

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

**Pricylla Santos Cavalcante do Nascimento¹, Adelson Santos da Silva Junior¹,
Camila Linhares Schulz¹, Maria Victória Rodrigues dos Santos¹, Alexandre Magno
Andrade Maciel¹, Rodrigo Lins Rodrigues², Robson Raabi do Nascimento²,
Fernanda Maria Ribeiro Alencar¹**

¹Universidade de Pernambuco – Recife – PE – Brazil.

²Universidade Federal Rural de Pernambuco – Recife – PE – Brazil.

pscn@ecom.poli.br; assj2@ecom.poli.br; cls@ecom.poli.br; mvrs2@ecom.poli.br;
alexandre.maciell@upe.br; rodrigomuribec@gmail.com; robsonraabi@gmail.com;
fernandaalenc@gmail.com

Abstract. *With technological advances in the educational environment, new challenges have also emerged. Among these, there is the difficulty in identifying factors that corroborate with a good academic performance of students in distance learning courses. Thus, this work aims to analyze the impacts of time management on students' academic performance. For this, the K-means technique was used to carry out the grouping of students in relation to their academic performance, an Artificial Neural Network to classify these groups based on time management variables, and the SHAP method to interpret the variables that most impacted on the performance rating, according to time management. For the construction of this research, data from distance learning courses extracted from the moodle platform of a public university in the state of Pernambuco were used. In conclusion, it was possible to observe which time management characteristics positively impact students' academic performance.*

Resumo. *Com os avanços tecnológicos no ambiente educacional, surgiram 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 as variáveis que mais impactaram na classificação do desempenho, de acordo com a gestão do tempo. 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 dos estudantes.*

1. Introdução

Os avanços da tecnologia permeiam por diversas áreas, entre essas a área de educação. Diante da evolução dos artefatos tecnológicos para a educação, ambientes computacionais de ensino e conteúdos digitais passaram a ser vistos como instrumentos essenciais no processo de ensino (BARKER; ISOTANI & CARVALHO, 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). Com a utilização desses processos surgiu a necessidade da utilização de técnicas para analisar dados que são gerados por plataformas de ensino. Assim, surge a mineração de dados educacionais ou Educational Data Mining (EDM), que é definida como uma área de pesquisa que busca analisar dados provindos de interações educacionais, com o foco em solucionar questões de pesquisa envolvendo aprendizagem (BARNES et.al, 2009). Deste modo, esta área 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; CRUZ JUNIOR & FAGUNDES, 2018).

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).

Os cursos EAD estão ganhando espaço e 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 (RODRIGUES & SANTOS, 2020). Deste modo, torna-se mais simples a acomodação das atividades educacionais dentro da rotina dos estudantes com diversas ocupações e ainda permite que o mesmo organize-se e planeje-se de uma maneira mais flexível. Contudo, esta flexibilidade traz consigo a necessidade de uma postura mais ativa para o planejamento e o gerenciamento no processo de ensino e 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.

A flexibilidade dos cursos EAD, criam 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 e locais fixos para a atividade. Mesmo tendo essa maior flexibilidade pode gerar 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.

2. Fundamentação Teórica

2.1. Mineração de Dados Educacionais

A área de mineração de dados educacionais busca desenvolver ou adaptar métodos e algoritmos de mineração de dados existentes, facilitando com isso a compreensão dos dados nos contextos educacionais, que são produzidos por estudantes, professores/tutores e ambiente educativo 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).

2.2. 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. De acordo com Santos e Rodrigues (2020) a gestão de tempo não deve ser considerada como um traço ou habilidade, mas como um processo em que o educando escolhe quando, onde e por quanto tempo eles passam realizando as atividades. Assim, os autores afirmam ainda que essa gestão é passível de mudanças ao longo do tempo por ter como influência as ações sociais e as experiências do sujeito.

3. Materiais e Métodos

3.1. Descrição da Base de Dados

A base de dados foi extraída da plataforma Moodle de cursos de graduação na modalidade EAD da Universidade de Pernambuco. Existiam 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 retidas. Assim, a análise foi desenvolvida com 4.521 linhas de dados, onde 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 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.

Atributo	Descrição	Intervalo	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_webquest	Média geral das duas atividades	0 a 10	Desempenho
Desempenho binár	Aprovado ou reprovado	0 ou 1	Desempenho
Desempenho	Desempenho Final	0 a 10	Desempenho
VAR24	Média semanal da quantidade de acessos do discente ao ambien	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

Os valores das variáveis foram todos normalizados para que pudessem ser aplicadas as técnicas/algoritmos de agrupamento.

3.2. Procedimento Experimental

Para a realização do processo de agrupamento, foi utilizada a técnica de clustering através do algoritmo K-means. O agrupamento foi baseado nas variáveis de desempenho, são eles: “PRIMEIRA_PROVA”, “SEGUNDA_PROVA”, “WEBQUEST01”, “WEBQUEST02”, “FORUM01”, “FORUM02”, “FORUM03” e “FORUM04”.

Utilizou-se a técnica de cotovelo para determinar o número k de clusters e a validação por 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 (RNA) 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 (BOULIC & REANULT, 1991), foi aplicada. Assim, tornou-se possível interpretar e identificar a influência da gestão do tempo no desempenho do educando, bem como quais das variáveis mais afetam este desempenho.

4. Análise e Discussão dos Resultados

4.1. Resultado do Agrupamento

A técnica de “cotovelo”, que tem como objetivo calcular o agrupamento k-means usando valores diferentes de clusters k, em seguida o wss, na soma do quadrado, é desenhado de acordo com o número de clusters, assim a localização de uma dobra de “cotovelo” no gráfico é geralmente considerada como um indicador do número adequado de cluster (WICKHAM, GROLEMUN; 2019). Segue o gráfico da Figura 1.

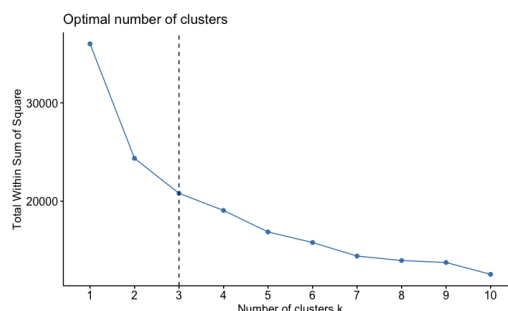


Figure 1 Gráfico de silhueta para o curso

A Figura 1 apresenta um valor de $k = 3$. Ainda foram aplicadas as técnicas de silhueta e o dendrograma e também chegamos ao mesmo número k .

Construiu-se o gráfico de dispersão dos clusters, apresentado na Figura 2.

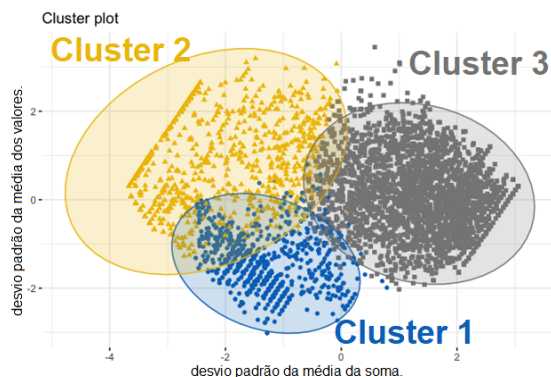


Figure 2 Clusters de Desempenho

Pode-se perceber 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, isto é, existem interseções entre os clusters.

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

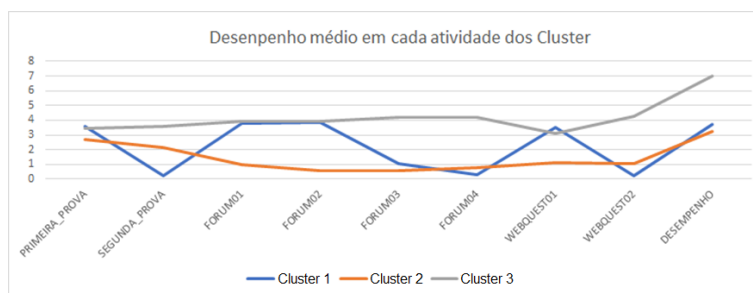


Figure 3 Desempenho médio dos clusters em cada atividade

Na figura 3, podemos verificar que os estudantes que compõem o Cluster 1 apresentam uma curva com muitos altos e baixos, vemos que os bons desempenhos deste Cluster estão nas atividades da primeira metade da disciplina. A curva do Cluster 2 mostra que esses estudantes apresentam um melhor desempenho nas provas e baixo desempenhos nas demais atividades, e na curva do Cluster 3 vemos que os estudantes deste Cluster mantêm bom desempenho ao longo de todas as atividades.

4.2. Resultado da Classificação com Rede Neural Artificial

Na etapa de classificação, as 10 variáveis de gestão do tempo compõem os inputs da Rede Neural Artificial (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 ReLU, bastante usual em classificadores por RNA. Na última camada usou-se 3 neurônios, um para cada cluster encontrado. Uma função logística é aplicada como ativação para os outputs como pertencimento a classe.

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

Tabela 2 Resultados do treinamento da RNA

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

O Cluster 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 no resultado da classificação. 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.

Pôde-se ver que os atributos Quantidade de Acessos (VAR31), Dias distintos de acesso à disciplina (VAR31b), Acessos ao ambiente pela manhã (VAR32a) e Dias distintos de acesso a plataforma (VAR31c) contribuindo positivamente enquanto Tempo médio de resposta a uma postagem (VAR25) contribui negativamente.

Assim vemos que os estudantes que acessam a plataforma e com mais frequência, mostrado com as variáveis VAR31, VAR31b e VAR31c, apresentam melhor desempenho, a demora em responder postagens (VAR25) apontam um baixo desempenho e os estudantes que podem acessar pela manhã (VAR32a) apresentam melhor desempenho.

O SHAP dá como saída um gráfico em que os atributos com maior relevância são listados em ordem de impacto, é mostrado como esse atributo contribui para o valor do SHAP a cor mais clara indica baixos valores e a cor mais escura altos valores enquanto o eixo horizontal indica a contribuição da variável.

Os resultados para o Cluster 1 - com desempenho na primeira metade da disciplina, representados na Figura 3.

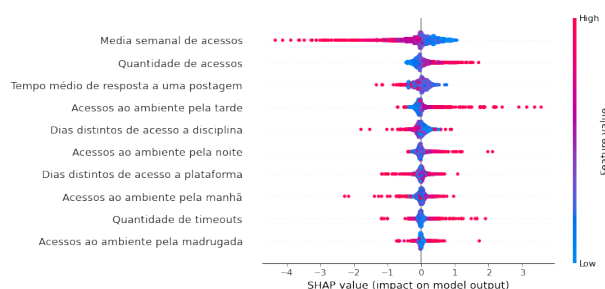


Figure 3 SHAP para o Cluster 1

Pode-se ver que os estudantes deste grupo apresentaram 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 Cluster 2 – estudantes com bom desempenho apenas em provas, têm seu comportamento mostrado na Figura 4.

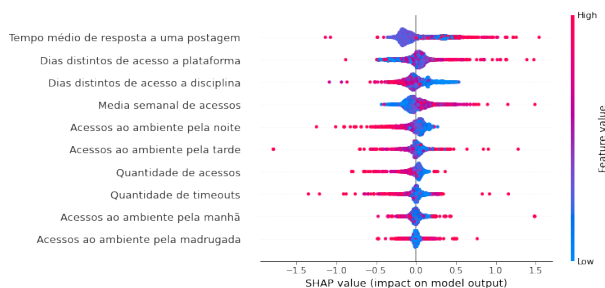


Figure 4 SHAP para o Cluster 2

Para a classificação como parte do Cluster 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 Cluster. E o Cluster apresenta baixos valores de timeout. Assim, pode-se inferir que os estudantes que fazem parte deste Cluster 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.

Os resultados das análises do Cluster 3 – estudantes com melhor desempenho em todas atividades, são apresentados na Figura 5.

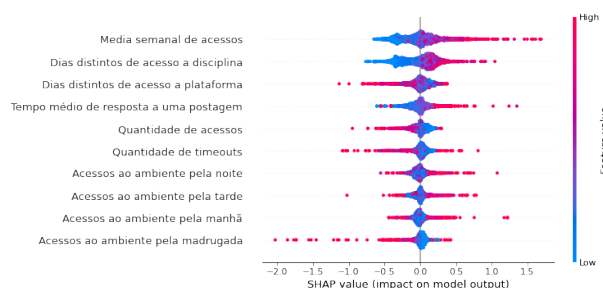


Figure 5 SHAP para o Cluster 3

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 Cluster 2 e 3, evidenciando uma grande distinção entre os alunos que tiveram bons resultados apenas em provas (Cluster 2) e os que tiveram bons desempenho em todas as atividades (Cluster 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 mostra 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 Cluster 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 Rede Neural Artificial também podem ser usadas para prever o desempenho final do aluno.

References

- RODRIGUES, R. L.; SANTOS, J. L. S. (2020) Revisão sistemática sobre características de gestão de tempo na realização de atividades educacionais em sistemas de gerenciamento de aprendizagem. RELATEC.
- BARNES, T.; DESMARAIS, M.; ROMERO, C.; VENTURA, S. (2009) Educational Data Mining. 2nd International Conference on Educational Data Mining, Proceedings. Cordoba, Spain.
- NASCIMENTO, R. L. S.; CRUZ JUNIOR, G. G.; FAGUNDES, R.A. (2018) Mineração de dados educacionais: um estudo sobre indicadores da educação em bases de dados do INEP. NOVA TECNOLOGIA NA EDUCAÇÃO. V.16.
- MELO, E. P. S.; FARIAS, A. P. S. (2015) 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.
- BROADBENT, J.; POON, W. L. (2015) Self regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. The Internet and Higher Education.
- SAS, (2021) Mineração de dados: O que é e qual a sua importância? Disponível em: https://www.sas.com/pt_br/insights/analytics/mineracao-de-dados.html Acessado em 14 de Abril de 2021
- KASSAMBARA, A. Cluster Analysis in R. Vol. 1, 2017.
- SANTOS, J. L. S.; RODRIGUES, L. R. (2020) Revisão sistemática sobre características de gestão de tempo na realização de atividades educacionais em sistemas de gerenciamento de aprendizagem. Revista Latinoamericana de Tecnología Educativa.
- SHAP (2021) Documentação oficial do SHAP. Disponível em: <https://shap.readthedocs.io/en/latest/index.html> Acessado em 11 de Maio de 2021
- WICKHAM, H.; GROLEMUND, G (2019). R para data science. Alta Books.