

Análise dos Impactos da Gestão do Tempo no Desempenho Acadêmico Através da Mineração de Dados Educacionais

Pricylla Santos Cavalcante do Nascimento, Rodrigo Lins Rodrigues

Resumo

Com os avanços tecnológicos, nasceram também novos desafios. Entre estes está a dificuldade de identificar fatores que corroboram com um bom desempenho acadêmico dos alunos de cursos EAD. Assim, este trabalho tem o objetivo de analisar os impactos da gestão do tempo no desempenho acadêmico dos discentes. Para isso, foram utilizadas a técnica *K-means* para a realização do agrupamento dos discentes em relação ao seu desempenho acadêmico, uma Rede Neural Artificial para classificar estes grupos com base nas variáveis de gestão do tempo, e o método SHAP para interpretar de forma eficiente os resultados dessa classificação. Para a construção desta pesquisa, utilizou-se dados de cursos EAD extraídos da plataforma *moodle* de uma universidade pública do estado de Pernambuco. Como conclusão, foi possível observar quais características da gestão do tempo impactam positivamente no desempenho acadêmico do aluno.

Palavras-Chave: *gestão do tempo; EAD; mineração de dados.*

Abstract

With technological advances, new challenges were born. Amongst them is the problem of identifying factors that corroborate with the good academic performance of students of distance learning courses. This work aims to analyze the impacts of time management on the academic performance of the student. For this, the K-means technique was used to group students in relation to their academic performance, a neural network was used to classify these groups according to the time management variables, and the SHAP method was used to interpret the classes obtained in an efficient way. The construction of this research uses data from distance learning courses extracted from the moodle platform of a public university in the state of Pernambuco. As a conclusion, it was possible to observe which characteristics of time management impact the student's academic performance positively.

1 Introdução

1.1 Contextualização

Os avanços da tecnologia permeiam por diversas áreas, entre essas a área de educação. Diante da evolução no desenvolvimento de artefatos tecnológicos para a educação, ambientes computacionais de ensino e conteúdos digitais passaram a ser vistos como instrumentos essenciais no processo de ensino (BAKER, et al, 2011).

Logo, com o crescimento e popularização da internet, tem-se elevado a adoção de ambientes online de aprendizagem em contextos distintos (RODRIGUES & SANTOS, 2020). Contudo, com a utilização desses processos surgiu a necessidade de usarmos técnicas para analisar os dados que são gerados por essas plataformas. Assim, surge a mineração de dados educacionais ou Educational Data Mining (EDM) que é definida por (BARNES, et al. , 2009) como uma técnica que analisa os dados educacionais com o foco em solucionar questões de pesquisa educacionais. Deste modo, esta técnica possibilita o conhecimento de fatores que melhorem a proposta educacional, além de prever o desempenho dos alunos e de fatores que influenciam no aprendizado (NASCIMENTO, 2018).

Através da técnica de mineração de dados, este trabalho busca analisar, tomando por base uma plataforma de Educação a Distância (EAD), o desempenho desses discentes em relação ao curso, tendo como foco a gestão do tempo. Aqui, entende-se gestão do tempo como algo que está relacionado com a gestão da rotina de um determinado trabalho, que tem como meta estabelecer, manter e melhorar padrões de qualidade e produtividade das atividades que serão executadas (MELO, FARIAS, 2015).

Sabendo que a mineração de dados educacionais é uma área de pesquisa que possui como foco o desenvolvimento de métodos para explorar conjuntos de dados coletados em ambientes educacionais (BARKER, et. al, 2011), este trabalho busca estudar dados que corroborem com a análise do desempenho do aluno e o impacto da gestão do tempo neste desempenho.

1.2 Descrição do Problema

Os cursos EAD estão ganhando espaço e, de acordo com (RODRIGUES & SANTOS, 2020), isso se justifica porque, diferente da modalidade de ensino presencial, a aprendizagem online permite uma maior flexibilidade em relação ao tempo e à localização física da sala de aula. Deste modo, torna-se mais fácil para os alunos que possuem pouco tempo disponível o acesso ao curso e ainda permite que o mesmo organize-se e planeje-se de

uma maneira mais flexível, já que as atividades são propostas de forma assíncrona. Contudo, esta flexibilidade traz consigo a necessidade de uma postura mais ativa para o planejamento e o gerenciamento no processo de ensino-aprendizagem (BROADBENT; POON, 2015) . Logo, o problema de pesquisa está relacionado com a postura que os discentes tomam, em relação a gestão do tempo, e os impactos desta postura no seu desempenho acadêmico.

1.3 Objetivo

Como a flexibilização das atividades interferem no planejamento e no gerenciamento do processo de ensino-aprendizagem dos discentes, este estudo tem por objetivo geral verificar qual o impacto da gestão do tempo no desempenho acadêmico dos educandos de cursos EAD. Como objetivo específico iremos analisar o desempenho dos estudantes; analisar as atividades que foram entregues; se essas atividades foram desenvolvidas dentro do período estipulado pelo professor; e participações em webquests.

1.4 Justificativa

A flexibilização dos cursos EAD cria oportunidades para os discentes que possuem o desejo de iniciar um curso, porém não conseguem atender as demandas de um curso presencial, como um tempo fixo de dedicação. Tendo como exemplo um estudo feito a partir do curso de Licenciatura em Matemática à distância pela UFPB, foi notado que as reprovações chegavam a mais de 60% no primeiro período e ela ia decrescendo ao longo dos semestres. Essas reprovações teriam como base duas variáveis correlacionadas: a evasão e a reprovação por nota (PINTO & MOREIRA, 2018). Mesmo essa flexibilização do tempo dando uma liberdade maior, nem sempre os estudantes estão preparados e adaptados para essa situação, e isso pode acarretar em um acúmulo de atividades que, conseqüentemente, impactará no seu desempenho acadêmico. Por isso, este trabalho visa identificar o impacto da gestão do tempo no desempenho do discente, para assim se fazer possível identificar quais aspectos da organização do tempo se vinculam com um bom desempenho.

1.5 Escopo negativo

Esta pesquisa irá utilizar a base de dados do Moodle, uma plataforma de Ambientes Virtuais de Aprendizagem. Este ambiente apresenta dados, ao longo de 7 anos, de 4 cursos (Administração, Biologia, Letras e Pedagogia). Por se tratar de cursos em EAD, não é possível monitorar as atividades acadêmicas fora do Ambiente Virtual de Aprendizagem.

2 Fundamentação Teórica

2.1 Área do Negócio

2.1.1 Educação a Distância

A educação a distância (ou EAD) surge como uma modalidade de ensino que possibilita diferentes formas de se encarar as realidades da sociedade, bem como de ensinar e aprender. Trazendo aspectos positivos ao contexto educacional, ela propicia a produção de conhecimento individual e coletivo, utilizando ambientes digitais e interativos para promovê-la (MARTINS, 2016).

Conforme a sociedade avança, novas tecnologias e técnicas são criadas e as antigas sofrem mudanças e melhorias. Por esta razão, é importante que profissionais se mantenham sempre atualizados e em constante busca por formas de continuar aprendendo.

Contudo, nem sempre é possível para as pessoas voltarem à sala de aula, seja por falta de tempo ou por outros fatores. Neste aspecto, a educação a distância se torna peça fundamental para auxiliar neste processo. Por ser uma forma de ensino que proporciona, a alunos que não podem, por qualquer razão, comparecer à escola diariamente, a oportunidade de continuar ou voltar a estudar, a EAD consegue inserir novamente o estudante no ambiente institucional (MARTINS, 2016).

O perfil do aluno que busca a educação a distância possui, geralmente, alguma das seguintes características: são adultos ou jovens aprendizes inseridos no mercado de trabalho, pessoas que residem em locais distantes de unidades de ensino, estudantes que não conseguiram aprovação em cursos regulares e/ou que não tenham tempo para se dedicar aos estudos no ensino presencial (MARTINS, 2016).

Por proporcionar um ensino de qualidade de forma flexível, o que permite ao aluno encaixá-lo em suas reais necessidades, a educação a distância é considerada um meio, uma ferramenta que permite a ampliação do acesso à escola, o atendimento ao adulto que deseja retornar à sala de aula e

possibilita o acesso e o uso de novas tecnologias de informação e comunicação (MARTINS, 2016).

2.1.2 Gestão do Tempo

A gestão do tempo consiste em um conjunto de técnicas utilizado para aprimorar a realização de tarefas, de modo que a produtividade e a eficiência sejam mantidas. Portanto, o gerenciamento do tempo nada mais é do que a adoção de uma série de processos e ferramentas que auxiliam no cumprimento de atividades e projetos dentro dos prazos determinados e que garantam que os resultados sejam satisfatórios (MELO, FARIAS, 2015).

Existem diversas formas de se gerir o tempo, considerando as pessoas envolvidas e as atividades a serem realizadas. Para que esta gestão seja feita de forma efetiva e satisfatória, é imprescindível haver um bom planejamento, definir as metas, prazos e prioridades dentro de um projeto e saber lidar com as situações imprevistas que possam vir a acontecer.

Apesar de essa preocupação ser mais comum no ambiente de trabalho, existe a necessidade de se gerir bem o tempo em todos os âmbitos da vida. Isso inclui, é claro, as atividades acadêmicas, objeto de pesquisa deste artigo.

Avaliando como os discentes organizaram seus acessos à plataforma online e como eles interagiram com as atividades lá postadas, espera-se conseguir traçar um paralelo entre uma boa gestão do tempo e um bom desempenho acadêmico.

2.2 Mineração de Dados

A Mineração de Dados (Data Mining) busca encontrar anomalias, padrões e correlações em grandes conjuntos de dados para que seja possível prever resultados a partir deles. Através de uma gama de técnicas, é possível usar as informações obtidas pela análise das bases de dados para aumentar a renda, reduzir custos, se antecipar a possíveis situações adversas, entre várias outras aplicabilidades (MARTINS, 2016).

Por meio de suas ferramentas e processos, a mineração de dados permite que dados ruidosos e repetidos sejam removidos das bases de dados, para que assim seja possível entender o que é relevante e o que não é. De posse destes dados valiosos, o pesquisador pode então fazer um bom uso desta informação para avaliar possíveis resultados e acelerar o ritmo de tomadas de decisão, por exemplo (MARTINS, 2016).

Na área da mineração de dados, encontram-se diversas técnicas e métodos, cada um podendo ser utilizado para diferentes fins (COSTA, et.al, 2021).

A modelagem descritiva dos dados busca por semelhanças compartilhadas ou agrupamento dos dados para determinar as razões para o sucesso ou o fracasso de determinada situação. Algumas técnicas comuns para realizá-la são o clustering, a detecção de anomalias e a separação dos dados em grupos de afinidade (COSTA, et.al, 2021).

A modelagem preditiva dos dados utiliza a análise das bases de dados para classificar eventos futuros ou estimar resultados desconhecidos. Usando as informações obtidas, é possível determinar, por exemplo, a probabilidade de uma situação acontecer. Dentre as muitas técnicas disponíveis, destacam-se a regressão e o uso de redes neurais e/ou árvores de decisão para fazer previsões (COSTA, et.al, 2021).

Por último, a modelagem prescritiva surge para analisar, filtrar e transformar dados não-estruturados de forma que seja possível incluí-los nos modelos preditivos, visando aumentar a precisão de suas previsões. Ela olha para as variáveis internas e externas, bem como quaisquer restrições, para recomendar um ou mais cursos de ação (COSTA, et.al, 2021).

2.2.1 Mineração de Dados Educacionais

A área de mineração de dados educacionais busca desenvolver ou adaptar métodos e algoritmos de mineração existentes, facilitando com isso a compreensão dos dados nos contextos educacionais, que são produzidos por estudantes e professores/tutores nos ambientes em que os mesmos interagem. Logo, tem por objetivo entender melhor o estudante no seu processo de aprendizagem, analisando a sua interação com o ambiente (SAS, 2021).

A seguir, são apresentadas e descritas as técnicas utilizadas pela mineração de dados educacionais conforme sua categorização e sub-área (BAKER, et al, 2011).

:

A primeira técnica é a predição, que possui como meta desenvolver modelos que fazem inferência sobre aspectos dos dados através da análise e associação dos diversos aspectos encontrados nos mesmos. Como sub-área, possui a classificação e a regressão (SAS, 2021).

A segunda técnica é a de agrupamento, que tem por objetivo dividir o conjunto de dados em grupos, em que os objetos agrupados são semelhantes entre si (SAS, 2021).

A terceira técnica é a mineração de relações, que consiste em derivar regras de conhecimento, referindo-se ao relacionamento entre objetos de um conjunto de dados com o objetivo de exibir características e tendências. Como sub-áreas, estão a mineração de regras de associação; mineração de correlações; mineração de padrões e sequências e mineração de causas (SAS, 2021).

A quarta técnica é a destilação de dados, utilizada para facilitar decisões humanas. Tem como foco apresentar os dados de uma maneira mais legível e visual para auxiliar na compreensão humana. Podendo, assim, apoiar decisões importantes baseadas nos dados coletados (SAS, 2021).

A quinta e última técnica são as descobertas com modelos, que partem de um modelo gerado através de um método de predição e, posteriormente, utiliza-se como componente em uma outra análise com técnicas de predição ou mineração de relações (SAS, 2021).

2.2.2 K-Means

O k-means clustering é uma técnica de agrupamento que consiste em dividir a base de dados em um número k de grupos (chamados clusters), em que k é um número especificado pelo analista. Ele classifica os objetos em diferentes grupos, de modo que os elementos dentro de um mesmo cluster possuem maior similaridade intra-classe, ou seja, eles são tão parecidos quanto possível. De forma análoga, os elementos que estão em grupos diferentes devem ter baixa similaridade intra-classe, o que significa que eles são pouco semelhantes entre si. No k-means, cada cluster é representado pelos seus centróides, que são os pontos médios de cada grupo (KASSAMBARA, 2017).

A ideia principal do k-means é definir uma quantidade de clusters de forma que a variação intra-cluster total seja mínima. Para isso, primeiramente o analista define em quantos grupos sua base de dados será dividida, o k . Após isso, o algoritmo seleciona aleatoriamente os centróides de cada cluster. Então, os elementos restantes são agrupados junto com seus centróides mais próximos, utilizando a distância Euclidiana para obtê-los. Por fim, depois que todos os elementos foram separados em seus clusters, o centróide é calculado novamente e atualizado. O agrupamento dos elementos em clusters e a atualização do centróide são realizados de forma iterativa até que os clusters parem de mudar, ou seja, até que os clusters da iteração atual sejam os mesmos da anterior (KASSAMBARA, 2017).

O algoritmo do k-means funciona de forma simples, mas uma pergunta fundamental a ser respondida é: como escolher o melhor número de clusters (k)? Para isto, existe uma técnica que utiliza a plotagem de um gráfico e suas possíveis quantidades de clusters e a observação de seu “cotovelo”, ou seja, ela busca encontrar uma curva no gráfico. O ponto em que esta curva está indicaria o melhor número de grupos a ser utilizado (KASSAMBARA, 2017).

Por fim, deseja-se visualizar os resultados após a execução do algoritmo do k-means. A melhor forma para tal é a plotagem de gráficos. O mais comum deles é o gráfico de dispersão, que mostra os clusters, a dispersão de seus elementos e a distância de um cluster para outro (KASSAMBARA, 2017). Além disso, outros gráficos podem ser utilizados para visualizar e validar os clusters encontrados, como o dendrograma e o gráfico do teste de silhueta.

2.2.3 Redes Neurais Artificiais

Uma Rede Neural Artificial (RNA) é uma estrutura complexa interligada por elementos de processamento simples, os chamados neurônios. Eles têm capacidade de realizar operações como cálculos em paralelo, para processar dados de forma mais eficiente e criar uma melhor representação do conhecimento extraído dos mesmos. A RNA ganhou popularidade após a introdução de algoritmos de treinamento, como o backpropagation, que permitem a realização de um treino posterior para melhorar os resultados do modelo (GRÜBLER, 2021). A RNA é uma abstração de uma rede neural biológica. Seu objetivo é utilizar o sistema nervoso biológico como modelo para o aprendizado e resolução de problemas computacionais de maior complexidade. Através de estudos para seu aperfeiçoamento, a RNA herdou características biológicas, como robustez, alto paralelismo, aprendizado através de padrões e tolerância a falhas (GRÜBLER, 2021).

No contexto deste trabalho, a RNA foi utilizada como um classificador de dados. Segundo SCHALKOFF [14], o reconhecimento de padrões pode ser visto como o particionamento dos dados, que são representados por pontos em um espaço multidimensional. Então, estes pontos se encontrarão dentro das regiões indicativas de suas respectivas classes, cuja quantidade costuma ser definida pelo analista.

Em sua utilização como um classificador de padrões, a RNA pode fornecer, em sua saída, não apenas o conjunto (classe) ao qual determinada entrada pertence, mas também a confiabilidade desta classificação. Com isso, padrões ambíguos podem ser rejeitados e, por consequência, o modelo como um todo será aprimorado (SHALKOFF, 1992).

2.2.4 SHAP

O SHAP (SHapley Additive exPlanations) é um algoritmo usado para explicar as saídas de qualquer modelo de aprendizado de máquina utilizando a teoria dos jogos. Seu funcionamento consiste basicamente em utilizar os valores clássicos de Shapley para conectar as melhores alocações de créditos às explicações locais buscadas (ROTH, 1988).

O valor de Shapley é uma função utilizada na teoria dos jogos para calcular a contribuição de cada jogador, em um modo cooperativo, na busca por um objetivo. A partir desse valor, é possível determinar o poder de um jogador, medido principalmente pela sua contribuição potencial com os demais jogadores (ROTH, 1988).

A ideia do SHAP é utilizar esse valor e, por consequência, a teoria dos jogos, para interpretar um modelo de aprendizado de máquina. As funcionalidades seriam os “jogadores”. Então, tenta-se predizer qual tarefa é fundamental (o que equivaleria ao “jogador com mais poder”) para alcançar determinado “objetivo”, ou seja, a predição em questão. O retorno desta análise é justamente a explicação e a interpretação dos resultados obtidos após a execução do modelo (Documentação SHAP).

2.3 Trabalhos Relacionados

Para melhor entendimento do contexto do problema em questão, buscou-se relacionar trabalhos já existentes na área com o tema deste artigo. Então, foram selecionados três artigos e papers que serão explanados a seguir.

O primeiro artigo analisado, chamado “Development and evaluation of a dashboard for monitoring learning activities in LMS” (PARK; JO, 2015), buscou desenvolver e validar um sistema de monitoramento de variáveis comportamentais de estudantes, em forma de dashboard, fazendo a utilização do algoritmo da regressão logística. O estudo desenvolveu uma solução de software com o objetivo de utilização no ensino a distância, se mostrando relevante para a visualização de indicadores comportamentais dos alunos por meio do monitoramento da aprendizagem.

O segundo artigo, chamado “Um Modelo para Gerenciamento de Múltiplas Trilhas Aplicado a Sistemas de Apoio à Educação” (SILVA, et. al, 2009), busca ampliar a utilização de dados de trilhas de ambientes educacionais, usando um modelo de gerenciamento chamado MultiTrail. Este atua com dados gerados pelas entidades acompanhadas na disponibilização e aplicação de sistemas de apoio à educação, em que cada entidade será armazenada em trilhas diferentes com um objetivo específico. O estudo identificou a possibilidade de composição de múltiplas trilhas com conjuntos de dados obtidos, como, por exemplo, sistemas de gerenciamento acadêmico e ambientes virtuais de aprendizagem.

O terceiro e último estudo visto foi o “Contributos dos Modelos da Auto-Regulação da Aprendizagem para a formação de Alunos e Professores no Ensino Superior” (Gonçalves, 2019). Este analisa o processo de Bolonha (que tem um conjunto de objetivos e iniciativas no espaço europeu de ensino superior) e o programa de tutorado. Seu foco é um sistema de ensino baseado no desenvolvimento das competências dos estudantes, em que os componentes de trabalho experimental e as competências transversais devem ter papel decisivo na formação. Foi observado que processos avaliativos como a auto-avaliação ou método de aprendizagem auto-regulada, ou seja, o

acompanhamento na aprendizagem, traziam resultados de desempenhos melhores nos estudantes.

3 Materiais e Métodos

3.1 Descrição da Base

A base de dados que será utilizada será extraída da plataforma Moodle de cursos superiores EAD da Universidade de Pernambuco. Existiam 26.737 linhas de dados referentes a discente, período, curso e disciplina dos seguintes cursos: Administração, Biologia, Letras e Pedagogia. Esses discentes frequentam entre o 1º e o 8º período, em que foram obtidos dados de sessenta e seis semestres e duzentas e duas disciplinas distintas. As bases totalizaram 18,9 gigabytes de dados provenientes de 12 plataformas diferentes, ao longo de 7 anos, com sistemas de gerenciamento de bancos de dados variando entre PostgreSQL e MySQL.

Após as análises dos clusters, o curso de Biologia apresentou uma concordância entre as técnicas de validação utilizadas, sendo então escolhido para dar prosseguimento a análise. Logo, sobraram um total de 5.235 linhas de dados referentes a discentes, período e disciplina. Depois da análise prévia verificou-se que 714 não apresentaram participação efetiva na plataforma e foram retiradas, assim a análise foi desenvolvida com 4.521 linhas de dados, em que estas linhas representam interações dos estudantes na plataforma.

Neste estudo foram utilizadas 23 variáveis comportamentais significativas armazenadas na plataforma, possibilitando a eliminação do questionário. Isso tornou possível utilizar apenas o conjunto de variáveis comportamentais registrados na plataforma para dar prosseguimento às análises.

Os dados que serão utilizados são referentes ao desempenho do aluno e à gestão do tempo do mesmo. Assim, a descrição dos mesmos pode ser vista na Tabela 1.

Tabela 1: Descrição das variáveis

Atributo	Descrição	Possíveis Variáveis	Relacionado
Primeira prova	Nota da primeira prova ou da segunda chamada	0 a 10	Desempenho
Segunda_prova	Nota da segunda prova ou da segunda chamada	0 a 10	Desempenho
Média_prova	Média geral das duas provas presenciais	0 a 10	Desempenho
Fórum01	Nota do primeiro fórum	0 a 10	Desempenho
Fórum02	Nota do segundo Fórum	0 a 10	Desempenho
Fórum03	Nota do terceiro Fórum	0 a 10	Desempenho
Fórum04	Nota do quarto Fórum	0 a 10	Desempenho
Media_fórum	Média geral dos quatro fóruns	0 a 10	Desempenho
Webquest01	Nota da primeira atividade (webquest)	0 a 10	Desempenho
Webquest02	Nota da segunda atividade (webquest)	0 a 10	Desempenho
Media_webques	Média geral das duas atividades	0 a 10	Desempenho
Desempenho binário	Aprovado ou reprovado	0 ou 1	Desempenho
Desempenho	Desempenho Final	0 a 10	Desempenho
VAR24	Média semanal da quantidade de acessos do discente ao ambiente	0 a 26	G. do tempo
VAR25	Tempo médio entre a criação de um tópico no fórum e a primeira postagem do discente	0 a 288.98	G. do tempo
VAR28	Quantidade de <i>time out</i>	-7 a 561	G. do tempo
VAR31	Quantidade de acesso do discente ao ambiente	0 a 687	G. do tempo

VAR31b	Quantidades de dias distintos que o discente entrou na disciplina	0 a 125	G. do tempo
VAR31c	Quantidade de dias distintos que o discente entrou na plataforma	0 a 178	G. do tempo
VAR32a	Quantidade de acessos do discente ao ambiente no turno da manhã	0 a 299	G. do tempo
VAR32b	Quantidade de acessos do discente ao ambiente no turno da tarde	0 a 378	G. do tempo
VAR32c	Quantidade de acessos do discente ao ambiente no turno da noite	0 a 350	G. do tempo
VAR32d	Quantidade de acessos do discente ao ambiente pela madrugada	0 a 77	G. do tempo

3.2 Metodologia Experimental

Durante o processo metodológico, foi utilizada a técnica de clustering K-means, que particiona um determinado conjunto de dados em um conjunto k de grupos onde k representa o número de grupos pré-especificados pelo analista. Esta técnica classifica os dados em vários grupos, de forma que os objetos contidos nestes clusters sejam o mais semelhantes possível. A validação por cluster é um procedimento para avaliação da qualidade dos resultados do algoritmo de agrupamento, evitando encontrar padrões aleatórios, como também comparar dois algoritmos de cluster. Neste estudo, os dados agrupados foram os relacionados ao desempenho. São eles: "PRIMEIRA_PROVA", "SEGUNDA_PROVA", "WEBQUEST01", "WEBQUEST02", "FORUM01", "FORUM02", "FORUM03" e "FORUM04".

Análises de validação foram feitas neste processo, sendo elas a análise de silhueta e a validação por índice Dunn. A análise de silhueta mede o quão boa uma observação está agrupada e estima a distância média entre os clusters. O gráfico de silhueta mostra uma medida de quão próximo cada ponto do cluster está próximo dos outros vizinhos. Com o número de clusters definidos se pode localizar os centróides dos clusters, plotando um gráfico de dispersão. O teste de silhueta apresenta um índice que pode variar de -1 a 1. Quanto maior o índice, melhor. As silhuetas devem ser mais curtas e ter a maior parte das áreas positivas (KASSAMBARA, 2017). Após esse processo é feito um dendrograma, que constrói uma árvore que apresenta os elementos com proximidades por suas similaridades estatísticas, as linhas de ligação mais altas indicam menos similaridades (KASSAMBARA, 2017).

Por fim, se pode extrair o valor do índice Dunn. Este é o índice de validação interna dos dados, assim, quanto maior o valor, melhor a escolha do número de clusters (KASSAMBARA, 2017).

Após a finalização do processo de agrupamento, utilizou-se então uma Rede Neural Artificial para classificar as variáveis de gestão do tempo (VAR24 a VAR32d) com relação aos grupos de desempenho obtidos após a utilização do k-means.

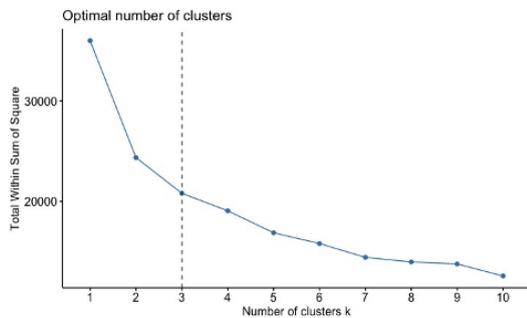
Então, a técnica SHAP, que explica as saídas dos modelos de aprendizagens de máquinas (KASSAMBARA, 2017), foi aplicada. Assim, tornou-se possível interpretar e identificar a influência da gestão do tempo no desempenho do educando, bem como qual(is) das variáveis mais afetam este desempenho.

4 Análise e Discussão dos Resultados

4.1 Resultados do Agrupamento com o K-Means

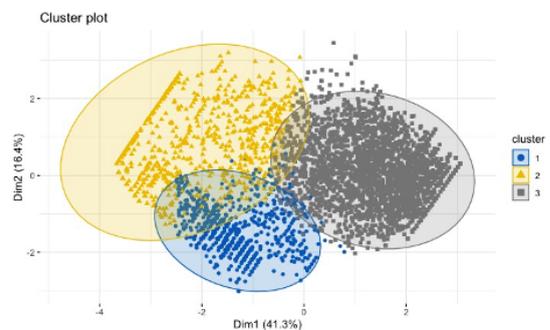
No curso de Biologia, no gráfico de silhueta é encontrado o $k = 3$, apresentado na Figura 1. O gráfico de dispersão dos clusters, a tabela que apresenta o tamanho dos clusters e a distância do centróide, o teste de silhueta, o dendrograma e o índice de Dunn estão representados nas Figuras 1, 2, 3, 4 e 5. É importante ressaltar ainda que os gráficos foram gerados através da linguagem R.

Figura 1: Gráfico de silhueta do curso de Biologia.



Fonte: autores.

Figura 2: Gráfico de dispersão do curso de Biologia.



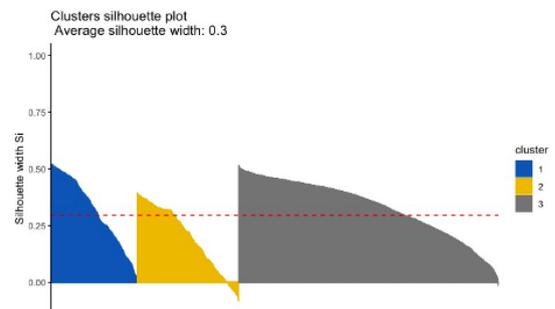
Fonte: autores.

Figura 3: Tamanho dos *clusters* e distância do centróide

```
cluster size ave.sil.width
1      1  867      0.31
2      2 1021      0.20
3      3 2613      0.33
> |
```

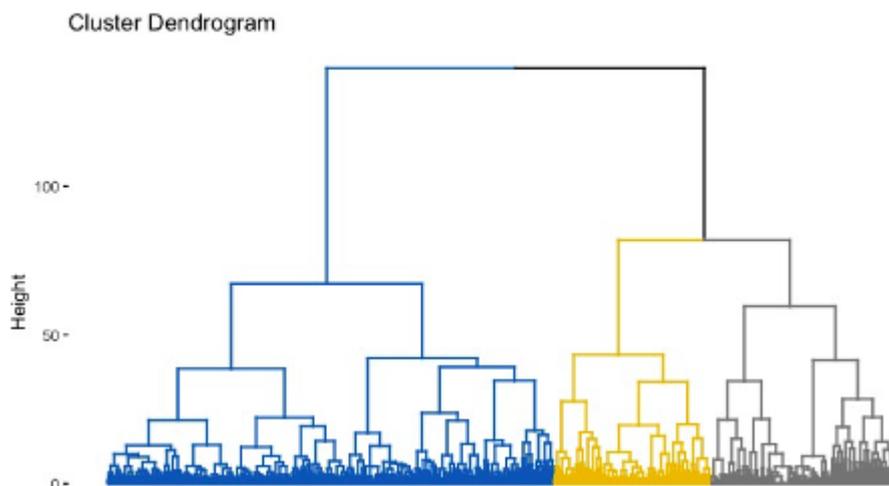
Fonte: autores.

Figura 4: Teste de silhueta para o curso de Biologia



Fonte: autores.

Figura 5: Dendrograma do curso de Biologia

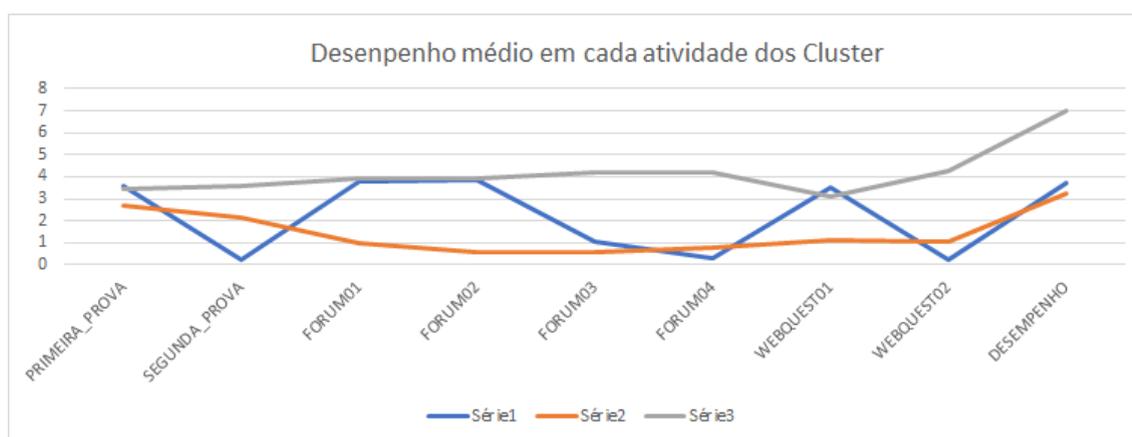


Fonte: autores.

Pode-se ver na Figura 2 que existem elementos nos clusters que estão distantes dos centróides. Mostrando que os limites dos clusters não são rígidos. Na validação usou-se o índice de Dunn, que foi testado para diversos valores de k e o melhor valor corroborou o valor k=3. Obtivemos o índice de Dunn de 0,03882083 confirmando que as bordas dos clusters são bem flexíveis.

Os três grupos identificados após o agrupamento são separados da seguinte forma: o grupo 1 compreende os alunos que apresentam melhor desempenho em atividades da primeira metade da disciplina; o grupo 2, definido por estudantes com bom desempenho em prova, mas com baixo desempenho nas demais atividades; e o grupo 3 é caracterizado pelos estudantes que possuem um melhor desempenho em todas as atividades.

Figura 6: Gráfico apresentando o desempenho médio



Fonte: autores.

Na figura 6, podemos ver que os estudantes que compõem o grupo 1 apresentam uma curva com muitos altos e baixos, vemos que os bons desempenhos deste grupo estão nas atividades da primeira metade da disciplina. A curva do grupo 2 mostra que esses estudantes apresentam um melhor desempenho nas provas e baixo desempenhos nas demais atividades, e na curva do grupo 3

vemos que os estudantes deste grupo mantêm bom desempenho ao longo de todas as atividades.

4.2 Resultados da Classificação com RNA

Na etapa de classificação, realizada através da linguagem de programação Python, as 10 variáveis de gestão do tempo compõem os inputs da RNA, seguidos então de 3 camadas do tipo Dense, onde todos os neurônios são conectados a todos os inputs dessa camada.

A primeira camada, com 120 neurônios, e a segunda camada, com 40 neurônios, são ativadas por uma função do tipo Rectified Linear Unit (ReLU), bastante usual em classificadores por RNA.

Após o treinamento, o modelo de classificação obteve os resultados apresentados no Tabela 2.

Tabela 2: Dados da classificação.

Classe	Precisão	Recall	F1-score	Suporte
Grupo 1	0,77	0,70	0,73	187
Grupo 2	0,71	0,68	0,69	203
Grupo 3	0,86	0,90	0,88	509

Fonte: autores.

O Grupo 3 mostra melhores métricas devido a maior quantidade de instâncias classificadas como tal que a rede considerou durante o treinamento e os testes. Além disso, o modelo de forma geral atendeu às expectativas, com uma acurácia de 81%.

4.3 Impacto das variáveis nos clusters

Na tentativa de trazer uma explicação para os resultados obtidos nas etapas anteriores, foi utilizado o método SHAP , através da linguagem de programação Python, no resultado da classificação do modelo aplicado à toda a base de dados do curso de Biologia.

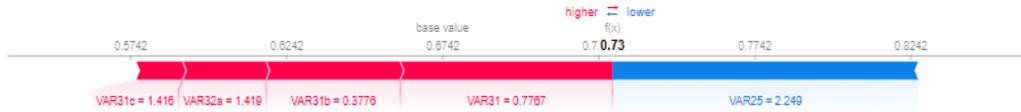
O valor Shap consiste no impacto que determinado atributo teve na saída do modelo e como ele influencia, positiva ou negativamente no resultado final. Os Explainers, métodos utilizados para cada tipo de modelo, são responsáveis por calcular esses valores para cada uma das saídas do classificador.

Para o cálculo dos valores Shap, foi utilizado o método KernelExplainer, que realiza uma regressão linear ponderada usando os dados de treinamento. Apesar de oferecer uma boa explicação, o método apresenta a desvantagem de ser bastante custoso computacionalmente

Alguns gráficos são gerados para uma maior compreensão dos resultados. O gráfico de força evidencia a contribuição individual de cada atributo para uma

determinada saída. A Figura 7 ilustra, para uma instância aleatória pertencente ao grupo 3, o quanto cada atributo contribui para esse resultado.

Figura 7: Impacto dos atributos para uma única saída pertencente ao grupo 3.



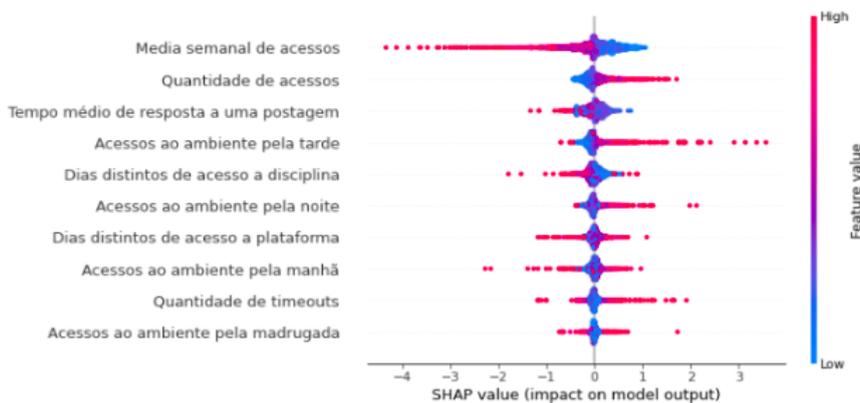
Fonte: autores.

Pode-se ver então os atributos VAR31 (Quantidade de Acessos), VAR31b (Dias distintos de acesso à disciplina), VAR32a (Acessos ao ambiente pela manhã) e VAR31c (Dias distintos de acesso a plataforma) contribuindo positivamente enquanto VAR25 (Tempo médio de resposta a uma postagem) contribui negativamente. Juntos, chegam ao valor de 73% de probabilidade de pertencimento a esse grupo.

É importante notar que o atributo de maior impacto não necessariamente é aquele decisivo para uma determinada classificação. O tempo médio de resposta a uma postagem teve maior impacto no resultado mesmo influenciando negativamente. No entanto, a soma das contribuições dos outros fatores fez a classificação correta para essa instância.

No gráfico de resumo, em primeiro lugar, é possível observar quais os atributos possuem maior relevância, listados por ordem de impacto. Em segundo lugar, é mostrado como esses atributos contribuem para o valor Shap. A temperatura da cor indica o valor do atributo: do ciano ao magenta, valores mais baixos a valores mais altos, respectivamente. Enquanto que o eixo horizontal indica a contribuição ou impacto da variável.

Figura 9:



Fonte: autora

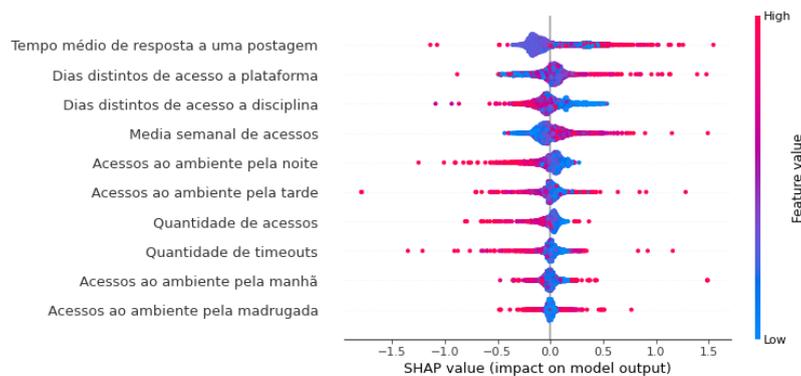
Pode-se ver que os estudantes deste grupo apresentam como atributos de maior impacto no resultado três principais variáveis de gestão do tempo: média semanal de acessos (VAR24), quantidade de acessos (VAR31) e tempo médio de resposta a uma postagem (VAR25).

A média semanal de acessos alta tem uma influência negativa nessa classificação. Mas é preciso observar que essa relação não é necessariamente proporcional, como vemos pelo gráfico de força. Ainda mais, os valores altos para a quantidade de acessos impactam positivamente na classificação da instância como parte deste grupo.

Então os estudantes que têm bom desempenho na primeira metade da disciplina são estudantes que fazem diversos acessos em poucos dias e respondem rapidamente às atividades. Geralmente entram no turno Vespertino. Assim pode-se inferir que são estudantes que entram no fim de semana, respondendo aos fóruns, que ficam abertos com datas mais flexíveis de participação, mostrando dificuldades na gestão do tempo.

O segundo cluster resultante do agrupamento, chamado de grupo 2 – estudantes com bom desempenho apenas em provas, têm seu comportamento mostrado na Figura 10.

Figura 10:



Fonte: autora

Para a classificação como parte do grupo 2, as três variáveis de gestão do tempo com maior relevância são: o tempo médio de resposta a uma postagem (VAR25) os dias distintos de acesso à plataforma (VAR31c), e os dias distintos de acesso à disciplina (VAR31b).

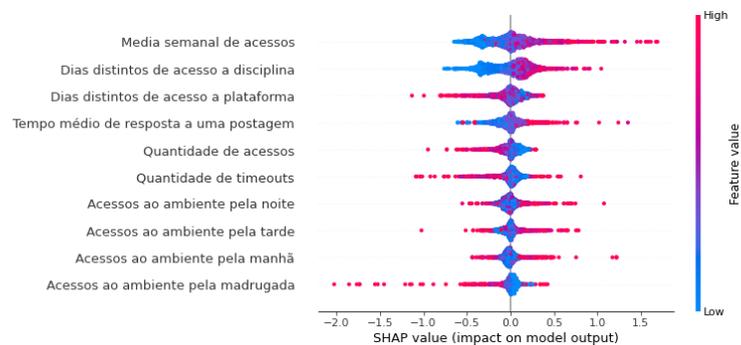
Vê-se que o acesso à plataforma em dias distintos apresenta valor alto. Infere-se daí que esses estudantes mantêm hábitos de regularidade de acesso na plataforma, apesar de os dias distintos de acesso à disciplina contribuírem negativamente.

A média semanal de acessos também é um valor alto para os estudantes deste grupo. E o grupo apresenta baixos valores de timeout. Assim, pode-se inferir que os estudantes que fazem parte deste grupo entram na plataforma, passam pouco tempo nela e depois entram novamente. Muito provavelmente, neste meio tempo, estão buscando informações em outras fontes.

Assim os estudantes com bom desempenho apenas em provas normalmente demoram a responder as atividades, mas acessam o ambiente com frequência, com acessos concentrados nas disciplinas com acessos geralmente no horário da tarde. Infere-se que são estudantes que criaram hábitos de entrar na disciplina e fazem pesquisas antes de responder. São estudantes que desenvolveram estratégias de gestão do tempo mais eficiente.

Os resultados das análises para o terceiro cluster, o grupo 3 – estudantes com melhor desempenho em todas atividades, são apresentados na Figura 11.

Figura 11:



Fonte: autores.

Pode-se verificar que este grupo apresenta, como principais características de gestão do tempo, a média semanal de acessos, a quantidade de dias distintos de acesso à disciplina e a quantidade de dias distintos de acesso à plataforma.

Os estudantes deste grupo apresentam um valor de média semanal de acesso alto. O valor de dias distintos de acesso à disciplina também é alto. Infere-se daí que esses estudantes entram diversas vezes nas disciplinas, mas não em dias específicos. Entretanto, esses estudantes têm um valor de acesso, em dias distintos, à plataforma baixo. Este fato indica que os estudantes acessam a plataforma em alguns dias específicos, mas não apresentam uma rotina de acesso às disciplinas.

Estudantes que apresentam melhor desempenho em todas atividades não apresentam rotinas específicas de acessos, mas acessam o ambiente com grande frequência e buscam responder as atividades com prontidão. Mostram uma boa gestão de tempo, porém com flexibilidade nos acessos.

Os acessos às disciplinas e à plataforma se comportam de forma inversa na classificação dos grupos 2 e 3, evidenciando uma grande distinção entre os alunos que tiveram bons resultados em webquests (grupo 2) e nas provas (grupo 3).

5. Conclusão

O trabalho apresenta que as técnicas utilizadas nesta pesquisa são promissoras, pois mostram um impacto relevante da gestão do tempo no desempenho acadêmico do aluno. Elas possibilitam a análise da problemática de uma maneira nova, utilizando parâmetros que não eram tão usuais e simplificados e que precisam de uma capacidade computacional significativa.

A utilização dos agrupamentos para as variáveis de desempenho mostram perfis diferentes de alunos. Já a classificação realizada pela RNA determina de maneira confiável uma melhor interpretação desses perfis baseados nas variáveis de gestão do tempo utilizando SHAP.

Desse modo, foi visto que a construção de um perfil comportamental eficiente em relação à gestão do tempo permite ao aluno um melhor desempenho ao longo

do curso. Além disso, é possível também identificar, através dos dados extraídos, qual é o perfil comportamental dos alunos que fazem parte de um mesmo grupo com características em comum.

Portanto, como trabalho futuro seria interessante analisar que tipos de intervenção professores e tutores podem realizar para incentivar os fatores referentes à gestão do tempo que impactam positivamente no desempenho do estudante. Além disso, as técnicas de RNA também podem ser usadas para prever o desempenho final do aluno.

Referências

BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. Mineração de dados educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, 2011.

RODRIGUES, R. L.; SANTOS, J. L. S. Revisão sistemática sobre características de gestão de tempo na realização de atividades educacionais em sistemas de gerenciamento de aprendizagem. RELATEC, 2020.

BARNES, T.; DESMARAIS, M.; ROMERO, C.; VENTURA, S. Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings. Cordoba, Spain, 2009.

NASCIMENTO, R. L. S.; CRUZ JUNIOR, G. G.; FAGUNDES, R.A. Mineração de dados educacionais: um estudo sobre indicadores da educação em bases de dados do INEP. NOVA TECNOLOGIA NA EDUCAÇÃO. V.16, 2018.

MELO, E. P. S.; FARIAS, A. P. S. Gerenciamento do tempo e da rotina de trabalho: um estudo de caso numa agência do banco Bradesco S/A. CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 2015.

BAKER, R. S. J.; ISOTANI, S.; CARVALHO, A. Mineração de dados educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, 2011.

BROADBENT, J.; POON, W. L. Self regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. The Internet and Higher Education, 27, 1-13, 2015.

PINTO, R. P. L. J. M.; MOREIRA, A. M. Análise da Reprovação em Disciplinas do Curso de Matemática a Distância da Universidade Federal da Paraíba, 2018.

MARTINS, K. A Importância da Educação a Distância na Sociedade Atual, 2016.

COSTA, E.; BARKER, R. S. J.; MAGALHÃES, J.; MARINHO, T. Mineração de dados educacionais: conceitos, técnicas, ferramentas e aplicações. Jornada de Atualização em Informática na Educação - JAIE , 2012.

SAS, Mineração de dados: O que é e qual a sua importância?, 2021. Acessado em 14 de Abril de 2021:

https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html

KASSAMBARA, A. Cluster Analysis in R. Vol. 1, 2017.

GRÜBLER, M. Entendendo o funcionamento de uma Rede Neural Artificial, 2018. Acessado em 11 de Maio de 2021:

<https://medium.com/brasil-ai/entendendo-o-funcionamento-de-uma-rede-neural-artificial-4463fcf44dd0>

SHALKOFF, R. Pattern Recognition - Statistical, Structural and Neural Approaches. John Wiley & Sons, Inc. USA, 1992.

ROTH, A. E. The Shapley Value: Essays in honor of Lloyd S. Shapley. Cambridge: Cambridge University Press, 1988.

PARK, Y. JO, I. Development and evaluation of a dashboard for monitoring learning activities in LMS. Journal of Universal Computer Science, 2015.

SILVA, J.R; ROSA, J.; BARBOSA, J.; FRANCO, L.; BARBOSA, D. N. F.; PALAZZO, L. A. M.. Um Modelo para Gerenciamento de Trilhas em Ambientes de Computação Móvel. In: SIMPÓSIO EM SISTEMAS COMPUTACIONAIS DE ALTO DESEMPENHO (WSCAD), 10. , 2009.

GONÇALVES, I. C. Contributos dos Modelos da Auto-Regulação da Aprendizagem para a formação de Alunos e Professores no Ensino Superior, 2010.