

## Ferramenta para Predição do Desempenho Acadêmico no Ensino Superior

Léo Manoel Lopes da Silva Garcia<sup>1</sup>, Daiany Francisca Lara<sup>1</sup>, Raquel Salcedo Gomes<sup>1</sup>, Sílvio Cezar Cazella<sup>2</sup>

<sup>1</sup>Faculdade de Ciências Exatas – Curso de Ciências da Computação – Universidade do Estado de Mato Grosso (UNEMAT) – Barra do Bugres – MT – Brasil

<sup>2</sup>Programa de Pós - Graduação em Informática na Educação – Universidade Federal do Rio Grande do Sul (UFRGS) – Porto Alegre – RS – Brasil

leoneto@unemat.br, dflara@unemat.br, raquel.salcedo@ufrgs.br, silvio.cazella@gmail.com

**Abstract.** *This paper presents the development of a tool for predicting academic performance in higher education, specifically in subjects to be studied. The predictive models are designed based on the course taken by the students. To assess the effectiveness of the developed tool, data from 15 disciplines of a Computer Science course in the academic period of 2022/1 were computed. The results demonstrated a high success rate, with an average of 86% of correct predictions for general performance, and 90% for specific prediction of failure cases. It is noteworthy that these averages of accuracy were obtained even in disciplines of the first academic phases.*

**Resumo.** *Este trabalho apresenta o desenvolvimento de uma ferramenta para a predição de desempenho acadêmico na educação superior, especificamente em disciplinas a serem cursadas. Os modelos preditivos são concebidos a partir do percurso curricular realizado pelos alunos. Para aferir a eficácia da ferramenta desenvolvida, foram computados dados de 15 disciplinas de um curso de Ciências da Computação no período letivo de 2022/1. Os resultados demonstraram uma alta taxa de acerto, com uma média de 86% de predições corretas para o desempenho em geral, e 90% para predição específica de casos de reprovação. Destaca-se que essas médias de precisão foram obtidas mesmo em disciplinas de fases letivas iniciais (a partir da segunda fase).*

### 1.Introdução

O desempenho dos alunos é um dos componentes centrais para avaliar qualquer sistema educacional, sendo crucial para lidar com questões do processo de aprendizagem [ALHAZMI e SHENEAMER, 2023]. Embora, em certa medida, essa seja uma constatação pragmática, este é um fenômeno complexo e dependente de várias condições internas e externas ao processo educacional [CZIBULA et al. 2019]. Destarte, pesa aos professores e gestores educacionais a responsabilidade de prescreverem estratégias de ensino que se sobreponham às dificuldades e favoreçam a aprendizagem [GARCIA et al., 2022]. Nesse cenário, dentre um aparato de recursos de gestão e metodologias de ensino à disposição, o uso de tecnologias da informação (TI) tem ganhado destaque não só como infraestrutura, mas também como apoio à tomada de decisão.

Cabe destacar a onipresença da TI em todas as etapas do sistema de ensino. Semelhante a muitas outras áreas da sociedade e da atividade humana, a educação foi significativamente impactada pelos avanços tecnológicos [RODRÍGUEZ-HERNÁNDEZ et al., 2021]. Um ponto em comum em todos os recursos de TI envolvidos é a miríade de registros e geração de dados. Perante a dificuldade de se analisar manualmente essa grande quantidade de dados gerados, tecnologias têm sido propostas e utilizadas para analisar automaticamente esses dados, de forma a permitir que estudantes, professores e administradores obtenham novas perspectivas sobre o comportamento dos atores do sistema educacional [SANTOS et al., 2021]. Emerge, dentre essas tecnologias, a Mineração de Dados Educacionais (MDE) como um dos recursos computacionais mais promissores, sendo capaz de coletar, processar e armazenar uma grande quantidade de informações no/do contexto educacional que, ao serem analisadas, podem ser valiosas para as instituições de ensino e seus agentes [RABELO et al., 2017]. A MDE implementa técnicas do campo da Inteligência Artificial (IA) que, utilizando modelos matemáticos e estatísticos, podem fornecer insights significativos sobre o processo de ensino-aprendizagem.

Há diferentes abordagens para o uso de MDE, todas relativas aos objetivos a serem alcançados. Todavia, é possível destacar a predominância de aplicações que buscam realizar a previsão do desempenho dos alunos [CZIBULA et al. 2019]. A grande demanda por essa abordagem é fomentada pela perspectiva de que, ao prever uma futura ocorrência de um caso de insucesso, talvez seja possível realizar intervenções para reverter essa tendência e evitar o fracasso e a evasão. Essa percepção é corroborada pelos trabalhos de [LI e LIU, 2021] [PALLATHADKA et al., 2023] [RODRÍGUEZ-HERNÁNDEZ et al., 2021] [PÁEZ, 2022] [CZIBULA et al. 2019], que descrevem em seus estudos que a previsão do desempenho pode subsidiar estratégias educacionais capazes reverter os casos eminentes de reprovação ou abandono, reduzindo as taxas de insucesso da instituição.

Nesse contexto, esse estudo busca contribuir com esse arcabouço teórico apresentando o desenvolvimento de uma ferramenta para previsão do desempenho acadêmico no ensino superior suportada por MDE. O recurso apresenta a previsão de aprovação ou reprovação dos alunos matriculados em uma determinada disciplina, sob uma interface simples, intuitiva e sem exigência de conhecimentos técnicos por parte dos usuários finais, alinhada com a perspectiva de [PHAN et al., 2023] que discorrem sobre a importância de tornar interpretáveis os resultados da MDE de forma que os humanos possam compreender a decisão recomendada pelo algoritmo. Neste aspecto reside uma importante motivação deste estudo, no que tange à finalidade de transformar os modelos preditores em algo palpável e alcançável aos principais interessados, como professores e gestores educacionais, de forma transparente, sem exigir um conhecimento acerca dos processos técnicos de descoberta do conhecimento. Tendo como destaque neste estudo a mitigação do insucesso dos alunos, caracterizado aqui como reprovação ao cursar as disciplinas do curso, visto que a reprovação em disciplinas tem significativa influência na definição dos resultados das trajetórias acadêmicas dos alunos (diplomação, diplomação tardia e evasão) [GARCIA et al., 2021]. Uma vez identificada precocemente a tendência de um aluno reprovar em uma determinada disciplina, o professor pode, estrategicamente, realizar intervenções pedagógicas para evitar que isso aconteça e, ao final, conseguir que o referido aluno tenha êxito na disciplina. O modelo predictor utilizado pela ferramenta é baseado no percurso curricular realizado pelo aluno antes de cursar uma determinada disciplina.

O desenvolvimento desse artigo segue a seguinte organização: na seção 2 são apresentados os trabalhos correlatos a este estudo, em seguida é apresentada a proposta e metodologia de concepção da ferramenta. Já na seção 4 é realizada a discussão acerca dos resultados obtidos e as contribuições do estudo para esse campo de pesquisa. Por fim, são tecidas as considerações finais sobre o trabalho realizado.

## 2. Trabalhos Relacionados

Conforme relatado previamente, é grande a popularidade do uso de MDE para a previsão do desempenho acadêmico. Todavia, há distinção nas possibilidades de abordagens a serem realizadas nesse campo. Em uma perspectiva, a previsão de desempenho remete-se à identificação dos casos de sucesso ou insucesso acadêmico. Substancialmente com foco na predição de casos de insucesso, há expectativa de que os agentes responsáveis possam, pontualmente, reverter esse resultado de alguma maneira. Especificando um pouco mais, esse insucesso pode referir-se à reprovação, baixos níveis de desempenho, evasão, diplomação tardia, insatisfação com a formação e, até mesmo, baixa empregabilidade.

Em [SANTOS et al., 2021], são aplicadas técnicas de classificação para a previsão de casos de evasão ou diplomação. Os autores buscam identificar alunos com comportamento semelhante a alunos que evadiram no passado e optam por uma abordagem de se utilizar somente dados acadêmicos durante o curso. Uma particularidade desse estudo é a geração de vários modelos preditores, gerando progressivamente um modelo para cada fase letiva, que deverão ser aplicados dependendo do período em que o aluno se encontra no curso. Com o mesmo objetivo de prever casos de evasão ou diplomação, [PHAN et al. 2023] utilizam uma abordagem híbrida com a utilização de clusterização e mineração de textos. São utilizados dados estruturados referentes à formação sociodemográfica dos alunos, informações sobre matrículas em programas e cursos, notas e situação de estudo, além de dados não estruturados na forma de feedback textual sobre cada uma de suas disciplinas matriculadas.

Contudo, são mais aderentes a este trabalho os estudos que abordaram o desempenho obtido em disciplinas. Como é o caso de [CZIBULA et al., 2019], em que é apresentado um modelo de classificação para prever o resultado final de um aluno em uma disciplina usando regras de associação relacional. São utilizados como atributos os resultados obtidos por alunos em semestres anteriores. Já em [RODRÍGUEZ-HERNÁNDEZ et al., 2021], é explorada a classificação de alunos nos possíveis níveis de rendimento que eles imprimirão no curso, sendo de alto nível ou baixo nível. A metodologia é baseada no uso de redes neurais artificiais, tendo como atributos previsores informações socioeconômicas, características do ensino médio e seu estado de trabalho, além de informações de gênero e idade. Com objetivos semelhantes [ALHAZMI e SHENEAMER, 2023], apresentam técnicas de agrupamento e classificação para identificar o impacto do desempenho dos alunos no estágio inicial no *Grade Point Average* – GPA do aluno ao final do curso. Em resumo, o objetivo do estudo é utilizar os dados dos registros e características dos alunos no estágio inicial para prever seu desempenho no estágio final (GPA acumulativo final).

No estudo de [MAIA et al., 2010], é simulada a previsão das notas a serem obtidas de alunos matriculados em um curso de graduação a partir de seus resultados passados nas disciplinas do curso. O histórico de desempenho dos alunos é disposto em

grafos, de forma que modelos de crescimento baseados em redes complexas foram capazes de prever a evolução do grafo e, assim, prever os próximos resultados. Em uma abordagem para previsão do desempenho em uma disciplina específica de Cálculo Diferencial e Integral, [MORAIS et al., 2022], utilizam algoritmos genéticos para a previsão do desempenho dos alunos matriculados na disciplina a partir de dados sobre o desempenho escolar obtidos anteriormente pelos estudantes na educação básica, além do perfil socioeconômico desses estudantes. Em [PAÉZ, 2022], o autor propõe a utilização de modelos preditivos progressivos baseados em algoritmos de classificação para prever o resultado final (Aprovação ou Reprovação) dos alunos em uma disciplina. Nessa abordagem, são utilizadas as notas obtidas em atividades durante a disciplina. A cada atividade, o modelo é avaliado, e observou-se que a partir da terceira atividade já é possível obter uma acurácia de 70% na previsão.

Aproximando-se dessa proposta, apresentando uma ferramenta para previsão do desempenho acadêmico com uma interface simplificada para usuário final, [PEREIRA et al., 2019] desenvolvem um ambiente que, a partir de dados de interação de uma turma na modalidade EaD, permite visualizar a previsão do desempenho nas categorias Alto Nível, Nível Médio e Baixo Nível. A ferramenta permite a inserção de novas instâncias por meio de uma interface gráfica simples e, com os dados inseridos, imediatamente são exibidas as previsões de desempenho do aluno em questão, indicando a categoria e a probabilidade sugerida pelos três algoritmos implementados pela ferramenta: Naive Bayes, Árvores de Decisão e SVM (*Support Vector Machine*). Nessa linha [STOLL et al., 2019], demonstram a concepção de uma ferramenta com diversas funcionalidades para análise de dados acadêmicos, dentre elas a predição do desempenho, distinguido em casos de sucesso (alto rendimento e diplomação) e alunos em risco acadêmico (baixo rendimento e evasão).

### **3. Proposta e Metodologia**

Este estudo é realizado no âmbito do desenvolvimento de um sistema integrado para análise de dados acadêmicos, sendo concebida aqui, especificamente, uma interface para previsão do desempenho de alunos matriculados em uma disciplina. Em resumo, findadas as matrículas dos estudantes nas disciplinas do curso, os professores poderão verificar, já no início do período letivo, as probabilidades de cada aluno ser aprovado ou reprovado em sua disciplina. Esta previsão é realizada por meio da utilização de modelos preditivos, especificamente classificadores, baseados em percurso curricular realizado previamente pelo aluno ao cursar uma determinada disciplina.

#### **3.1. Descrição do modelo preditivo**

É importante destacar que as avaliações dos modelos preditivos que serão utilizados nesta ferramenta já foram realizadas previamente em [GARCIA et al., 2022], onde diferentes classificadores foram avaliados, bem como diferentes configurações dos atributos preditores e técnicas de balanceamento. Deste estudo, resultaram como destaque os algoritmos J48, Random Forest e IBk (nas configurações de 1 e 8 vizinhos) com os melhores resultados para os objetivos propostos. Quanto à configuração da base de treinamento, foram superiores os resultados alcançados ao se utilizar somente os dados acadêmicos, excluindo-se os atributos de gênero e faixa etária. Sendo assim, para a concepção da ferramenta foram utilizados os modelos preditivos baseados nesses algoritmos e com a referida configuração da base de treinamento.

Contudo, convém discorrer aqui acerca da concepção dos atributos preditores, visto que esta abordagem utilizando o percurso curricular configura uma importante caracterização desse estudo. Esse percurso pode ser descrito como a interação dos alunos com os componentes curriculares e nos resultados destas interações, e é assim descrito em [GARCIA et al., 2022, p. 1080]:

[...] quando o aluno se matricula em uma disciplina X na sua 3ª fase letiva em um curso, por exemplo, seu percurso curricular é retratado por todas as disciplinas cursadas em sua 1ª e 2ª fase letiva (período anterior à matrícula na disciplina) e nos respectivos resultados, definidos em Reprovação ou Aprovação. Desse modo, deseja-se obter, por meio de um modelo preditivo, a previsão deste aluno ser aprovado ou reprovado na disciplina X.

Para a realização de testes na ferramenta desenvolvida, foram utilizados dados de um curso de Ciências da Computação da Universidade do Estado de Mato Grosso (UNEMAT) do campus de Barra do Bugres<sup>1</sup>. Desse modo, cabe esclarecer que as decisões acerca de quais algoritmos utilizar e qual tratamento de dados aplicar são norteadas pelos testes realizados nesse curso. No entanto, constata-se que é necessária a realização da mesma bateria de testes para cada curso no momento de estender a aplicação para toda instituição. Assim, diferentes algoritmos podem ser recomendados para cada curso.

### 3.2. Desenvolvimento da Ferramenta

Esta ferramenta foi desenvolvida na linguagem Java, utilizando-se a IDE Eclipse e banco de dados MySQL. Embora a interface de acesso dos professores dar-se-á por meio de um sistema web, toda classificação é realizada em uma aplicação local para evitar sobrecarga do servidor. Após ser realizada a previsão, os dados são inseridos no banco de dados compartilhado e poderá ser visualizado pelo sistema web. Os algoritmos de classificação foram utilizados a partir da biblioteca Java do Weka 3.8.5. Uma vez que já foram escolhidos quais algoritmos a serem utilizados, tal como a configuração da base de treinamento. Foram gerados os modelos previsores, já treinados, dos algoritmos J48, Random Forest e IBk com 1 e 8 vizinhos, configurando assim 4 modelos previsores a serem utilizados pela ferramenta.

Quanto à implementação da principal funcionalidade da ferramenta, que é a classificação de novas instâncias, cabe destacar que essa pesquisa não tem acesso direto ao banco de dados do sistema acadêmico da instituição. Desse modo, a inserção de novos alunos para os quais se deseja prever o desempenho é realizada via um arquivo csv (*comma-separated-values*), incluindo somente o nome da disciplina, nome dos alunos matriculados nela no período letivo e suas respectivas matrículas. Estas informações estão contidas na lista de matriculados que podem ser obtidas no sistema acadêmico. Com o upload da lista de matriculados de uma disciplina, a aplicação gera o vetor de características preditivas de cada aluno. Para tanto, é preciso percorrer o banco de dados do histórico dos alunos (replicado do sistema acadêmico) construindo o percurso realizado por eles antes do momento de cursar a disciplina para a qual se deseja fazer a previsão. Para isso, é necessário computar em qual fase letiva se encontra cada aluno. Assim, quando o aluno que for cursar uma disciplina no período letivo de 2022/2, por exemplo, e ingressou no curso em 2021/1, estará em sua 4ª fase letiva e seu

---

<sup>1</sup> Pesquisa aprovada no comitê de ética na Plataforma Brasil, sob número CAAE 50431321.0.0000.5347

percurso será composto por todas as disciplinas cursadas em suas 1ª, 2ª e 3ª fases letivas e, obviamente, dos resultados obtidos. Para toda disciplina já cursada é coletado o resultado obtido, podendo ser Reprovação ou Aprovação e, quando ele não cursou uma determinada disciplina é atribuído o valor “?”.

Após construído o vetor de característica, cada um deles é submetido aos 4 modelos preditores utilizados. Os resultados não apresentam simplesmente uma indicação definitiva do resultado a ser obtido pelo aluno. Mas, por meio do método *distributionForInstance* da API Weka, para cada classificador, é obtida a probabilidade de o aluno resultar em Aprovação ou Reprovação na disciplina. Por fim, é calculada a média da probabilidade que cada modelo preditivo indicou para cada classe. E, a partir dessa média, o resultado da previsão é apresentado. É dada a possibilidade de visualizar a lista de alunos matriculados somente com a indicação da previsão de cada um deles em ser aprovado ou reprovado, conforme demonstrado na Figura 1. E, também, o detalhamento dos resultados de cada modelo, como exposto na Figura 2.

The screenshot shows the 'Previsão de Desempenho' interface for the discipline 'Sistemas Embarcados 2022/1'. It features a table with two columns: 'Nome' and 'PREVISÃO'. The names are redacted with red brush strokes. The predicted results are listed as follows:

Nome	PREVISÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	APROVAÇÃO
[Redacted]	APROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	APROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO
[Redacted]	REPROVAÇÃO

Figura 1. Interface de previsão de desempenho

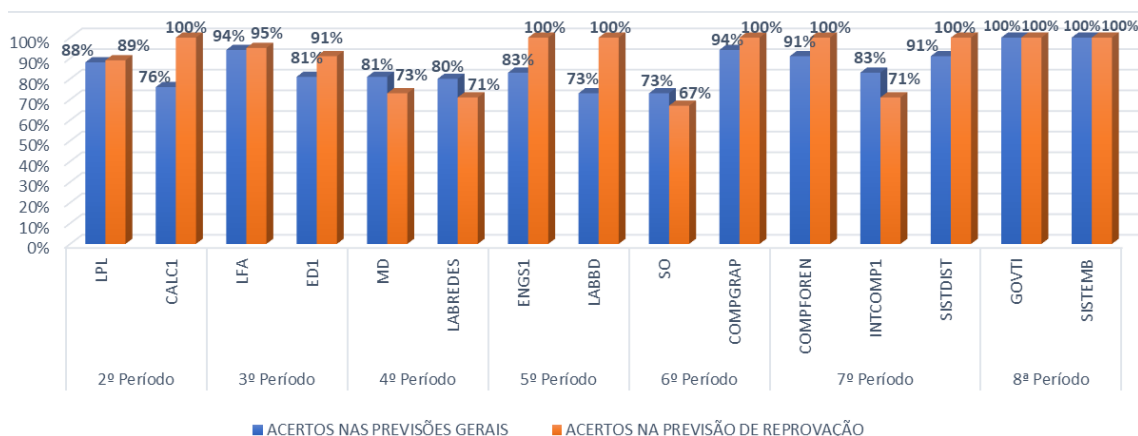
The screenshot shows the 'Detalhamento' view of the 'Previsão de Desempenho' interface for the discipline 'Sistemas Embarcados 2022/1'. It features a table with 12 columns: 'Nome', 'J48 APROVADO', 'J48 REPROVAD', 'Random Forest APROVADO', 'Random Forest REPROVAD', 'IBk1 APROVADO', 'IBk1 REPROVAD', 'IBk8 APROVADO', 'IBk8 REPROVAD', 'MÉDIA APROVADO', 'MÉDIA REPROVAD', and 'PREVISÃO'. The names are redacted with red brush strokes. The data is as follows:

Nome	J48 APROVADO	J48 REPROVAD	Random Forest APROVADO	Random Forest REPROVAD	IBk1 APROVADO	IBk1 REPROVAD	IBk8 APROVADO	IBk8 REPROVAD	MÉDIA APROVADO	MÉDIA REPROVAD	PREVISÃO
[Redacted]	0,75	0,25	0,06	0,94	0	1	0,22	0,78	25,78%	74,22%	REPROVAÇÃO
[Redacted]	0,15	0,85	0,22	0,78	0	1	0	1	9,34%	90,66%	REPROVAÇÃO
[Redacted]	0,83	0,17	0,07	0,93	0	1	0	1	22,67%	77,33%	REPROVAÇÃO
[Redacted]	0,88	0,12	0,53	0,47	0,62	0,38	0,62	0,38	66,55%	33,45%	APROVAÇÃO
[Redacted]	0,61	0,39	0,62	0,38	1	0	0,5	0,5	68,07%	31,93%	APROVAÇÃO

Figura 2. Interface de detalhamento da previsão

Para avaliar a eficácia da ferramenta, foram utilizados dados de 15 disciplinas do período letivo de 2022/1 do curso de Computação, contemplando disciplinas previstas para serem cursadas da 2ª à 8ª fase letiva, de acordo com o currículo do curso. As listas de matrículas foram carregadas na ferramenta, os vetores de características foram gerados e foi realizada a predição de desempenho de todos os matriculados. Por fim, as previsões foram comparadas com os resultados reais obtidos pelos alunos. Cabe destacar que nenhum dos alunos fez parte da base de treinamento. A base de treinamento foi composta por ingressantes dos períodos letivos de 2013/2 a 2017/1, enquanto para este teste foram contemplados os ingressantes de 2017/2 em diante.

Durante o treinamento dos modelos preditivos, o melhor desempenho da média dos algoritmos escolhidos foi de 72,60% de acurácia e 80% de recall para a classe Reprovados. Com os testes com os dados reais do período letivo de 2022/1, os resultados foram ainda melhores, apresentando a média de 86% de acurácia e 90% de recall para a previsão de reprovados. O Gráfico 1 apresenta os resultados obtidos para cada disciplina, em que as barras azuis representam o percentual de acertos gerais, enquanto as barras laranjadas representam o percentual de acertos somente para os casos de reprovações. Ou seja, para todos os alunos que foram reprovados em uma disciplina, é indicado o percentual de previsões que indicaram corretamente que o aluno iria reprovar.



**Gráfico 1. Acertos na previsão de desempenho**

Ainda que [PAÉZ, 2022] e [SANTOS et al., 2021], ao utilizarem modelos progressivos, relatem que, conforme acumulam-se as informações ao longo do tempo ocorre a melhora na previsão, observa-se, no Gráfico 1, que o percentual de acertos é semelhante à média geral já no segundo período. Todavia, no 8º período, o percentual de acerto é de 100%. Como já citado previamente, prever alunos em risco de reprovação é o objetivo principal desse tipo de empreendimento. Pois, de posse dessa informação, as intervenções pedagógicas poderão reverter essa tendência de reprovação. Nessa perspectiva, os resultados são altamente satisfatórios, ao imprimir uma média de 90% de acerto nessa previsão, apresentando, inclusive, 100% de acertos em 8 das 15 disciplinas.

#### 4. Discussão dos Resultados

Norteados pelos avanços alcançados em trabalhos correlatos, destacam-se duas principais contribuições do presente estudo, primeiramente acerca dos atributos preditores utilizados para compor o modelo preditivo e, seguidamente, pelas características da

ferramenta desenvolvida. A abordagem alternativa para geração de atributos previsores, baseando-se no percurso curricular, é motivada pelo entendimento de que a motivação dos alunos, bem como suas características cognitivas relacionadas à aprendizagem não são inatas e nem imutáveis [APA, 2015].

Nessa perspectiva, embora os atributos demográficos e socioeconômicos sejam reconhecidamente poderosos para predição, como demonstrado nos trabalhos correlatos, esta pesquisa propende a considerar experiência do aluno durante o curso, correlacionando os resultados obtidos nas disciplinas. Ainda que em estudos como [SANTOS et al., 2021], [CZIBULA et al., 2019], [ALHAZMI e SHENEAMER, 2023], [MAIA et al., 2010] e [PAÉZ, 2022] sejam considerados os resultados obtidos durante o curso ou disciplina como notas ou média cumulativa (GPA), este enfoque na média de desempenho pode deixar de considerar importantes relações entre os componentes curriculares. Ao indicar especificamente cada disciplina cursada ou não cursada durante o percurso, conjectura-se neste estudo que o arranjo de disciplinas cursadas pelo aluno impõe uma correlação lógica que reflete nos resultados obtidos. Por exemplo, quando o aluno reprova em uma disciplina X, de acordo com as competências exigidas nesta disciplina, é possível que ele tenha um mal desempenho em disciplinas que exigirem as mesmas competências, da mesma forma isso pode ser aplicado para um bom desempenho. Noutra perspectiva, é possível que o algoritmo reconheça padrões que indiquem que, por exemplo, quando há a combinação de resultados negativos nas disciplinas X1, X6, e X9 é grande a possibilidade de reprovação na disciplina X19. Outrossim, também pode reconhecer uma tendência de que ninguém nunca é aprovado na disciplina X29, sem antes ser aprovado nas disciplinas X4, X21 e X17.

Com uma análise mais detalhada dos resultados obtidos nos testes, é possível inferir que os modelos preditivos assimilaram essa correlação entre o percurso e o desempenho em disciplinas específicas. Cita-se como exemplo, a identificação de 12 alunos que cursaram mais de uma disciplina dentre as 15 analisadas e obtiveram diferentes resultados, isto é, foram aprovados em algumas disciplinas e reprovados em outras. Apesar disso, a ferramenta acertou a previsão em todos os casos (destes 12 alunos), demonstrando que, com essa abordagem é possível é realizar a predição para cada disciplina específica. Ainda nesse contexto, destaca-se o caso de um aluno, aqui denominado como Aluno118, que cursou 5 disciplinas dentre as analisadas, obtendo 2 aprovações e 3 reprovações, e em todos os casos houve a predição correta pela ferramenta. Para certificar-se de que os modelos preditivos não estão tendenciosos às taxas de aprovação da disciplina, classificando como aprovados todos os alunos matriculados em uma disciplina com altas taxas de aprovação, ou do aluno, predizendo que ele será aprovado em qualquer disciplina somente por ter muitas aprovações no passado, observou-se as taxas de reprovações de cada disciplina. Foi identificado que o Aluno118 foi aprovado na disciplina de Estrutura de Dados I, que imprimiu uma taxa de reprovação de 69%. Ou seja, mesmo sendo reprovado em disciplinas com menores taxas de reprovação, ele foi aprovado em uma com alta taxa, e em todos os casos as previsões de desempenho foram corretas. Com isso, é possível inferir que os modelos preditivos foram acertadamente sensíveis à correlação entre o percurso realizado pelo Aluno118 e sua capacidade de ter um bom desempenho em disciplinas específicas, e desempenho inferior em outras. Certamente, as metodologias de ensino empregadas pelos professores têm forte impacto nesses resultados. Todavia, há de se reconhecer que os modelos preditivos apresentaram um alto desempenho mesmo sem ter a informação de quais seriam as metodologias de ensino empregadas no período letivo de 2022/1.



Reitera-se a aderência dessa metodologia à flexibilidade do currículo no ensino superior, a despeito da abordagem de [SANTOS et al., 2021], por exemplo, onde são utilizados modelos preditivos para cada semestre letivo em que cada aluno se encontra, não sendo levado em consideração que cada aluno constrói seu próprio percurso, cursando diferentes disciplinas. Isto é, 2 alunos que estão na mesma 5ª fase letiva, podem ter cursado diferentes disciplinas, pois o currículo é flexível. Estes alunos tiveram experiências distintas que imprimiram diferentes características em seu perfil, as quais podem refletir em resultados opostos mesmo que eles curse as mesmas disciplinas nas próximas fases letivas. Esse apontamento é ratificado pela observação da heterogeneidade de fases letivas dos alunos de todas as disciplinas analisadas, e isso significa que os alunos matriculados em uma disciplina são provenientes de diferentes turmas, ingressaram no curso em períodos diferentes e, conseqüentemente, estão em diferentes fases letivas. Tome-se, por exemplo, a disciplina de Linguagens Formais e Autômatos que, de acordo com o currículo, é prevista para ser ofertada na 2ª fase letiva. Porém, dentre os matriculados de 2022/1, havia alunos de 8 fases letivas distintas, desde a 2ª até a 9ª fase letiva.

Quanto à ferramenta desenvolvida, sustenta-se a perspectiva de [PHAN et al., 2023] no que concerne à necessidade de que um recurso baseado em MDE necessite estar acessível aos educadores e gestores de educação por meio de uma interface amigável, do contrário, serão reduzidos os impactos positivos na instituição. O que se constata como característico dos estudos em MDE, é que há uma prevalência em delimitar os resultados na avaliação dos modelos preditivos, em geral concluindo que os modelos preditivos possuem uma boa acurácia. O que se estende-se nesse estudo é o “passo seguinte”, isto é, a utilização desses modelos preditivos, já avaliados, em ferramentas que possam ser utilizadas por educadores na predição de novas instâncias (alunos), e subsidiando suas estratégias educacionais. Ressalta-se que dos 10 trabalhos relacionados apresentados, apenas 2 trazem uma ferramenta com essas características, enquanto as demais tratam da validação dos modelos preditivos ou, como em [MAIA et al., 2010], [MORAIS et al., 2022] e [CZIBULA et al., 2019], restringem-se a nível de código. Em outros termos, é possível utilizar os modelos na predição de novas instâncias via código de programação. Assim, exige-se a disponibilidade do especialista. Por fim, salienta-se o aporte realizado pela ferramenta quanto ao método de classificar as novas instâncias e apresentar os resultados. Em [PEREIRA et al., 2019], por exemplo, é preciso inserir manualmente os atributos de cada aluno para o qual se deseja fazer a previsão, enquanto na ferramenta aqui desenvolvida é necessário apenas inserir a lista de alunos matriculados em uma determinada disciplina para ser apresentado à previsão do desempenho (reprovação ou aprovação) de cada aluno. Desse modo, a ferramenta desonera o professor da dispendiosa tarefa de coleta de dados, bem como das dificuldades de análise técnica dos resultados da MDE.

## **5. Considerações Finais**

Neste estudo é apresentada uma ferramenta para previsão do desempenho acadêmico na educação superior, sob a dicotomia de aprovação ou reprovação. Os resultados alcançados nos testes foram satisfatórios, sendo factível sua utilização para subsidiar intervenções pedagógicas que favoreçam o desempenho dos alunos nas disciplinas. Destaca-se que o momento de realização da previsão é em tempo hábil, pois é no início da execução da disciplina, de forma que os professores ainda tenham a oportunidade de implementar estratégias pedagógicas para mitigar casos de risco de reprovação.

As altas taxas de acerto não validam somente a ferramenta ou os modelos preditivos mas, também, o uso dos percursos curriculares como atributos previsores. Os resultados demonstraram que esse método foi capaz de considerar as particularidades de cada disciplina, e como isso se relaciona com a história do aluno até o momento de cursá-la. Ademais, nesta abordagem, os percursos curriculares são atualizados a cada período letivo sem que seja preciso alterar a estrutura dos dados para treinamento ou previsão. Dessa forma, o modelo preditivo pode ser mantido sempre atualizado e mais próximo da realidade corrente do curso.

Ainda que a análise do desempenho acadêmico não seja o foco deste estudo, é possível ponderar que a amostragem é uma limitação, pois uma maior quantidade de disciplinas e cursos avaliados podem tornar mais expressiva a comprovação da eficácia da ferramenta. Desse modo, como trabalhos futuros é planejado estender os testes para mais disciplinas e cursos, e avaliar o uso da ferramenta em intervenções de fato realizadas com base em seus resultados. Também elenca-se como limitação, a impossibilidade de realizar a predição na primeira fase letiva dos alunos, uma vez que ainda não há percurso anterior. Essa condição deve ser contornada, futuramente, com a utilização de atributos pré-ingresso.

## Referências

- Alhazmi, E., Sheneamer, A. (2023). Early Predicting of Students Performance in Higher Education. in IEEE Access, vol. 11, pp. 27579-27589, 2023, doi: 10.1109/ACCESS.2023.3250702.
- American Psychological Association. (2015). 2015 survey of psychology health service providers. Washington, DC: Author.
- Czibula G., Mihai A., Crivei L. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction, Procedia Computer Science, Volume 159. <https://doi.org/10.1016/j.procs.2019.09.156>.
- Garcia, L. M. L. S., Daiany F. L., e Antunes, F. (2021). “Análise da Retenção no Ensino Superior: um Estudo de Caso em um Curso de Sistemas de Informação”. Revista da Faculdade de Educação 34 (2):15-38. <https://doi.org/10.30681/21787476.2020.34.1538>.
- Garcia L. M. L. S., Lara D. F., Gomes R. S., Cazella, S. C. (2022) Mineração de Dados Educacionais na Predição do Desempenho Acadêmico: um prognóstico a partir do percurso curricular realizado. In: SBIE, 33., 2022, Manaus. Anais. Porto Alegre: SBC. DOI: <https://doi.org/10.5753/sbie.2022.225221>.
- Li S., Liu T. Performance Prediction for Higher Education Students Using Deep Learning, Complexity, vol. 2021, Article ID 9958203, 10 pages, 2021. <https://doi.org/10.1155/2021/9958203>
- Maia R.F., Spina E.M., Shimizu S.S. (2010). Sistema de Previsão de Desempenho de Alunos para Auxílio a Aprendizagem e Avaliação de Disciplinas. In Anais do XXI SBIE-XVI WIE, 2010.
- Morais D. M. G., Souza H. R., Silva V. B. (2023) Modelo preditivo do desempenho acadêmico: um estudo para a disciplina de cálculo diferencial e integral para Engenharias. FTT Journal of Engineering and Business. N.8. P. 47 - 68.

- Páez A. R. (2022) Modelos predictivos progresivos del rendimiento académico de estudiantes universitarios. *RIDE Revista Iberoamericana Para La Investigación Y El Desarrollo Educativo*, Vol. 12, Núm. 24. <https://doi.org/10.23913/ride.v12i24.1196>
- Pereira M., Oliveira A., Teixeira M. M., Neto C. S. (2019). Uma aplicação web para predição de desempenho de alunos. In: *Escola Regional De Computação Do Ceará, Maranhão e Piauí*, 7., 2019, São Luís. Anais. Porto Alegre: SBC, 2019. p. 198-205.
- Phan M., Caigny A., Coussement K. (2023) A decision support framework to incorporate textual data for early student dropout prediction in higher education, *Decision Support Systems*, Volume 168. <https://doi.org/10.1016/j.dss.2023.113940>.
- Rabelo H., Burlamaqui A., Valentim R., Souza Rabelo, D. S., Medeiro S. (2017). Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de ead em ambientes virtuais de aprendizagem. In *SBIE*, volume 28. DOI: <http://dx.doi.org/10.5753/cbie.sbie.2017.1527>
- Rodríguez-Hernández C. F., Musso M., Kyndt E., Cascallar E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation, *Computers and Education: Artificial Intelligence*, Volume 2, 2021, 100018, ISSN 2666-920X, <https://doi.org/10.1016/j.caeai.2021.100018>.
- Santos C. H. D. C., Martins, S. L., Plastino, A. (2021). É Possível Prever Evasão com Base Apenas no Desempenho Acadêmico? In: *Simpósio Brasileiro de Informática na Educação*, 32. Online. Anais. Porto Alegre: SBC. 2021. p. 792-802. DOI: <https://doi.org/10.5753/sbie.2021.218105>.
- Stoll B. B., Cury D., De Lira Tavares, O., Silva De Menezes, C. (2019) Análise de dados acadêmicos baseado em previsão, recomendação e visualização. *Revista Novas Tecnologias na Educação*, Porto Alegre, v. 17, n. 1. DOI: 10.22456/1679-1916.95794.