

Análise comportamental das notas dos alunos do Ensino Superior do Instituto Federal de Minas Gerais - *campus* Sabará - com a adoção do Ensino Remoto Emergencial (ERE)

GONÇALVES, A. J. V.¹, TARGA, C. N.¹, GOMES, B. N.¹

¹Bacharelado em Sistemas de Informação – Instituto Federal de Minas Gerais (IFMG)
CEP – 34590-390 – Sabará – MG – Brasil

anajulia99@gmail.com, {cristiane.targa,bruno.nonato}@ifmg.edu.br

Abstract. *Emergency remote teaching (ERE) was a didactic and pedagogical strategy developed by the Federal Institutes to reduce the damage caused by the pandemic, a remedy for the SARS-CoV-2 virus. This paper purpose is to compare grades of subjects from IFMG Sabara higher education, before, between and after the adoption of ERE. To do that, was used mining algorithms of data from Weka tool. The results found show us there was an increase in the grades of higher education students during the pandemic.*

Resumo. *O ensino remoto emergencial (ERE) foi uma estratégia didática e pedagógica adotada pelos Institutos Federais para diminuir os danos provocados pela pandemia, causada pelo vírus SARS-CoV-2. O objetivo deste trabalho é fazer um comparativo de notas das disciplinas dos cursos superiores do IFMG - Campus Sabará antes, durante e depois da adoção do ERE. Para isso foram utilizados algoritmos de Mineração de dados da ferramenta Weka. Os resultados mostram que houve aumento do valor das notas tiradas pelos alunos do ensino superior durante a pandemia.*

1. Introdução

Em 31 de dezembro de 2019, a Organização Mundial da Saúde (OMS) foi alertada sobre a Covid-19, uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global¹. Segundo a Organização Pan-Americana da Saúde, em 11 de março de 2020, a COVID-19 foi caracterizada pela Organização Mundial da Saúde (OMS)² como uma pandemia. Com o surgimento desse vírus, implementou-se um regime de quarentena onde todos os serviços não essenciais ficaram fechados por meses para evitar maior transmissão do vírus.

Um dos setores mais afetados com a com a pandemia do coronavírus, a partir de 2020, foi o educacional. Segundo [Oliveira and Silva 2022], cerca de 1,6 bilhão de alunos no mundo ficaram sem aulas presenciais em 2020 e no Brasil, estima-se que cerca de 20 milhões de alunos tiveram as aulas suspensas. Os Institutos Federais de Educação, Ciência e Tecnologia (IFMG) também tiveram suas aulas presenciais suspensas a partir

¹<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>

²<https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19>

de março de 2020 [Castilho and da Silva 2020]. Para atenuar os danos escolares causados pelo novo coronavírus, o Ministério da Educação (MEC) autorizou a substituição do ensino presencial por aulas que utilizassem meios e Tecnologias de Informação e Comunicação (TIC's) em cursos que já estivessem em andamento (PORTARIA N° 343, DE 17 DE MARÇO DE 2020³). Os Institutos Federais adotaram o Ensino Remoto Emergencial (ERE) como uma forma de manter as atividades didáticas-pedagógicas, nos diferentes níveis de ensino. No Instituto Federal de Minas Gerais - *campus* Sabará, além dos cursos técnicos integrados, também são ofertados os cursos superiores: Bacharelado em Sistemas de Informação (SABSINF), Engenharia de Controle e Automação (SABENCA), Logística (SATLOGI) e Bacharelado em Administração (SABADMI).

Mesmo antes da pandemia, consequentemente antes do ERE, o ensino superior já enfrentava desafios. [Casanova et al. 2020] fez uma análise a partir de uma mostra de 650 estudantes e identificou algumas possíveis dificuldades, entre elas, adaptação à instituição, aprendizagem, relacionamento interpessoal e autonomia. Os resultados mostram que, exceto na aprendizagem, os níveis de dificuldade sentida pelos estudantes diminuem à medida que decorrem as semanas de aulas. [Bardagi and Hutz 2012] defende a importância da relação entre alunos e professores para a permanência do aluno no curso superior. Ele demonstra como o bom relacionamento com colegas de classe ajuda a postergar a evasão do curso. Durante o período de pandemia, não apenas houve uma transformação significativa na forma de ensino, mas também uma mudança nos relacionamentos dentro da comunidade acadêmica, o que poderia ter aumentado ainda mais as dificuldades sentidas pelos discentes.

O presente trabalho se propõe a fazer uma análise comportamental das notas dos alunos dos cursos superiores do IFMG - *campus* Sabará com a adoção do ERE. O objetivo é fazer um comparativo de desempenho dos estudantes em algumas disciplinas dos cursos superiores antes, durante e depois da pandemia e analisar através das notas como foi a adaptação dos alunos diante dessas mudanças. Para isso foram utilizados algoritmos de Mineração de dados da ferramenta weka⁴. Os resultados mostram que o desempenho dos estudantes foi melhor durante a pandemia em comparação com os períodos pré e pós a quarentena.

O artigo está organizado da seguinte forma. A seção 1 apresenta a introdução ao problema de pesquisa, sua relevância e a proposta de solução adotada. A fundamentação teórica e revisão bibliográfica é descrita na seção 2. A metodologia empregada, bem como a descrição detalhada dos passos do desenvolvimento é apresentada na seção 3. Na seção 4 são discutidos os resultados alcançados, e por fim, na seção 5 é apresentada a conclusão do trabalho.

2. Fundamentação Teórica

Segundo [Tan et al. 2009], Mineração de dados (em inglês *Data Mining*) é um conjunto de técnicas que possibilita o aprendizado prático de padrões a partir de dados, possibilitando explicações sobre a natureza destes dados e previsões a partir dos padrões encontrados. A partir dos padrões descobertos, têm-se condições de gerar conhecimento útil para um processo de tomada de decisão. Trata-se, portanto, da aplicação de técnicas, im-

³<https://www.in.gov.br/en/web/dou/-/portaria-n-343-de-17-de-marco-de-2020-248564376>

⁴cs.waikato.ac.nz/ml/weka/

plementadas por meio de algoritmos computacionais, capazes de receber como entrada, um conjunto de fatos ocorridos no mundo real e devolver um padrão de comportamento como saída. Este padrão pode ser expresso, por exemplo, como uma regra de associação, que busca encontrar padrões frequentes entre conjuntos de dados para realizar previsões.

Como exemplo, suponha que instituições de ensino com milhares de de discentes resolva fazer uma pesquisa de satisfação com a comunidade acadêmica. Nessa pesquisa aborda-se qualidade das aulas dos professores, as condições das salas de aula, a alimentação da cantina e tenha como resultado uma planilha em excel com cem mil respostas. Para se analisar todos esses dados seria necessário selecionar uma equipe que se disporia a ler essa planilha e interpretar esses dados. Com a mineração de dados é possível encontrar, de maneira mais rápida e precisa, padrões que, talvez, a olhos humanos não seriam perceptíveis.

A mineração de dados possui diversas tarefas. Conforme [Camilo and Silva 2009], as mais comuns são:

- **Análise de Agrupamento**, também conhecida como *Cluster* ou clusterização, é uma técnica utilizada para encontrar a existência de diferentes grupos dentro de um ou mais conjuntos de dados e, em caso de sua existência, determinar quais são esses grupos. No artigo de [Fonseca and Beltrame 2010] pode-se ver aplicações práticas dos algoritmos *K-means* e *Bisecting K-means*.
Suponha que uma escola deseje entender melhor seus alunos e oferecer intervenções personalizadas para melhorar o desempenho acadêmico. Utilizando técnicas de clusterização, como o algoritmo K-means, esses dados podem ser agrupados em diferentes grupos: um grupo de alunos com notas consistentemente altas em todas as disciplinas, alta frequência escolar e participação ativa na escola e outro cluster com alunos com notas variáveis, frequência escolar irregular e comportamento inadequado em sala de aula. Com essa segmentação, a escola pode direcionar recursos e intervenções de maneira mais eficaz, oferecendo suporte diferente para cada grupo de alunos com base em suas necessidades específicas.
- **Regras de Associação** representa um padrão de relacionamento entre os itens de dados do domínio da aplicação que ocorrem com determinada frequência na base de dados [Targa 2002]. O algoritmo Apriori, destaque da regra de associação, trabalha com um número grande de atributos, gerando várias alternativas combinatórias entre eles, para confirmar o potencial do algoritmo como instrumento de gestão de Ciência e Tecnologia (C&T), através de regras de associação consistentes e nichos específicos descobertos [Romão et al. 1999].
Um exemplo muito utilizado para demonstrar a regra de associação é uma análise realizada pela rede varejista norte-americana Walmart, através do uso de inteligência artificial, onde se encontrou uma relação entre a venda de cervejas e fraldas. O Walmart utilizou essa informação para reorganizar a localidade de seus produtos, gerando assim um aumento de 30% nas vendas de ambos.
- **Sumarização** determina uma descrição compacta, um resumo, para um dado subconjunto. É frequentemente utilizada na análise exploratória de dados com

geração automatizada de relatórios, sendo responsável pela descrição de um conjunto de dados. Este método é aplicado nos agrupamentos obtidos através da clusterização, sendo a Lógica Indutiva e Algoritmos Genéticos exemplos que podem implementar a sumarização. [Coradine et al. 2011] ressalta que a sumarização não é usada para a resolução de problemas, mas possibilita identificar características no conjunto de dados que possa estar contaminadas por ruídos, que interfiram no processo de análise, ou redundantes, gerando uma tendência errônea à análise.

Um exemplo de aplicação deste seria utilizar para descobrir o perfil de um cliente de assinatura de uma revista e direcionar a oferta desta revista para o público alvo correto.

- A classificação associa ou classifica um item a uma ou várias classes categóricas pré-definidas. O objetivo desta técnica é realizar comparações entre itens e categorias de forma a encontrar semelhanças ou diferenças entre tais itens e classificar uma nova observação a uma classe já rotulada. Dentre os métodos de classificação, o trabalho de [Pereira Junior 2020] demonstra vários algoritmos como a árvore de decisão (J48), JRip, PART, K-NN e o *Random Forest*. A Árvore de decisão, também conhecida com *Random Tree*, é uma das técnicas utilizadas na tarefa de Classificação. Essa técnica organiza o conjunto de dados de treinamento em nós e folhas, unidos por arestas para navegar através dos nós.

Na Figura 1 pode-se ver um exemplo de árvore de decisão onde verifica a possibilidade de ir a praia. Se estiver com sol, então analisar a próxima condição que é analisar se há vento. Se não houver sol, a ida a praia já é descartada e nenhuma outra condição é analisada.



Figura 1. Exemplo árvore de decisão

Fonte: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>

Na Figura 2 pode-se visualizar a estrutura por trás desta mesma árvore: o nó raiz seria verificar se há sol, os nós intermediários são condições com valores para atributos específicos, neste caso, verificar se há vento e os nós folhas representam as classes que são a decisão final de ir ou não a praia.

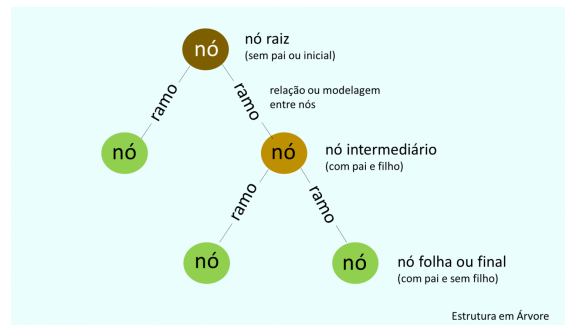


Figura 2. Estrutura árvore de decisão

Fontes:

<https://colaborae.com.br/blog/2023/07/19/arvore-de-decisao/>

Diversos trabalhos de pesquisa tem abordado técnicas de mineração de dados. [Gonçalves et al. 2018] utilizou três algoritmos de classificação, Naive Bayes, Support Vector Machine e J48, para prever a evasão dos estudantes de graduação do Instituto Federal do Maranhão. Com o algoritmo do J48 chegou a 98,08% de previsão, permitindo assim que a instituição direcione seus esforços no combate a evasão ao público correto. Como por exemplo, tentar auxiliar o possível aluno com alta probabilidade de evasão antes que isso aconteça, aumentando assim o número de estudantes que se formam. No trabalho de [de Souza and dos Santos 2021] foram utilizadas técnicas de Aprendizagem Profunda (AP) e Aprendizagem de Máquina (AM) com o objetivo de realizar a previsão do desempenho de alunos. Neste trabalho foi constatado que os modelos gerados a partir dos algoritmos tradicionais de AM têm um bom desempenho, mas inferior ao modelo AP que teve uma acurácia de 94%. Adicionalmente, foi identificado que atributos relacionados às atividades escolares são mais preditores para o desempenho dos alunos do que os dados de características demográficas e socioeconômicas.

O trabalho de [BELENKE DOS SANTOS 2021] leva em consideração parte do período que envolve a Covid 19. Este projeto buscou prever a potencial evasão de estudantes do ensino superior do Instituto Federal de Santa Catarina. Para tal, foram utilizadas as técnicas de Árvore de Decisão e Redes Neurais. A técnica de Árvore de Decisão apresentou um desempenho melhor, onde a acurácia alcançou 84%, correspondendo a 87% de precisão na detecção de evasão, enquanto que Redes Neurais chegaram a 82% de acurácia com 78% de precisão.

Até o momento da escrita deste trabalho, não foram identificados na literatura estudos que busquem comparar notas de alunos no ensino superior utilizando árvores de decisão, conforme proposto neste documento. Este estudo propõe uma análise comportamental das notas dos alunos do IFMG - campus Sabará, considerando a implementação do Ensino Remoto Emergencial (ERE), e realiza um comparativo das notas em algumas disciplinas relevantes dos cursos superiores, abrangendo períodos antes, durante e após a pandemia.

3. Metodologia

A presente seção demonstra a metodologia usada neste trabalho, sendo ilustrada cada etapa de desenvolvimento na Figura 3. Inicialmente os dados foram retirados do sis-

tema acadêmico do IFMG - campus Sabará e registrados em planilhas do formato Excel. As planilhas foram importadas para uma única tabela no SQL Server. A partir dos dados armazenados em somente um local, criou-se filtros e consultas SQL para facilitar a visualização da massa de dados. O SQL Server foi escolhido por ser uma ferramenta própria para armazenamento de dados e geração de consultas dos dados armazenados. A partir dos resultados gerados com consultas SQL, utilizou-se a ferramenta da Weka para geração de Árvores de Decisão e posterior análise dos dados.



Figura 3. Etapas de desenvolvimento
Fonte: De autoria própria

3.1. Pré-processamento da Base de Dados

Nesta subseção, aprofunda-se a explicação sobre como os dados foram tratados. Esta etapa foi de suma importância, pois tornou-se evidente que analisar os dados em sua forma bruta, repletos de informações desnecessárias e desprovidos de um padrão específico, seria uma tarefa consideravelmente mais complexa para qualquer algoritmo utilizado. O processo de pré-processamento dos dados, portanto, desempenhou um papel importante na preparação adequada dos dados para análise.

Este trabalho foi desenvolvido a partir de uma base de dados fornecida pelo IFMG campus Sabará, recebidos em planilhas excel separadas por curso. Cada planilha conteve dados a respeito das notas, disciplinas, código das disciplinas, código de status do aluno, número de créditos de tais disciplinas, carga horária e outros dados.

Dentre os dados recebidos, algumas informações foram excluídas da análise por não serem consideradas relevantes ao trabalho, são elas: código do status do aluno, número de créditos, carga horária, status do aluno, conceito, número de faltas e id da turma. Das vinte colunas da tabela original foram selecionadas apenas seis, são elas: código da disciplina, nome da disciplina, código do curso, código do semestre, nota final e matrícula. Depois que os dados foram selecionados, realizou-se consultas para o entendimento e particularidades de cada curso, conforme ilustrado na Figura 4. O filtro apresentado na Figura 4 mostra o curso, a quantidade de disciplinas cursadas e a quantidade de alunos matriculados nestes cursos e disciplinas, no período analisado. Com as funções do próprio SQL, como por exemplo o COUNT(), automaticamente os alunos foram agrupados em uma informação apenas, protegendo assim a integridade dos dados.

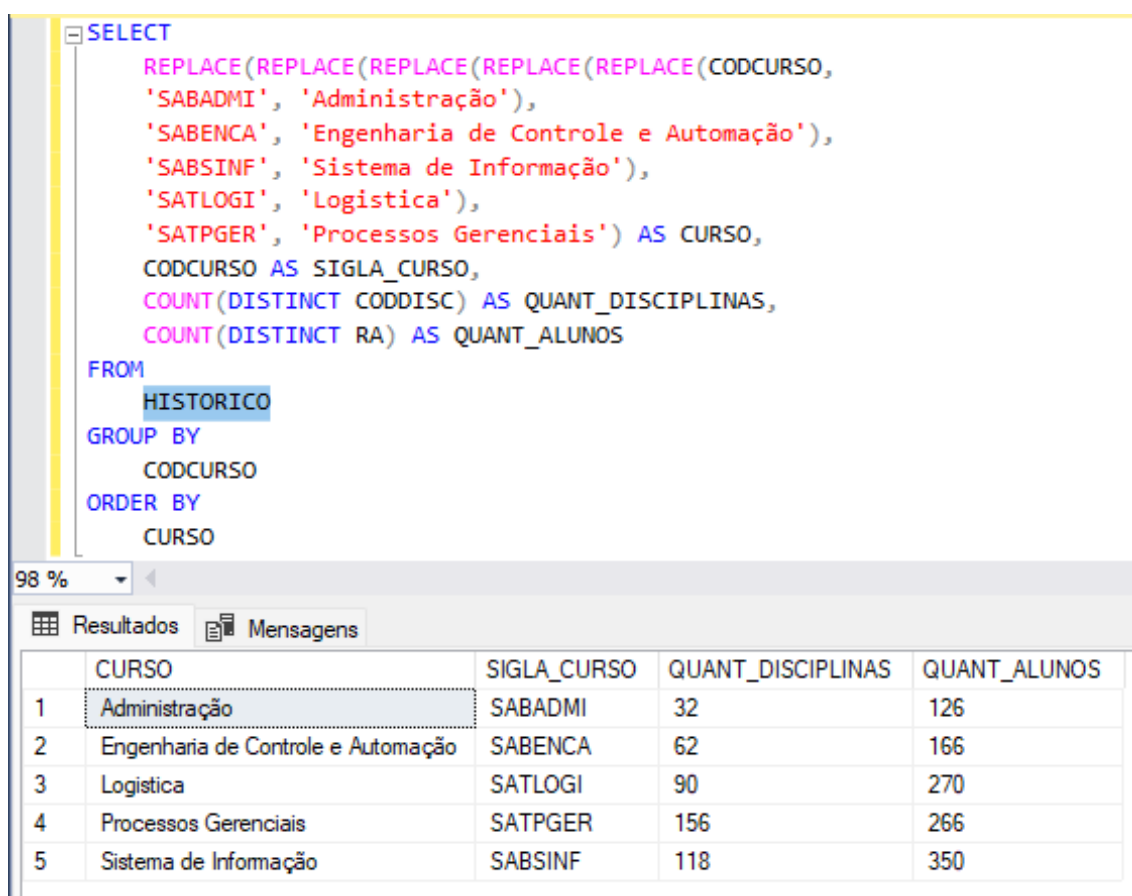


Figura 4. Resumo dos dados dos cursos analisados

Fonte: De autoria própria

Outra consulta realizada, demonstrada na Figura 5, mostra também os dados recebidos. Por exemplo em 2013.1 foram 6 disciplinas diferentes cursadas, em um total de 162 matrículas para 27 alunos diferentes e uma média de 6 disciplinas por aluno.

Para cada um dos cursos superiores existentes no campus Sabará, gerou-se uma planilha no formato excel. Estas planilhas foram extraídas através do Assistente de Importação e Exportação do SQL Server e salvas em uma única tabela do banco de dados criado para este trabalho. Todas as informações pessoais recebidas como nomes e código da matrícula foram mascarados para proteção dos dados dos alunos, professores e da instituição envolvida.

Para a utilização dos algoritmos da ferramenta Weka, a discretização foi realizada com o objetivo de agrupar muitas informações variadas, para facilitar o funcionamento/desempenho do algoritmo. Os períodos de aulas foram divididos em "Antes da Pandemia", "Durante a pandemia" e "Após a pandemia", conforme a Tabela 1.

As notas dos alunos também sofreram discretização. Elas foram alteradas para conceitos seguindo intervalos, sendo descritos na tabela 2.

Outro procedimento realizado diz respeito aos nomes das disciplinas. Na base de dados original foram encontradas variações nos nomes das disciplinas, mesmo se referindo à mesma matéria. Por exemplo, em um semestre, poderia constar "Programação

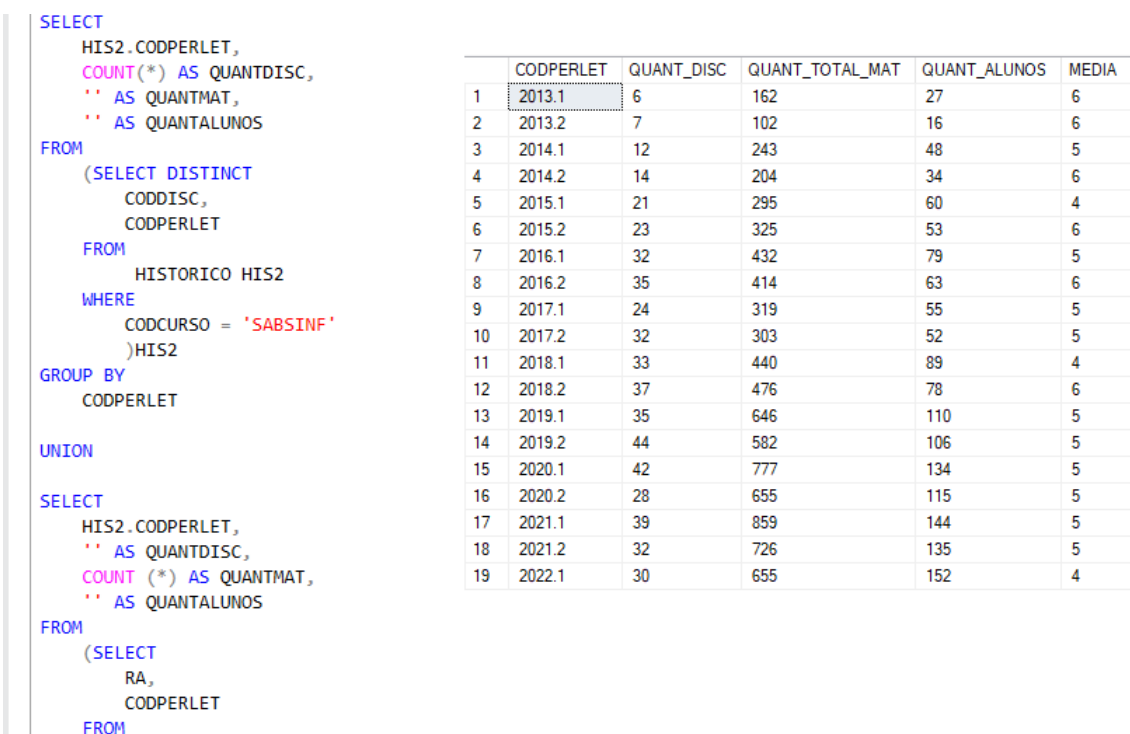


Figura 5. Demonstrativo de quantidade de alunos e disciplinas

Fonte: De autoria própria

Tabela 1. Discretização dos períodos de aula.

Título	Descrição	Referência
ANT_PAND	Antes da pandemia	Até 2º semestre 2020
DUR_PAND	Durante a pandemia	1º semestre 2020 até 1º semestre de 2021
POS_PAND	Após a pandemia	A partir do 2º semestre 2021

Fonte: De autoria própria

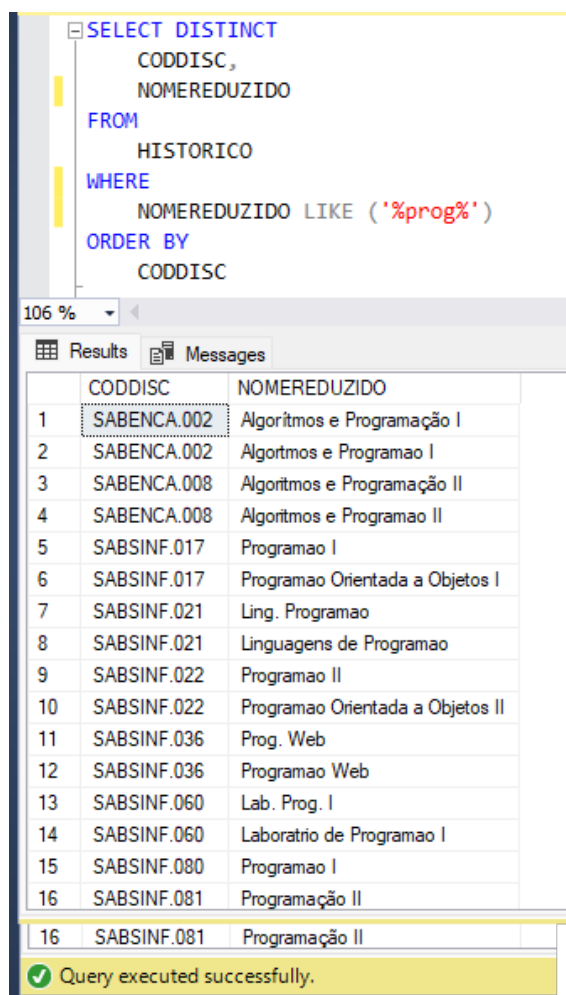
Tabela 2. Discretização das Notas.

Conceito	Intervalo
A	Maior que 90
B	Entre 80 e 90
C	Entre 70 e 80
D	Entre 60 e 70
E	Menor que 60

Fonte: De autoria própria

I”, enquanto em outro, ”Algoritmos e Programação I”. Para tratar isso, assim como para tratar a variação do código de disciplina de um curso para outro, foi realizado um filtro a parte, visto na Figura 6. Esse filtro foi utilizado para identificar todas as disciplinas, agrupar manualmente e só então utilizar esse filtro na consulta principal. Neste caso, por exemplo, para filtrar as informações relacionadas a Programação I, foram utilizados

os códigos da disciplina: SABENCA.002, SABSINF.017, e SABSINF.080. Este agrupamento foi necessário e aplicado para todas as disciplinas apresentadas neste trabalho, pois o algoritmo trataria as disciplinas de forma distintas, sendo que se trata da mesma informação.



```

SELECT DISTINCT
  CODDISC,
  NOMEREDUZIDO
FROM
  HISTORICO
WHERE
  NOMEREDUZIDO LIKE ('%prog%')
ORDER BY
  CODDISC

```

	CODDISC	NOMEREDUZIDO
1	SABENCA.002	Algoritmos e Programação I
2	SABENCA.002	Algoritmos e Programao I
3	SABENCA.008	Algoritmos e Programação II
4	SABENCA.008	Algoritmos e Programao II
5	SABSINF.017	Programao I
6	SABSINF.017	Programao Orientada a Objetos I
7	SABSINF.021	Ling. Programao
8	SABSINF.021	Linguagens de Programao
9	SABSINF.022	Programao II
10	SABSINF.022	Programao Orientada a Objetos II
11	SABSINF.036	Prog. Web
12	SABSINF.036	Programao Web
13	SABSINF.060	Lab. Prog. I
14	SABSINF.060	Laboratrio de Programao I
15	SABSINF.080	Programao I
16	SABSINF.081	Programação II
16	SABSINF.081	Programação II

Query executed successfully.

Figura 6. Agrupamento de código de disciplina
Fonte: De autoria própria

3.2. A Ferramenta Weka

A Ferramenta *Waikato Environment for Knowledge Analysis* - Weka ⁵ - é um software livre, do tipo *open source*, desenvolvido em Java pela Universidade Waikato, Nova Zelândia, dentro das especificações da GPL (General Public License).

Essa ferramenta proporciona um conjunto poderoso de algoritmos de aprendizado de máquina. Além disso, também inclui estatísticas básicas e ferramentas de visualização de gráficos, assim como ferramentas de pré-processamento, classificação e clusterização, todas disponíveis através de uma interface simples e amigável [Markov and Russell 2006].

⁵<https://www.cs.waikato.ac.nz/ml/weka/>

Muito utilizada no meio acadêmico, como nos trabalhos de [Santos and Pereira 2020] e [Pinto 2023], a Weka trabalha preferencialmente com arquivos no formato ARFF e CSV para as tarefas de mineração, porém, é preciso inserir os atributos de cada instância. A Figura 7 é um exemplo de arquivo no formato ARFF com dados acadêmicos.

```
1 @relation BancoDeDados
2 @attribute Curso { SABSINF }
3 @attribute Epoca { ANT_PAND, DUR_PAND, POS_PAND }
4 @attribute Nota { A, B, C, D, E }
5 @data
6 SABSINF,ANT_PAND,D
7 SABSINF,DUR_PAND,E
8 SABSINF,POS_PAND,E
9 SABSINF,ANT_PAND,E
10 SABSINF,ANT_PAND,E
11 SABSINF,DUR_PAND,B
12 SABSINF,DUR_PAND,B
13 SABSINF,ANT_PAND,B
14 SABSINF,ANT_PAND,B
15 SABSINF,POS_PAND,C
16 SABSINF,DUR_PAND,B
17 SABSINF,ANT_PAND,E
18 SABSINF,POS_PAND,C
19 SABSINF,DUR_PAND,A
20 SABSINF,ANT_PAND,B
```

Figura 7. Exemplo de Arquivo no formato ARFF com dados acadêmicos.

Fonte: De autoria própria

O predicado @relation na Figura 7 representa a identificação do arquivo. As linhas iniciadas por @attribute definem os atributos do arquivo, especificando o seu nome e o seu tipo. Os tipos que os dados podem assumir são numérico, booleano, string, data ou especificação nominal. As linhas abaixo de @data representam itens do conjunto de dados com os valores dos atributos separados por vírgula, sempre seguindo a mesma ordem da declaração dos atributos.

4. Análises dos Resultados

A presente seção apresenta os principais resultados encontrados durante o desenvolvimento do trabalho. Inicialmente descreve-se análises iniciais realizadas através de consultas sql para agrupamento, contagem e ordenação dos dados. Em seguida, são apresentadas as análises realizadas a partir do algoritmo Weka, assim como os resultados encontrados.

4.1. Análises inicial dos dados

Após limpeza e tratamento dos dados no SQL, foram realizadas consultas utilizando as funções para agrupamento de dados, contagem, junção dos resultados das consultas e ordenação das mesmas. A partir dos resultados dessas consultas, foram criados gráficos

utilizando bibliotecas da linguagem Python para proporcionar uma visualização mais clara e detalhada do desempenho dos alunos nos diferentes cursos de Ensino Superior do IF de Sabará, conforme apresentado na Figura 8.

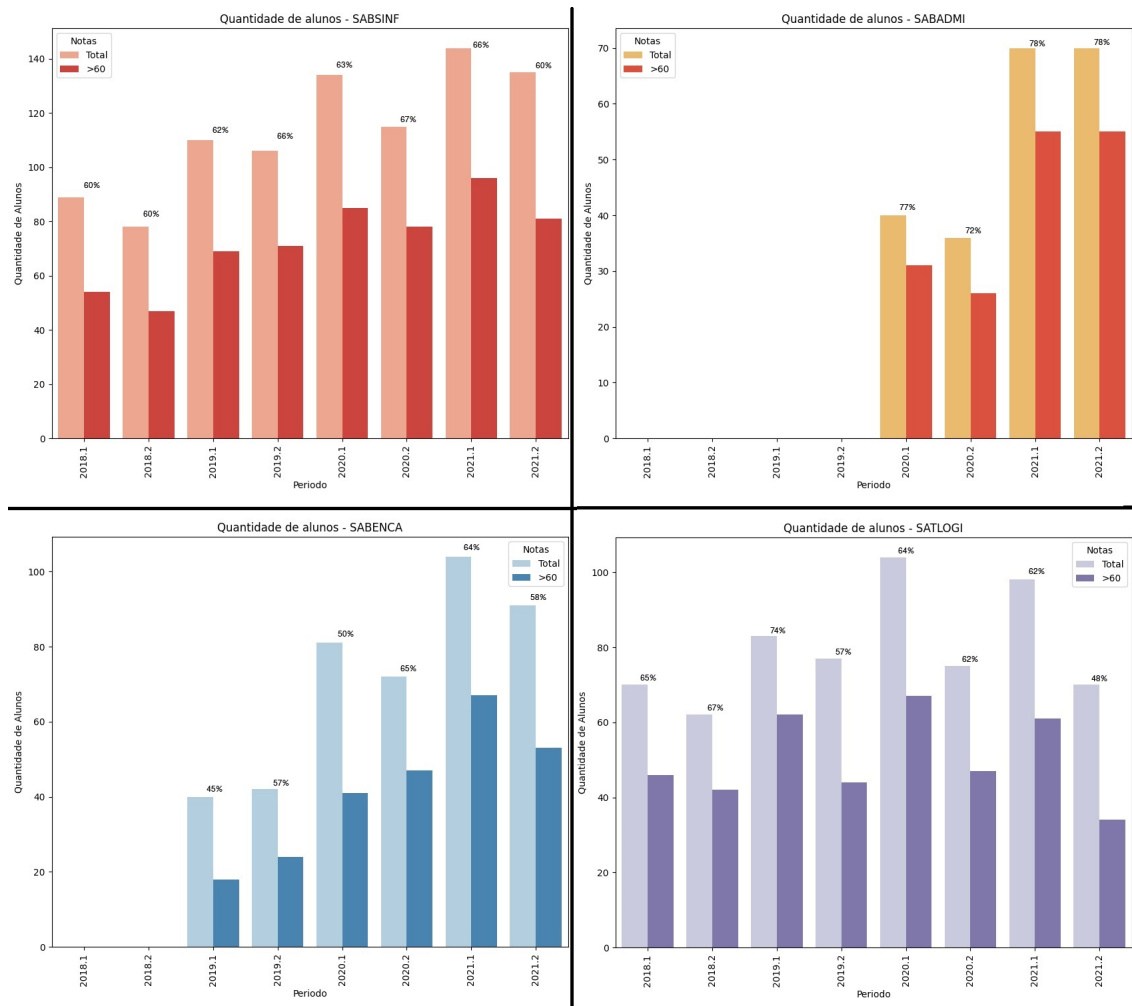


Figura 8. Comparação dos cursos

Fonte: De autoria própria

Uma das premissas deste trabalho foi baseada na hipótese de que durante o período da pandemia houve um aumento na quantidade de disciplinas que os alunos se matricularam. Durante o ERE todos os pré-requisitos de disciplinas foram quebrados e os semestres divididos em módulos. As disciplinas durante o ERE eram executadas de forma assíncrona. Assim, os alunos podiam se matricular em várias disciplinas ao mesmo tempo, pois não havia mais conflito de horários entre as disciplinas, o que não acontece no presencial.

A Figura 8 apresenta indicativos dessa afirmativa e mostra uma comparação entre a quantidade total de alunos, representada no eixo x, por cada período, representado no eixo y. As colunas representam a quantidade de alunos com "notas maior que 0", ou seja, a quantidade total de alunos matriculados no curso, em comparação com a quantidade de alunos com nota maior que 60, na coluna ao lado, sendo tais colunas identificadas na

legenda. As porcentagens acima das colunas representam a porcentagem total de aluno com nota maior que 60 em cada semestre, ou seja, quantidade de alunos aprovados por semestre. O curso de engenharia teve sua abertura em 2019 e o de administração em 2020, por isso não tem alunos matriculados nos anos anteriores. Essa análise evidencia uma mudança significativa no comportamento de matrícula dos alunos durante o período de implementação do ERE, refletindo uma maior flexibilidade no processo de seleção de disciplinas e possivelmente uma resposta às novas condições impostas pela pandemia.

Seguindo a análise das notas nos períodos relacionados a pandemia, observa-se também as informações representadas nas tabelas 3, 4 e 5, onde pode-se verificar de forma mais abrangente a quantidade de alunos que tirou determinado conceito em cada período.

Tabela 3. Conceitos Sistema de Informação por período.

Conceito	ANT PAND	DUR PAND	POS PAND
A	264	641	141
B	336	519	151
C	339	320	119
D	315	267	97
E	438	507	873

Fonte: De autoria própria

Na tabela 3, que representa o curso de Sistema de Informação, pode-se ver que antes da pandemia 264 alunos tiraram nota A, ou seja, acima de 90. Já durante a pandemia, a quantidade de alunos aumenta significativamente, sendo 641 alunos com conceito A, quantidade essa que cai novamente pós pandemia, com 141 alunos com este conceito. Essas análises detalhadas nos permitem compreender melhor o impacto das mudanças nos métodos de ensino durante a pandemia nos resultados acadêmicos dos alunos.

O mesmo padrão de aumento de notas durante a pandemia pode ser observado nos cursos de Engenharia e Controle de Automação, descrito na Tabela 4, e no curso de Logística, Tabela 5, onde a quantidade de alunos que tiraram conceito A também aumenta no período durante a pandemia, no curso de Engenharia aumentando de 56 alunos para 470 e no curso de Logística aumentando de 110 alunos para 327.

Tabela 4. Conceitos Engenharia e Controle de Automação.

Conceito	ANT PAND	DUR PAND	POS PAND
A	56	470	122
B	47	294	111
C	82	182	69
D	114	122	52
E	195	446	870

Fonte: De autoria própria

Tabela 5. Conceitos de Logística por período.

Conceito	ANT PAND	DUR PAND	POS PAND
A	110	327	32
B	247	256	61
C	250	204	43
D	191	195	49
E	303	450	632

Fonte: De autoria própria

4.2. Análises dos dados com a Ferramenta Weka

A presente subseção contém as análises realizadas por meio da plataforma Weka. Esta contém vários algoritmos para análise e classificação dos dados, dentre eles encontramos o *Random Tree* e o *Random Forest*. Devido a grande semelhança entre estes algoritmos, é importante ressaltar a diferença entre eles, sendo que *Random Forest* é um método que combina múltiplas *Random Trees* em um grande classificador, usando ainda mais randomização. O algoritmo *Random Forest*, assim como outros foram testados empiricamente, mas os resultados encontrados não foram satisfatórios quanto o *Random Tree*.

O *Random Tree* foi escolhido devido por ser conhecido pela capacidade de agrupamento dos dados. Ele cria aleatoriamente várias árvores de decisão e as combina para ter maior precisão. Após a execução desse algoritmo, aplicado a um arquivo ARFF, explicado na seção 3.2, uma árvore de decisão como da Figura 9 é gerada, neste caso, uma para cada disciplina.

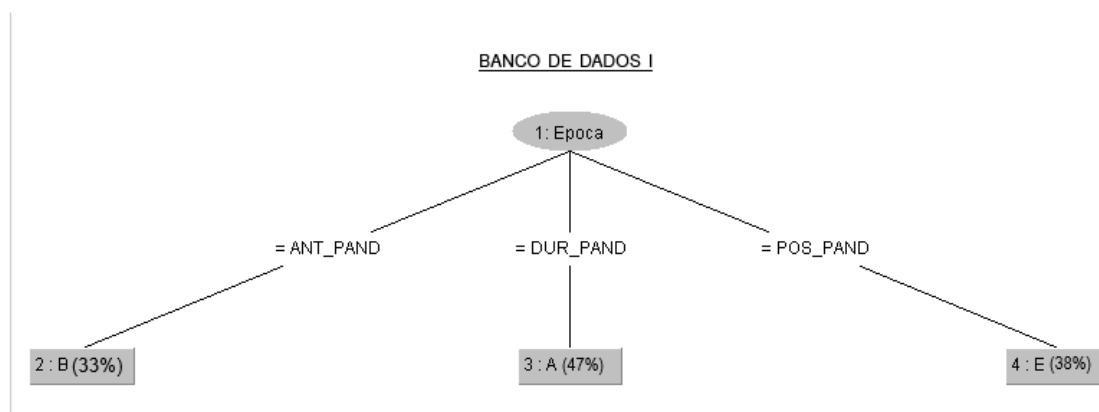


Figura 9. Árvore de Decisão - Banco de Dados I

Fonte: De autoria própria

Nos nós folha da árvore de decisão representada na Figura 9, a porcentagem ao lado do conceito representa a frequência que tal conceito foi encontrado naquele período. Para a disciplina de Banco de Dados I, no curso de Sistema de Informação, 33% dos alunos tiveram nota B antes da pandemia, 47% tiveram nota A durante a pandemia e 38% nota E após a pandemia. Tal fato mostra indícios do princípio descrito nesse trabalho, de que durante a pandemia houve um aumento médio nas notas dos alunos.

Na Figura 10 vê-se outro exemplo da aplicação da técnica de classificação da mineração de dados para uma disciplina específica, nesse caso, Programação I. O nó raiz é o período em que essas notas foram tiradas pelos alunos, divididas entre antes, durante e depois da pandemia, o nó intermediário é o curso onde a disciplina foi ministrada e os nós folhas classificam as notas mais frequentes.

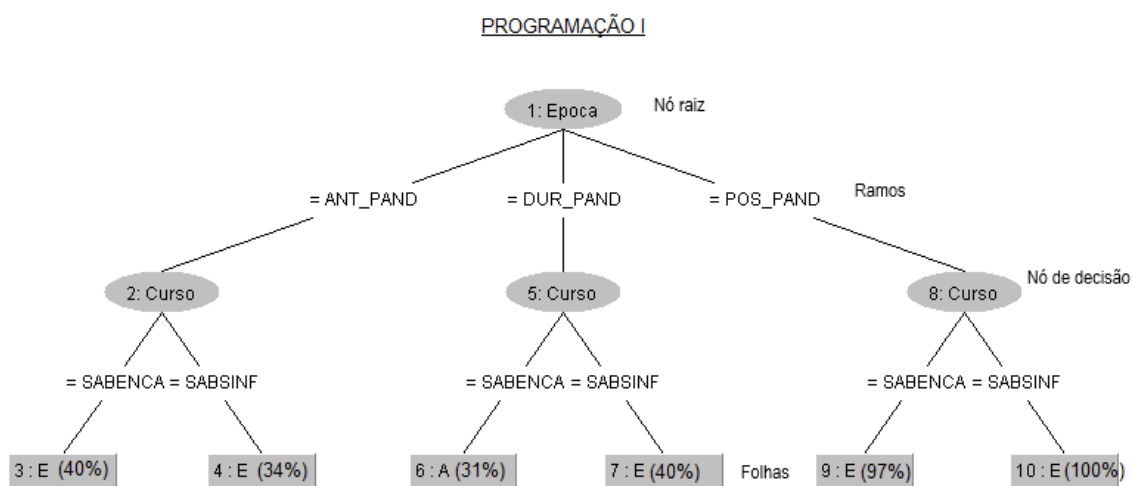


Figura 10. Árvore de Decisão - Programação I

Fonte: De autoria própria

A partir da árvore de decisão é possível derivar regras do tipo SE-ENTÃO. Considerando a árvore de decisão da Figura 10 é possível extrair várias regras, como por exemplo:

- SE as aulas da disciplina de Programação I foram lecionadas antes da pandemia para o curso de Engenharia de Controle e Automação ENTÃO o conceito médio das notas tiradas pelos alunos é E.
- SE as aulas da disciplina de Programação I foram lecionadas durante a pandemia para o curso de Engenharia de Controle e Automação ENTÃO o conceito médio das notas tiradas pelos alunos é A.

As Tabelas 6, 7 e 8 apresentam um resumo dos resultados significativos ao se executar o algoritmo *Random Tree*. Na Tabela 6 destacam-se as disciplinas que seguem o mesmo padrão visto na Figura 9. Nesta tabela foi descrito novamente a informação de Banco de Dados I, para uma representação diferente da informação, assim como das demais disciplinas analisadas, onde pode-se ver o mesmo padrão das notas médias aumentando durante a pandemia. Dentre estas disciplinas, destaca-se a de Cálculo II onde é possível perceber também uma variação nas notas durante o período de pandemia, com o conceito médio subindo para B e A, nos cursos de Engenharia e Sistema de Informações respectivamente, e depois da pandemia, se mantendo em B no curso de Engenharia, mas caindo novamente para E no curso de Sistema de Informações. Em Banco de dados II e Redes II, fica nítido também a queda das notas após a pandemia, sendo 100% dos alunos da base de dados com conceito E, ou seja, foram reprovados ou desistiram da disciplina.

Tabela 6. Disciplinas com aumento médio de nota durante a pandemia.

Disciplina	Curso	ANT PAND	DUR PAND	POS PAND
Banco de Dados I	SABSINF	B(33%)	A(47%)	E(38%)
Banco de Dados II	SABSINF	D(29%)	A(35%)	E(100%)
Calculo II	SABSINF	E(40%)	A(37%)	E(57%)
Matemática Discreta	SABSINF	C(40%)	B(32%)	E(100%)
Metodologia de Pesquisa	SABSINF	E(40%)	A(31%)	B(97%)
Redes II	SABSINF	E(55%)	A(46%)	E(100%)
Estatística	SABSINF	E(38%)	A(50%)	A(43%)
Calculo II	SABENCA	E(49%)	B(30%)	B(38%)
Programação I	SABENCA	E(40%)	A(31%)	B(97%)
Metodologia de Pesquisa	SABENCA	E(31%)	A(35%)	E(39%)
Estatística	SABENCA	E(0%)	A(57%)	C(45%)
Programação II	SABENCA	E(56%)	A(26%)	E(67%)
Estatística Aplicada	SATLOGI	E(55%)	D(34%)	E(44%)
Introdução a Logistica	SATLOGI	C(38%)	B(29%)	E(48%)
Economia	SATLOGI	D(33%)	A(40%)	E(96%)
Marketing	SATLOGI	B(64%)	A(45%)	A(46%)

Fonte: De autoria própria

Mesmo que as circunstâncias do ERE tenham influenciado no aumento de notas dos alunos, elas não determinaram isso. Na Tabela 7 pode-se ver disciplinas que mantiveram as notas durante a pandemia e na Tabela 8, vê-se as disciplinas que as notas diminuíram durante a pandemia, sendo estas uma quantidade bem menor de disciplinas tanto em relação a quantidade de disciplinas em que a nota se manteve, quanto em relação a quantidade de disciplinas que houve aumento das notas na pandemia.

Nem todos os alunos do ensino superior tiveram experiências positivas com a adoção do ERE, como demonstra também [Nunes 2021], que relata que muitos estudantes relacionam a falta de tempo como razão para a desistência. Também como visto no trabalho de [Simon et al. 2021], que encontrou resultados que evidenciaram a importância do fortalecimento do contato institucional direto com os estudantes para a redução dos índices de evasão.

Tabela 7. Disciplinas que a nota se manteve durante a pandemia.

Disciplina	Curso	ANT PAND	DUR PAND	POS PAND
Calculo I	SABSINF	E(50%)	E(32%)	E(36%)
Redes I	SABSINF	B(33%)	B(44%)	E(38%)
Calculo I	SABENCA	E(50%)	E(32%)	E(36%)
Empreendedorismo e Inovação	SATLOGI	E(25%)	E(43%)	E(52%)
Ética e Cidadania	SATLOGI	A(72%)	A(44%)	E(100%)
Fundamento de Administração	SATLOGI	A(55%)	A(39%)	E(100%)
Gestão de processos	SATLOGI	A(31%)	A(38%)	E(100%)

Fonte: De autoria própria

Tabela 8. Disciplinas que a nota diminuiu durante a pandemia.

Disciplina	Curso	ANT PAND	DUR PAND	POS PAND
Empreendedorismo e Inovação	SABSINF	A(38%)	B(50%)	A(59%)
Programação II	SABSINF	B(26%)	E(54%)	E(57%)
Redes I	SABENCA	A(67%)	E(59%)	E(85%)

Fonte: De autoria própria

Mesmo com essa evasão, é possível elencar algumas razões para o aumento de notas encontrado nesse trabalho: ensino online com horário de estudos flexível e, para alguns alunos, diminuindo a pressão da sala de aula presencial; o ensino online que proporciona o acesso de vários recursos didáticos, como vídeos e materiais disponibilizados na internet; a falta de preparo e adaptação dos professores e instituições na elaboração das avaliações remotas de ensino-aprendizagem. Essas razões apontam para a complexidade do ambiente de ensino durante a pandemia e ressaltam a necessidade de uma abordagem multifacetada para compreender os impactos das mudanças educacionais.

5. Conclusão

A pandemia afetou várias áreas do nosso país e a área da educação não foi diferente. O IFMG *campus* Sabará adotou o ERE com estratégia de ensino durante a pandemia. O presente estudo trabalhou com dados retirados do sistema acadêmico de cursos superiores do IFMG *campus* Sabará. Esses dados foram tratados, discretizados, analisados e entendidos para depois serem submetidos aos algoritmos da ferramenta Weka.

A análise dos dados mostrou que os alunos se matricularam em um número maior de disciplinas. Isso pode ser explicado pelo fato da retirada de pré-requisito e choque de horários. Os resultados obtidos com o algoritmo *Random Tree* da ferramenta Weka foi que os alunos, de uma forma geral, conseguiram desempenho superior mesmo cursando mais créditos. Ao analisar as notas pós pandemia, percebe-se que esse aumento de notas não reflete necessariamente um aprendizado real, visto que muitas vezes as notas caíram novamente no retorno ao presencial. Esse aumento das notas durante o ERE pode ser atribuído a uma combinação de fatores, incluindo a flexibilidade de horários, o acesso a recursos adicionais de aprendizagem e a falta de preparo e adaptação dos professores na elaboração das avaliações remotas também pode ter influenciado esse aumento das notas.

Como trabalhos futuros, planejamos acrescentar dados socioeconômicos, os quais tem um grande impacto no desempenho dos estudantes. Além da análise de evasão dos alunos dos ensino superior no IFMG *campus* Sabará durante a pandemia. Quantos que se matricularam em disciplinas, mas não concluíram e quantos matricularam em menos disciplinas são alguns dos fatores que também poderiam ser analisados.

Espera-se que estes resultados sirvam como suporte a decisão no contexto do ensino dos Institutos Federais, e que se possa aproveitar o conhecimento gerado para eventuais adversidades no futuro.

Referências

- Bardagi, M. P. and Hutz, C. S. (2012). Rotina acadêmica e relação com colegas e professores.
- BELENKE DOS SANTOS, J. C. (2021). Usando mineração de dados para predição da evasão escolar.
- Camilo, C. O. and Silva, J. C. d. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFG)*, 1(1):1–29.
- Casanova, J. R., Araújo, A. M., and Almeida, L. S. (2020). Dificuldades na adaptação acadêmica dos estudantes do 1º ano do ensino superior.
- Castilho, M. L. and da Silva, C. N. N. (2020). A covid-19 e a educação profissional e tecnológica: um panorama das ações de acompanhamento e enfrentamento da pandemia nos institutos federais. *Revista Nova Paideia-Revista Interdisciplinar em Educação e Pesquisa*, 2(3):18–34.
- Coradine, L. C., Lopes, R. V. V., and Maciel, A. F. (2011). Mineração de dados: Uma introdução. *Journal of the Brazilian Neural Network Society*, 9:168–184.
- de Souza, V. F. and dos Santos, T. C. B. (2021). Processo de mineração de dados educacionais aplicado na previsão do desempenho de alunos: Uma comparação entre as técnicas de aprendizagem de máquina e aprendizagem profunda. *Revista Brasileira de Informática na Educação*, 29:519–546.
- Fonseca, F. C. S. and Beltrame, W. A. R. (2010). Aplicações práticas dos algoritmos de clusterização k-means e bisecting k-means. *UFES, Vitória*.
- Gonçalves, T. C., da Silva, J. C., and Cortes, O. A. C. (2018). Técnicas de mineração de dados: um estudo de caso da evasão no ensino superior do instituto federal do maranhão. *Revista Brasileira de Computação Aplicada*, 10(3):11–20.
- Markov, Z. and Russell, I. (2006). An introduction to the weka data mining system. *ACM SIGCSE Bulletin*, 38(3):367–368.
- Nunes, R. C. (2021). Um olhar sobre a evasão de estudantes universitários durante os estudos remotos provocados pela pandemia do covid-19. *Research, Society and Development*, 10(3):e1410313022–e1410313022.
- Oliveira, E. A. M. and Silva, C. A. (2022). *Ensino Remoto Emergencial (ERE): múltiplas visões e vivências no ensino técnico e tecnológico em tempos de pandemia*, volume 1, chapter Os desafios da gestão pedagógica na vigência do Ensino Remoto Emergencial em 2020, pages 37–51. Editora Fi, Porto Alegre, 1 edition.
- Pereira Junior, L. (2020). Análise comparativa de algoritmos de aprendizagem de máquina: um estudo de caso na área educacional. B.S. thesis, Universidade Tecnológica Federal do Paraná.
- Pinto, B. d. R. (2023). Análise comparativa da eficiência de três modelos de previsão de séries temporais do software weka. B.S. thesis, Universidade Tecnológica Federal do Paraná.

- Romão, W., Niederauer, C. A., Martins, A., Tcholakian, A., Pacheco, R. C., and Barcia, R. M. (1999). Extração de regras de associação em c&t: O algoritmo apriori. *XIX Encontro Nacional em Engenharia de Produção*, 34:37–39.
- Santos, K. F. d. C. d. and Pereira, S. J. M. (2020). Análise dos processos de descoberta de conhecimentos com as ferramentas de mineração de dados: Orange e weka.
- Simon, L. W., Gotardo, R. C. d. C., Amorim, J. H., and Bautitz, I. R. (2021). A busca ativa como estratégia de contenção da evasão no ensino superior.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2009). *Introdução ao datamining: mineração de dados*. Ciência Moderna.
- Targa, C. (2002). *Mineração eficiente de regras de associação através da indexação de conjuntos candidatos*. PhD thesis, Dissertação de Mestrado, Universidade Federal Fluminense.