas variáveis correlacionadas do conjunto de dados. Dessa forma, foi possível identificar o modelo que melhor conciliou robustez, capacidade de generalização e desempenho no cenário proposto.

Usando o método de validação cruzada, as seguintes métricas foram considerados para avaliação do modelo:

Média de Precision, Recall e F1-Score: Para analisar o equilíbrio entre previsões corretas e erros em cada classe.

Cohen's Kappa: Para avaliar a consistência entre as previsões e os rótulos reais.

Mean Squared Error (MSE): Para identificar desvios médios nas previsões.

Após isso e ter todos os resultados das métricas, foi feito um levantamento com o resultado da melhor combinação de cada algoritmo, embora todos os algoritmos tenham apresentado resultados válidos em algum aspecto, o Gradient Boosting Classifier destacou-se em termos gerais. Ele demonstrou consistentemente as melhores métricas de desempenho, especialmente em cenários que exigiam o refinamento iterativo e a captura de padrões não lineares, e também apresentou melhores testes empíricos, onde foram feitas simulações com entradas guiados pseudo-aleatórios, para observar se as saídas realmente faziam sentido, o que foi testado para todos os algoritmos.

Abaixo é apresentado o resultado através das métricas supracitadas do treinamento com cada algoritmo e a combinação dos seus melhores hiperparâmetros, capturados com o auxílio do GridSearch:

MÉTRICAS	MLP	Rand For	GBC	SVM	Log Reg	Naive
Precisão	0.94	0.97	0.981	0.91	0.92	0.94
Recall	0.94	0.96	0.987	0.91	0.93	0.95
F1-Score	0.93	0.97	0.991	0.92	0.93	0.94
MSE Treino	3.2	1.3	0.15	5.141	3.932	1.782
MSE Teste	3.78	1.5123	0.09	5.912	4.734	2.194
Cohen's Kappa	0.923	0.975	0.989	0.895	0.925	0.941

Tabela 3.5 – Comparação das métricas dos algoritmos de classificação testados.

O Gradient Boosting Classifier (GBC) destacou-se pelas métricas como o algoritmo mais eficiente e alinhado aos objetivos do projeto devido à sua capacidade de capturar padrões complexos e não lineares presentes no conjunto de dados, além das razões teóricas citados na sessão de fundamentação teórica. Embora seu treinamento tenha demandado mais tempo, cerca de 42 horas ao todo, com 1200 fits (combinações), isso se deve à natureza iterativa do modelo e à necessidade de ajustar hiperparâmetros como o número de estimadores, profundidade máxima das árvores e taxa de aprendizado. Essa abordagem detalhada permitiu ao GBC refinar progressivamente suas previsões, minimizando erros e evitando tanto o overfitting quanto o underfitting. Conforme Friedman (2001)[Fri01b], essa flexibilidade e robustez tornam o GBC ideal para cenários de alta dimensionalidade e para projetos que exigem precisão e generalização, características fundamentais no sistema de recomendação proposto.

Essa abordagem foi essencial para explorar todo o potencial do modelo, com diferentes algoritmos, garantindo um equilíbrio entre precisão e capacidade de generalização para o modelo final.

Hiperparâmetros no uso do algoritmo GBC

No treinamento do modelo, o algoritmo Gradient Boosting Classifier (GBC) escolhido foi ajustado utilizando um conjunto de hiperparâmetros cuidadosamente definidos após uma exaustiva combinação de hiperparâmetros diferentes, onde a configuração mostrada abaixo foi obtida como a melhor entre todas as 1200 combinações testadas. Esses hiperparâmetros foram configurados para garantir um equilíbrio entre a capacidade do modelo de aprender padrões complexos e sua generalização para novos dados. Cada hiperparâmetro desempenha um papel crítico no comportamento do modelo, conforme descrito abaixo:

- **n_estimators (500)**: Define o número de árvores no modelo. Um valor de 500 foi escolhido para garantir um aprendizado profundo, mas sem exceder o tempo de treinamento ou introduzir overfitting excessivo.
- **max_depth (5)**: Limita a profundidade máxima de cada árvore. Um valor de 5 permite capturar relações importantes sem sobreajustar os dados.
- **learning_rate (0.01)**: Controla a taxa de aprendizado do modelo. Um valor baixo de 0.01 foi selecionado para obter um modelo robusto e reduzir a sensibilidade a ruídos nos dados.
- **subsample (0.8)**: Determina a proporção de amostras utilizadas para treinar cada árvore. Usar 80% dos dados em cada árvore adiciona uma regularização natural, ajudando a reduzir overfitting.
- **min_samples_split (5)**: Especifica o número mínimo de amostras necessárias para dividir um nó. Isso evita que divisões ocorram em nós com poucos dados, mantendo a estabilidade do modelo.
- **min_samples_leaf (3)**: Define o número mínimo de amostras em cada folha. Com 3 amostras por folha, o modelo evita a criação de nós muito específicos.
- **max_features ('sqrt')**: Controla o número máximo de features consideradas em cada divisão. O uso da raiz quadrada do número total de features é uma escolha comum, pois equilibra desempenho e eficiência.

A escolha desses hiperparâmetros foi baseada em experimentos preliminares e na literatura, garantindo que o modelo fosse capaz de lidar com os desafios do conjunto de dados, como dimensionalidade alta e variáveis correlacionadas.

Tabela 3.6 – Hiperparâmetros do GBC utilizados no modelo

Hiperparâmetro	Valor	Descrição e Interpretação
n_estimators	500	Número de árvores no modelo. Um número mais elevado de árvores, que garante aprendizado detalhado sem exceder o tempo de treinamento.
max_depth	5	Limita a profundidade das árvores, evitando overfitting e capturando relações importantes.
learning_rate	0.1	Taxa de aprendizado alta para maior aprendizado do algoritmo.
subsample	0.9	Sub-amostra integral de 90% dos dados em cada árvore para regularizar o modelo e evitar overfitting.
min_samples_split	5	Garante que divisões ocorrem apenas em nós com dados suficientes, melhorando a estabilidade.
min_samples_leaf	3	Previne a criação de folhas com poucos dados, garantindo nós mais generalizáveis.
max_features	sqrt	Considera apenas a raiz quadrada do total de features em cada divisão, equilibrando desempenho e eficiência.

A configuração desses hiperparâmetros foi essencial para garantir que o GBC capturasse padrões complexos no conjunto de dados, mantendo robustez e generalização. O aprendizado progressivo do modelo, aliado ao ajuste desses parâmetros, permitiu uma melhoria significativa nas métricas de desempenho, como precisão, recall, f1-score e erro médio quadrático (MSE). Essas escolhas também minimizaram problemas de overfitting, promovendo um modelo balanceado e eficaz para o contexto da aplicação.

3.2.3 Testes do Modelo de Machine Learning

O modelo de Machine Learning passou também por um processo rigoroso de validação. Os testes foram realizados utilizando entradas reais, com dados fornecidos por usuários reais e que estavam testando também a usabilidade do aplicativo, para avaliar a precisão das recomendações geradas pelo modelo e a sua satisfação, e também com entradas guiadas, com um conjunto de entradas cuidadosamente projetados para simular diferentes perfis de usuários, testando a robustez do modelo em diversos cenários, com as suas respostas.

O objetivo principal desses testes foi verificar se o modelo entregava respostas confiáveis e alinhadas às características dos usuários, e foi de fundamental importância para ajustes no algoritmo e modelo ao decorrer do desenvolvimento do aplicativo. Os resultados finais, com o algoritmo e hiperparâmetros já definidos, confirmaram que o modelo é eficaz em identificar padrões complexos e gerar recomendações personalizadas e relevantes.

3.2.4 Aspectos Técnicos do Desenvolvimento do Modelo

O modelo de Machine Learning foi desenvolvido utilizando a linguagem Python, devido à sua ampla utilização e suporte para tarefas de aprendizado de máquina. As bibliotecas Scikit-Learn e Pandas foram escolhidas como principais ferramentas para o desenvolvimento, permitindo tanto a manipulação de dados quanto a aplicação de algoritmos de classificação.

O processo de desenvolvimento do modelo envolveu várias etapas fundamentais:

-Pré-processamento dos Dados: Utilizando a biblioteca Pandas, o conjunto de dados foi organizado, limpo e preparado para o treinamento. As features foram padronizadas e as labels categorizadas, garantindo compatibilidade com os algoritmos escolhidos.

-Treinamento com Scikit-Learn: A biblioteca Scikit-Learn foi utilizada para implementar e treinar o modelo Gradient Boosting Classifier (GBC), conforme descrito na seção de metodologia. Diversos testes e ajustes de hiperparâmetros foram realizados para otimizar o desempenho do modelo.

Após o treinamento e validação em Python, o modelo final foi convertido para o formato CoreML, compatível com a biblioteca CreateML da Apple. Essa conversão foi necessária para integrar o modelo ao código do aplicativo, desenvolvido em Swift e SwiftUI. O processo seguiu os seguintes passos:

-Exportação do Modelo Treinado: O modelo treinado em Python foi exportado para um formato intermediário compatível, utilizando a biblioteca CoreMLTools, que facilita a conversão de modelos entre diferentes frameworks.

-Validação no CreateML: No ambiente de desenvolvimento da Apple, o modelo foi ajustado para atender aos requisitos da biblioteca CreateML, garantindo que mantivesse o mesmo comportamento e precisão observados em Python.

4. PROJETO DO APLICATIVO CARREIRA PRA TI

4.1 Objetivo

4.2 Desenvolvimento do Aplicativo

O *Carreira pra TI* é a interface e o sistema central do projeto, desenvolvido para integrar o modelo de classificação e oferecer aos usuários uma experiência interativa, intuitiva e informativa. O aplicativo permite que os usuários explorem os resultados gerados pelo modelo, naveguem por recomendações personalizadas de carreira e acessem descrições detalhadas das áreas técnicas e emocionais relacionadas ao seu perfil, tudo de maneira prática e acessível.

Por meio de uma interface amigável, o aplicativo coleta dados do usuário através de sliders, nos quais é possível atribuir valores de 0 a 100% para indicar o nível de afinidade, interesse e conhecimento em cada área técnica ou habilidade social. Com essas informações de entrada, o modelo de aprendizado de máquina processa os dados e retorna recomendações personalizadas, destacando as áreas de atuação em TI que melhor se alinham às características e competências do usuário.

Além das recomendações, o aplicativo também oferece ferramentas para visualizar gráficos detalhados, explorar carreiras sugeridas e compreender como as diferentes áreas de estudo se conectam aos interesses e habilidades do usuário. Esse sistema busca facilitar a tomada de decisão, promovendo um maior autoconhecimento e direcionamento profissional para os usuários no mercado de TI.

4.3 Interface do Usuário e Fluxo do aplicativo

O aplicativo será composto por seções principais, cada uma delas projetada para oferecer uma experiência de usuário fluida e eficaz. Parte dessas seções dizem respeito ao back-end, que são questões ligadas ao processamento dos dados do aplicativo, e outra parte estão ligadas ao front-end, ou seja, a parte visual do aplicativo. A seguir, detalho as funcionalidades e componentes de cada seção:

4.3.1 Tela de Boas-vindas e Introdução



Figura 4.1 – Tela 1 do OnBoarding



Figura 4.2 – Tela 2 do OnBoarding

Descrição: Ao iniciar o aplicativo, o usuário é recepcionado por uma série de telas de boas-vindas, conhecidas como onboarding, que apresentam de forma breve e clara as principais funcionalidades do aplicativo e seu objetivo central: recomendar carreiras personalizadas na área de Tecnologia da Informação (TI). Essas telas também incluem uma solicitação amigável para que o usuário insira seu nome, permitindo uma experiência mais personalizada durante a navegação.

Funcionalidades: O onboarding inclui botões interativos para avançar na navegação, permitindo que novos usuários explorem o aplicativo de forma guiada e progressiva, conhecendo o que o aplicativo faz. A interface é projetada para garantir uma navegação intuitiva e fluida, proporcionando uma primeira impressão acolhedora e profissional.

4.3.2 Formulário de Perfil do Usuário



Figura 4.3 – Tela do questionário sobre áreas de estudo e habilidades sociais



Figura 4.4 – Tela para tirar a dúvida sobre a área ou habilidade

Descrição:

Após o passar pelas telas de boas-vindas o usuário vai para as telas de questionário, uma etapa fundamental do aplicativo, onde o usuário fornecerá informações detalhadas sobre suas afinidades, habilidades, interesses e conhecimento. A interface utiliza um slider interativo que permite ao usuário avaliar seu nível de familiaridade e interesse para cada uma das 27 perguntas entre 0 a 100%, divididas em dois grupos:

Áreas de estudo e conhecimento técnico: Compostas por 17 perguntas, onde o usuário ajusta o nível de familiaridade com a área utilizando um slider de porcentagem para indicar o quanto possui, variando entre 0% e 100%, como ilustrado na Figura 4.3.

Habilidades sociais: Compostas por 10 perguntas, nas quais o usuário também utiliza um slider de porcentagem para indicar o quanto acredita possuir de determinada habilidade social, variando entre 0% e 100%.

Para maior acessibilidade, a tela conta com um recurso adicional: um pop-up informativo (Figura 4.4), que pode ser acionado para exibir explicações detalhadas sobre o tema abordado em cada questão, ajudando o usuário a tomar decisões mais precisas ao preencher o questionário.

Funcionalidades: Sliders para avaliar o nível de interesse e conhecimento em diferentes áreas da TI e conteúdos que lhe foram apresentados durante a sua graduação na área da TI e suas habilidades e características sociais. Botão informativo que abre um Pop-up para responder uma dúvida sobre do que se trata a questão.

4.3.3 Recomendação da carreira



Figura 4.5 – Tela do resultado

Descrição: Após preencher todas as áreas, o usuário vem para essa tela, que é onde o modelo de aprendizado de máquina de classificação é acionado, e após 11 interações e seguir a fórmula de pesos, dá o score de cada uma das áreas de atuação, então são escolhidas as três com maior score e então é mostrado ao usuário as suas recomendações 4.5, mostrando as carreiras que mais combinem com ele. E ao clicar no gráfico acima ou no nome em uma que ele queira saber mais detalhadamente das que lhe foram recomendadas, é encaminhado para uma tela com maior detalhes sobre a área de atuação e suas aplicações práticas.

Também há um gráfico abaixo onde é mostrado a média de cada campo de estudo e emocional da TI que foram segmentados. Como por exemplo a média do campo de Software, de Dados e IA ou a de habilidades pessoais, e ao selecionar é mostrado numa

label abaixo a informação com o nome do campo e a sua média, informando o usuário sobre os seus melhores campos na TI.

Também mais abaixo, há um botão que redireciona o usuário para responder o formulário sobre o que achou do aplicativo e do modelo de aprendizado de máquina, que garantiu o feedback no decorrer do processo de como estava a satisfação com o aplicativo e com o modelo, para saber onde ir melhorando ao decorrer do processo.

Funcionalidades: Gráfico em forma de um Donut Chart, que ao ser clicado leva para uma tela de detalhes da respectiva área de atuação. Mais abaixo gráfico onde mostra a média de cada campo da TI e das habilidades sociais, que ao ser clicado mostra o nome do campo e a média. E mais abaixo um botão que leva o usuário a responder a o formulário sobre a experiência de uso do aplicativo e do modelo.

4.3.4 Tela de detalhes sobre a carreira



Figura 4.6 – Tela de detalhes da carreira

Descrição: Ao clicar no gráfico ou no nome da carreira, o usuário vem para essa tela onde é descrito o que a carreira faz, mais detalhadamente, para o usuário entender melhor o seu resultado. Ela conta com uma imagem animada que representa o profissional da área, para maior representabilidade, acompanhado da porcentagem de recomendação dessa carreira vindo da tela, logo abaixo aparece o nome e resumidamente o que se espera do profissional dessa área, e mais abaixo está a lista com 4 aplicações práticas dentro da área.

Funcionalidades: Imagem no topo com a porcentagem de recomendação da área, e abaixo a área de texto onde é descrito detalhadamente as informações sobre a carreira e mais detalhes e suas aplicações práticas.

4.4 Desenvolvimento

O desenvolvimento do aplicativo "Carreira pra TI" foi guiado pelo modelo de Challenge Based Learning (CBL), uma abordagem educacional multidisciplinar que conecta aprendizagem a desafios do mundo real. O CBL incentiva as pessoas a utilizar tecnologia para solucionar problemas relevantes em suas vidas e comunidades, buscando uma solução a partir da percepção de um problema do convívio do desenvolvedor ou a sua volta.

No contexto deste projeto, o desafio escolhido foi baseado em uma dificuldade pessoal: encontrar uma carreira ideal na área de Tecnologia da Informação (TI) com base em habilidades técnicas e sociais. Partindo dessa premissa, a ideia do trabalho e do aplicativo foi ter concebido o Carreira pra TI como uma ferramenta prática e acessível para auxiliar estudantes e profissionais em situação semelhante. Essa abordagem garantiu um processo criativo que alinhou aprendizado e inovação durante o desenvolvimento e foi desenvolvido uma solução que é esperado ajudar outras pessoas que também passam por isso.

4.4.1 Uso do Modelo no aplicativo

Após o treinamento e avaliação do modelo de classificação, foi necessário integrá-lo ao aplicativo de forma eficiente para gerar recomendações personalizadas baseadas nos dados fornecidos pelos usuários. Essa integração envolveu a preparação das entradas, o uso do modelo treinado e a geração de respostas instantâneas. Nesta seção, descrevemos metodologicamente o processo de uso do modelo no aplicativo.

Preparação dos Dados de Entrada

A entrada dos usuários no aplicativo consiste em valores entre 0 e 100% para cada uma das 27 features (17 áreas de estudo e 10 habilidades sociais). Para que esses dados fossem compatíveis com o modelo treinado, foi necessário realizar uma conversão para o intervalo de 0.0 a 6.0, correspondente à escala utilizada no conjunto de dados original. Essa conversão foi feita utilizando a fórmula:

$$\text{Valor Convertido} = \frac{\text{Valor de Entrada}}{100} \times 6 \quad (4.1)$$

Esse processo garantiu consistência entre os dados de entrada do usuário e o formato esperado pelo modelo, permitindo que os cálculos subsequentes fossem realizados corretamente.

Processamento pelo Modelo de Classificação

Após a conversão das entradas, os dados do usuário são submetidos ao modelo de classificação treinado com o algoritmo Gradient Boosting Classifier (GBC), onde ao final é gerado o score para cada uma das 16 profissões possíveis, como saídas.

Geração da Resposta Final

Após a interação com o modelo GBC utilizando os hiperparâmetros previamente apresentados na tabela supracitada. é apresentado ao usuário na tela de resultados do aplicativo, as três saídas (áreas de atuação) com os maiores valores (scores), que foram selecionadas como as recomendações mais relevantes para o usuário.

Esse processo, além de eficiente, possibilita que os resultados sejam apresentados de forma instantânea no aplicativo após o envio das informações, reforçando a experiência interativa e responsiva do sistema.

4.4.2 Testes de Usabilidade do Aplicativo

Para garantir uma experiência fluida e intuitiva, foram realizados testes rigorosos de usabilidade no aplicativo, internamente no aplicativo e com usuários. Esses testes envolveram:

Testes Unitários: Avaliação de componentes específicos do aplicativo, como telas, botões e fluxos de navegação, para verificar sua funcionalidade isolada.

Testes Não Unitários: Simulação de cenários de uso real, com foco na interação completa do usuário, desde o onboarding até a entrega das recomendações finais.

Os testes foram conduzidos com grupos de usuários em potencial, que em outro momento haviam respondido a pesquisa no começo do desenvolvimento da ideia, como estudantes e profissionais da área de Tecnologia da Informação (TI). Durante esses testes, foram analisados fatores como clareza das instruções, responsividade das interfaces e tempos de resposta. Feedbacks coletados foram usados para ajustar detalhes e aprimorar a experiência geral ao decorrer do processo de desenvolvimento.

4.4.3 Aspectos Técnicos do Desenvolvimento do aplicativo

O aplicativo foi desenvolvido exclusivamente para a plataforma iOS, utilizando a IDE Xcode e a linguagem Swift, devido à sua robustez e ampla capacidade de integração com o ecossistema da Apple. Para o front-end, optou-se pelo uso do SwiftUI, um framework

moderno que oferece alta eficiência e facilidade no design de interfaces de usuário interativas e responsivas. Foi usado também o CoreML e CreateML para fazer uso do modelo de classificação treinado.

A arquitetura escolhida foi o padrão MVC (Model-View-Controller), conhecida por sua simplicidade e eficácia em projetos móveis. Essa escolha garantiu uma estrutura de código organizada, facilitando manutenções futuras e melhorando a escalabilidade do aplicativo durante o desenvolvimento.

Cabe lembrar que todas as imagens usadas dentro do aplicativo foram retiradas do site FreePik, um repositório de imagens online de uso livre, porém ainda assim os devidos direitos ao site e aos autores da foto estão atribuídos dentro do aplicativo

Integração do modelo com o Aplicativo:

O modelo gerado em python foi convertido através do CoreMLTools do python e integrado ao projeto iOS no Xcode, permitindo sua utilização diretamente no código Swift para realizar previsões com base nos dados fornecidos pelos usuários no aplicativo.

Esse processo de treinamento e também transição entre plataformas permitiu aproveitar as capacidades avançadas de desenvolvimento e validação oferecidas por Python e Scikit-Learn, ao mesmo tempo que garantiu a compatibilidade e desempenho necessários para a execução eficiente no ambiente iOS e o Swift, com a biblioteca SwiftUI de desenvolvimento de interfaces. A combinação dessas ferramentas proporcionou uma solução robusta e adaptável, alinhada aos objetivos do projeto e às necessidades dos usuários finais, mesmo em duas linguagens e plataformas diferentes.

4.4.4 Publicação e Disponibilidade

Após o término do desenvolvimento e testes rigorosos, o aplicativo foi publicado na App Store e está disponível para download. Com isso, usuários no mundo todo podem explorar suas funcionalidades, beneficiando-se de uma experiência intuitiva e soluções personalizadas para encontrar carreiras alinhadas a seus perfis técnicos e sociais.

Link para download do Carreira pra TI: <https://apps.apple.com/br/app/carreira-pra-ti/id6738338660>

5. RESULTADOS GERAIS

O projeto "Carreira pra TI" trouxe resultados significativos tanto no âmbito acadêmico quanto na aplicação prática, refletindo o sucesso na integração de diferentes áreas do conhecimento em uma solução computacional robusta. Este capítulo analisa os principais resultados alcançados, destacando suas contribuições e implicações acadêmicas

5.1 O Conjunto de Dados e o Modelo De Aprendizado de Máquina

Um dos principais destaques foi o desempenho do modelo de classificação baseado no Gradient Boosting Classifier (GBC). Com uma precisão de 98,1%, F1-Score de 99,1% e um Cohen's Kappa de 98,9%, o modelo demonstrou alta eficácia na tarefa de recomendação de carreiras. Esses resultados superaram expectativas iniciais e indicam que as técnicas de treinamento, ajustes de hiperparâmetros e integração de dados foram aplicadas de forma consistente com as melhores práticas descritas na literatura acadêmica.

A aplicação de uma pesquisa direta para coletar dados reais e aprimorar o conjunto de dados inicial foi um aspecto central e bem-sucedido do projeto. Essa abordagem permitiu identificar lacunas e inconsistências presentes no conjunto original, além de trazer maior representatividade e equilíbrio ao modelo. Ao incorporar dados provenientes de respostas realistas e contextualizadas, a pesquisa ajudou a reduzir vieses que estavam comprometendo o modelo de classificação.

A utilização de dados reais coletados diretamente de profissionais e estudantes de TI também adiciona relevância prática e acadêmica ao trabalho, pois reflete uma preocupação do presente projeto em criar um modelo que represente fielmente a diversidade e complexidade das carreiras na área.

5.2 Validação e Aplicabilidade do Aplicativo

Do ponto de vista acadêmico, o desenvolvimento do aplicativo iOS foi uma oportunidade bem-sucedida de demonstrar a aplicação prática de conceitos como interação humano-computador (HCI), usabilidade e design de interfaces. O aplicativo, além de oferecer resultados gerados pelo modelo, promove uma navegação fluida pelos dados e insights, o que valida teorias de design centrado no usuário, como toda a área de UX estuda. Os gráficos interativos e as explicações sobre as áreas de TI destacam a capacidade do sistema de traduzir dados técnicos em informações acessíveis, um objetivo frequentemente buscado na área acadêmica e prática de sistemas de recomendação.

5.3 Contribuição Acadêmica

O "Carreira pra TI" também representa uma contribuição ao debate sobre o uso de inteligência artificial para orientação de carreiras. O projeto aplicou conceitos avançados de aprendizado de máquina, como otimização de hiperparâmetros e validação cruzada, entre outros, enquanto abordava desafios de escalabilidade e variabilidade em um conjunto de dados com 27 features e mais de 9 mil amostras que apresentou certo viés no início, que foi contornado após a pesquisa. Essas práticas não apenas garantiram resultados robustos, mas também agregaram valor acadêmico ao explorar soluções para problemas como overfitting e viés de dados.

Além disso, o projeto introduz uma reflexão sobre como ferramentas computacionais, como um aplicativo e um modelo de classificação, podem ser usadas para integrar habilidades técnicas e sociais em um contexto de orientação profissional, um tema de crescente interesse acadêmico.

5.4 Impacto Geral

Os resultados gerais do "Carreira pra TI" confirmam a tentativa bem-sucedida do projeto em achar uma convergência entre teoria e prática. Ele contribui academicamente ao explorar métodos inovadores de integração de dados e treinamento de modelos, ao mesmo tempo em que apresenta uma solução prática na orientação de carreiras. Essa dualidade entre rigor acadêmico e aplicabilidade prática reflete a relevância do trabalho como uma contribuição significativa para o campo de TI e o da Ciência da Computação.

5.5 Satisfação do Usuário

A aplicação de um formulário para coletar feedback dos usuários revelou insights valiosos sobre a aceitação do aplicativo. Os resultados destacaram altos índices de satisfação em relação à interface, utilidade das recomendações e facilidade de uso. Esses dados, além de validar o impacto prático do trabalho, indicam que o projeto atingiu seus objetivos acadêmicos e práticos de promover escolhas mais informadas e alinhadas para estudantes e profissionais, como mostrado na sessão a seguir.

5.6 Avaliação

5.6.1 Pesquisa sobre a experiência ao usar o aplicativo e o modelo

Como avaliação, foi feito um formulário com 91 respondentes que avaliaram a sua experiência com o aplicativo e sobre os resultados modelo de aprendizado de máquina, ao decorrer do desenvolvimento, mas principalmente nas etapas finais. Como dito, enquanto o projeto estava em desenvolvimento e ao final, foram feitos testes de usuários de forma presencial com possíveis usuários do aplicativo para testar a qualidade e intuitividade do aplicativo, além da satisfação das respostas dadas pelo modelo de classificação, no fim do teste o link do formulário era disponibilizado para o usuário que testou o aplicativo, no qual ele respondia de forma anônima após. Quando o aplicativo subiu para a loja da App Store, na página das repostas estava disponibilizado um botão que encaminhava para a página do mesmo formulário a ser respondido, além de que junto com a divulgação do aplicativo para o público-alvo, também foi encaminhado para o usuário testar após usar. Todos esses usuários puderam responder o mesmo formulário, que no total, como dito, obteve 91 respostas, mostradas a seguir:

Pergunta 1: de 1 a 10, qual sua satisfação geral com o aplicativo como uma ferramenta de orientação de carreira e entender melhor suas opções na TI?

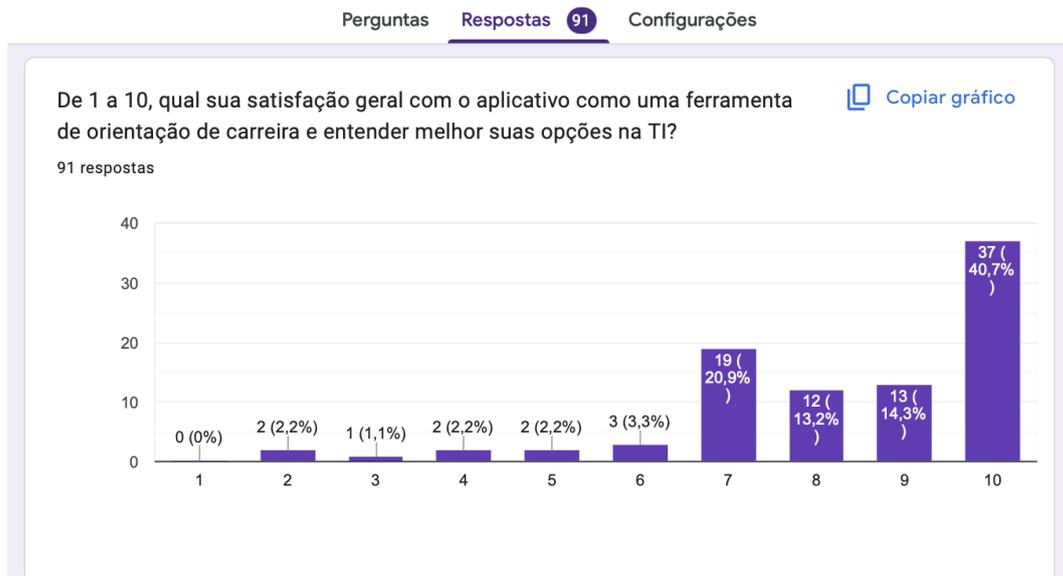


Figura 5.1 – Gráfico da primeira pergunta sobre qual a satisfação geral com o aplicativo como uma ferramenta de orientação de carreira e entender melhor suas opções na TI

Para a pergunta "Qual sua satisfação geral com o aplicativo como uma ferramenta de orientação de carreira e entendimento das opções na TI?", os gráficos de análise de

feedback indicaram resultados significativamente positivos. A maioria dos usuários classificou o aplicativo como altamente satisfatório, enfatizando tanto sua eficácia em oferecer recomendações personalizadas quanto sua capacidade de promover uma compreensão mais profunda das possibilidades de carreira na área de TI.

Esses resultados não apenas reforçam o impacto prático do aplicativo, mas também validam a abordagem metodológica adotada no desenvolvimento deste trabalho do ponto de vista acadêmico. O alinhamento entre os objetivos do projeto e a percepção dos usuários sugere que o modelo de classificação e a interface do aplicativo foram bem-sucedidos em traduzir teorias e práticas de aprendizado de máquina em uma aplicação acessível e relevante. Esse reconhecimento pelos usuários reflete a contribuição acadêmica do projeto, demonstrando como metodologias sólidas podem gerar soluções práticas eficazes para problemas reais de orientação de carreira.

Pergunta 2: De 1 a 10, o quanto você considera o aplicativo intuitivo, agradável e com uma boa experiência de usuário?

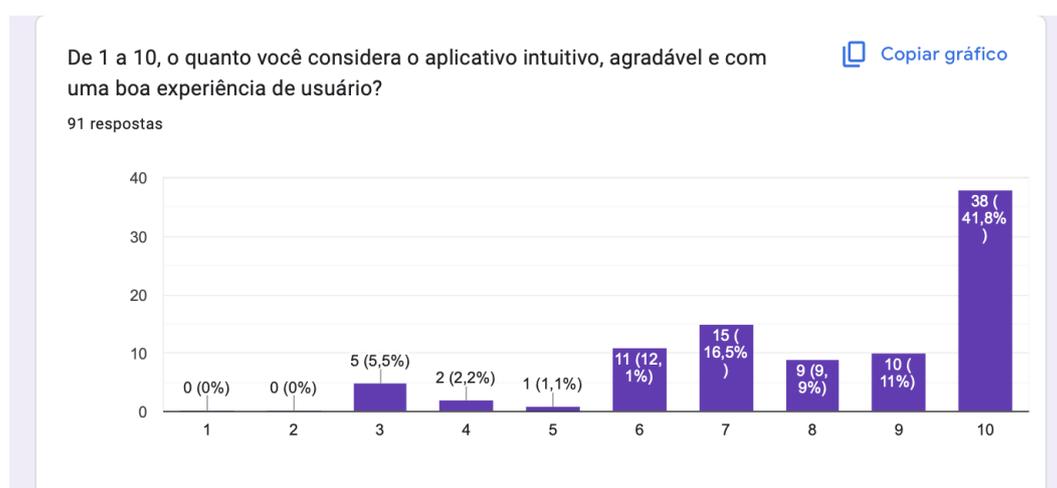


Figura 5.2 – Gráfico da primeira pergunta sobre o quanto você considera o aplicativo intuitivo, agradável e com uma boa experiência de usuário

Os gráficos de análise de feedback apresentaram resultados expressivos, com a maioria dos usuários atribuindo notas altas à intuitividade, ao design agradável e à experiência de usuário oferecida pelo aplicativo. Esses aspectos foram amplamente elogiados, destacando o sucesso da interface em proporcionar uma interação acessível e envolvente.

Do ponto de vista acadêmico, esses resultados confirmam a eficácia das práticas de design, engenharia de software e usabilidade aplicadas no desenvolvimento do aplicativo. A aplicação de princípios de design centrado no usuário, aliados a uma abordagem iterativa no desenvolvimento da interface, demonstrou como conceitos teóricos podem ser transformados em soluções práticas de alta qualidade. A satisfação dos usuários, medida por meio de métricas objetivas, evidencia que o aplicativo não só cumpre sua proposta funcional, mas também eleva a experiência do usuário a um nível exemplar, validando sua

contribuição como ferramenta inovadora e bem projetada no campo de orientação de carreira na TI.

Pergunta 3: De 1 a 10, o quanto o aplicativo ajudou você a entender melhor suas habilidades e interesses nas áreas de TI?

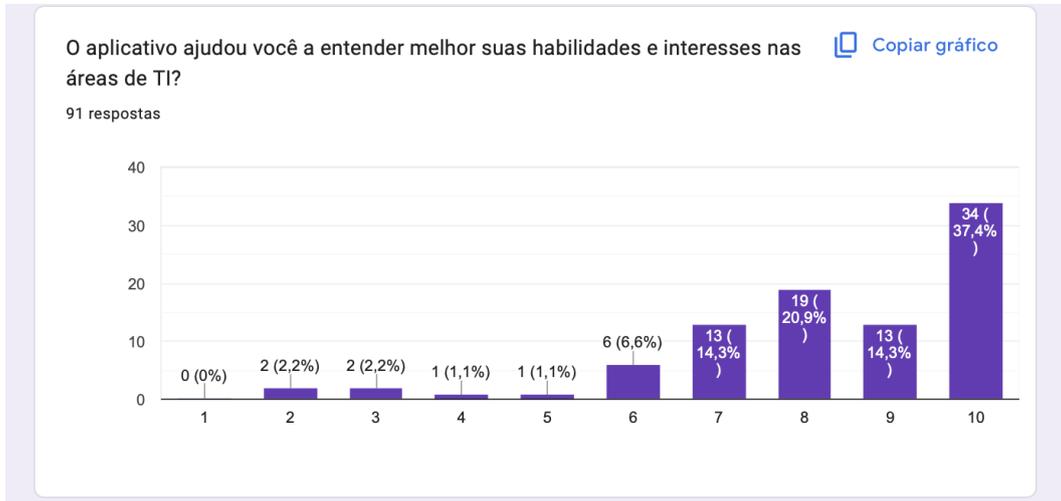


Figura 5.3 – Gráfico da pergunta 3

Os gráficos de análise de feedback apresentaram resultados positivos, com a maioria dos usuários atribuindo notas altas. Esses números refletem que o aplicativo cumpriu de maneira eficaz seu propósito de auxiliar os usuários a compreender melhor suas habilidades técnicas e sociais, além de ajudá-los a correlacionar essas características com potenciais carreiras na área de TI.

Sob a perspectiva acadêmica, esses resultados destacam a relevância do modelo de machine learning integrado ao aplicativo. A análise detalhada das 27 áreas de conhecimento e habilidades sociais, combinada com uma apresentação clara dos resultados, demonstrou a eficácia da aplicação de metodologias teóricas para resolver problemas práticos de autoconhecimento e orientação profissional. Essa validação prática, respaldada por altos índices de satisfação dos usuários, reforça a contribuição do projeto para os estudos de sistemas de recomendação e sua aplicabilidade no campo de orientação de carreira na TI.

Pergunta 4: Você acredita que o aplicativo te ajudou a entender melhor as opções de carreira para você na TI?

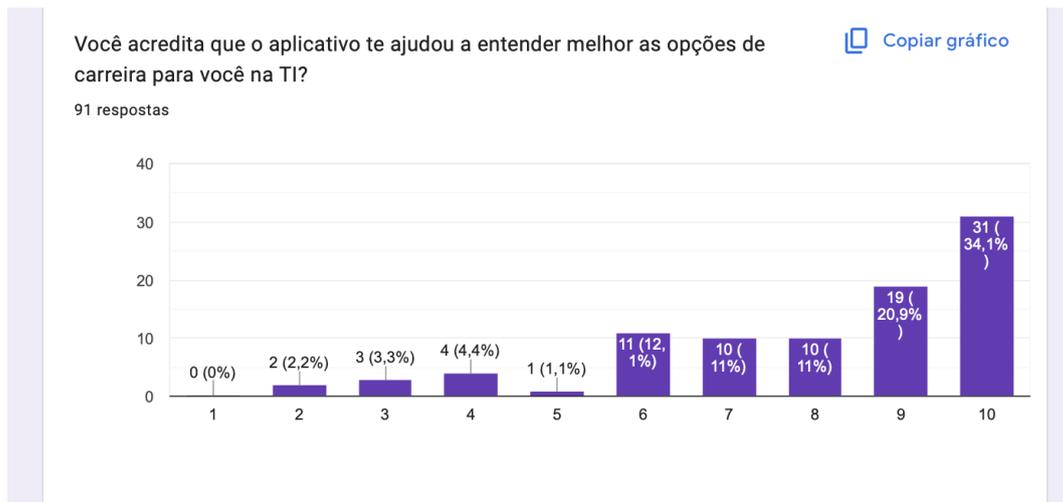


Figura 5.4 – Gráfico da pergunta 4

Os gráficos de análise de feedback indicaram resultados positivos, com a maioria dos usuários afirmando que o aplicativo os ajudou a entender melhor suas opções de carreira na área de TI. Esses números destacam o sucesso do aplicativo como uma ferramenta de orientação profissional, cumprindo seu propósito de apresentar recomendações personalizadas e de fácil compreensão.

Sob o ponto de vista acadêmico, esses resultados refletem a eficácia do sistema de recomendação baseado em machine learning. O modelo não apenas identificou áreas de carreira alinhadas às características individuais dos usuários, como também apresentou essas informações de forma clara e prática. Esse impacto positivo evidencia o potencial do projeto para contribuir tanto para a orientação profissional quanto para estudos futuros na aplicação de inteligência artificial em sistemas de recomendação, especialmente no contexto da TI. Além disso, o alto índice de aceitação e satisfação dos usuários reforça a relevância acadêmica e prática do projeto, posicionando-o como uma contribuição significativa na interface entre tecnologia e educação.

Pergunta 5: Você acredita que o modelo de machine learning analisou bem suas habilidades técnicas e sociais para gerar as recomendações?

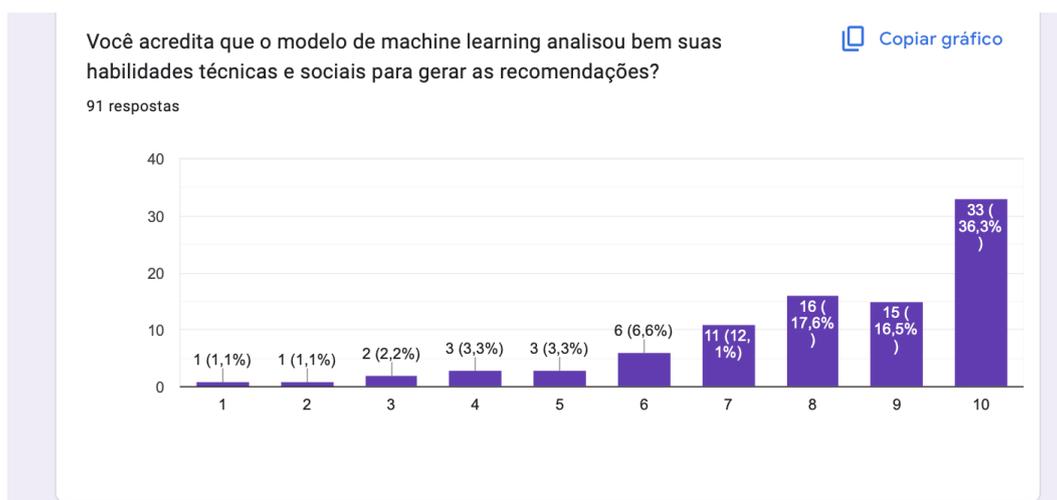


Figura 5.5 – Gráfico da pergunta 5

Os resultados apresentaram ótimos números, com a maioria dos usuários atribuindo notas elevadas, especialmente na faixa de 7 a 10. Esses valores indicam que o modelo de machine learning foi amplamente eficaz em analisar as habilidades técnicas e sociais dos usuários para gerar recomendações precisas e alinhadas aos seus perfis.

Sob a perspectiva acadêmica, esses resultados reforçam a relevância da abordagem metodológica adotada no desenvolvimento do modelo. A escolha do Gradient Boosting Classifier (GBC) como algoritmo central, aliada ao pré-processamento rigoroso dos dados e ao ajuste refinado dos hiperparâmetros, demonstrou ser uma estratégia adequada para lidar com a complexidade do problema. O uso de validação cruzada e a análise criteriosa de métricas como precisão, recall e F1-score garantiram que o modelo apresentasse alta capacidade de generalização, minimizando problemas como overfitting.

Além disso, a resposta positiva dos usuários valida que o sistema foi capaz de interpretar com precisão informações técnicas e sociais variadas, destacando a eficácia do modelo em traduzir essas características em recomendações personalizadas e confiáveis. Do ponto de vista acadêmico, esse feedback comprova o potencial da aplicação de técnicas avançadas de machine learning para resolver problemas complexos no campo de orientação de carreira, marcando uma contribuição significativa tanto em termos práticos quanto teóricos.

Pergunta 6: As profissões sugeridas ao final pelo modelo de Machine Learning fazem sentido para o seu perfil, afinidades e interesses?

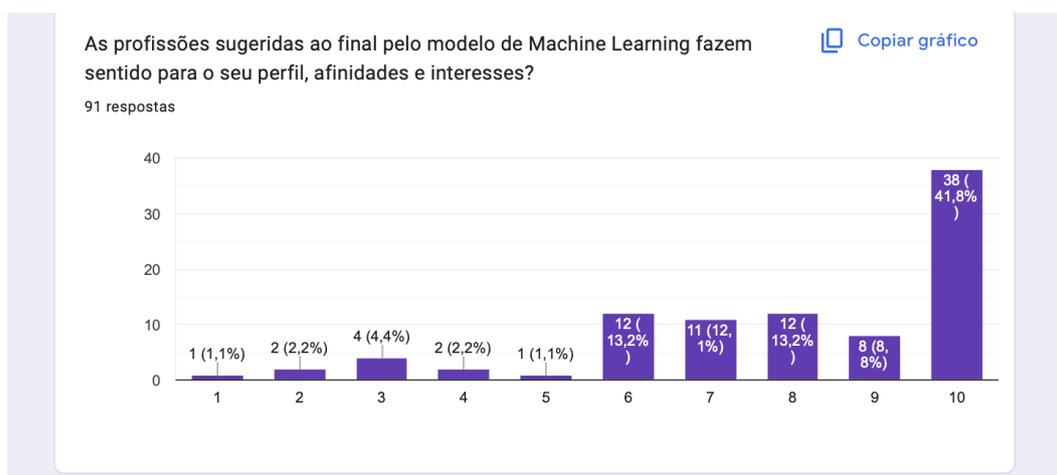


Figura 5.6 – Gráfico da pergunta 6

Os resultados mostraram valores expressivos, com a maioria das respostas concentradas na faixa de 6 a 10, o que confirma a precisão do modelo de machine learning e a eficácia da metodologia aplicada. Esses dados refletem que o sistema foi capaz de fornecer recomendações coerentes e alinhadas ao perfil, afinidades e interesses dos usuários.

Do ponto de vista acadêmico, esses resultados reforçam a validade do Gradient Boosting Classifier (GBC) como escolha central para o modelo de classificação. A habilidade do algoritmo em lidar com conjuntos de dados complexos e correlacionar variáveis técnicas e sociais com profissões específicas destaca sua robustez e aplicabilidade. Além disso, a abordagem rigorosa adotada para o treinamento, incluindo validação cruzada e otimização de hiperparâmetros, garantiu um alto grau de generalização, minimizando erros e oferecendo previsões consistentes.

A resposta positiva dos usuários não apenas valida a capacidade técnica do modelo, mas também sublinha a relevância de sistemas personalizados no campo de orientação profissional. Academicamente, isso demonstra como a aplicação de machine learning pode contribuir significativamente para resolver problemas complexos, proporcionando recomendações que não apenas fazem sentido, mas também têm potencial para gerar impacto real no processo de escolha de carreira. Esses resultados destacam o sucesso do projeto em integrar tecnologia de ponta a um problema prático, entregando valor aos usuários e fortalecendo a contribuição do trabalho no campo de machine learning e orientação vocacional.

Pergunta 7: De 1 a 10, O quão eficaz o aplicativo foi para ajudar você a entender o que cada determinada área de atuação da TI faz?



Figura 5.7 – Gráfico da pergunta 7

Os resultados dessa pergunta mostraram uma avaliação predominantemente positiva, com a maioria das respostas concentradas na faixa de valores acima de 5. Esses dados refletem a eficácia do aplicativo como uma ferramenta educativa, destacando sua capacidade de esclarecer as diferentes áreas de atuação na Tecnologia da Informação (TI).

Do ponto de vista acadêmico, esses resultados reforçam a importância de integrar conteúdo explicativo e acessível em ferramentas de orientação profissional. A média elevada nas respostas indica que o aplicativo não apenas atingiu seu objetivo de orientar sobre carreiras, mas também agregou valor ao fornecer descrições detalhadas e contextuais de cada área de atuação. Os usuários relataram que essas informações foram úteis para entender as demandas, habilidades associadas e possibilidades práticas de cada campo, promovendo uma compreensão mais abrangente das opções disponíveis.

Essa abordagem didática, aliada ao uso do modelo de machine learning para personalizar as recomendações, demonstra como o aplicativo combina tecnologia avançada com princípios educacionais sólidos para gerar impacto positivo. Academicamente, esse resultado evidencia o sucesso do projeto em alinhar inovação tecnológica com objetivos práticos, como a educação e a orientação profissional, consolidando seu papel como uma contribuição relevante no campo de machine learning e aplicativos educacionais voltados para a TI.

Pergunta 8: Você pretende explorar mais as profissões sugeridas pelo aplicativo?

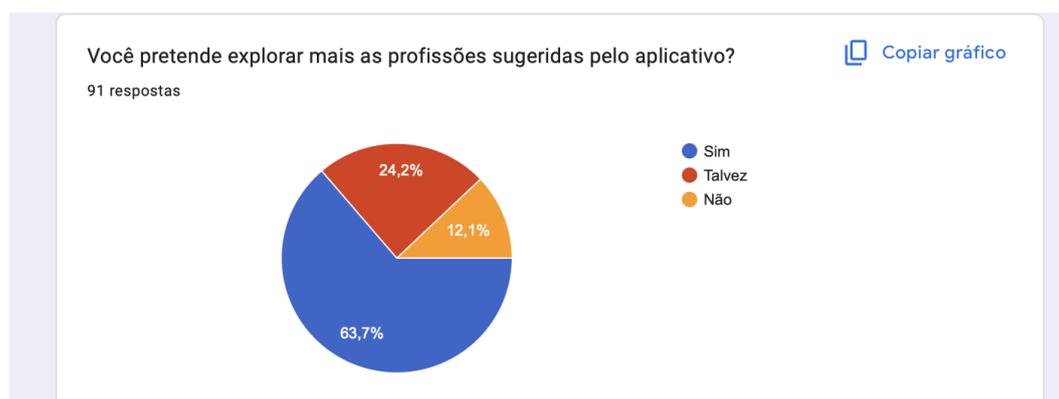


Figura 5.8 – Gráfico da pergunta 8

Os resultados dessa pergunta revelaram que a maioria dos usuários demonstrou interesse em explorar as profissões sugeridas pelo aplicativo, validando sua relevância como ferramenta de orientação profissional. Esses números destacam o impacto positivo do sistema de recomendação em despertar curiosidade e engajamento com as áreas indicadas, demonstrando que as sugestões fornecidas foram percebidas como pertinentes e úteis para os usuários.

Do ponto de vista acadêmico, esses dados reforçam a eficácia do modelo de machine learning e do design do aplicativo em promover a interação e a reflexão sobre possibilidades de carreira. O interesse dos usuários em investigar mais profundamente as áreas sugeridas reflete não apenas a qualidade das recomendações, mas também a capacidade do aplicativo de atuar como um ponto de partida confiável e informativo para decisões futuras. Esse resultado indica que o projeto atingiu seu objetivo de não apenas fornecer recomendações, mas também de inspirar os usuários a explorar e entender melhor suas opções profissionais. Academicamente, a capacidade de despertar esse interesse e engajamento é uma contribuição significativa para o campo de sistemas de recomendação aplicados à orientação de carreira. O sucesso do aplicativo em conectar dados analíticos com a experiência do usuário destaca sua relevância como um caso de estudo para futuras pesquisas e desenvolvimentos na área de Tecnologia da Informação e educação profissional.

Pergunta 9: O aplicativo e o modelo de machine learning trouxeram alguma perspectiva ou insight que você não havia considerado antes?

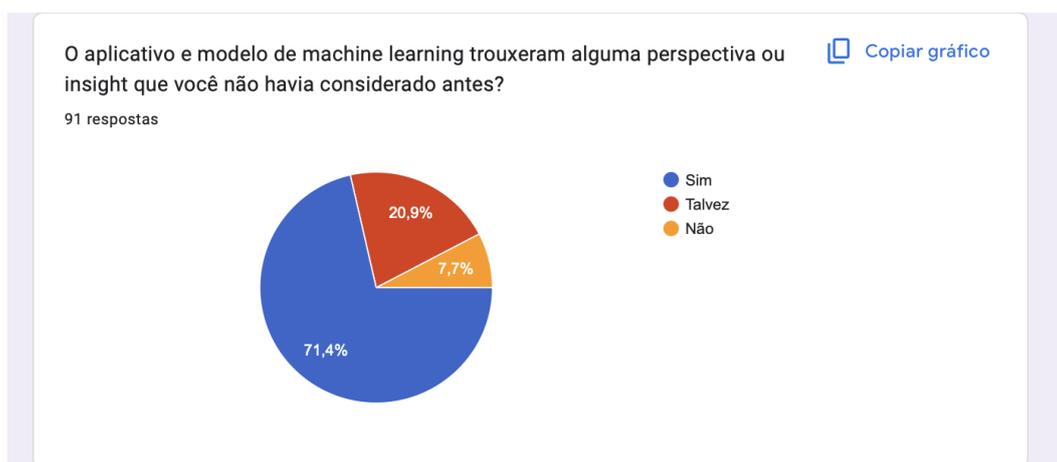


Figura 5.9 – Gráfico da pergunta 9

s resultados dessa pergunta mostraram que a maioria dos usuários reconheceu ter obtido novas perspectivas ou reflexões a partir das análises e recomendações do aplicativo. Esses números destacam a capacidade do modelo de machine learning de identificar padrões e sugerir possibilidades que os usuários não haviam considerado anteriormente, agregando valor real ao processo de orientação. Do ponto de vista acadêmico, esses resultados são particularmente relevantes, pois demonstram que o projeto conseguiu ir além de uma ferramenta de recomendação básica, proporcionando insights personalizados baseados nas características únicas de cada usuário. Essa capacidade de ampliar o entendimento dos usuários sobre suas próprias habilidades, interesses e potenciais caminhos de carreira é um diferencial importante, validando tanto o modelo desenvolvido quanto o design do aplicativo.

Além disso, a resposta positiva dos usuários reflete a eficácia da abordagem metodológica empregada, incluindo o pré-processamento de dados e a escolha criteriosa do Gradient Boosting Classifier (GBC). O impacto de oferecer recomendações que geram reflexões e novos pontos de vista não apenas reafirma o sucesso do projeto, mas também contribui significativamente para o campo acadêmico, evidenciando o potencial dos sistemas de recomendação baseados em inteligência artificial como ferramentas educativas e de autoconhecimento.

Pergunta 10: De 1 a 10, após o uso do aplicativo, você sente mais noção e confiança para seguir caminhos e tomar decisões sobre sua carreira na TI?

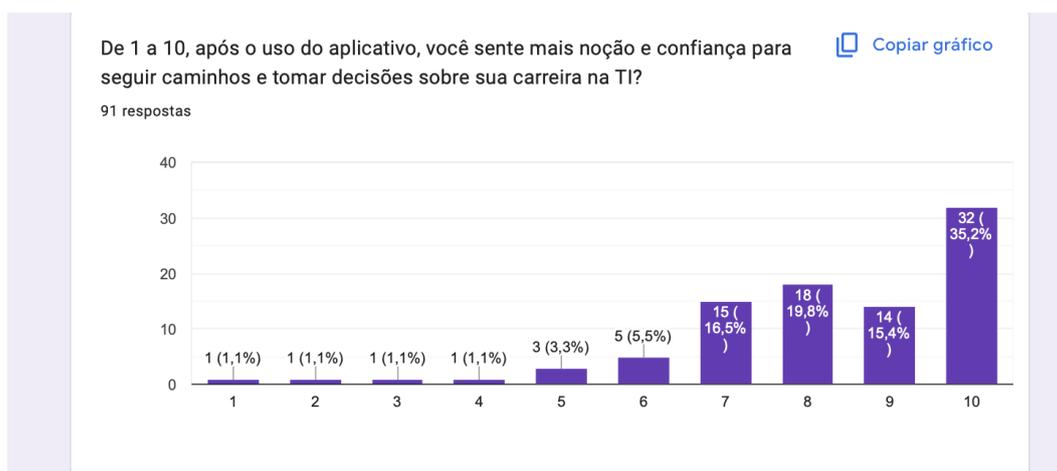


Figura 5.10 – Gráfico da pergunta 10

Os resultados desta pergunta indicaram que a maioria dos usuários relatou um impacto positivo em sua clareza e confiança para tomar decisões relacionadas à carreira em TI. Embora os números não sejam excepcionalmente altos, apresentam uma média sólida, destacando que o aplicativo atingiu seu objetivo de proporcionar orientações úteis e contribuir para a tomada de decisões mais seguras. Do ponto de vista acadêmico, esses resultados refletem a capacidade do aplicativo e do modelo de machine learning de ir além da simples recomendação, atuando como uma ferramenta de apoio educacional e emocional para os usuários. A confiança gerada demonstra que o sistema consegue alinhar as características e interesses individuais às oportunidades do mercado, fornecendo informações concretas que reforçam a segurança dos usuários ao considerar novos caminhos profissionais. Além disso, o impacto do aplicativo no fortalecimento da autoconfiança reforça o sucesso da abordagem metodológica do projeto, que combina um modelo de aprendizado de máquina robusto com uma interface intuitiva. Esse alinhamento entre tecnologia e usabilidade foi essencial para alcançar resultados que vão além de números, promovendo uma experiência significativa para os usuários e contribuindo para a relevância acadêmica do trabalho.

5.6.2 Métricas dos números de Downloads, visualizações da página e falhas

A App Store disponibiliza métricas detalhadas do desempenho do aplicativo desde seu lançamento até o momento atual, abrangendo dados como número de downloads, visualizações da página, sessões iniciadas e eventuais falhas de uso. Essas informações são apresentadas no painel mostrado na figura a seguir:



Figura 5.11 – Painel com as métricas do aplicativo

O painel revela resultados altamente positivos, com destaque para a ausência de falhas em todas as sessões iniciadas, evidenciando a robustez e a estabilidade do aplicativo. Essa confiabilidade reflete a eficácia das escolhas técnicas adotadas no desenvolvimento, alinhadas às recomendações da literatura. A escolha de frameworks e linguagens adequadas, combinadas com boas práticas de programação, garantiu a entrega de um sistema sólido e livre de problemas críticos, fator essencial para a experiência positiva do usuário.

Além disso, o aplicativo registrou um desempenho expressivo, com quase 200 downloads e mais de 1,5 mil visualizações da página. Esses números demonstram não apenas o interesse do público pela proposta do projeto, mas também a capacidade do aplicativo de se comunicar bem com seus potenciais usuários, gerando curiosidade e engajamento. Essa resposta inicial do mercado valida a proposta do aplicativo como relevante e acessível para sua audiência-alvo.

Do ponto de vista acadêmico, esses resultados corroboram a eficiência das escolhas metodológicas e técnicas realizadas ao longo do projeto. A decisão de utilizar o Gradient Boosting Classifier (GBC) como o algoritmo central do modelo de machine learning demonstrou-se apropriada, proporcionando análises precisas e confiáveis que sustentaram a experiência do usuário. Além disso, o cuidado na arquitetura da interface e no desenvolvimento do sistema contribuiu para a alta usabilidade, aspecto frequentemente destacado na literatura como crucial para a adoção de ferramentas tecnológicas.

Por fim, as métricas fornecidas pela App Store refletem a relevância do aplicativo enquanto ferramenta prática de orientação de carreira na área de TI. A combinação de um modelo preditivo eficaz com uma interface robusta e intuitiva foi decisiva para alcançar esses resultados. A ausência de falhas e o feedback positivo dos usuários reforçam a qualidade técnica do projeto e seu impacto significativo, tanto em termos práticos quanto acadêmicos, estabelecendo-o como uma solução viável e bem fundamentada para o problema proposto.

6. CONCLUSÃO E TRABALHOS FUTUROS

6.1 Conclusão

O presente trabalho teve como principais objetivos desenvolver um modelo de classificação de aprendizado de máquina e um aplicativo inovador e eficiente para a recomendação personalizada de carreiras na área de Tecnologia da Informação (TI), utilizando, como dito, machine learning como base tecnológica para análise de dados. Desde a concepção da ideia, guiada pela metodologia Challenge Based Learning (CBL), até a implementação final, o projeto buscou atender a uma demanda real: oferecer aos usuários uma ferramenta prática para alinhar suas habilidades técnicas e sociais às oportunidades de carreira mais adequadas. Por meio de um processo rigoroso de pesquisa, desenvolvimento e avaliação, foi possível criar uma solução que combina usabilidade, precisão técnica e relevância prática, demonstrando o impacto positivo que a tecnologia pode ter na orientação de decisões importantes, como a escolha de uma carreira.

No contexto acadêmico, o projeto contribuiu significativamente ao propor uma abordagem robusta para o uso de machine learning na orientação de carreiras. A escolha do Gradient Boosting Classifier (GBC) como modelo de machine learning foi um dos pilares do sucesso do sistema de recomendação, destacando-se por sua capacidade de lidar com dados de alta dimensionalidade e padrões complexos. Essa escolha foi respaldada pela literatura, que aponta o GBC como uma técnica avançada e confiável para problemas de classificação. O treinamento do modelo envolveu validação cruzada e otimização de hiperparâmetros, assegurando um desempenho consistente e adaptado ao conjunto de dados. A integração do modelo ao aplicativo por meio do framework CreateML representou uma solução técnica eficiente e inovadora, garantindo a compatibilidade entre o modelo treinado e a experiência final do usuário.

Os testes rigorosos, realizados com entradas baseadas em dados reais, validaram o modelo, demonstrando sua capacidade de capturar padrões significativos nas informações dos usuários e traduzir essas informações em recomendações alinhadas aos perfis individuais. O feedback obtido a partir de pesquisas quantitativas reforçou a eficácia do modelo e do aplicativo, apontando para a assertividade das recomendações e para o impacto positivo na percepção dos usuários sobre suas próprias habilidades e interesses. Esses resultados comprovam não apenas a viabilidade técnica do projeto, mas também sua relevância acadêmica, ao demonstrar como modelos de machine learning podem ser aplicados para resolver problemas complexos em orientação profissional.

Por fim, os resultados obtidos evidenciam que o aplicativo não apenas atendeu aos objetivos propostos, mas também se destacou como uma ferramenta intuitiva, agradável

vel e impactante para o público-alvo. Além de oferecer orientações de carreira confiáveis, o aplicativo promoveu o autoconhecimento e uma maior clareza sobre as possibilidades de atuação na TI, contribuindo para o entendimento do papel que a tecnologia desempenha na orientação profissional. Acadêmica e tecnicamente, o trabalho agrega valor ao campo de estudo, exemplificando como técnicas avançadas de aprendizado de máquina podem ser aplicadas de maneira prática e eficaz. Como perspectivas futuras, destaca-se a possibilidade de ampliar o escopo do modelo para outras áreas de conhecimento, bem como de incorporar novas funcionalidades ao aplicativo, aumentando ainda mais a personalização e a eficácia das recomendações.

O desenvolvimento deste trabalho foi uma jornada de aprendizado enriquecedora e desafiadora, que me proporcionou um profundo crescimento acadêmico, técnico e pessoal. Ao longo do processo, tive a oportunidade de explorar conceitos avançados de machine learning, aplicar metodologias de desenvolvimento de software e compreender de maneira mais ampla a importância de alinhar tecnologia e usabilidade para resolver problemas reais. Sou profundamente grato(a) pelo apoio recebido durante esta caminhada, desde os professores e colegas que compartilharam seus conhecimentos e sugestões, até os usuários que participaram das pesquisas e testes, contribuindo para a evolução do projeto. Este trabalho não apenas reflete os objetivos acadêmicos alcançados, mas também simboliza a dedicação e o entusiasmo que me guiaram ao longo dessa jornada.

6.2 Trabalhos Futuros

6.2.1 Expansão do Conjunto de Dados Reais para Aprimoramento do Modelo

Como um dos trabalhos futuros prioritários, propõe-se a expansão do conjunto de dados reais utilizados no treinamento do modelo de machine learning. Embora o modelo atual tenha demonstrado resultados satisfatórios em termos de precisão e personalização das recomendações, um conjunto de dados mais amplo e diversificado pode contribuir significativamente para melhorar sua capacidade preditiva e sua robustez. Essa expansão permitirá incorporar cenários e perfis que ainda não foram completamente contemplados, aumentando a representatividade e a generalização do modelo.

O foco será na coleta de dados reais, provenientes de diferentes perfis de usuários e contextos, ampliando a diversidade de amostras relacionadas às áreas de estudo e habilidades sociais. Isso pode incluir, por exemplo, a participação de profissionais em estágios diferentes da carreira, estudantes de outras regiões e indivíduos com backgrounds variados em TI. A coleta desses dados será guiada por métodos rigorosos de pesquisa,

assegurando qualidade e integridade, além de possibilitar a análise de novas tendências e comportamentos no setor de Tecnologia da Informação.

Adicionalmente, com um conjunto de dados mais amplo, será possível aplicar ainda mais as técnicas avançadas de aprendizado de máquina, como o fine-tuning do modelo e a validação em novos subgrupos, para garantir que o sistema se mantenha relevante e eficaz ao longo do tempo. Essa etapa de expansão reforça o compromisso do projeto com a melhoria contínua e a contribuição para estudos futuros que explorem a interseção entre machine learning, orientação de carreira e análise de perfis profissionais.

Outro importante trabalho futuro é a publicação de um artigo científico detalhando os aspectos técnicos e metodológicos do projeto, incluindo o desenvolvimento do modelo de classificação baseado em machine learning e sua integração ao aplicativo *Carreira pra TI*. Esse artigo servirá como uma contribuição relevante para a comunidade acadêmica, apresentando as escolhas metodológicas, os desafios enfrentados e os resultados obtidos, bem como promovendo a discussão sobre o papel da tecnologia na orientação de carreiras.

6.2.2 Publicação e Divulgação do Projeto para a Comunidade Acadêmica

Outro importante trabalho futuro é a publicação de um artigo científico detalhando os aspectos técnicos e metodológicos do projeto, incluindo o desenvolvimento do modelo de classificação baseado em machine learning e sua integração ao aplicativo *Carreira pra TI*. Esse artigo servirá como uma contribuição para a comunidade acadêmica, apresentando as escolhas metodológicas, os desafios enfrentados e os resultados obtidos, bem como promovendo a discussão sobre o papel da tecnologia na orientação de carreiras.

Além do artigo, o material relacionado ao projeto, como o conjunto de dados refinado, a documentação do modelo e o código-fonte do aplicativo, será estruturado para possível compartilhamento em repositórios acadêmicos e plataformas de código aberto. Essa iniciativa visa fomentar a colaboração, possibilitando que outros pesquisadores utilizem, adaptem e aprimorem o modelo e as ideias apresentadas neste trabalho. O compartilhamento do material pode ser especialmente útil para estudos futuros que busquem expandir o uso de machine learning em contextos educacionais e profissionais.

Por fim, planeja-se divulgar o aplicativo e o modelo em eventos acadêmicos, conferências e fóruns especializados, com o intuito de alcançar tanto o público acadêmico quanto profissionais interessados no tema. A promoção do trabalho em ambientes especializados permitirá mostrar sua relevância, receber feedback construtivo e explorar potenciais parcerias para continuidade e expansão do projeto. Essa etapa reforça o compromisso de contribuir para o avanço do conhecimento acadêmico e tecnológico, além de posicionar o *Carreira pra TI* como uma referência em ferramentas baseadas em machine learning para orientação de carreira.

6.2.3 Estudo de Impacto a Médio e Longo Prazo para Estudantes e Profissionais de TI

Um dos trabalhos futuros mais relevantes está no acompanhamento e avaliação do impacto do *Carreira pra TI* a médio e longo prazo, visando compreender como o sistema influencia efetivamente a trajetória de estudantes e profissionais da área de Tecnologia da Informação. Este estudo será fundamental para medir não apenas a satisfação imediata dos usuários, mas também os resultados concretos em termos de decisões de carreira, avanços profissionais e requalificação no setor.

A proposta envolve a aplicação de pesquisas longitudinais, conduzidas em intervalos regulares, para avaliar como os usuários do aplicativo progrediram em suas escolhas e trajetórias de carreira ao longo do tempo. Questões como a adequação das recomendações fornecidas pelo modelo de machine learning às experiências reais dos usuários, o impacto no direcionamento acadêmico e profissional, e o desenvolvimento de habilidades técnicas e sociais estarão no centro dessa análise. O objetivo é obter uma visão detalhada do papel do sistema na melhoria da autoconfiança, clareza profissional e adaptação às demandas do mercado.

Além disso, o estudo poderá explorar métricas quantitativas, como o aumento no número de profissionais atuando em áreas correlatas às recomendações, e qualitativas, como relatos sobre como o aplicativo ajudou os usuários a tomar decisões mais embasadas e estratégicas. Essa abordagem permitirá uma análise holística do impacto gerado, contribuindo para identificar oportunidades de aprimoramento e novas funcionalidades que atendam ainda melhor às necessidades do público-alvo.

Com o foco no impacto a médio e longo prazo, essa frente de estudo reforça o compromisso do projeto em gerar valor sustentável para a comunidade acadêmica e profissional de TI, fortalecendo o propósito de promover escolhas de carreira mais informadas e alinhadas às habilidades e interesses dos indivíduos.

6.2.4 Acessibilidade no Aplicativo

Outro trabalho futuro essencial para o *Carreira pra TI* é o aprimoramento de sua acessibilidade, garantindo que pessoas com diferentes necessidades e condições possam utilizar o aplicativo de maneira eficiente e inclusiva. Esse objetivo está alinhado ao compromisso do projeto em democratizar o acesso à orientação de carreira na área de Tecnologia da Informação, promovendo igualdade de oportunidades para todos os interessados.

O plano inclui a implementação de melhorias baseadas nas diretrizes de acessibilidade do *Web Content Accessibility Guidelines (WCAG)* e nas recomendações da Apple

para o desenvolvimento de aplicativos acessíveis no *iOS*. Algumas das ações planejadas incluem:

Suporte a Leitores de Tela: Garantir que todos os elementos da interface sejam compatíveis com leitores de tela, como o *VoiceOver*, permitindo que pessoas com deficiência visual possam navegar e interagir com o aplicativo de forma autônoma.

Ajustes de Contraste e Tamanhos de Fonte: Introduzir opções para aumentar o contraste de cores e ajustar o tamanho das fontes, facilitando a leitura para usuários com baixa visão ou dificuldades visuais específicas.

Gestos e Navegação Simplificados: Adaptar a navegação para incluir gestos simplificados e controles alternativos, atendendo usuários com limitações motoras.

Textos Alternativos e Áudio Descritivo: Adicionar textos alternativos para imagens e, quando necessário, descrições em áudio para os elementos mais importantes do aplicativo.

Tradução e Localização: Expandir o suporte linguístico para alcançar usuários que não falam português, tornando o aplicativo mais inclusivo globalmente.

Essas ações visam proporcionar uma experiência mais acessível para pessoas com diferentes necessidades, incluindo aquelas com deficiências visuais, auditivas, motoras ou cognitivas. Além disso, a acessibilidade beneficiará usuários que, em diferentes contextos, possam encontrar dificuldades temporárias, como ambientes com baixa iluminação ou problemas de concentração, afinal o uso do aplicativo é de interesse de toda a comunidade acadêmica envolvida com a Tecnologia da Informação.

Ao tornar o *Carreira pra TI* mais acessível, o projeto reafirma seu compromisso com a inclusão e amplia seu alcance, permitindo que um público ainda maior se beneficie da tecnologia e das recomendações personalizadas para orientação de carreira na área de TI.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AA23] Author, F.; Author, S. “Disrupting misaligned values and actions in hrd: A critical perspective”. In: *Stretching the Concept*, Springer, 2023, pp. 45–60.
- [Agg16] Aggarwal, C. C. “Recommender Systems: The Textbook”. Springer International Publishing, 2016.
- [Aut10] Autor, D. “The polarization of job opportunities in the us labor market: Implications for employment and earnings”, *Community Investments*, vol. 23, 01 2010, pp. 11–16.
- [Aut14] Autor, D. H. “Skills, education, and the rise of earnings inequality among the “other 99 percent””, *Science*, vol. 344–6186, 2014, pp. 843–851, <https://www.science.org/doi/pdf/10.1126/science.1251868>.
- [Bat22] Bateman, T. “Wearable technology applications in football: A review”, *Office of Career Services | Harvard University – Faculty of Arts and Science Mignone Center for Career Success logo*, 2022.
- [BB12] Bergstra, J.; Bengio, Y. “Random search for hyper-parameter optimization”, *Journal of Machine Learning Research*, vol. 13–10, 2012, pp. 281–305.
- [Bis06] Bishop, C. M. “Pattern Recognition and Machine Learning (Information Science and Statistics)”. Berlin, Heidelberg: Springer-Verlag, 2006.
- [BL05] Brown, S. D.; Lent, R. W. “Career Development and Counseling: Putting Theory and Research to Work”. Hoboken, NJ: Wiley, 2005.
- [Bre01] Breiman, L. “Random forests”, *Machine Learning*, vol. 45, 10 2001, pp. 5–32.
- [Cas09] “The Transformation of Work and Employment: Networkers, Jobless, and Flex-Timers”. John Wiley Sons, Ltd, 2009, cap. 4, pp. 216–354, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781444319514.ch4>.
- [CG16] Chen, T.; Guestrin, C. “Xgboost: A scalable tree boosting system”. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2016, pp. 785–794.
- [CNM06] Caruana, R.; Niculescu-Mizil, A. “An empirical comparison of supervised learning algorithms”. In: Proceedings of the 23rd International Conference on Machine Learning (ICML), 2006, pp. 161–168.

- [Dem17] Deming, D. J. “The Growing Importance of Social Skills in the Labor Market*”, *The Quarterly Journal of Economics*, vol. 132–4, 06 2017, pp. 1593–1640, https://academic.oup.com/qje/article-pdf/132/4/1593/30637899/qjx022_supp.pdf.
- [DM14] Dong, X.; McIntyre, S. “The second machine age: Work, progress, and prosperity in a time of brilliant technologies”, *Quantitative Finance*, vol. 14, 11 2014.
- [Dom12] Domingos, P. “A few useful things to know about machine learning”, *Commun. ACM*, vol. 55–10, oct 2012, pp. 78–87.
- [DvW11] Davis, S. J.; von Wachter, T. M. “Recessions and the cost of job loss”, *Brookings Papers on Economic Activity*, vol. 43–2, 2011, pp. 1–72.
- [ELL24] ELLGomes. “Guia-me: Um sistema de recomendação para ajudar a iniciar ou migrar de carreira na Área de TI”. Acessado em: 08 de novembro de 2024, Capturado em: <https://repositorio.ifpe.edu.br/xmlui/handle/123456789/1053>, 2024.
- [FKA24] Faruque, S.; Khushbu, S.; Akter, S. “Unlocking futures: A natural language driven career prediction system for computer science and software engineering students”, 05 2024.
- [FO13] Frey, C. B.; Osborne, M. A. “The future of employment: How susceptible are jobs to computerization?”, *Technological Forecasting and Social Change*, vol. 114, 2013, pp. 254–280.
- [Fri01a] Friedman, J. “Greedy function approximation: A gradient boosting machine”, *Annals of Statistics*, vol. 29, 10 2001, pp. 1189–1232.
- [Fri01b] Friedman, J. H. “Greedy function approximation: A gradient boosting machine”, *The Annals of Statistics*, vol. 29–5, 2001, pp. 1189–1232.
- [G121] G1. “Apenas 15% conseguem emprego na área em até 3 meses após formatura, diz pesquisa”. Accessed: 2024-12-13, Capturado em: <https://g1.globo.com/economia/concursos-e-emprego/noticia/2021/04/18/apenas-15percent-conseguem-emprego-na-area-em-ate-3-meses-apos-formatura-diz-p-1.html>, 2021.
- [GFWA17] Graziotin, D.; Fagerholm, F.; Wang, X.; Abrahamsson, P. “Consequences of unhappiness while developing software”, *arXiv preprint arXiv:1701.05789*, 2017.

- [GHL13] Gortmaker, S.; Hosmer, D.; Lemeshow, S. “Applied logistic regression”, *Contemp Sociol.*, vol. 23, 01 2013.
- [Hea17] Heaton, J. “Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618”, *Genetic Programming and Evolvable Machines*, vol. 19, 10 2017.
- [HTF09] Hastie, T.; Tibshirani, R.; Friedman, J. “The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)”. 2009.
- [Hub23] Hubstaff Blog. “Employee turnover statistics and trends for 2024”. Accessed: 2024-12-13, Capturado em: <https://hubstaff.com/blog/employee-turnover-statistics/>, 2023.
- [Koh01] Kohavi, R. “A study of cross-validation and bootstrap for accuracy estimation and model selection”, vol. 14, 03 2001.
- [LJT+17] Li, L.; Jing, H.; Tong, H.; Yang, J.; He, Q.; Chen, B.-C. “Nemo: Next career move prediction with contextual embedding”, 2017, pp. 505–513.
- [MN01] Mccallum, A.; Nigam, K. “A comparison of event models for naive bayes text classification”, *Work Learn Text Categ*, vol. 752, 05 2001.
- [Mur13] Murphy, K. P. “Machine learning : a probabilistic perspective”. Cambridge, Mass. [u.a.]: MIT Press, 2013.
- [NTMEB+18] Nadjem, A.; Torres-Moreno, J.-M.; El-Bèze, M.; Marrel, G.; Bonte, B. “Predicting personalized academic and career roads: First steps toward multi-uses recommender systems”, 2018, pp. 1–4.
- [RRS10] Ricci, F.; Rokach, L.; Shapira, B. “Recommender Systems Handbook”. 2010, vol. 1-35, pp. 1–35.
- [Sav05] Savickas, M. “Career construction theory and practice”, *Career development and counseling: Putting theory and research to work*, 01 2005.
- [Sen23] Senac Piauí. “Oportunidades e desafios nas carreiras de tecnologia”. Accessed: 2024-12-13, Capturado em: <https://blog.pi.senac.br/oportunidades-e-desafios-nas-carreiras-de-tecnologia/>, 2023.
- [SSWRB24] Siswipraptini, P.; Spits Warnars, H. L. H.; Ramadhan, A.; Budiharto, W. “Personalized career-path recommendation model for information technology students in indonesia”, *IEEE Access*, vol. PP, 03 2024, pp. 1–1.

- [Ter23] Terra Notícias. “Apenas 23% dos funcionários estão engajados no trabalho”. Accessed: 2024-12-13, Capturado em: <https://www.terra.com.br/noticias/apenas-23-dos-funcionarios-estao-engajados-no-trabalho%2C9ca3578a165be71c1c3f06e187a42a6dnnioev7u.html>, 2023.
- [Vap00] Vapnik, V. “The Nature of Statistical Learning Theory”. 2000, vol. 8, pp. 1–15.
- [wef20] “The future of jobs report 2020”. Accessed: 2024-12-13, Capturado em: <https://www.weforum.org/publications/the-future-of-jobs-report-2020/>, 2020.
- [Wes23] West Monroe Partners. “It soft skills: Why tech skills alone aren’t enough”. Accessed: 2024-12-13, Capturado em: <https://www.westmonroe.com/perspectives/signature-research/it-soft-skills?>, 2023.