

## **Aplicação de Ciência de Dados em um Ambiente Virtual de Aprendizagem: um estudo de caso na disciplina de Programação Orientada a Objetos**

Mateus de Sousa Araújo, Universidade Federal do Ceará, mateusaraujo@alu.ufc.br,  
<https://orcid.org/0009-0007-2775-6237>

Marcos Vinicius de Andrade Lima, Universidade Federal do Ceará, marcos.vinicius@ufc.br  
<https://orcid.org/0000-0002-5797-0222>

**Resumo:** Este trabalho apresenta a análise de desempenho de alunos na disciplina de programação orientada a objetos utilizando modelos descritivos baseados em técnicas de ciência de dados de uma instituição pública de ensino superior no município de Russas no estado do Ceará. O objetivo foi categorizar os alunos em grupos de desempenho, identificar perfis de aprendizado e recomendar conteúdos específicos. Foram aplicadas regras de classificação para agrupar os alunos em categorias como bom, regular, alerta e crítico. Além disso, *outliers* foram identificados e removidos para garantir a precisão dos resultados. A análise revelou padrões de comportamento e desempenho, permitindo recomendações mais precisas para apoio educacional. Conclui-se que o modelo proposto contribui significativamente para melhorar o suporte ao aprendizado.

**Palavras-chave:** análise de aprendizagem, programação orientada a objetos, ciência de dados, recomendação de conteúdos.

### **Application of Data Science in a Virtual Learning Environment: a case study in Object-Oriented Programming**

**Abstract:** This paper presents an analysis of student performance in the subject of object-oriented programming using descriptive models based on data science techniques at a public higher education institution in the municipality of Russas in the state of Ceará. The aim was to categorize students into performance groups, identify learning profiles and recommend specific content. Classification rules were applied to group students into categories such as good, fair, alert and critical. In addition, *outliers* were identified and removed to ensure the accuracy of the results. The analysis revealed patterns of behavior and performance, allowing for more accurate recommendations for educational support. It is concluded that the proposed model contributes significantly to improving learning support.

**Keywords:** learning analytics, object-oriented programming, data science, content recommendation.

## **1. Introdução**

O interesse pelo aprendizado de programação de computadores tem crescido de forma significativa nas últimas décadas, especialmente com a expansão da oferta de cursos e escolas especializadas na área (Scaico, 2018). Programação, no entanto, traz desafios complexos para os estudantes, que precisam desenvolver diversas habilidades simultaneamente, como a compreensão da sintaxe e da semântica de linguagens de programação, além de raciocínio lógico para a resolução de problemas (Gomes, 2015). Dentre os paradigmas de programação, a Programação Orientada a Objetos (POO) se destaca pela capacidade de modelar sistemas complexos, sendo fundamental em cursos como Engenharia de Software e Ciência da Computação (Farinelli, 2007).

Em uma disciplina de POO, oferecida nos anos de 2023 e 2024 no ensino superior de uma instituição pública no município de Russas no estado do Ceará, com o suporte de um Ambiente Virtual de Aprendizagem (AVA) Moodle, observou-se que

os professores ainda enfrentam dificuldades em identificar precisamente as lacunas no aprendizado dos alunos, o que compromete o desenvolvimento de atividades pedagógicas mais direcionadas. Diante desse cenário, a presente pesquisa investigou como a análise de dados gerados no AVA pode ser utilizada para identificar essas lacunas no ensino da disciplina de POO, permitindo uma recomendação mais eficaz de recursos de aprendizagem, possibilitando ainda que o docente dê maior assistência aqueles alunos que apresentam maiores dificuldades.

Este trabalho tem como principal objetivo realizar uma Análise de Aprendizagem na disciplina de POO, utilizando técnicas de Ciência de Dados (CD) para melhorar o processo de ensino-aprendizagem. A análise seguiu diretrizes do processo descrito por Cielen, Meysman e Ali (2016), sendo realizada sobre os dados gerados pelo AVA. A intenção foi detectar possíveis lacunas no aprendizado e, com base nisso, propor recomendações de conteúdo e atividades pedagógicas mais assertivas, visando aprimorar o desempenho dos estudantes.

Para tanto, foi desenvolvido um modelo descritivo baseado em CD que auxilia a identificação das deficiências no aprendizado dos alunos de forma mais clara. A partir desse modelo, foi possível categorizar os estudantes em diferentes grupos de desempenho, o que facilita a criação de intervenções mais direcionadas às necessidades específicas de cada grupo, promovendo um processo de ensino mais eficaz.

Este trabalho está estruturado em cinco seções. A Seção 1 introduz o tema, destacando a importância da POO e os desafios enfrentados no ensino dessa disciplina em uma instituição pública de ensino superior no município de Russas no estado do Ceará, além de apresentar os objetivos e a justificativa para a realização da pesquisa. Em seguida, a Seção 2 revisa trabalhos relacionados, destacando pesquisas anteriores e suas contribuições. Depois, a Seção 3 detalha os procedimentos metodológicos utilizados na coleta e análise dos dados, as fases da pesquisa, a aplicação das técnicas de CD e o processo de construção do modelo. Na Seção 4, são apresentados os resultados obtidos. Por fim, a Seção 5 traz as conclusões em relação aos objetivos estabelecidos e aos resultados alcançados.

## **2. Trabalhos relacionados**

O trabalho de Silva, Carvalho e Maciel (2021) propôs o desenvolvimento de um *Learning Analytics Dashboard* (LAD) como módulo do *Framework* de Mineração de Dados Educacionais (FMDEV). O LAD foi projetado para criar visualizações a partir de dados educacionais coletados no Moodle, com o objetivo de oferecer suporte ao ensino e aprendizado. A pesquisa envolveu a análise de dados de 30.217 registros de alunos, agrupados em construtos de autorregulação da aprendizagem, usados para treinar modelos preditivos. A validação do LAD com professores do ensino superior demonstrou sua utilidade e aceitação, com sugestões de futuras melhorias baseadas no *feedback* dos usuários.

Já o estudo de Shrestha e Pokharel (2021) utilizou técnicas de mineração de dados educacionais para prever o desempenho de alunos com base em dados coletados no Moodle. Após o pré-processamento e a seleção de recursos relevantes por meio do método Boruta, modelos de classificação como *Random Forest*, *SVM* e *Decision Tree* foram avaliados com métricas como precisão e *F1-Score*. Os resultados mostraram que as atividades online dos alunos possuíam uma forte relação com o desempenho acadêmico, permitindo identificar precocemente dificuldades e melhorar processos de ensino.

Este trabalho, assim como as pesquisas descritas acima, utilizou dados do AVA Moodle para obtenção dos resultados. Mas, diferentemente dos anteriores, este trabalho

aplicou um processo e técnicas da CD para explorar padrões e características específicas em conjunto de dados dos alunos.

### 3. Metodologia

#### 3.1. Caracterização da pesquisa

Este estudo foi realizado sob a abordagem de pesquisa quantitativa, seguindo o paradigma pragmático descrito por Creswell (2010), que oferece flexibilidade aos pesquisadores na escolha dos métodos, técnicas e procedimentos que melhor atendam às suas necessidades e aos seus objetivos. A abordagem quantitativa permite a avaliação de teorias objetivas, examinando as interações entre variáveis que podem ser mensuradas por meio de instrumentos específicos. Esses dados numéricos são então analisados por meio de procedimentos estatísticos (Creswell, 2010).

#### 3.2. Fases da pesquisa

Os procedimentos da pesquisa foram divididos em duas fases principais: o processo da CD e a recomendação de conteúdos a partir dos resultados obtidos desse processo. A Figura 1 ilustra a sequência de passos adotados para a execução da pesquisa. Os passos numerados de 1 a 6 fazem parte do processo da CD, enquanto o passo 7 se refere à recomendação de conteúdos e possíveis intervenções pedagógicas.

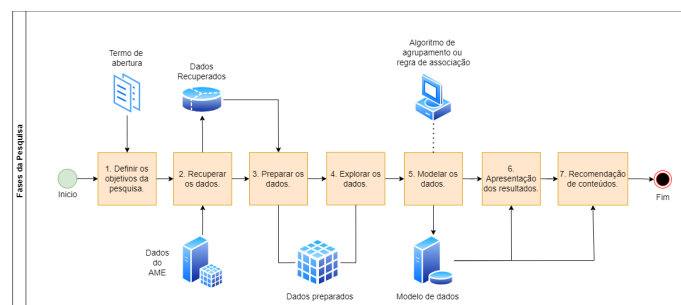


Figura 1. Fases da pesquisa.

O processo de construção do modelo foi realizado combinado com o processo de CD proposto por Cielen, Meysman e Ali (2016) que traz uma abordagem estruturada, que segundo o autor oferece a vantagem de maximizar as chances de sucesso em um projeto, reduzindo custos. Além disso, viabiliza a execução de projetos em equipe, com cada membro concentrando-se em suas áreas de interesse.

A seguir, os procedimentos em cada passo são detalhados:

- 1. Definição dos objetivos de pesquisa:** O primeiro passo consiste em criar uma carta de projeto que assegure o entendimento claro sobre o quê, porque e como o projeto será desenvolvido. No contexto deste trabalho, os principais objetivos definidos foram:
  - (a) Registrar as principais atividades utilizadas na disciplina de POO, através do AVA de uma instituição pública de ensino superior;
  - (b) Aplicar um processo de CD para desenvolver um modelo que forneça *insights* relevantes para o apoio ao aprendizado;
  - (c) Apoiar a criação de conteúdos e atividades personalizadas educacionais com base na análise do modelo gerado no processo de CD.
- 2. Recuperação dos dados:** esta etapa visa garantir que os dados necessários estejam disponíveis, com qualidade e acessíveis para análise. Neste trabalho, foi concedido o acesso a duas cópias da turma do AVA de dois períodos diferentes

em 2023, fornecida pelo professor responsável pela disciplina de POO. Esse banco de dados contém informações valiosas, como detalhes sobre as atividades realizadas pelos alunos, frequência de acessos ao sistema, participação em fóruns, notas obtidas em avaliações, e o *status* de conclusão da disciplina. Esses dados foram fundamentais para aplicar técnicas de CD e realizar uma análise profunda e precisa.

3. **Preparação dos dados:** nesta etapa, os dados são manipulados para garantir sua adequação à análise. Nesta fase, com as duas cópias de *backup* do Moodle, obtidas na fase anterior, uma com 92,1 Megabytes e outra com 53,1 Megabytes, ambas no formato *Moodle Backup File (MBZ)*. Foi necessária a instalação do Moodle juntamente com o banco de dados MariaDB, na mesma versão do sistema utilizada atualmente pelo professor de POO. Após importar os *backups*, para visualizar as tabelas do banco de dados foi necessário a instalação de um *workbench*. O banco de dados padrão do Moodle contém 461 tabelas, das quais foram analisadas cautelosamente na próxima fase.
4. **Exploração dos dados:** A exploração dos dados é a quarta fase, com a limpeza dos dados e a obtenção de um formato adequado para utilização. Foi necessário definir uma estratégia, e definir os dados mais importantes a serem extraídos da base para serem submetidos a modelagem de dados. A partir disso, adotou-se uma abordagem de categorização dos alunos em diferentes listas, de acordo com critérios específicos definidos durante a próxima fase de modelagem. As categorias criadas foram: **alerta, regulares, críticos e bons**. Além disso, para garantir maior precisão no modelo, foram desenvolvidos perfis de alunos que serviram como referência para determinar a inclusão nas respectivas listas, assegurando que as classificações fossem mais consistentes e alinhadas com os padrões estabelecidos.

Sendo assim, foi vista a necessidade de extrair dois arquivos *Comma Separated Values (CSV)*. O primeiro arquivo, nomeado de "tentativas", contendo 303 Kilobytes, registrava todas as tentativas dos alunos em atividades e provas com as seguintes colunas: atividade, tentativas realizadas, tentativas permitidas, identificador da atividade, nota da tentativa, nome do usuário, identificador do usuário, estado, tipo, seção, tempo em segundos e e-mail do usuário. O segundo arquivo, nomeado de "resultados", registrava os resultados finais da disciplina e era composto pelas colunas: nota final, identificador do usuário e nome do usuário. Durante o processamento, valores nulos foram eliminados de ambos os arquivos, garantindo que apenas informações completas e consistentes fossem utilizadas.

Na criação de perfis, o foco foi entender melhor o comportamento dos alunos nas diferentes categorias, para definir os limites e parâmetros que separam um grupo de outro. A criação dos perfis inclui:

**Definição dos Perfis por Categoria:** Os perfis para cada categoria (Bom, Regular, Alerta e Critico) são definidos a partir de valores médios de desempenho observados nas variáveis: nota média, tempo médio gasto, e número médio de tentativas. Esses perfis funcionam como parâmetros de comparação para classificar os alunos.

Com base nesses parâmetros, as regras para cada perfil foram construídas, conforme descrito a no Quadro 1.

Quadro 1: Descrição dos perfis dos estudantes

Perfil	Descrição
Perfil Bom	Notas altas, gastam pouco tempo nas atividades e realizam poucas tentativas. Eles representam o grupo de melhor desempenho.
Perfil Regular	Desempenho intermediário, que têm notas medianas e um número de tentativas e tempo gasto também medianos.
Perfil Alerta	Desempenho abaixo do esperado em pelo menos uma das variáveis (nota, tempo ou tentativas). Eles exigem uma atenção especial para identificar possíveis problemas.
Perfil Crítico	Desempenho significativamente baixo em múltiplos fatores, caracterizados por notas baixas, muitas tentativas e/ou tempos muito elevados. Eles estão em situação crítica e precisam de intervenção urgente.

**Cálculo das Médias para Comparação:** Para realizar o cálculo das médias foi desenvolvido um *script* na linguagem Python que acessa o arquivo de "resultados" e filtra os resultados do curso segundo a nota que define cada perfil, exceto o perfil Alerta que não observa a nota, sendo assim: alunos bons acima de 7, alunos regulares abaixo de 7 e acima de 5 e alunos críticos abaixo de 5, e retorna listas de alunos que finalizaram o curso como Bons, Regulares e Críticos. A partir disso, o *script* acessa o arquivo de "tentativas" e soma a nota, tempo, e tentativa para cada tentativa dos alunos de cada lista e calcula as médias de cada perfil.

Com as médias gerais de cada perfil calculadas, o processo de classificação foi automatizado ao comparar o desempenho individual de cada aluno com os parâmetros dos perfis definidos. Dependendo de como as médias de um aluno se alinham com os perfis, ele é alocado na categoria correspondente. Esses perfis não só ajudam a identificar os alunos em cada categoria, mas também servem como base para analisar as relações entre as variáveis (nota, tempo e tentativas), oferecendo uma visão mais profunda sobre como o desempenho varia entre diferentes grupos de estudantes.

Somente tentativas concluídas foram analisadas, garantindo confiabilidade. As atividades, realizadas presencialmente sob supervisão, garantiram integridade do processo. A cada nova tentativa novas questões e alternativas eram randomizadas pelo Moodle. O tempo de realização foi destacado para identificar inconsistências e comportamentos fora do padrão.

O K-Means (Jain; Murty e Flynn, 1999) foi utilizado nesta fase de exploração dos dados, para auxiliar na identificação dos padrões e agrupamentos entre as tentativas dos estudantes. Ele foi aplicado para gerar 4 *clusters*, visto que são 4 categorias de alunos definidas. Ele gerou as 4 classes, que foram adicionadas ao arquivo de "tentativas". Sendo assim, foi possível entender melhor a relação entre esses dados ao aplicar esse algoritmo de *clustering*, sendo possível segmentar os alunos com base em características como notas, tempo gasto e número de tentativas. Essa análise revelou relações importantes entre as variáveis principais e contribuiu para a compreensão do comportamento dos estudantes. Os grupos formados pelo K-Means serviram como base para a categorização dos alunos e a definição dos critérios de classificação, tornando-se uma ferramenta fundamental para a construção do modelo final.

5. **Modelagem dos dados:** após compreender os dados e adquirir conhecimento

sobre o domínio na etapa anterior, os modelos são empregados para abordar a questão de pesquisa.

O modelo de categorização de estudantes foi desenvolvido com base em critérios claros para avaliar o desempenho dos alunos a partir de três variáveis principais: nota, tempo gasto nas atividades e número de tentativas. Esses três fatores foram usados para calcular médias que definem perfis de desempenho, servindo como referência para a classificação dos alunos em quatro categorias: bons, regulares, críticos e alerta.

Inicialmente, as médias de desempenho de cada estudante foram calculadas somando-se a nota, o tempo total gasto e o número de tentativas realizadas em suas atividades e dividindo esses valores pelo número de tentativas efetuadas. Com essas médias, o aluno foi comparado aos perfis pré-definidos de cada categoria.

A categorização foi feita de forma automatizada com um *script* em Python que seguiu as diretrizes estabelecidas no Quadro 2.

**Quadro 2: Critérios de classificação nas categorias dos estudantes**

<b>Categoria</b>	<b>Critérios</b>
Bom	O aluno precisa apresentar uma média de notas superior ou igual à média do perfil Bom, além de ter um tempo gasto e número de tentativas abaixo dos valores estabelecidos para esse mesmo perfil.
Regular	Estudante precisa manter suas médias de nota abaixo do perfil Bom e acima do perfil Crítico, ter um tempo gasto e número de tentativas abaixo dos valores do perfil.
Crítico	Os estudantes que seus indicadores estejam abaixo das médias de nota e acima dos limites de tempo e tentativas para o perfil Crítico.
Alerta	Foi projetada para incluir todos os estudantes que, inicialmente, não atendem a nenhum padrão específico, ou seja, que não se encaixam diretamente nos perfis estabelecidos de Bom, Regular ou Crítico. Esses alunos foram sinalizados como estando fora do padrão, e a partir daí, foram reavaliados com base apenas em suas notas para decidir sua classificação final.

### **3.3. Recomendação de conteúdos**

Durante o processo de cálculo das médias, o *script* desempenhou um papel fundamental ao identificar as atividades nas quais os alunos apresentaram desempenho insatisfatório, com notas inferiores a 7. Essa análise não apenas destacou as dificuldades enfrentadas por certos estudantes, mas também ofereceu uma visão abrangente das áreas no qual a intervenção pedagógica era mais necessária. Com base nessas informações, é possível desenvolver recomendações personalizadas de conteúdos e estratégias pedagógicas que atendam às necessidades específicas de cada aluno. Assim, a identificação das atividades de baixo desempenho se torna uma ferramenta poderosa para orientar a prática docente e promover um ensino mais eficaz e direcionado.

### **3.4. Aspectos éticos da pesquisa**

Neste trabalho, é fundamental ressaltar a importância de considerar os aspectos éticos relacionados à proteção de dados pessoais, conforme estabelecido pela Lei Nº 13.709, de 14 de agosto de 2018, conhecida como Lei Geral de Proteção de Dados (LGDP). Conforme mencionado na Subseção 2, os dados utilizados foram fornecidos pelo professor da disciplina de POO, por meio do preenchimento do Termo de Consentimento Livre e Esclarecido (TCLE). É essencial seguir rigorosamente as diretrizes estabelecidas

pela legislação para assegurar a privacidade e a segurança dos dados dos indivíduos envolvidos, sendo assim, os nomes dos estudantes estão anonimados, impedindo qualquer tipo de identificação.

A LGPD regulamenta o tratamento de dados pessoais, inclusive nos meios digitais, por pessoas naturais ou jurídicas, públicas ou privadas. Seu objetivo é proteger os direitos fundamentais de liberdade, privacidade e o livre desenvolvimento da personalidade da pessoa natural. Dessa forma, a conformidade com a LGPD garante que todas as atividades de coleta, armazenamento e análise de dados sejam realizadas de maneira ética e legal, respeitando os direitos dos indivíduos. Todos os dados colhidos nesta pesquisa foram utilizados apenas para treinar os algoritmos, portanto, não foram divulgados nenhum dos dados individuais dos estudantes.

## 4. Resultados e Discussão

### 4.1. Apresentação de resultados

Após obter o modelo de dados, como foi descrito na fase de Modelagem dos Dados, chegamos à sexta fase do processo da CD proposto por (Cielen; Meysman e Ali, 2016), que é a apresentação dos resultados. O modelo desenvolvido recebeu dados de tentativas de alunos de 2 turmas de POO, totalizando 98 alunos. Após a limpeza de dados restaram apenas 86 alunos, tendo sido submetidos ao modelo todas as tentativas, compreendendo um total de 10 seções. Foram categorizados 51 estudantes como Bom, 5 como Críticos, 30 como Regular e 73 como alerta. O modelo categorizou corretamente 59 alunos de 86, obtendo uma acurácia de 68,60%. É importante destacar que o modelo aumenta sua porcentagem de acerto na classificação gradativamente ao receber mais dados, por exemplo, quando submetido dados de 5 seções para o modelo, a taxa de acerto é de 60,97%, quando submetido a 8 seções a taxa de acerto sobe para 65,88% de acerto. Para compreender melhor os resultados, foram construídos gráficos que mostram as relações dos dados analisados.

No Gráfico 1 é possível perceber um padrão inversamente proporcional, nos quais alunos que realizaram mais tentativas tendem a apresentar notas mais baixas. Isso sugere que esses alunos encontram dificuldades nas atividades e dependem de múltiplas tentativas para melhorar seu desempenho. Por outro lado, alunos que realizaram menos tentativas geralmente obtiveram notas mais altas, indicando maior domínio do conteúdo. Esse comportamento reforça a necessidade de intervenções pedagógicas para os alunos com maior número de tentativas, visando melhorar a retenção e compreensão do conteúdo abordado.

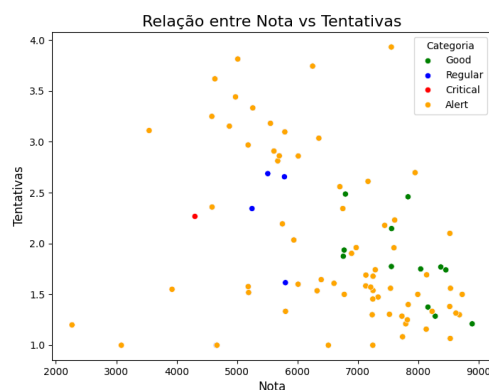


Gráfico 1. Relaciona as notas dos alunos com a quantidade de tentativas realizadas, separado por categoria.

O Gráfico 2 revela uma correlação positiva entre essas duas variáveis. Isso indica que, à medida que os alunos dedicam mais tempo às atividades, suas notas tendem a aumentar. Esse padrão sugere que o investimento de tempo nas tarefas e estudos está associado a um melhor desempenho acadêmico. A implicação disso é que encorajar os alunos a se dedicarem mais ao tempo de estudo pode ser uma estratégia eficaz para melhorar suas notas.

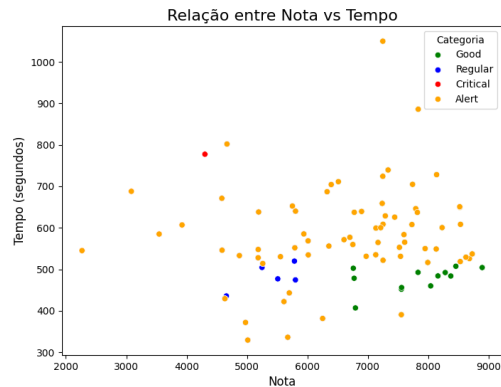


Gráfico 2. Relaciona as notas dos alunos com o tempo gasto nas atividades, separado por categoria.

O Gráfico 3 representa a classificação geral, e uma visão clara da distribuição dos alunos nas diferentes categorias de desempenho: Bom, Regular, Alerta e Crítico. A partir dos dados, é possível observar que a maioria dos alunos se encontra na categoria 'Bom', indicando um bom nível de aproveitamento. No entanto, a presença de um número significativo de alunos nas categorias 'Alerta' e 'Crítico' sinaliza a necessidade de intervenções pedagógicas direcionadas. Essa categorização permite identificar os grupos que podem se beneficiar de apoio adicional, ajudando a direcionar esforços educacionais para aqueles que mais precisam. Em suma, a classificação geral não apenas reflete o desempenho acadêmico dos alunos, mas também fornece *insights* valiosos para a personalização do ensino e a implementação de estratégias de melhoria.

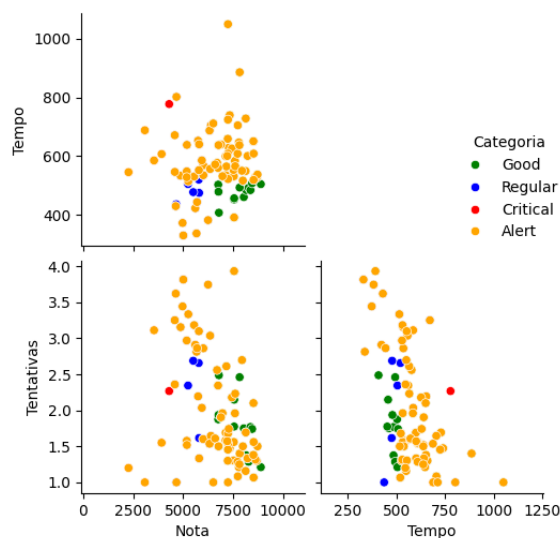


Gráfico 3. Visão geral do desempenho dos alunos, destacando as áreas com mais estudantes.

Além das análises mencionadas, um aspecto importante abordado na limpeza dos



dados foi a identificação e remoção de *outliers*, que poderiam distorcer as conclusões da pesquisa. Para isso, foi implementada uma função específica que categorizou os alunos como *outliers* com base em critérios pré-estabelecidos. Alunos foram considerados *outliers* se apresentassem notas médias inferiores a 1, mais de 4 tentativas em uma atividade ou um tempo médio gasto superior a 1200 segundos. A remoção desses *outliers* foi fundamental para assegurar que as análises subsequentes fossem mais representativas e que os perfis de desempenho gerados refletissem com precisão a realidade do grupo de alunos. Essa abordagem rigorosa na limpeza dos dados não só aumentou a confiabilidade dos resultados, mas também possibilitou uma análise mais refinada e focada nos alunos que realmente necessitam de suporte.

#### 4.2. Sobre a Identificação de Conteúdos Problemáticos

O Gráfico 4 fornece uma visão panorâmica das tentativas de atividades realizadas pelos alunos e suas respectivas notas. Essa visualização serve como uma ferramenta poderosa para identificar quais conteúdos estão apresentando maiores dificuldades para os alunos.

Ao analisar as tentativas e as notas, é possível observar claramente quais seções e atividades geraram um desempenho insatisfatório. Através dessa análise, conseguimos identificar padrões que destacam os conteúdos que mais necessitam de apoio pedagógico.

Além disso, o Gráfico 4 não apenas ilustra a relação entre as tentativas e as notas, mas também serve como base para recomendações específicas de conteúdos. Com essa informação, os educadores podem direcionar intervenções pedagógicas mais eficazes, como revisões de conteúdo, materiais de apoio ou estratégias de ensino diferenciadas, para abordar as áreas identificadas como problemáticas.

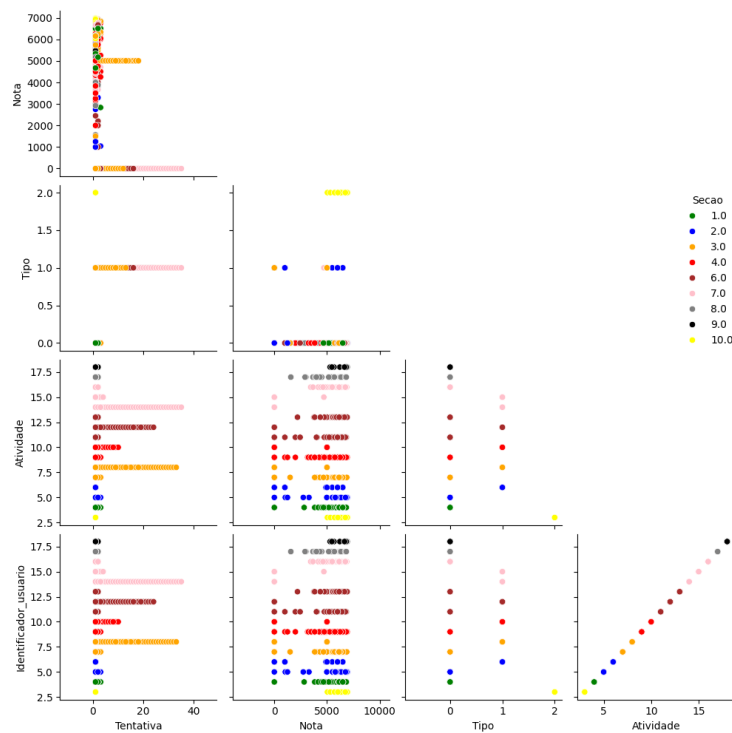


Gráfico 4. Visão geral dos conteúdos a recomendar.

## 5. Conclusões

Esta pesquisa teve como principal objetivo analisar os dados gerados no AVA Moodle na disciplina de POO e propor um modelo que fornecesse *insights* relevantes para

o apoio ao aprendizado dos alunos, facilitando o acompanhamento por parte do docente.

Por meio da aplicação de um processo de CD, foi possível identificar padrões significativos nas interações dos alunos com as atividades do curso. O primeiro objetivo, que consistia em registrar as principais atividades utilizadas na disciplina, foi alcançado com sucesso, fornecendo uma visão detalhada das interações dos alunos. Esse registro se mostrou crucial para compreender onde os alunos estavam enfrentando dificuldades, como, por exemplo, no tópico 10, que se tratava de uma atividade prática.

Em relação ao segundo objetivo, a análise dos dados permitiu a criação de perfis de desempenho, categorizando os alunos em grupos: bons, regulares, críticos e em alerta. Essa categorização foi fundamentada em variáveis obtidas no AVA, como notas, tempo gasto nas atividades e número de tentativas. Com isso, foi possível identificar aqueles alunos que necessitavam de apoio adicional, possibilitando intervenções mais direcionadas e eficazes.

O terceiro objetivo, que envolvia apoiar a criação de conteúdos e atividades personalizadas, também foi atendido. Com base nas análises, foram elaboradas recomendações para os educadores, indicando quais conteúdos e seções apresentaram maior necessidade de intervenção pedagógica. Os gráficos gerados serviram como ferramentas visuais que exemplificam a relação entre tentativas e desempenho, facilitando a identificação de áreas problemáticas.

Observa-se que todos os objetivos da pesquisa foram alcançados, trazendo uma contribuição para a melhoria do processo de ensino-aprendizagem na disciplina de POO, oferecendo um modelo que não só analisa dados, mas que também orienta ações pedagógicas fundamentadas em evidências. As implicações desses resultados sugerem que a adoção de abordagens baseadas em dados pode ser uma estratégia eficaz para promover um aprendizado mais significativo e adaptado às necessidades dos alunos.

É importante destacar algumas limitações no Moodle relacionadas à sua capacidade de detectar fraudes e comportamentos irregulares durante as avaliações. Embora a plataforma forneça ferramentas para monitorar o progresso dos alunos, ela não possui recursos robustos para identificar colaborações indevidas em questionários ou verificar a autenticidade das respostas. Essa falta de monitoramento detalhado pode comprometer a confiabilidade dos dados, podendo dificultar a identificação de comportamentos atípicos ou tentativas de burlar o sistema.

Em trabalhos futuros, esperamos desenvolver um *plugin* do modelo para o AVA Moodle, visando proporcionar aos docentes acesso simplificado e direto às informações sobre o desempenho dos alunos. Além disso, buscaremos aprimorar a acurácia do modelo, refinando os métodos de análise e classificação para gerar resultados ainda mais precisos e úteis para o acompanhamento do aprendizado em programação.

## Referências

- Cielen, D.; Meysman, A.; Ali, M. **Introducing Data Science: Big Data, Machine Learning, And More, Using Python Tools**. Shelter Island: Manning, 2016. Acesso em: 29 de abr. 2024.
- Creswell, J. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. Artmed, v. 3, 2010.
- Farinelli, F. **Conceitos básicos de programação orientada a objetos**. ResearchGate, v. 1, 2007. Disponível em: [https://www.researchgate.net/publication/268362210\\_Conceitos\\_basicos\\_de\\_programacao\\_orientada\\_a\\_objetos](https://www.researchgate.net/publication/268362210_Conceitos_basicos_de_programacao_orientada_a_objetos). Acesso em: 18 mar. 2024.
- Gomes, M. **Os benefícios do ensino de linguagem de programação no currículo regular**. Administradores, 2015. Disponível em: <https://administradores.com.br/artigos/>

os-beneficios-do-ensino-de-linguagem-de-programacao-no-curriculo-regular). Acesso em: 07 de ago. 2024.

Jain, A. K.; Murty, M. N.; Flynn, P. J. **Data clustering: a review**. [S.l.: s.n.], 1999. v. 31.

Scaico, P. **Um Estudo Sobre o Desenvolvimento de Interesse Pela Aprendizagem de Programação**. Universidade Federal de Pernambuco, 2018. Disponível em: <https://repositorio.ufpe.br/handle/123456789/30620>. Acesso em: 17 de jun. 2024.

Shrestha, S.; Pokharel, M. **Educational data mining in moodle data**. The International Journal of Informatics and Communication Technology (IJ-ICT), v. 10, n. 1, 2021. Disponível em: <https://ijict.iaescore.com/index.php/IJICT/article/view/20292>. Acesso em: 27 de abr. 2024.

Silva, G.; Carvalho, J.; Maciel, A. **Desenvolvimento de um *Learning Analytics Dashboard* a partir de Modelos de Mineração de Dados Educacionais**. Revista de Engenharia e Pesquisa Aplicada, v. 6, n. 3, 2021. Disponível em: <http://revistas.poli.br/index.php/rep/article/view/1688>. Acesso em: 26 de abr. 2024.