# Introdução

A evasão de alunos de graduação é um desafio persistente e complexo que afeta as instituições de ensino superior no Brasil. Trata-se de um fenômeno multifacetado que envolve a interrupção prematura dos estudos por parte dos estudantes, acarretando consequências negativas para os indivíduos e para a sociedade. Compreender as causas e os fatores relacionados à evasão é fundamental para o desenvolvimento de estratégias efetivas de prevenção e intervenção, visando aumentar a retenção e promover o sucesso acadêmico dos alunos.

Diversos estudos realizados por pesquisadores brasileiros têm abordado a evasão de alunos de graduação, fornecendo contribuições significativas para a compreensão desse fenômeno em um contexto nacional. Segundo Alves e Lopes (2019), a evasão é um problema complexo influenciado por fatores individuais, socioeconômicos e institucionais. É necessário considerar aspectos como a falta de preparação acadêmica, dificuldades financeiras, falta de motivação e adaptação ao ambiente universitário.

A pesquisa de Bastos et al. (2021) destaca a importância do apoio socioemocional no combate à evasão. O suporte emocional, a orientação acadêmica e o fortalecimento da autoconfiança dos alunos são aspectos fundamentais para a promoção da permanência e conclusão dos cursos de graduação.

Além disso, Santos et al. (2018) enfatizam a relevância de políticas institucionais que visem à retenção dos alunos, como programas de monitoria, tutoria e nivelamento acadêmico. Essas estratégias contribuem para o fortalecimento do vínculo entre os estudantes e a instituição, bem como para o desenvolvimento de habilidades acadêmicas necessárias para o sucesso no ensino superior.

Nesse contexto, esta dissertação tem como objetivo analisar as principais causas e consequências da evasão de alunos de graduação no contexto brasileiro, bem como propor estratégias contemporâneas e baseadas em evidências para prevenção e intervenção. Por meio de uma revisão atualizada da literatura científica brasileira, serão explorados estudos recentes e pesquisas relevantes, fornecendo subsídios teóricos e práticos para o desenvolvimento de políticas e práticas institucionais efetivas de retenção e sucesso dos alunos de graduação no Brasil.

Espera-se que esta pesquisa contribua para o avanço do conhecimento sobre a evasão de alunos de graduação no contexto brasileiro, fornecendo subsídios fundamentais para a implementação de abordagens mais eficazes na prevenção e combate a esse fenômeno, promovendo, assim, a conclusão dos cursos de graduação e o desenvolvimento educacional e social dos estudantes brasileiros.

* 1. Justificativa

Sabemos que o problema é complexo e multifacetado e que afeta instituições de ensino superior em todo o mundo. Embora as taxas de evasão variem entre as instituições e regiões, elas podem ter um impacto significativo na qualidade da educação, bem como na sustentabilidade financeira das instituições de ensino. Além disso, a evasão de alunos também pode afetar negativamente o desenvolvimento pessoal e profissional dos estudantes, bem como sua capacidade de contribuir positivamente para a sociedade (Tinto, 1993; Pascarella e Terenzini, 1991).

No entanto, apesar da relevância do tema, ainda há uma falta de consenso sobre as principais causas e fatores que levam à evasão de alunos de graduação. A compreensão desses fatores é crucial para o desenvolvimento de políticas e práticas eficazes para prevenir e combater a evasão de alunos, bem como para promover uma educação de qualidade e inclusiva (Bean e Metzner, 1985; Cabrera et al., 1992).

Nesse sentido, esta pesquisa tem como objetivo analisar através de aprendizado de máquina as principais variáveis de fatores que levam à evasão de alunos de graduação, bem como identificar estratégias eficazes para prevenir e combater esse problema. Espera-se que os resultados desta pesquisa contribuam para o desenvolvimento de políticas e práticas educacionais mais eficazes, bem como para a compreensão mais aprofundada dos desafios enfrentados pelos estudantes de graduação e pelas instituições de ensino superior (Braxton et al., 2004; Tinto, 2006).

* 1. Causas da Evasão de Alunos

As principais causas da evasão de alunos de graduação no Brasil têm sido objeto de estudo e pesquisa por muitos anos, e há uma ampla literatura sobre o assunto. Algumas das principais causas de evasão de alunos de graduação no Brasil, com referências, são:

Dificuldades financeiras: O alto custo da educação superior no Brasil pode ser uma barreira significativa para muitos estudantes. De acordo com a Pesquisa Nacional por Amostra de Domicílios (PNAD), em 2019, mais da metade dos estudantes universitários brasileiros vinham de famílias com renda per capita de até 1,5 salário mínimo. A falta de recursos financeiros pode levar os estudantes a atrasar ou interromper seus estudos ou até mesmo abandonar completamente a universidade (Silva e Lins, 2016; Oliveira e Santos, 2018).

Desafios acadêmicos: As demandas acadêmicas da universidade, como a carga de trabalho e as expectativas do corpo docente, podem ser um desafio significativo para alguns estudantes. Estudantes que não estão preparados para a transição da escola para a universidade ou que enfrentam dificuldades acadêmicas podem ser mais propensos a abandonar os estudos (Fischer e Bidarra, 2019; Santos, 2019).

Falta de apoio social e psicológico: A transição para a universidade pode ser um momento estressante e isolado para muitos estudantes, especialmente aqueles que são os primeiros em sua família a frequentar a universidade. A falta de apoio social e psicológico pode contribuir para a evasão (Oliveira e Santos, 2018; Santos, 2019).

Escolha da carreira: Muitos estudantes ingressam na universidade sem uma compreensão clara de suas preferências e habilidades, o que pode levar a uma escolha equivocada de carreira ou curso. Os estudantes que percebem que escolheram a carreira ou curso errado podem ser mais propensos a abandonar a universidade (Silva e Lins, 2016; Oliveira e Santos, 2018).

Problemas pessoais: Problemas pessoais, como problemas de saúde, familiares ou emocionais, podem levar os estudantes a abandonar a universidade (Fischer e Bidarra, 2019; Oliveira e Santos, 2018).

* 1. Objetivos

Desenvolver e avaliar um modelo de machine learning para predição de evasão de alunos de graduação, com o intuito de apoiar a tomada de decisão institucional para prevenir a evasão e promover a retenção dos alunos.

"O uso de técnicas de aprendizado de máquina tem se mostrado promissor em diversas áreas, incluindo a educação, e pode auxiliar na identificação precoce de fatores que levam à evasão, além de fornecer insights valiosos para a tomada de decisão institucional" (Guan et al., 2018).

# REFERENCIAL TEÓRICO

## Evasão de Alunos

A evasão de alunos é um fenômeno complexo e de grande preocupação para o sistema educacional brasileiro. Diversos estudos têm investigado os fatores associados à evasão e buscado desenvolver modelos preditivos capazes de identificar os alunos em risco. Nesta seção, serão apresentados conceitos e teorias relevantes para compreender e abordar a evasão de alunos.

## Fatores de Risco para a Evasão

No contexto brasileiro, alguns estudos têm destacado a importância de fatores socioeconômicos como preditores de evasão de alunos. Santos e Souza (2017) encontraram uma associação significativa entre a evasão escolar e a baixa renda familiar em escolas públicas de ensino fundamental. Almeida et al. (2019) apontaram que a falta de suporte financeiro também é um fator relevante que influencia a permanência dos estudantes no ensino superior.

Além disso, fatores acadêmicos têm sido frequentemente investigados como determinantes da evasão no contexto brasileiro. Um estudo conduzido por Ferreira e Oliveira (2018) em uma instituição de ensino técnico mostrou que o desempenho acadêmico insatisfatório e a repetência estão relacionados à maior probabilidade de evasão dos alunos.

## Modelos de Previsão da Evasão

No Brasil, a utilização de técnicas de machine learning para prever a evasão de alunos tem ganhado destaque. De Souza et al. (2017) desenvolveram um modelo preditivo baseado em regressão logística para identificar estudantes com maior risco de evadir em uma universidade pública. Resultados promissores também foram obtidos por Silva et al. (2020) ao aplicarem algoritmos de aprendizado de máquina, como Random Forest e SVM, para prever a evasão em cursos de graduação.

## Estratégias de Intervenção e Prevenção

Além da previsão da evasão, é fundamental desenvolver estratégias eficazes de intervenção e prevenção para reduzir o abandono escolar no Brasil. Um estudo conduzido por Pereira et al. (2018) analisou a efetividade de programas de tutoria e mentorias na redução da evasão em escolas públicas de ensino médio. Eles identificaram que a adoção dessas estratégias contribuiu para a maior permanência dos alunos na escola.

Outra abordagem relevante foi apresentada por Costa e Silva (2019), que propuseram um programa de acompanhamento acadêmico individualizado em uma universidade privada, focado na identificação precoce de alunos em risco. A implementação desse programa resultou em uma redução significativa das taxas de evasão.

# Aprendizado de Maquina (Machine Learning)

O aprendizado de máquina é uma área da inteligência artificial que se dedica a desenvolver algoritmos e modelos capazes de aprender a partir de dados e tomar decisões ou realizar tarefas de forma autônoma. Essa área tem sido cada vez mais utilizada em uma variedade de aplicações, como reconhecimento de voz, visão computacional, processamento de linguagem natural, análise de dados, entre outras.

O aprendizado de máquina é um subcampo da inteligência artificial que permite que as máquinas aprendam a partir de dados e experiências passadas para realizar tarefas específicas. Segundo Tom Mitchell, professor de ciência da computação da Universidade Carnegie Mellon, "o aprendizado de máquina é o campo de estudo que dá aos computadores a capacidade de aprender sem serem explicitamente programados".

Para implementar o aprendizado de máquina, são utilizadas diversas bibliotecas e frameworks, como TensorFlow, Scikit-learn, PyTorch, Keras, entre outras. Essas bibliotecas fornecem ferramentas para criar, treinar e avaliar modelos de aprendizado de máquina em diferentes linguagens de programação, como Python e R.

3.1 Inteligencia Artificial

A inteligência artificial (IA) é uma área da ciência da computação que busca desenvolver sistemas capazes de realizar tarefas que normalmente exigem inteligência humana, como reconhecimento de fala, visão computacional, processamento de linguagem natural e tomada de decisões. O desenvolvimento da IA tem sido impulsionado pelo aumento na disponibilidade de dados, avanços em algoritmos de aprendizado de máquina e aumento na capacidade de processamento de computadores.

Existem diversas técnicas de IA, incluindo o aprendizado de máquina, a lógica fuzzy, as redes neurais artificiais e os sistemas especialistas. O aprendizado de máquina é uma técnica que permite que os sistemas de IA aprendam a partir de dados e melhorem sua performance ao longo do tempo. Existem três tipos principais de aprendizado de máquina: supervisionado, não supervisionado e por reforço (Goodfellow et al., 2016).

A IA tem sido utilizada em diversas áreas, incluindo medicina, finanças, transporte, educação e segurança. Na área da medicina, por exemplo, a IA tem sido utilizada para diagnóstico médico, identificação de riscos de doenças e desenvolvimento de tratamentos personalizados (Topol, 2019). Na área da educação, a IA tem sido utilizada para personalizar o ensino, identificar alunos em risco de evasão e desenvolver sistemas de tutoria inteligente (Baker \& Inventado, 2014).

No entanto, a utilização da IA também apresenta desafios e preocupações, incluindo questões de privacidade, segurança, vieses algorítmicos e impacto na força de trabalho (Brynjolfsson & Mitchell, 2017). A ética da IA também tem sido objeto de discussão, com a necessidade de desenvolver sistemas de IA responsáveis, transparentes e justos (Floridi et al., 2018).

# Técnicas de Machine Learning

De acordo com Alpaydin (2010), as técnicas de Machine Learning podem ser divididas em três grandes categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. O aprendizado supervisionado envolve o treinamento de um modelo com um conjunto de dados rotulados, onde cada exemplo de entrada tem uma saída correspondente associada. O objetivo do algoritmo é aprender a mapear os inputs para os outputs corretos. O aprendizado não supervisionado, por sua vez, envolve o treinamento de um modelo com um conjunto de dados não rotulados, onde não há saídas conhecidas associadas aos exemplos de entrada. O objetivo do algoritmo é encontrar estruturas ou padrões nos dados que possam ser úteis. Finalmente, o aprendizado por reforço é uma abordagem em que o algoritmo é treinado através de interações com um ambiente dinâmico, onde o algoritmo deve aprender a escolher ações que maximizem uma recompensa numérica.

# Metodologia

5.1 Técnicas Utilizadas

O StackingClassifier é uma técnica de aprendizado de máquina que combina múltiplos modelos para melhorar a precisão das previsões. Ele foi proposto pela primeira vez em um artigo de 1996 intitulado "Stacked Generalization" por David Wolpert.

Segundo Wolpert, a ideia por trás do StackingClassifier é usar vários modelos de aprendizado de máquina para gerar previsões, que são então combinadas em um modelo de nível superior que usa as previsões como entrada para gerar uma previsão final mais precisa.

O processo de StackingClassifier envolve duas etapas. Na primeira etapa, vários modelos de aprendizado de máquina são treinados em um conjunto de dados. Cada modelo produz uma previsão para o conjunto de dados. Na segunda etapa, as previsões dos modelos de aprendizado de máquina são usadas como entrada para um modelo de nível superior, que é chamado de meta-modelo ou Stacker. O meta-modelo é treinado com as previsões de saída dos modelos de nível inferior. O objetivo do meta-modelo é combinar as previsões dos modelos de nível inferior e gerar uma previsão final mais precisa.

Existem várias variações do StackingClassifier, incluindo o Stacking com k-fold cross-validation e o Stacking em cascata. Em uma abordagem de Stacking com k-fold cross-validation, o conjunto de dados é dividido em k partições. O processo de treinamento e teste é repetido k vezes, com cada partição sendo usada uma vez para teste e as outras k-1 partições sendo usadas para treinamento. Em uma abordagem de Stacking em cascata, as previsões do primeiro nível de modelos são usadas como entrada para um segundo conjunto de modelos que geram previsões adicionais que são então usadas para treinar o meta-modelo.

O StackingClassifier é uma técnica poderosa que pode melhorar significativamente a precisão das previsões em problemas de aprendizado de máquina. No entanto, é importante ter cuidado ao usá-lo, pois a combinação incorreta de modelos de aprendizado de máquina pode levar a resultados imprecisos. É importante experimentar diferentes combinações de modelos de aprendizado de máquina e analisar os resultados para determinar qual combinação funciona melhor para um determinado problema.

5.1.1 Parâmetros de Ajuste das Técnicas

5.2 Compreensão dos dados

Esta base de dados fornece uma visão abrangente dos alunos matriculados em vários cursos de graduação oferecidos em uma instituição de ensino superior e inclui dados demográficos, fatores socioeconômicos e informações de desempenho acadêmico que podem ser usados para analisar os possíveis preditores de evasão e sucesso acadêmico. Neste mesmo conjunto de dados contém vários bancos disjuntos que consistem em informações relevantes disponíveis no momento da inscrição, como modo de inscrição, estado civil, curso escolhido e muito mais. Adicionalmente, estes dados podem ser utilizados para estimar o desempenho global dos alunos no final de cada semestre, através da avaliação das unidades curriculares creditadas/inscritas/avaliadas/aprovadas, bem como das respetivas classificações. Por fim, temos a taxa de desemprego, a taxa de inflação e o PIB da região, que podem nos ajudar a entender melhor como os fatores econômicos influenciam as taxas de evasão estudantil ou os resultados de sucesso acadêmico. Esta poderosa ferramenta de análise fornecerá informações valiosas sobre o que motiva os alunos a permanecer na escola ou abandonar seus estudos para uma ampla gama de disciplinas, como agronomia, design, educação, enfermagem, jornalismo, gerenciamento, serviço social ou tecnologias

Usando esse conjunto de dados, os pesquisadores podem investigar duas questões principais:

Quais fatores preditivos específicos estão ligados ao abandono ou conclusão do aluno?

como diferentes recursos interagem uns com os outros?

Por exemplo, os pesquisadores podem explorar se existem características demográficas (por exemplo, gênero, idade na matrícula, etc.) resultados educacionais. Ao responder a essas perguntas, é gerado um insight de pesquisa que pode fornecer informações críticas para os administradores na formulação de estratégias que promovam a conclusão bem-sucedida do diploma entre alunos de diversas origens em suas instituições.

Para usar este conjunto de dados de forma eficaz, é importante que os cientistas se familiarizem com todas as variáveis fornecidas no conjunto de dados, incluindo variáveis categóricas (qualitativas), como gênero ou modo de aplicação; variáveis numéricas como número de unidades curriculares no início dos semestres ou idade de ingresso; variáveis de tipo de medição de dados ordinais, como estado civil; tendências estudadas ao longo do tempo, como taxa de inflação ou PIB; variáveis de medição de frequência como percentual de bolsistas; etc. Além disso, os cientistas devem certificar-se de que estão cientes de todos os possíveis vieses incluídos nos dados antes de executar a análise - por exemplo, entender se uma população está sub-representada em comparação com outra - já que esse fenômeno pode levar a resultados inesperados se não for levado em consideração ao conduzir pesquisas realizadas usando este conjunto de dados. Finalmente, seria importante para os profissionais perceberem que este conjunto de dados Kaggle atual contém apenas informações de um semestre em cada admissão, enquanto estudos adicionais conduzidos por um período de tempo mais longo podem fornecer resultados mais precisos relacionados à área de tópico selecionada devido coeficientes de realização de retenção de deterioração adicionais obtidos a partir desses experimentos gradualmente precisos que se desenrolam em diferentes temporadas de admissão ao longo do ano.

Estado civil: O estado civil do aluno. (Categórico)

Modo de aplicação: O método de aplicação utilizado pelo aluno. (Categórico)

Ordem de inscrição: a ordem em que o aluno se inscreveu. (Numérico)

Curso: O curso feito pelo aluno. (Categórico)

Frequência diurna/noturna: Se o aluno frequenta as aulas durante o dia ou à noite. (Categórico)

Qualificação anterior: Qualificação obtida pelo aluno antes de ingressar no ensino superior.

(Categórico)

Nacionalidade: A nacionalidade do aluno. (Categórico)

Qualificação da mãe: A qualificação da mãe do aluno. (Categórico)

Qualificação do pai: A qualificação do pai do aluno. (Categórico)

Ocupação da mãe: A ocupação da mãe do aluno. (Categórico)

Ocupação do pai: A ocupação do pai do aluno. (Categórico)

Deslocado: se o aluno é uma pessoa deslocada. (Categórico)

Necessidades educacionais especiais: se o aluno tem alguma necessidade educacional especial. (Categórico)

Devedor: Se o aluno é um devedor. (Categórico)

Propinas em dia: Se as propinas do aluno estão em dia. (Categórico)

Sexo: O sexo do aluno. (Categórico)

Bolsista: Se o aluno é bolsista. (Categórico)

Idade na matrícula: a idade do aluno no momento da matrícula. (Numérico)

Internacional: Se o aluno é um estudante internacional. (Categórico)

Unidades curriculares 1º semestre (creditadas): Número de unidades curriculares creditadas pelo aluno no primeiro semestre. (Numérico)

Unidades curriculares 1º semestre (inscritos): Número de unidades curriculares inscritas pelo aluno no primeiro semestre. (Numérico)

Unidades curriculares 1º semestre (avaliações): Número de unidades curriculares avaliadas pelo aluno no primeiro semestre. (Numérico)

Unidades curriculares 1º semestre (aprovadas): Número de unidades curriculares aprovadas pelo aluno no primeiro semestre. (Numérico)

# Modelo Proposto

A metodologia proposta neste estudo pretende elaborar um planejamento para

realizar a probabilidade de evasão para posterior ação dos orgãos competentes em relação aos

estudantes de graduação. Deste modo, foi produzido um modelo de aprendizados de máquina no qual foi aplicado dados de uma universidade fictícia. Os dados aplicados

foram utilizados como parâmetro para o modelo, que permite simular o comportamento

dos estudantes.

A metodologia aplicada neste trabalho é baseada na metodologia para modelagem

XXXXX. O processo é composto por cinco etapas que vão desde o entendimento do problema

até a análise dos resultados, conforme representado na Figura 5. Essas etapas permitem

guiar a criação de um modelo que irá gerar métricas para realizar

análises de sobrevivência de estudantes de graduação.

passo 1 - Entendimento do problema

passo 2 - Implementação do Modelo

passo 3 - Definição dos Parâmetros de Entrada

passo 4 - Execução do modelo e extração das métricas

passo 5 - Analise dos Resultados

SANTOS, Juliana Ferreira dos. Análise do impacto da evasão e retenção no ensino superior utilizando cadeias de Markov absorventes. 2022. 77 f. Dissertação (Programa de Pós-Graduação em Informática Aplicada) - Universidade Federal Rural de Pernambuco, Recife.

# Resultados

# Conclusão

# Propostas para trabalhos futuros

# Referências bibliográficas

Almeida, M. A., Santos, L. P., & Oliveira, R. S. (2019). Fatores determinantes da evasão no ensino superior: um estudo em uma instituição de ensino privada. Revista Brasileira de Gestão e Desenvolvimento Regional, 15(2), 297-317.

Costa, D. R., & Silva, E. C. (2019). Programa de Acompanhamento Acadêmico: estratégia para redução da evasão em uma universidade privada. Revista Eletrônica de Educação, 13(3), 81-101.

Ferreira, M. R., & Oliveira, R. S. (2018). Determinantes da evasão escolar em instituições de ensino técnico: um estudo de caso. Revista Brasileira de Pesquisa em Educação em Ciências, 18(3), 1015-1033.

Pereira, A. P., Santos, D. L., & Lima, J. R. (2018). Programas de tutoria e mentorias no combate à evasão escolar: uma revisão sistemática da literatura. Educação e Pesquisa, 44, e192638.

Silva, G. A., Oliveira, M. V., Nogueira, L. L., & Santos, R. M. (2020). Previsão da evasão em cursos de graduação utilizando aprendizado de máquina. Revista Brasileira de Informática na Educação, 28(3), 383-397.

Santos, R. A., & Souza, A. R. (2017). Evasão escolar: análise das causas e consequências em escolas públicas de ensino fundamental. Revista Ibero-Americana de Estudos em Educação, 12(1), 54-71.

De Souza, L. M., Duarte, R. G., & Rodrigues, M. L. (2017). Um modelo preditivo para identificação de estudantes em risco de evasão em uma universidade pública. Revista Brasileira de Computação Aplicada, 9(2), 88-96.