

# Aplicação de Inteligência Artificial para Análise e Predição do Desempenho Acadêmico em Faculdades Brasileiras: Uma Abordagem Baseada em Dados do Censo

Guilherme Fernandes Medeiros

UNIFESP - Universidade Federal do Estado de São Paulo

São José dos Campos, Brasil

medeiros.guilherme@unifesp.br

**Abstract**—Este estudo explora a aplicação de conceitos e técnicas de inteligência artificial na previsão e diagnóstico do desempenho acadêmico de alunos do ensino superior. Através de experimentos que envolvem análises de correlação e classificação, buscamos identificar tendências e fatores que influenciam o desempenho dos estudantes. Para isso, os dados foram processados para eliminar strings e manter apenas variáveis numéricas, permitindo uma análise mais precisa. Através da aplicação de algoritmos, foi possível identificar as causas subjacentes às variações no desempenho acadêmico no contexto do ensino superior.

**Index Terms**—inteligência artificial, performance acadêmica, algoritmos, previsão de performance.

## I. INTRODUÇÃO

A inteligência artificial (IA) tem se tornado uma peça fundamental e cada vez mais influente na sociedade contemporânea, gerando grandes revoluções tecnológicas em diversos setores. Desde a área da saúde até a indústria educacional, a IA proporciona inovações reveladoras que, em sua essência, têm o potencial de melhorar a eficiência e qualidade dos serviços prestados. Dentre as áreas impactadas pela tecnologia, a educação se destaca como um setor vital que pode se beneficiar enormemente das aplicações da IA. Nos últimos anos, a IA tem demonstrado um crescimento notável em uma ampla gama de aplicações. Esse crescimento tem sido alimentado pela disponibilidade cada vez maior de dados e informações sobre pessoas, telemetrias e processamentos e pelo desenvolvimento de novos algoritmos mais sofisticados e eficientes, que são capazes de automaticamente aprender e se adaptar a inúmeros cenários. No âmbito da educação, especificamente no contexto do ensino superior brasileiro, a IA oferece uma oportunidade valiosa para otimizar o desempenho acadêmico. As faculdades brasileiras, como muitas instituições ao redor do mundo, estão repletas de dados, desde as informações demográficas dos estudantes até suas notas e taxas de frequência. Esses dados, se devidamente analisados e interpretados, podem oferecer insights valiosos sobre o desempenho acadêmico dos alunos e ajudar a prever e melhorar os resultados futuros. Neste trabalho, exploraremos como a IA pode ser usada para analisar e prever o desempenho acadêmico nas faculdades brasileiras,

usando dados do Censo de 2019. Ao aplicar métodos de IA a esses dados, esperamos identificar padrões e tendências que possam servir como indicadores preditivos de desempenho acadêmico. Ao fazer isso, podemos desenvolver novas estratégias para melhorar a qualidade do ensino e ajudar os estudantes a alcançarem seu potencial máximo. Ao mesmo tempo, essa abordagem baseada em dados pode ajudar a tornar o setor de educação mais eficiente e eficaz, contribuindo para uma formação de uma sociedade mais bem educada e preparada para enfrentar os desafios do século XXI.

## II. CONCEITOS FUNDAMENTAIS

### A. O que é a Inteligência Artificial (IA)?

A Inteligência Artificial é uma área da ciência da computação que se concentra na criação e desenvolvimento de máquinas e softwares capazes de exibir formas de inteligência semelhantes às humanas. Isso pode incluir a capacidade de aprender, raciocinar, resolver problemas, perceber o ambiente, interpretar a linguagem e até mesmo exibir criatividade.

### B. O que é o Aprendizado de Máquina (Machine Learning)?

O Aprendizado de Máquina, também conhecido como Machine Learning, é um setor da inteligência artificial que se concentra no desenvolvimento de algoritmos e modelos computacionais capazes de aprender e tomar decisões automáticas a partir de dados, sem serem explicitamente programados para cada tarefa em específico. Ao invés disso, o algoritmo de aprendizado de máquina é projetado para analisar padrões e informações dos dados de treinamento e fazer previsões, tomar decisões ou realizar tarefas.

### C. O que são Algoritmos de Aprendizado Supervisionado e Não Supervisionado?

Os algoritmos de aprendizado de máquina, um subcampo da Inteligência Artificial (IA), são classificados em duas categorias principais: aprendizado supervisionado e aprendizado não supervisionado. Essas categorias diferem na maneira como os dados são utilizados para treinar o modelo. Ambos os tipos de aprendizado de máquina têm suas aplicações e podem

ser utilizados de maneira complementar. A escolha entre aprendizado supervisionado e não supervisionado depende da natureza dos dados disponíveis e do problema específico que se deseja resolver.

#### *D. O que é a mineração de dados?*

Mineração de dados, ou data mining em inglês, é o processo de descoberta de padrões, informações e conhecimentos úteis em grandes conjuntos de dados. É uma etapa importante na análise de dados, que visa extrair insights e tomar decisões baseadas nos dados disponíveis. A mineração de dados envolve a aplicação de técnicas estatísticas e algoritmos de aprendizado de máquina para explorar os dados e identificar padrões significativos. Esses padrões podem incluir relações entre variáveis, tendências, agrupamentos de dados semelhantes e até mesmo previsões de eventos futuros.

#### *E. O que é a programação genética (GP)?*

A Programação Genética (GP, Genetic Programming) é uma técnica de inteligência artificial que utiliza princípios inspirados na evolução biológica para resolver problemas computacionais complexos. Ela faz parte do campo mais amplo da programação evolucionária. Na programação genética, um conjunto de programas candidatos é gerado inicialmente em forma de árvores sintáticas. Essas árvores representam soluções potenciais para o problema em questão, e cada nó da árvore corresponde a uma função ou operação.

#### *F. O que são redes neurais?*

Redes neurais são modelos matemáticos inspirados no funcionamento do cérebro humano. Elas são compostas por unidades básicas chamadas de neurônios artificiais, que são interconectados entre si. Cada neurônio artificial recebe entradas, realiza um processamento e produz uma saída. As redes neurais são treinadas com conjuntos de dados para aprender padrões e fazer previsões ou tomar decisões. Durante o treinamento, os pesos das conexões entre os neurônios são ajustados para minimizar a diferença entre as saídas desejadas e as saídas produzidas pela rede. Esse processo é chamado de aprendizado.

#### *G. O que é o algoritmo Random Forest (RF)?*

O algoritmo RF é um algoritmo de aprendizado de conjunto homogêneo desenvolvido com base na árvore de decisão. Ele seleciona amostras e atributos de recursos através da ideia de seleção aleatória. O RF pode lidar com dados com ruído e valores ausentes e tem um processo de ajuste de dados rápido. A importância dos recursos no RF pode fornecer uma base para a seleção de recursos.

#### *H. O que é o scikit-learn?*

O scikit-learn é uma biblioteca em Python amplamente utilizada para aprendizado de máquina. Ele fornece um conjunto de ferramentas eficientes e fáceis de usar para realizar várias tarefas relacionadas à análise de dados e modelagem preditiva.

#### *I. O que são regras de associação?*

As regras de associação são uma técnica popular de mineração de dados que visa descobrir relações interessantes, padrões ou associações entre conjuntos de itens em grandes bases de dados. Essas regras são geralmente usadas para identificar padrões de comportamento do cliente em compras, uma vez que permitem identificar produtos frequentemente comprados juntos.

### III. TRABALHOS RELACIONADOS

#### *A. Prever e Melhorar o Desempenho dos Alunos com Uso Combinado de Aprendizagem de Máquina e GPT [1]*

A aplicação de inteligência artificial para análise e previsão do desempenho acadêmico em faculdades brasileiras apresentada neste texto possui uma abordagem inovadora e promissora. A utilização de algoritmos de aprendizado de máquina em conjunto com o GPT (um modelo de linguagem avançado) demonstra o potencial de melhorar a experiência de aprendizado dos alunos e prever seu desempenho de forma mais precisa. Um dos pontos positivos dessa abordagem é a natureza quantitativa e experimental do estudo. Ao processar dados de 900 alunos utilizando 21 algoritmos diferentes, os pesquisadores foram capazes de obter resultados significativos. Essa abordagem ampla e abrangente permite uma análise mais completa e confiável, que pode beneficiar a comunidade acadêmica como um todo. Os resultados indicaram que a combinação dos algoritmos de aprendizado de máquina com o GPT superou outros métodos de previsão e melhoria do desempenho dos alunos. Isso sugere que a abordagem apresentada neste estudo pode ser uma ferramenta poderosa e efetiva para identificar as lacunas de conhecimento dos alunos e fornecer feedback personalizado, levando a uma formação mais efetiva. Além disso, a utilização de inteligência artificial nesse contexto educacional pode ter um impacto transformador. A capacidade de prever o desempenho acadêmico dos alunos com base em dados históricos e características individuais pode auxiliar professores e instituições a direcionar seus esforços de ensino de maneira mais eficiente. Isso pode resultar em um ensino mais personalizado e adaptado às necessidades de cada aluno, melhorando o aprendizado e aumentando as chances de sucesso acadêmico. No entanto, é importante considerar algumas questões éticas e práticas relacionadas à aplicação de inteligência artificial na educação. A privacidade dos dados dos alunos deve ser respeitada e garantida, bem como a transparência dos algoritmos utilizados. É fundamental que os resultados obtidos sejam interpretados e usados com cuidado, evitando qualquer forma de discriminação ou estigmatização dos estudantes. Em resumo, a aplicação de inteligência artificial para análise e previsão do desempenho acadêmico em faculdades brasileiras, como apresentada neste texto, possui um potencial significativo para melhorar a educação. No entanto, é essencial garantir uma abordagem ética, transparente e responsável, para que os benefícios sejam maximizados e quaisquer desafios ou limitações sejam adequadamente abordados.

### *B. Predição do Desempenho Acadêmico de Graduandos Utilizando Mineração de Dados Educacionais [2]*

O trabalho apresentado neste texto aborda a aplicação de Mineração de Dados Educacionais (EDM) para a predição do desempenho acadêmico de graduandos em universidades públicas brasileiras. A proposta visa fornecer informações úteis aos gestores educacionais, permitindo a identificação dos graduandos em risco de abandonar o sistema de ensino. A arquitetura EDM WAVE abrange todo o processo de descoberta de conhecimento em dados, desde o pré-processamento até o pós-processamento. Uma característica relevante dessa abordagem é que ela se baseia exclusivamente em dados acadêmicos que variam ao longo do tempo e são armazenados no sistema de gestão acadêmica. Nenhuma informação social ou econômica é considerada nas análises realizadas. Esse enfoque tem a vantagem de simplificar a coleta de dados e tornar o processo mais escalável, utilizando as informações já disponíveis no contexto educacional. Os resultados experimentais obtidos através da aplicação da arquitetura proposta são encorajadores. A precisão da predição do desempenho acadêmico dos graduandos a cada semestre letivo alcançou cerca de 80%. Esses resultados são importantes para o desenvolvimento de estratégias de intervenção e suporte adequadas aos estudantes em risco de abandono, bem como para aprimorar a eficácia do sistema de ensino em geral. A possibilidade de identificar as variáveis-chave relacionadas ao desempenho acadêmico permite uma compreensão mais aprofundada dos fatores que influenciam o sucesso dos estudantes, proporcionando informações valiosas para o planejamento de ações e políticas educacionais. No entanto, é importante mencionar que a abordagem apresentada neste trabalho possui algumas limitações. Por se basear exclusivamente em dados acadêmicos, pode não capturar completamente as complexidades e influências externas que podem afetar o desempenho dos estudantes, como fatores socioeconômicos ou aspectos emocionais. Portanto, é fundamental considerar esses aspectos e complementar as análises com outras fontes de dados relevantes. Em conclusão, o trabalho apresentado demonstra uma proposta promissora de aplicação de Mineração de Dados Educacionais para a predição do desempenho acadêmico de graduandos. A abordagem baseada na arquitetura EDM WAVE, utilizando apenas dados acadêmicos ao longo do tempo, mostra resultados significativos e permite identificar variáveis-chave relacionadas ao sucesso dos estudantes. Essa abordagem tem o potencial de apoiar gestores educacionais na tomada de decisões e no desenvolvimento de estratégias para melhorar o desempenho dos estudantes e reduzir o abandono escolar. No entanto, é importante considerar as limitações e complementar a análise com outros aspectos relevantes para uma compreensão mais completa dos fatores que influenciam o sucesso acadêmico.

### *C. A Inteligência Artificial como Ferramenta de Apoio no Ensino Superior [3]*

A utilização da inteligência artificial na educação superior tem se expandido significativamente, pois ela oferece diver-

sas possibilidades de aprimoramento do ensino. A aplicação dessa tecnologia como ferramenta de apoio visa melhorar a eficiência e a eficácia do ensino, personalizar o aprendizado, aumentar o acesso ao conhecimento e reduzir os custos. Uma das vantagens da inteligência artificial na educação é a capacidade de tornar o processo de aprendizagem mais dinâmico. Sistemas de tutoria inteligente podem ser adotados para auxiliar os alunos, proporcionando um aprendizado mais rápido e eficiente. Com base nas necessidades individuais de cada estudante, esses sistemas podem fornecer suporte personalizado, adaptando-se ao ritmo e estilo de aprendizado de cada um. No entanto, assim como em qualquer tecnologia, a implementação da inteligência artificial no ensino superior também apresenta desafios e limitações. A capacitação dos professores é um aspecto crucial, pois eles precisam compreender e utilizar efetivamente as ferramentas e recursos oferecidos pela inteligência artificial. É necessário investir em treinamento e desenvolvimento profissional para que os educadores possam aproveitar ao máximo os benefícios dessa tecnologia. Questões éticas e legais também surgem com o uso da inteligência artificial na educação. É fundamental garantir a privacidade e a segurança dos dados dos alunos, bem como garantir a transparência dos algoritmos utilizados. Além disso, é necessário considerar o impacto potencial da inteligência artificial na equidade e na justiça educacional, evitando a reprodução de vieses e discriminação. A dependência tecnológica é outro desafio a ser considerado. A utilização da inteligência artificial requer infraestrutura adequada, como acesso à internet e dispositivos compatíveis. Além disso, é preciso estar preparado para lidar com possíveis falhas técnicas e atualizações constantes dos sistemas. Aspectos financeiros também podem ser uma limitação na implementação da inteligência artificial na educação superior. Os investimentos necessários para adquirir e manter a tecnologia podem representar um desafio para as instituições de ensino, especialmente as de recursos limitados. Para garantir o sucesso da utilização da inteligência artificial no ensino superior, é crucial que sejam considerados cuidadosamente tanto os benefícios quanto os desafios. O desenvolvimento de políticas e práticas que promovam uma utilização responsável e ética da tecnologia é essencial. É necessário equilibrar a inovação tecnológica com os princípios pedagógicos, garantindo que a inteligência artificial seja uma ferramenta valiosa que contribua para uma educação de qualidade e acessível a todos.

### *D. Avaliação de Desempenho Acadêmico Utilizando Redes Neurais: Uma Análise Exploratória de Dados de Rankings Universitários [4]*

O texto aborda a importância da avaliação universitária e sua influência no cenário educacional global. O processo de avaliação é destacado como um meio de proporcionar reputação, prestígio social e acadêmico às instituições de ensino, além de fornecer um benchmark para comparação entre as instituições e identificação de pontos fortes e fracos. No entanto, o método atual de ranqueamento acadêmico, baseado em indicadores e notas, é criticado por não apresentar

consenso na formulação dos critérios e por reduzir o desempenho das instituições a um único número. Nesse contexto, a dissertação propõe uma abordagem alternativa utilizando técnicas de agrupamento com redes neurais, mais especificamente os mapas auto-organizáveis (SOM). Os mapas auto-organizáveis são modelos de redes neurais competitivas que realizam um mapeamento entre dados multidimensionais para uma representação bidimensional, aproximando a densidade original das informações. Essa técnica é comumente utilizada em análise de dados e reconhecimento de padrões. No trabalho apresentado, é realizada uma análise transversal dos dados das universidades brasileiras, utilizando os rankings universitários da Folha de 2014 e 2019 como dados de treinamento dos mapas. A partir dos perfis de agrupamento identificados nos mapas, é possível identificar os pontos positivos e negativos de cada agrupamento, permitindo uma análise mais detalhada do desempenho das instituições. Além disso, é realizada uma análise das transições entre os agrupamentos nos anos de 2014 e 2019, buscando compreender as mudanças e caracterizar o comportamento das instituições nesse período. Esse estudo oferece uma nova alternativa de análise dos dados de desempenho das Instituições de Ensino Superior (IES), que vai além do ranqueamento tradicional. Ao utilizar os mapas auto-organizáveis, é possível visualizar as disparidades existentes entre as regiões do Brasil e identificar tendências e padrões de desempenho ao longo do tempo. No entanto, é importante considerar as limitações dessa abordagem. O uso de técnicas de agrupamento com redes neurais pode fornecer uma perspectiva diferente sobre o desempenho das IES, mas também possui desafios, como a seleção adequada dos indicadores a serem considerados e a interpretação correta dos resultados. Além disso, é importante ter em mente que o desempenho acadêmico de uma instituição envolve diversos aspectos que vão além dos dados analisados, como a qualidade do corpo docente, a infraestrutura disponível e a produção científica. Em conclusão, o estudo apresentado propõe uma abordagem alternativa ao ranqueamento acadêmico tradicional, utilizando técnicas de agrupamento com redes neurais. Por meio dos mapas auto-organizáveis, é possível analisar o desempenho das IES, identificar pontos fortes e fracos e compreender as tendências ao longo do tempo. Essa abordagem oferece uma nova perspectiva de análise dos dados e contribui para a compreensão das disparidades regionais no contexto das universidades brasileiras. No entanto, é necessário considerar as limitações e desafios inerentes a essa técnica e complementá-la com outras formas de avaliação para uma compreensão abrangente do desempenho acadêmico.

#### *E. O Perfil Docente no Ensino Superior Privado e o Desempenho no Enade [5]*

O artigo apresenta um estudo descritivo que visa investigar a relação entre as características do perfil da docência no ensino superior privado e o desempenho acadêmico dos estudantes, medido pela média do Exame Nacional de Desempenho dos Estudantes (ENADE). Os dados utilizados são provenientes do ENADE (2016 e 2017), assim como os microdados dos

docentes e instituições de ensino superior do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. A análise foi realizada por meio de uma regressão linear múltipla, buscando identificar as variáveis que mais impactam a média do desempenho dos estudantes no ENADE por instituição. Os resultados obtidos revelaram que o fator mais significativo e positivo está relacionado ao grau de escolaridade dos docentes. Ou seja, quanto maior o nível de formação acadêmica dos professores, maior tende a ser o desempenho dos estudantes. Além disso, o estudo constatou que o regime de trabalho dos docentes não demonstrou ser uma variável relevante para agregar valor ao desempenho dos estudantes no ENADE. Isso indica que o tempo de trabalho dos professores não apresentou uma relação direta com o desempenho acadêmico dos estudantes nas instituições de ensino superior privadas analisadas. Esses resultados destacam a importância da formação acadêmica dos docentes como um fator determinante para o desempenho dos estudantes. Isso ressalta a necessidade de investimentos na qualificação e atualização dos professores, visando melhorar a qualidade do ensino superior. No entanto, é importante ressaltar que esse estudo se baseia em dados específicos do ENADE e dos microdados dos docentes e instituições de ensino superior privadas, limitando sua generalização para outras realidades e contextos educacionais. Além disso, é possível que outros fatores não abordados no estudo também influenciam o desempenho acadêmico dos estudantes. Em conclusão, o artigo demonstra que o grau de escolaridade dos docentes é um fator relevante para o desempenho dos estudantes no ensino superior privado, enquanto o regime de trabalho dos docentes não apresenta relação direta com o desempenho acadêmico. Essas informações podem fornecer insights importantes para políticas educacionais e práticas pedagógicas visando a melhoria da qualidade do ensino superior. No entanto, mais pesquisas são necessárias para uma compreensão abrangente dos diversos fatores que influenciam o desempenho dos estudantes nesse contexto.

#### *F. Artificial Intelligence-enabled Prediction Model of Student Academic Performance in Online Engineering Education [6]*

O artigo apresenta um modelo de previsão baseado em inteligência artificial (IA) para o desempenho acadêmico de estudantes em educação online de engenharia. Ele destaca que a programação genética (GP) pode ser usada para desenvolver modelos de previsão quantitativa para estabelecer as relações exatas entre as variáveis de entrada e a resposta de saída na previsão do desempenho de aprendizagem dos alunos. Além disso, o artigo também destaca que a aplicação de GP na educação online para obter os modelos de previsão quantitativa ainda não foi explorada, especialmente ao estabelecer certos critérios para analisar e converter processos de aprendizagem em dados de entrada. O artigo também enfatiza a importância de coletar conjuntos de dados com amplas faixas para as variáveis preditoras para desenvolver modelos mais robustos. Em adição, o estudo foi realizado em um curso de engenharia online - "Smart Marine Metastructures" - oferecido na Ocean College na Zhejiang University na China. Os participantes

eram 35 estudantes de pós-graduação em tempo integral da Ocean College na universidade; todos eram chineses com idades entre 22 e 27 anos. Por fim, conclui-se que o modelo de previsão baseado em IA pode ser usado para avaliar e prever o desempenho de aprendizagem na educação online de engenharia. Os principais achados indicaram que as variáveis dominantes no curso de engenharia online eram a aquisição de conhecimento, seguida pela participação em aula e o desempenho somativo, enquanto o conhecimento pré-requisito tendia a não desempenhar um papel fundamental.

#### *G. Modelling, Prediction and Classification of Student Academic Performance Using Artificial Neural Networks [7]*

O artigo de explora a aplicação de Redes Neurais Artificiais (ANNs) na modelagem, previsão e classificação do desempenho acadêmico dos alunos. O autor utiliza ANNs para prever o Coeficiente de Rendimento Acadêmico (CGPA) dos alunos e classificar os dados através de observações de entrada. Este processo é realizado com base em aprendizado de máquina supervisionado, onde o modelo ANN é expresso como uma função matemática. Os vetores de saída e entrada são representados por  $Y$  e  $X$ , respectivamente, e  $W$  é um vetor de parâmetros de peso que representa as conexões dentro da ANN. Para medir o desempenho da ANN, o autor utiliza um histograma de erro, que demonstra como os erros são distribuídos, com a maioria dos erros ocorrendo perto de zero. Além disso, uma matriz de confusão, também conhecida como matriz de erro, é usada para verificar o desempenho da ANN em termos de classificações. A matriz de confusão permite medidas de taxas como precisão de previsão, taxa de erro, sensibilidade, especificidade e precisão. O artigo também revisa várias aplicações anteriores de ANNs na educação. Por exemplo, ANNs foram usadas para modelar e simular a diversidade de estilos de aprendizagem entre os alunos através de dois paradigmas de aprendizagem: aprendizagem supervisionada (aprendizagem com professor) e aprendizagem não supervisionada (aprendizagem sem professor e através do autoestudo dos alunos). Com base nos resultados, os autores acreditam fortemente que a modelagem educacional usando ANNs pode ajudar a avaliar o desempenho dos alunos. Eles sugerem que modelos futuros podem incluir outros atributos, como o papel dos professores, classes curriculares adicionais, feedbacks de cursos e sistemas de e-learning, a fim de complementar uma modelagem ANN completa, enquanto se concentra na precisão da previsão da ANN. Em resumo, o artigo, apresenta uma visão abrangente da aplicação de ANNs na previsão e classificação do desempenho acadêmico dos alunos, destacando seu potencial para melhorar a avaliação do desempenho dos alunos e sugerindo direções para futuras pesquisas.

#### *H. Student Academic Performance Prediction using Artificial Neural Networks: A Case Study [8]*

O artigo aborda o problema significativo de abandono escolar e atraso na formação no Katsina State Institute of Technology and Management (KSITM). O autor identifica o

desempenho dos alunos durante o primeiro ano como um dos principais fatores que contribuem para esses problemas. O estudo visa prever o desempenho acadêmico fraco que pode levar ao abandono escolar ou atraso na formatura, permitindo que a instituição desenvolva programas estratégicos para melhorar o desempenho do aluno.

Para atingir esse objetivo, o autor utiliza Redes Neurais Artificiais (ANNs) para prever o GPA (Grade Point Average) dos alunos. O modelo leva em consideração as informações pessoais dos alunos, informações acadêmicas e local de residência. O conjunto de dados utilizado para treinar e testar o modelo consiste em informações de 61 alunos do curso de Redes de Computadores, e o modelo foi implementado no software WEKA.

O modelo demonstrou uma precisão promissora, sendo capaz de prever corretamente 73,68% do desempenho dos alunos e, especificamente, 66,67% dos alunos que provavelmente abandonariam a escola ou teriam atraso antes de se formar. A precisão do modelo foi avaliada usando vários critérios de avaliação conhecidos, incluindo a porcentagem de instâncias corretamente/incorrectamente classificadas, a estatística Kappa, a taxa de verdadeiros positivos (TP), a taxa de falsos positivos (FP), precisão (especificidade), recall (sensibilidade) e a medida F.

No entanto, o autor também observa que o uso de um pequeno conjunto de dados com atributos limitados e poucas instâncias de resultados de alunos do departamento de redes de 2018 torna o modelo menos confiável e pode falhar ou produzir um resultado enganoso quando aplicado a alunos de outros departamentos ou até mesmo a diferentes alunos do mesmo departamento. Para trabalhos futuros, o autor sugere que a pesquisa pode ser estendida com mais atributos distintos e um conjunto de dados maior para desenvolver um modelo mais preciso. Além disso, o tipo de rede neural, o número de camadas ocultas e o número de neurônios podem ser estudados intensivamente para determinar seus efeitos no treinamento de um modelo.

Em resumo, o artigo apresenta uma visão abrangente da aplicação de ANNs na previsão do desempenho acadêmico dos alunos, destacando seu potencial para melhorar a avaliação do desempenho dos alunos e sugerindo direções para futuras pesquisas.

#### *I. Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods [9]*

O artigo explora a aplicação do algoritmo Random Forest (RF) na previsão do desempenho acadêmico dos alunos em um ambiente de educação online. O artigo começa destacando que o comportamento dos alunos durante o processo de aprendizagem pode afetar seu desempenho acadêmico. A previsão e análise do desempenho dos alunos são realizadas principalmente através da capacidade do aluno, do processo de pesquisa, do propósito da pesquisa, do algoritmo e do conjunto de dados, bem como do suporte à análise de dados. O autor descreve o processo de construção de uma árvore

de decisão, que classifica os dados de acordo com uma série de regras. O algoritmo da árvore de decisão leva o melhor atributo no conjunto de treinamento como o nó raiz e, em seguida, encontra cada subnó através de recursão. A entropia da informação é usada para medir a incerteza de uma categoria, bem como a incerteza de um recurso. O autor então descreve o processo de construção do modelo de previsão de desempenho acadêmico online baseado no algoritmo RF. O modelo é construído com a ajuda do kit de ferramentas Scikit-learn baseado na linguagem Python. Os dados para a pesquisa são principalmente de alguns dados de aprendizado de alunos no banco de dados de uma plataforma de aprendizado de curso online. O autor conclui que o algoritmo RF pode melhorar o efeito de previsão do algoritmo e melhorar a precisão de reconhecimento da análise do comportamento de aprendizagem. No entanto, o autor também observa que a análise da relação entre as características de aprendizagem dos alunos é insuficiente, por isso não pode demonstrar de forma abrangente o estado de aprendizagem dos alunos online. Para pesquisas futuras, o autor sugere que o algoritmo de previsão pode ser analisado e selecionado profundamente para otimizar o método de previsão de desempenho acadêmico da educação online.

#### *J. Using Artificial Intelligence to Predict Students' Academic Performance in Blended Learning [10]*

O artigo aborda a aplicação de inteligência artificial (IA) para prever o desempenho acadêmico dos alunos em um ambiente de aprendizado híbrido. Os autores destacam a popularidade crescente dos algoritmos de otimização meta-heurística em campos como IA, otimização, aplicações de engenharia, mineração de dados e aprendizado de máquina. Esses algoritmos são flexíveis e podem lidar com diversas funções objetivas, sejam elas discretas, contínuas ou mistas. Além disso, são simples de aplicar devido à sua capacidade de resolver problemas complexos. No estudo, os autores utilizaram um novo modelo de Rede Neural Artificial (RNA), especificamente uma Rede Neural Perceptron Multicamadas (MLPNN) com Algoritmo de Otimização Firefly (FFA), para analisar a previsibilidade do desempenho dos alunos. Eles verificaram a generalidade da previsão através de novos fatores, que incluem exames intermediários, tarefas, presenças e interações virtuais/presenciais. Os autores também discutem a aplicação de três modelos estatísticos: mínimos quadrados ordinários (OLS), efeitos fixos e efeitos aleatórios. Esses modelos foram usados para analisar a relação entre uma única variável de saída e as variáveis de entrada. Os autores observaram que as variáveis de entrada (interações virtuais/presenciais, exames intermediários, tarefas) afetam positivamente os exames finais com um nível de significância  $p < 0.05$ . Os autores concluíram que todos os três modelos de previsão criados demonstraram mais de 80% de precisão. Em resumo, este artigo apresenta uma abordagem inovadora para a previsão do desempenho acadêmico dos alunos utilizando IA e otimização meta-heurística, demonstrando a eficácia dessas técnicas na análise e previsão do desempenho acadêmico.

## IV. OBJETIVOS DO TRABALHO

O objetivo principal deste trabalho é explorar a aplicação de técnicas de mineração de regras de associação para analisar e interpretar os padrões presentes nos dados do censo educacional. Pretendemos identificar e analisar as associações entre variáveis presentes nos dados, focando na natureza das instituições de ensino, o acesso à informática e à comunicação, e as taxas de ingresso. Através deste estudo, buscamos não apenas compreender melhor os fatores que podem influenciar as características das instituições e o acesso dos estudantes a recursos essenciais, mas também fornecer insights que possam ser utilizados para otimizar as estratégias de administração educacional. Com isso, pretendemos fornecer orientações embasadas em dados para instituições de ensino que buscam melhorar suas práticas, potencialmente beneficiando o ambiente de aprendizado dos estudantes.

## V. METODOLOGIA EXPERIMENTAL

Nossa metodologia experimental envolve a aplicação de técnicas de análise de dados e mineração de regras de associação para interpretar os dados do Censo de 2023. Inicialmente, procederemos com a limpeza e o pré-processamento dos dados para assegurar que estejam em um formato apropriado para a análise. Isso envolve a remoção de dados irrelevantes ou redundantes, a correção de erros e inconsistências, e a transformação de dados categóricos em numéricos, se necessário. Após o pré-processamento dos dados, utilizaremos o algoritmo Apriori, uma popular técnica de mineração de regras de associação, para identificar relações significativas entre diferentes variáveis. Este algoritmo nos ajudará a descobrir regras interessantes que não são imediatamente aparentes nos dados brutos. Cada regra de associação gerada será avaliada com base em métricas como suporte, confiança e lift. O suporte nos ajudará a entender a frequência com que uma regra aparece nos dados, a confiança mostrará a probabilidade de ocorrência de um item dado que outro item já ocorreu, e o lift indicará a força da associação entre dois itens. Realizaremos também uma análise de relevância das características para identificar quais variáveis têm o maior impacto nas regras geradas. Isso nos permitirá compreender melhor quais fatores são mais influentes na determinação de determinados padrões. Com base nos resultados de nossa análise, desenvolveremos recomendações para as universidades e instituições de ensino. Essas recomendações serão baseadas em evidências e guiadas por dados, e visarão ajudar as instituições de ensino a tomar decisões mais informadas e eficazes. As recomendações podem incluir, por exemplo, estratégias para melhorar o acesso à informática e à comunicação, ou considerações sobre políticas de cobrança de ingressos, dependendo das regras de associação que identificarmos.

## VI. RESULTADOS E DISCUSSÃO

Nossa pesquisa envolveu a análise minuciosa de um conjunto de dados extremamente extenso fornecido pelo Censo de 2019. Originalmente composto por mais de 250 colunas, aplicamos uma rigorosa abordagem de pré-processamento de

dados para reduzi-lo a 16 colunas relevantes. Essa etapa foi essencial para garantir a relevância e a clareza dos resultados de nosso estudo. Ao aplicar a mineração de regras de associação no conjunto de dados tratado, revelamos várias regras interessantes. Essas regras fornecem informações valiosas sobre os fatores que mais contribuem para o desempenho acadêmico nas instituições de ensino superior. Por exemplo, a regra de associação "(Ingresso Pago, Informatica Semi)  $\Rightarrow$  (Privada com fins lucrativos)" com um lift de 1.46 e leverage de 0.10 mostra que universidades que cobram ingresso e oferecem acesso intermediário à informática são mais propensas a serem privadas e com fins lucrativos do que seria esperado caso esses atributos fossem independentes. Isso evidencia o papel do modelo de negócio na oferta de infraestrutura tecnológica e a política de cobrança de ingresso. Outra regra importante encontrada é "(Privada com fins lucrativos)  $\Rightarrow$  (Ingresso Pago)", com um lift de 1.41 e leverage de 0.11, indicando que universidades privadas com fins lucrativos têm uma tendência maior de cobrar ingressos do que o esperado caso esses dois atributos fossem independentes. Isso destaca a natureza comercial dessas instituições. Interessantemente, a regra "(Comunicacao Semi)  $\Rightarrow$  (Informatica Semi)" com um lift de 1.30 e leverage de 0.13 sugere que universidades que têm um nível intermediário de acesso à comunicação tendem a ter também um nível intermediário de acesso à informática, mais do que seria esperado caso esses dois atributos fossem independentes. Isso sugere uma correlação entre as capacidades de comunicação e informática nas instituições de ensino. Além disso, a regra "(Privada com fins lucrativos)  $\Rightarrow$  (Ingresso Pago, Informatica Semi)" com um lift de 1.46 e leverage de 0.10 sugere que as universidades privadas com fins lucrativos têm uma probabilidade maior de cobrar ingresso e oferecer um nível intermediário de acesso à informática do que seria esperado se esses atributos fossem independentes. Isso indica que políticas de cobrança e de acesso à tecnologia podem estar inter-relacionadas nesse tipo de instituição. Por último, a regra "(Proporção Ingressantes-Concluintes muito abaixo da média)  $\Rightarrow$  (Informatica Semi)" com um lift de 1.29 e leverage de 0.06 indica que universidades com uma proporção de ingressantes-concluintes muito abaixo da média têm uma tendência maior de oferecer um nível intermediário de acesso à informática do que seria esperado se esses dois atributos fossem independentes. Este pode ser um indício do impacto da infraestrutura de informática na retenção e conclusão dos estudantes. Nossa pesquisa ilustra a importância do acesso à informática e comunicação em relação ao desempenho acadêmico, e a diferença significativa no desempenho entre instituições privadas com fins lucrativos e aquelas sem fins lucrativos. Mais importante, fornece um argumento baseado em dados para priorizar políticas e intervenções que melhorem o acesso à informática e à comunicação e incentivem práticas de gestão eficazes nas instituições de ensino.

## VII. CONCLUSÃO

Ao longo deste estudo, exploramos a aplicação de técnicas de aprendizado de máquina e mineração de regras

de associação para analisar o desempenho acadêmico nas instituições de ensino superior, usando dados do Censo de 2019. Após um rigoroso processo de pré-processamento de dados, fomos capazes de extrair regras significativas que lançam luz sobre os fatores que influenciam esse desempenho. Descobrimos que aspectos como o modelo de negócio das instituições, as políticas de cobrança de ingresso, e o acesso à informática e comunicação têm um impacto significativo no desempenho acadêmico. Em particular, encontramos fortes associações entre o tipo de instituição (privada com fins lucrativos) e as práticas de cobrança de ingresso, bem como o nível de acesso à informática. Este estudo é um passo importante para compreender melhor os fatores que afetam o sucesso acadêmico e identificar onde as intervenções podem ser mais eficazes. Com base nos resultados, recomendamos que as políticas de melhoria do desempenho acadêmico levem em consideração não apenas a natureza do currículo, mas também o modelo de negócio da instituição, as políticas de cobrança e o acesso à tecnologia. Esperamos que as descobertas deste estudo informem políticas educacionais eficazes e orientadas por dados, ajudando instituições de ensino a criar ambientes de aprendizagem mais eficientes e inclusivos. No entanto, lembramos que, apesar dos insights significativos obtidos, há limitações no uso de dados do Censo para este propósito, uma vez que existem muitos outros fatores que também podem influenciar o desempenho acadêmico e que não são capturados neste conjunto de dados. Pesquisas futuras podem buscar incluir mais variáveis relacionadas a fatores socioeconômicos, à qualidade do ensino, ao envolvimento dos alunos, entre outros. Além disso, o uso de outras técnicas de aprendizado de máquina pode revelar ainda mais insights sobre este complexo tema.

## REFERENCES

- [1] Benevento, M., & Meirelles, F. de S. (2023). Prever e melhorar o desempenho dos alunos com o uso combinado de aprendizagem de máquina e GPT. *Revista De Gestão E Avaliação Educacional*, e74348, p. 1–22. <https://doi.org/10.5902/2318133874348>
- [2] Manhães, Laci Mary Barbosa; Cruz, Sérgio Manuel Serra da; "PREDIÇÃO DO DESEMPENHO ACADÊMICO DE ALUNOS DA GRADUAÇÃO UTILIZANDO MINERAÇÃO DE DADOS", p. 2050-2064 . In: *Anais do XIX Simpósio de Pesquisa Operacional & Logística da Marinha*. São Paulo: Blucher, 2020. ISSN 2175-6295, DOI 10.5151/spolm2019-148
- [3] Costa Júnior, J. F., Uilliane Faustino de Lima, Mário Domingos Leme, Leonardo Silva Moraes, Jonas Bezerra da Costa, Diogo Magalhães de Barros, Maria Aparecida de Moura Amorim Sousa, & Luis Carlos Ferreira de Oliveira. (2023). A inteligência artificial como ferramenta de apoio no ensino superior. *Rebena - Revista Brasileira De Ensino E Aprendizagem*, 6, 246–269. Recuperado de <https://rebena.emnuvens.com.br/revista/article/view/111>
- [4] MAIA JÚNIOR, Manoel Isac. Avaliação de desempenho acadêmico utilizando redes neurais: uma análise exploratória de dados de rankings universitários. 2020. 134f. Dissertação (Mestrado em Engenharia de Produção) - Centro de Tecnologia, Universidade Federal do Rio Grande do Norte, Natal, 2020.
- [5] Vieira, Alboni Marisa Dudeque Pianovski, & Schneiker, Daniel. (2021). O perfil docente no ensino superior privado e o desempenho no Enade. *Educação& Formação*, 6(2), e4194. Epub 20 de abril de 2021. <https://doi.org/10.25053/redufor.v6i2.4194>
- [6] Jiao, P., Ouyang, F., Zhang, Q. *et al.* Artificial intelligence-enabled prediction model of student academic performance in on-

- line engineering education. *Artif Intell Rev* **55**, 6321–6344 (2022). <https://doi.org/10.1007/s10462-022-10155-y>
- [7] Lau, E.T., Sun, L. & Yang, Q. Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Appl. Sci.* **1**, 982 (2019). <https://doi.org/10.1007/s42452-019-0884-7>
  - [8] Umar, M. A. (2019). Student Academic Performance Prediction using Artificial Neural Networks: A Case Study. *International Journal of Computer Applications*, 178(48), 24–29. <https://doi.org/10.5120/ijca2019919387>
  - [9] Yu, J. (2021). Academic Performance Prediction Method of Online Education using Random Forest Algorithm and Artificial Intelligence Methods. *International Journal of Emerging Technologies in Learning (iJET)* <https://doi.org/10.3991/ijet.v16i05.20297>
  - [10] Hamadneh NN, Atawneh S, Khan WA, Almejalli KA, Alhomoud A. Using Artificial Intelligence to Predict Students' Academic Performance in Blended Learning. *Sustainability*. 2022; 14(18):11642. <https://doi.org/10.3390/su141811642>