

Classificação de Interações com Indicadores de Engajamento dos Estudantes no Aprendizado On-line

Aluisio José Pereira¹, Alex Sandro Gomes¹, Tiago Thompsen Primo²

¹Centro de Informática (CIn) – Universidade Federal de Pernambuco (UFPE) Caixa Postal 50740-560 – Recife – PE – Brasil.

²Centro de Engenharias (CEng) – Universidade Federal de Pelotas (UFPel) – Pelotas – RS – Brasil.

{ajp3, asg}@cin.ufpe.br, tiago.primo@inf.ufpel.edu.br

Abstract. *This study addresses the difficulty of analyzing student engagement indicators in online teaching and learning activities. We analyzed the performance of different Machine Learning (ML) algorithms, combined with heterogeneous and homogeneous ensemble classifier strategies, in order to identify the most effective approaches in predicting student interaction levels. The results indicate that the Boosting ensemble algorithms with Support Vector Machine (SVM) and Decision Tree (DT) performed better, suggesting that this ML strategy can help identify indicators of engagement in e-learning activities. In this sense, the combinations of classifiers were applied for analysis and presentation of interaction indicators to support human tutors in promoting student engagement.*

Resumo. *Este estudo aborda a dificuldade de acompanhar indicadores do engajamento dos estudantes em atividades de ensino-aprendizagem on-line. Foi analisado o desempenho de diferentes algoritmos de Aprendizagem de Máquina (AM), combinados com estratégias de comitês de classificadores heterogêneos e homogêneos, para identificar as abordagens mais eficazes na previsão dos níveis de interação dos estudantes. Os resultados indicam que o comitê Boosting com os algoritmos Máquina de Vetor de Suporte (MVS) e Árvore de Decisão (AD) apresentaram melhor desempenho. Esta estratégia de AM pode ajudar a identificar indicadores de engajamento em atividades no aprendizado on-line. Neste sentido, as combinações dos classificadores foram aplicadas na análise e apresentação dos indicadores de interação para apoiar tutores humanos na promoção do engajamento estudantil.*

1. Introdução

No contexto do aprendizado, o engajamento dos estudantes é crucial para a compreensão do conhecimento. Estudantes engajados interagem ativamente, participam das atividades, aprendem os conteúdos, fazem perguntas e se motivam a atingir os objetivos de aprendizagem, o que pode resultar em um melhor desempenho acadêmico. Para Almarghani, Mijatovic (2017) e Werang, Leba (2022) o engajamento dos estudantes pode ser influenciado por diversos fatores. Por exemplo, no ambiente de ensino presencial (tradicional), os tutores podem observar a dinâmica da sala de aula e ponderar sobre os níveis de engajamento de cada estudante (Salta *et al.*, 2022). Já no contexto do aprendizado on-line, a tarefa de identificar os indicadores de engajamento dos estudantes não é trivial, uma vez que, os indicadores estão dispersos em diversos cenários de interação, que muitas vezes, não são transparentes aos tutores e professores.

Apesar disto, o contexto digital permite coletar dados que servem para compreender a dinâmica das interações (Pereira *et al.* 2023). E, por conseguinte, o engajamento dos estudantes (Salta *et al.*, 2022). Entretanto, os tutores humanos têm limitações (Kim; Davis; Hong, 2022), isto é, dificuldades na análise de grandes volumes de dados, ou com elevadas cargas cognitivas (Sweller, 1994) e (Mayer, 2009). Ou seja, é difícil para um tutor humano realizar “busca ativa”, acompanhar individualmente e em tempo real, as interações (mensagens enviadas, comentários, pedidos de ajuda, turno

que estão disponíveis) de cada estudante no ambiente virtual de aprendizagem, especialmente em programas e cursos com muitos estudantes. Além disto, é complexo distinguir as práticas didático-pedagógicas que podem dar certo no ensino-aprendizado on-line ou presencial. Contudo, a apropriação da tecnologia adequada torna a pedagogia on-line única (Archambault; Leary; Rice, 2022).

Neste sentido, a Inteligência Artificial na Educação (IAEd) desempenha um papel fundamental (Srinivasa; Kurni; Saritha, 2022), (Alhazmi *et al.*, 2023), (Zafari *et al.*, 2023). Com a utilização de abordagens de Aprendizado de Máquina (AM) é possível automatizar e aperfeiçoar o processo de análise de dados a ser utilizado, por exemplo, por Sistemas de Tutoria Inteligentes (STIs) (St-Hilaire *et al.*, 2022). Neste caso, os STIs podem ser capazes de personalizar a aprendizagem dos estudantes, ao serem atrelados a abordagens de recomendações (Pereira; Gomes; Primo, 2022). Isto pode ser feito em cooperação com os tutores humanos, o que nos permite explorar abordagens de *design*, uso e avaliação de sistemas híbridos (humanos e IA) no aprendizado (Echeverria *et al.*, 2022). E assim, possibilitar que tutores humanos se concentrem em necessidades mais específicas, flexíveis e adaptáveis.

Embora existam diferentes abordagens de AM que podem ser utilizadas na análise de indicadores de engajamento no ensino-aprendizagem dos estudantes, sistemas baseados em comitês de classificadores, que constroem um modelo preditivo integrando vários modelos (Rokach, 2010), têm mostrado resultados favoráveis em diversos domínios, quando comparados a sistemas baseados em um único modelo para classificar informações. Sendo assim, este estudo tem o objetivo de analisar o desempenho de classificadores na busca por respostas para a seguinte indagação: “*Quais as melhores combinações de estratégias de aprendizagem de máquina para classificar interações dos estudantes no aprendizado on-line?*”. Para isto, utilizou-se de dados históricos de um contexto de interação entre estudantes e tutores do aprendizado on-line, mediado por uma plataforma social de aprendizagem.

Além desta introdução (Seção 1), este artigo apresenta os antecedentes teóricos nos trabalhos relacionados (Seção 2), a metodologia seguida (Seção 3), os resultados e discussões (Seção 4), as considerações finais, lacunas a ser exploradas por estudos futuros e limitações (Seção 5). Também, ressalta as conformidades éticas e lista os trabalhos referenciados.

2. Trabalhos relacionados

No processo de ensino-aprendizado, diferentes modelos de AM têm sido eficazes em problemas que envolvem a previsão de desempenho de estudantes (De Bruin, 2022). As estratégias vêm sendo adotadas na análise de indicadores de engajamento dos estudantes ao combinar classificadores, como: *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Decision Tree* (DT) e *Artificial Neural Network* (ANN) (Chang *et al.*, 2018) e (Buono *et al.*, 2022). Estas técnicas combinadas resultam em comitês (*ensembles*) de classificadores que podem ser aplicados para melhorar a eficácia da classificação (Kamath *et al.*, 2022) e (Saleh; YU; Chen, 2022).

Ao agregar diferentes abordagens individuais, é formado *Ensemble Based Systems*, também conhecidos como *Multiple Classifier Systems*, *Committee of Classifiers*, ou ainda de *Mixture of Experts*. Estas abordagens são utilizadas em diversas situações para solucionar problemas difíceis de serem resolvidos por um único classificador, independentemente da quantidade de dados. Estes sistemas combinam diferentes algoritmos de AM para obter resultados mais precisos e/ou robustos. O

objetivo na formação dos comitês é encontrar a combinação adequada de diferentes modelos fracos e melhorar o desempenho do modelo final. Podem ser utilizadas algumas estratégias, a saber: *Bagging*, *Boosting*, *Stacking* e *Voting* (Muhlbaier; Topalis; Polikar, 2008) e (Wang *et al.*, 2011). O *Bagging* usa aprendizes fracos homogêneos, treinando-os independentemente uns dos outros em paralelo, e combinando-os seguindo algum tipo de processo de média determinística. O *Boosting* também usa aprendizes fracos homogêneos, mas os treina de forma sequencial e adaptativa, onde cada modelo base depende dos anteriores, e os combina por estratégia determinística (Džeroski; Ženko, 2004), (Nti; Adekoya, Weyori, 2020) e (Taser, 2021). O *Stacking* usa aprendizes fracos heterogêneos, treina um meta-modelo e produz uma previsão com base nas diferentes previsões dos modelos fracos (Adejo; Connolly, 2018). Já o *Voting* também usa aprendizes fracos heterogêneos, a classe com mais votos (*hard*) ou maior probabilidade (*soft*) é escolhida na previsão final (Moraitis *et al.*, 2022). Os trabalhos relacionados usam modelos de AM para predição do desempenho dos estudantes. Neste artigo, foi explorado o engajamento dos estudantes, porém em um arcabouço de interações que se constitui como indicador do engajamento no aprendizado on-line.

Para analisar o engajamento dos estudantes, podem ser considerados modelos de *ensemble*, tanto homogêneos quanto heterogêneos, a partir de indicadores como interação, motivação, desempenho, entre outros. Em Evangelista e Sy (2022), os *ensembles* heterogêneos apresentaram um desempenho um pouco superior em relação aos homogêneos ao analisar o desempenho dos estudantes no ensino-aprendizagem. Saleem *et al.* (2021) utilizaram dados gerados do ambiente virtual para avaliar o desempenho dos estudantes por meio de algoritmos de AM. Concluíram que o modelo poderia ajudar os estudantes a tomar decisões, realizando recomendações, notificações e informando-os de maneira proativa. Evangelista (2021) também propôs algo semelhante, usando métodos *ensemble* e obteve resultados mais precisos do que os classificadores individuais, auxiliados por técnicas de seleção de características. Já Brohi *et al.* (2019) abordaram técnicas semelhantes, que mostrou no contexto do ensino superior que os comitês se mostraram mais efetivos.

3. Método

A análise experimental desta pesquisa foi constituída em etapas de: coleta, análise exploratória com pré-processamento dos dados, avaliação prévia dos algoritmos, e avaliação dos comitês (*Bagging*, *Boosting*, *Stacking* e *Voting*). Para a execução dos experimentos, foi adotada a linguagem *Python* no *Google Colab*, através do qual foi desenvolvido um *framework*¹ (que integra este artigo) utilizando a biblioteca *scikit-learn* para obter algoritmos de AM, técnicas de *ensemble*, esquemas e medidas de validação. A Figura 1 esquematiza as etapas para a realização dos experimentos. As etapas envolveram a coleta dos dados, divisão da base de dados em treino e teste, comparação prévia dos classificadores, comparação dos classificadores heterogêneos e homogêneos, resultados e conclusões. No mencionado repositório virtual encontra-se também os algoritmos desenvolvidos e a base de dados utilizada para conduzir as análises experimentais desta pesquisa.

¹ Disponível em: https://github.com/AluisioPereira/artigo_ensemble

procedimentos experimentais envolveram técnicas de normalização das variáveis na escala entre 0 (zero) e 1 (um). Ocorreu a utilização de abordagens de Regressão Linear (RL) para verificar valores ausentes e enriquecer a base de treinamento. Nos casos em que os estudantes não contabilizaram interações, as consultas retornavam o valor 0 (zero), por isso, não houve valores ausentes ($\Phi = 0$).

Tabela 1. Descritiva das métricas de interações dos estudantes

	Ensino	N	Φ	μ	Me	σ	SE	Min.	Max.	Intervalo de confiança (95%)	
										Mais baixo	Superior
VAR01	E.F.	504	0	34,740	21	53,715	23,927	0	628	30,051	39,430
	E.M.	459	0	8,950	6	11,299	0,5274	0	119	7,916	9,984
VAR02	E.F.	504	0	26,821	16	45,397	20,221	0	541	22,858	30,785
	E.M.	459	0	4,839	2	7,553	0,3525	0	80	4,148	5,530
VAR03	E.F.	504	0	0,796	0	2,161	0,0963	0	17	0,607	0,984
	E.M.	459	0	0,185	0	0,666	0,0311	0	5	0,124	0,246
VAR04	E.F.	504	0	12,970	2	58,053	25,859	0	675	7,902	18,038
	E.M.	459	0	1,841	0	5,848	0,2730	0	62	1,306	2,376
VAR05	E.F.	504	0	9,155	0	56,633	25,226	0	673	4,211	14,099
	E.M.	459	0	0,32	0	1,171	0,0547	0	11	0,213	0,427
VAR06	E.F.	504	0	3,675	0	8,859	0,3946	0	82	2,901	4,448
	E.M.	459	0	1,377	0	5,464	0,2551	0	62	0,877	1,877
VAR07	E.F.	504	0	1,214	0	1,891	0,0842	0	7	1,049	1,379
	E.M.	459	0	1,658	2	1,097	0,0512	0	4	1,558	1,758
VAR08	E.F.	504	0	3,089	0	7,186	0,3201	0	61	2,462	3,717
	E.M.	459	0	0,954	0	3,048	0,1423	0	37	0,675	1,233
VAR09	E.F.	504	0	54,909	29	78,665	35,040	0	579	48,041	61,776
	E.M.	459	0	17,301	11	20,465	0,9552	0	188	15,428	19,173
VAR10	E.F.	504	0	17,819	9	25,420	11,323	0	217	15,600	20,039
	E.M.	459	0	7,837	4	11,814	0,5514	0	113	6,756	8,917

Nota: VAR01 a VAR10 - variáveis coletadas; E.F. - Ensino Fundamental; E.M. - Ensino Médio; N - número de estudantes observados; Φ - Quantidade de valores vazios; μ - Média; Me - Mediana; σ - Desvio padrão; Min. - Mínimo; Max. - Máximo; SE - erro padrão médio.

A divisão da base de dados e a conversão de dados qualitativos em quantitativos, para permitir operações numéricas (abordagem utilizadas pelos algoritmos de AM) foi realizada com a adoção da biblioteca LabelEncoder². Ao final da etapa de pré-processamento, foram obtidos a divisão da base de dados, realizada aleatoriamente em 70% para Treino e 30% para Teste, com a função *k-fold*³ para realizar a validação cruzada repetida (Refaeilzadeh; Tang; Liu, 2009). No caso, o número de *folds* foram definidos em 5 (cinco), isto é, o conjunto de dados foi dividido em cinco partes iguais. A validação cruzada foi repetida 10 vezes, com novas divisões aleatórias do conjunto de dados em cada repetição.

3.3. Comparações Prévia dos Classificadores

No experimento, foram usados quatro algoritmos básicos para a formação de comitês (Seção 2). No KNN, o número *k* dos vizinhos mais próximos foi variado *n_neighbors*: {3, 5, 7, 9, 11, 13}, e por padrão, permitiu que a própria função decidisse o melhor algoritmo a ser usado para calcular os vizinhos mais próximos com base nos valores passados. Para SVM, os parâmetros de regularização foram variados *C*: {1, 10, 100 e 1000}, o tipo de *kernel*: {"linear", "poly", "rbf", "sigmoid"}, para o *kernel*: "rbf" a ser usado no algoritmo, foram adotados os coeficientes *gamma*: [0.001, 0.0001]. No DT, os parâmetros do critério variaram *criterion*: {"gini", "entropy", "log_loss"}, a profundidade máxima da árvore *max_depth*: [2, 3, 5, 7, 10], o número mínimo de

² Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>. Acesso em: 15 mar. 2023.

³ Disponível em: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html. Acesso em: 31 mar. 2023.

amostras necessário para dividir um nó interno *min_samples_split*: [2, 3, 4, 5, 10, 15], e o número mínimo de amostras necessário para estar em um nó folha *min_samples_leaf*: [2, 5, 10, 15, 20]. Já para ANN, a função de ativação da camada oculta foi variada *activation*: ['identity', 'logistic', 'tanh', 'relu'], além da força do termo de regularização L2 *alpha*: [0.0001, 0.001, 0.01, 0.1, 1.0], e a tolerância para otimização *tol*: [0.001, 0.01, 0.1]. As melhores configurações de parâmetros foram utilizadas nos comitês, conforme (Seção 3.4).

3.4. Comparações dos Comitês de Classificadores

Os melhores ajustes de classificadores foram aplicados a cada um dos comitês correspondentes. Isto permitiu avaliar o comportamento dos modelos após a aplicação dos métodos *Bagging* (que utilizou: KNN, SVM, DT e ANN), *Boosting* (que utilizou: SVM e DT), *Stacking* (que utilizou: KNN, SVM, DT e ANN) e *Voting* (que utilizou: KNN, SVM, DT e ANN) na base de dados das interações. Ao avaliar os resultados dos modelos, foram utilizadas quatro métricas de desempenho: *accuracy*, *precision*, *recall*, *f1-score* (ou seja: acurácia, precisão, revocação e F-medida). *Accuracy* é dada pela proporção de previsões corretas em relação a todos os casos avaliados pelo modelo. *Precision* representa a proporção de verdadeiros positivos em relação a todos os resultados positivos previstos pelo modelo de classificação. *Recall* evidencia a proporção de verdadeiros positivos em relação a todos os casos positivos reais na base de dados. Já *f1-score* é a média harmônica entre a *precision* e *recall* combinada para avaliar o desempenho do modelo de classificação. Não foi adotada somente a acurácia, pois embora seja uma métrica importante, ela pode ser enganosa em algumas situações, especialmente quando as classes estão desbalanceadas. No contexto real, há tendência de existirem níveis de interação muito altos e outros muito baixos. Por exemplo, os estudantes que **M. I.**: E.F. = 42 (8,3%) e E.M. = 31 (6,7%); **I. E.**: E.F. = 164 (32,5%) e E.M. = 234 (50,9%); **I. R.**: E.F. = 298 (59,1%) e E.M. = 194 (42,3%). Ou seja, uma abordagem que prever os estudantes que interagem esporadicamente e interagem raramente teria melhor precisão, porém não seria adequado para prever os estudantes que mais interagem. Neste caso, *precision* e *recall* são apropriados para avaliar os modelos, ao fornecerem informações sobre como o modelo está se comportando na detecção dos exemplos positivos. Os resultados destas análises são apresentados a seguir (Seção 4).

4. Resultados e Discussão

Esta seção apresenta os resultados da avaliação dos classificadores. A partir das métricas analisadas é possível depreender os modelos com melhor desempenho.

4.1. Resultados das Avaliações Prévias

A comparação preliminar dos classificadores básicos foi crucial na formação de comitês, ao permitir identificar quais configurações dos algoritmos apresentam melhor desempenho no conjunto de dados. Os algoritmos KNN, SVM, DT e ANN foram avaliados previamente e, então, selecionadas as configurações adequadas a serem utilizadas pelos comitês (*ensembles*) de classificadores. Ao analisar as configurações de parâmetros dos algoritmos mais adequados para classificação dos dados das interações dos estudantes, os resultados das avaliações indicaram que o KNN com número *k* de vizinhos mais próximos *n_neighbors* = 3 obteve no teste uma acurácia média de 91%,

uma precisão média de 93%, *recall* médio de 75% e *f1-score* médio de 79%. O SVM com os parâmetros de regularização $C = 100$ e o tipo de *kernel* = “linear”, apresentou no teste uma acurácia média de 92%, uma precisão média de 93%, *recall* médio de 91% e *f1-score* médio de 91%. Já a DT com critério *criterion* = “gini”, a profundidade máxima da árvore *max_depth* = 5, o número mínimo de amostras necessário para dividir um nó interno *min_samples_split* = 4, e o número mínimo de amostras necessário para estar em um nó folha *min_samples_leaf* = 2, obteve no teste uma acurácia média de 89%, uma precisão média de 89%, *recall* médio de 74% e *f1-score* médio de 78%. Por fim, a ANN, com a função de ativação da camada oculta *activation* = “identity”, a força do termo de regularização L2 *alpha* = 0.01, e a tolerância de otimização *tol* = 0.001, alcançou uma acurácia média de 84%, uma precisão média de 56%, *recall* médio de 61% e *f1-score* médio de 58%. Estes resultados indicam que o SVM apresentou o melhor desempenho geral, seguido pelo KNN, enquanto os algoritmos de DT e ANN apresentaram desempenhos intermediários.

4.2. Avaliação das combinações dos Classificadores

Ao utilizar os melhores ajustes de parâmetros dos algoritmos (KNN, SVM, DT e ANN) para formar os comitês de classificadores (*Bagging*, *Boosting*, *Stacking* e *Voting*), os resultados permitiram avaliar as estratégias que obtiveram maiores desempenhos. Entre os algoritmos individuais, o SVM apresentou um desempenho superior. O método *Boosting* apresentou melhor acurácia média de 94% (Tabela 2). A proporção de previsões dos níveis de interação corretas em relação a todos os estudantes avaliados pelo modelo nos *ensembles* homogêneos (*Bagging* e *Boosting*) destacam-se em média de 1%, em relação aos *ensembles* heterogêneos (*Stacking* e *Voting*).

Tabela 2. Resultados dos algoritmos e modelos *ensembles* a partir da acurácia

Algoritmos	Classificador	<i>Bagging</i>	<i>Boosting</i>	<i>Stacking</i>	<i>Voting</i>	$\mu\alpha$
KNN	0,91	0,90	--	0,92	0,89	0,91
SVM	0,92	0,89	0,92	0,93	0,89	0,91
DT	0,89	0,90	0,95	0,90	0,88	0,90
ANN	0,84	0,82	--	0,89	0,89	0,86
μm	0,89	0,88	0,94	0,91	0,89	0,90

Nota: μm - média dos modelos; $\mu\alpha$ - média dos algoritmos.

A combinação dos algoritmos SVM e DT no método *Boosting* melhorou a acurácia do algoritmo de DT (Tabela 2), e a precisão de ambos os algoritmos (Tabela 3). Chegando a alcançar um aumento na proporção de previsões verdadeiras positivas das interações dos estudantes em relação a todos os resultados classificados como positivamente previstos pelo modelo em média de 1% (Tabela 3). É possível perceber que houve uma mudança de comportamento da precisão em relação à acurácia, ou seja, a precisão média dos *ensembles* heterogêneos (*Stacking* e *Voting*) é maior em média de 2%, em relação aos *ensembles* homogêneos (*Bagging* e *Boosting*), apesar do método *Boosting* permanecer com a melhor precisão em média de 94%.

Tabela 3. Resultados a partir das médias de *precision*

Algoritmos	Classificador		<i>Bagging</i>		<i>Boosting</i>		<i>Stacking</i>		<i>Voting</i>		$\mu\alpha$
	μma	μp	μma	μp	μma	μp	μma	μp	μma	μp	
KNN	0,93	0,91	0,93	0,91	--	--	0,91	0,92	0,92	0,90	0,92
SVM	0,93	0,92	0,90	0,89	0,94	0,93	0,92	0,93	0,92	0,90	0,92
DT	0,89	0,89	0,89	0,90	0,95	0,95	0,89	0,90	0,92	0,89	0,91
ANN	0,56	0,78	0,55	0,76	--	--	0,90	0,90	0,92	0,90	0,78
μm	0,85		0,84		0,94		0,91		0,91		0,89

Nota: μma - média macro; μp - média ponderada; μm - média dos modelos; $\mu\alpha$ - média dos algoritmos

Por outro lado, a combinação dos algoritmos SVM e DT no método *Boosting*, piorou o *recall* médio *macro* do SVM em um decréscimo de 4% (Tabela 4).

Tabela 4. Resultados a partir das médias de *recall*

Algoritmos	Classificador		<i>Bagging</i>		<i>Boosting</i>		<i>Stacking</i>		<i>Voting</i>		μ_a
	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	
KNN	0,75	0,91	0,80	0,90	--	--	0,87	0,92	0,80	0,89	0,86
SVM	0,91	0,92	0,85	0,89	0,87	0,92	0,89	0,93	0,78	0,89	0,89
DT	0,74	0,89	0,84	0,90	0,89	0,95	0,85	0,90	0,77	0,88	0,86
ANN	0,61	0,84	0,59	0,82	--	--	0,84	0,89	0,80	0,89	0,79
μ_m	0,82		0,82		0,91		0,89		0,84		0,86

Nota: μ_{ma} - média macro; μ_p - média ponderada; μ_m - média dos modelos; μ_a - média dos algoritmos

Mesmo assim, manteve-se com o melhor *recall* para na previsão da proporção dos níveis de interação dos estudantes tidos como verdadeiros positivos em relação a todos os casos positivos reais das interações dos estudantes. Ao analisar o *f1-score* (Tabela 5), que combina a precisão e *recall* com valor que varia entre 0 (zero) e 1 (um), sendo que, um limiar mais próximo de 1 indica o modelo com melhor desempenho.

Tabela 5. Resultados a partir das médias de *f1-score*

Algoritmos	Classificador		<i>Bagging</i>		<i>Boosting</i>		<i>Stacking</i>		<i>Voting</i>		μ_a
	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	μ_{ma}	μ_p	
KNN	0,79	0,90	0,84	0,90	--	--	0,89	0,92	0,84	0,89	0,87
SVM	0,91	0,92	0,87	0,89	0,90	0,92	0,90	0,93	0,82	0,88	0,89
DT	0,78	0,88	0,86	0,90	0,91	0,95	0,87	0,90	0,81	0,88	0,87
ANN	0,58	0,80	0,57	0,79	--	--	0,86	0,89	0,84	0,89	0,78
μ_m	0,82		0,83		0,92		0,90		0,86		0,86

Nota: μ_{ma} - média macro; μ_p - média ponderada; μ_m - média dos modelos; μ_a - média dos algoritmos

É possível notar que o método *Boosting* permanece com o melhor desempenho, seguido do método *Stacking* que obteve também bom desempenho ao utilizar todos os algoritmos testados. Porém, percebe-se que os *ensembles* heterogêneos (*Stacking* e *Voting*) tiveram desempenho superior em média de 1% em comparação aos *ensembles* homogêneos (*Bagging* e *Boosting*). Além disso, é importante destacar que o SVM, apesar de se manter em destaque, tem perdas que variam em termos de *f1-score* em média macro de 1% até 9% de desempenho quando combinado nas diferentes estratégias de comitês. Estes resultados se mostram significativos a partir de um grau de significância estabelecido em 1% ou menos. A significância reforça a validade dos resultados e fornece uma base sólida para fundamentar futuras decisões e intervenções com base em um escopo maior de dados. A partir disto, aplicaram-se as melhores combinações dos classificadores (isto é, o comitê *Boosting* e os algoritmos SVM e DT), em uma estratégia de análise e visualização dos indicadores de interação.

4.3. Aplicação das combinações dos Classificadores

Para apoiar tutores humanos na promoção do engajamento estudantil, expuseram-se, por meio de visualizações, as classificações dos níveis de interação dos estudantes, conforme apresentado na Figura 2. As visualizações encontram-se associadas a funcionalidades que permite gerir grupos de estudantes por níveis de interação, visualizar grupos que mais, esporadicamente, ou raramente interagem. Algo que se assemelha ao proposto por Pereira, Gomes, Primo (2023). Ao classificar as interações entre os estudantes e respectivas similaridades de interações, permite-se formar e recomendar pares de ajuda. Neste caso, aplicou-se a classificação, para manter uma lista

atualizada de estudantes por níveis de interação e visível aos tutores humanos. Ao visualizar as interações entre os estudantes, os tutores conseguem acompanhar os três perfis de estudantes (mais, esporadicamente e raramente interagem) a partir de períodos específicos conseguem entender com quem interagem e tomar iniciativas didático-pedagógicas para minimizar situações de desengajamento, e consecutiva evasão e abandono dos cursos. O acompanhamento dos indicadores de engajamento, representado de forma visual, possibilita aos tutores entender que quanto mais um estudante interage, mais ele encontra-se a partir do filtro centralizados ou vice-versa. Além disto, os pares de ajuda resultam da representação das classificações e permitem que os tutores formem grupos, de modo que, os estudantes com baixos níveis de interação possam melhorar seus indicadores de engajamento ao receber ajuda de estudantes que esporadicamente ou mais interagem.



Figura 2. Captura de tela do perfil de um tutor - Tela de visualização da classificação dos níveis de interações de estudantes do aprendizado mediado on-line.

Na aplicação desenvolvida (conforme Figura 2), foi possível identificar e apresentar os padrões de interação dos estudantes a partir das informações coletadas e analisadas pelos classificadores, quanto à quantidade de contatos e amigos, mensagens, pedidos de ajuda, comentários, exercícios realizados no ambiente virtual de aprendizagem. Tal forma de visualizar os padrões pode permitir intervenções como alerta e recomendações didático-pedagógicas para apoiar decisões de tutoria personalizadas a grupos de estudantes, por exemplo, com baixos níveis de engajamento.

5. Considerações Finais

Neste estudo, foram analisados diferentes algoritmos de AM para identificar a abordagem mais eficaz na classificação de indicadores de engajamento dos estudantes no aprendizado on-line. Utilizando um conjunto de dados de interações de estudantes do ensino-aprendizado com atividades mediadas on-line, foram avaliados os parâmetros

dos algoritmos KNN, SVM, DT e ANN, bem como as combinações de *ensembles* heterogêneos (*Stacking* e *Voting*) e homogêneos (*Bagging* e *Boosting*), mais adequados para prever o nível de interação dos estudantes que mais interagem, interagem esporadicamente ou raramente. Os resultados mostraram, em algumas métricas, melhor desempenho dos classificadores heterogêneos em comparação aos homogêneos, porém no geral o método *Boosting* com os algoritmos SVM e DT apresentou desempenho superior em todas as métricas analisadas, superando os demais desempenhos dos algoritmos. Estes resultados sugerem que a combinação de diferentes estratégias de AM pode permitir uma melhora na classificação dos níveis de interações dos estudantes e apoiar tutores na visualização de indicadores do engajamento no aprendizado on-line. Uma aplicação dos resultados foi apresentada, para permitir que tutores humanos visualizem os padrões descobertos, porém se faz necessários futuras avaliações da efetividade da aplicação.

5.1. Limitações

Não foram exploradas exaustivamente as diferentes abordagens de AM. Além disto, fatores como usabilidade da plataforma e coleta de dados eventuais da atuação humana não foram explorados neste estudo. A compreensão do papel dos tutores humanos e a percepção destes sobre o apoio proporcionado pelo sistema, quanto ao incentivo às interações e a consequente promoção do engajamento dos estudantes no processo de ensino-aprendizagem on-line, é algo que pode ser mais explorado. Estas limitações podem ser supridas em estudos futuros ao envolver diferentes contextos de uso da plataforma social de aprendizagem em diferentes modalidades de ensino-aprendizado.

5.2 Estudos Futuros

Com base nas conclusões e limitações, as possibilidades futuras desta pesquisa incluem realizar estudos no contexto dos participantes e diversificar as amostras para verificar a generalização dos resultados encontrados. Podem ser exploradas outras técnicas de AM e outras métricas de avaliação de desempenho para melhorar a precisão na classificação de interações dos estudantes no aprendizado on-line. As abordagens utilizadas para concepção das estratégias de apoio a atuação de tutores humanos cabe futuros testes de efetividade da aplicação na classificação e prevenção de baixos indicadores de engajamento dos estudantes. Além disto, a possibilidade de identificar quais as efetivas contribuições para o processo de ensino-aprendizagem em longo prazo.

Conformidade com os padrões éticos: não há potenciais conflitos de interesse nesse estudo. Confirmamos seguir preceitos morais e éticos e cumprir com a Lei n.º 13.709/2018 vigente sobre proteção de dados. A utilização dos dados ocorreu de forma devidamente concedida por termos fornecidos pela instituição e envolvidos.

Referências

- Adejo, O. W.; Connolly, T. Predicting student academic performance using multi-model heterogeneous ensemble approach. **Journal of Applied Research in Higher Education**, v. 10, n. 1, p. 61-75. 2018. <https://doi.org/10.1108/JARHE-09-2017-0113>
- Alhazmi, A. K. *et al.* AI's Role and Application in Education: Systematic Review. **Intelligent Sustainable Systems: Selected Papers of Worlds4 2022**, v. 1, p. 1-14. 2023. https://doi.org/10.1007/978-981-19-7660-5_1

- Almarghani, E. M.; Mijatovic, I. Factors affecting student engagement in HEIs-it is all about good teaching. **Teaching in higher education**, v. 22, n. 8, p. 940-956. 2017. <https://doi.org/10.1080/13562517.2017.1319808>
- Archambault, L.; Leary, H.; Rice, K. Pillars of online pedagogy: A framework for teaching in online learning environments. **Educational Psychologist**, v. 57, n. 3, p. 178-191. 2022. <https://doi.org/10.1080/00461520.2022.2051513>
- Brohi, S. N. *et al.* Accuracy comparison of machine learning algorithms for predictive analytics in higher education. In: **Emerging Technologies in Computing: Second International Conference, iCETiC 2019, London, UK, August 19–20, 2019, Proceedings 2**. Springer International Publishing. p. 254-261. 2019. https://doi.org/10.1007/978-3-030-23943-5_19
- Buono, P. *et al.* Assessing student engagement from facial behavior in on-line learning. **Multimedia Tools and Applications**, p. 1-19. 2022. <https://doi.org/10.1007/s11042-022-14048-8>
- Chang, c. *et al.* An ensemble model using face and body tracking for engagement detection. In: **Proceedings of the 20th ACM international conference on multimodal interaction**. p. 616-622. 2018. <https://doi.org/10.1145/3242969.3264986>
- De Bruin, L. R. Collaborative learning experiences in the university jazz/creative music ensemble: Student perspectives on instructional communication. **Psychology of Music**, v. 50, n. 4, p. 1039-1058. 2022. <https://doi.org/10.1177/03057356211027651>
- Džeroski, S.; ŽENKO, B. Is combining classifiers with stacking better than selecting the best one?. **Machine learning**, v. 54, p. 255-273. 2004. <https://doi.org/10.1023/b:mach.0000015881.36452.6e>
- Echeverria, V. *et al.* Designing Hybrid Human-AI Orchestration Tools for Individual and Collaborative Activities: A Technology Probe Study. **IEEE Transactions on Learning Technologies**. 2023. <https://doi.org/10.1109/TLT.2023.3248155>
- Evangelista, E. A hybrid machine learning framework for predicting students' performance in virtual learning environment. **International Journal of Emerging Technologies in Learning (iJET)**, v. 16, n. 24, p. 255-272. 2021. <https://doi.org/10.3991/ijet.v16i24.26151>
- Evangelista, E.; SY, B. D. An approach for improved students' performance prediction using homogeneous and heterogeneous ensemble methods. **International Journal of Electrical and Computer Engineering**, v. 12, n. 5, p. 5226. 2022. <https://doi.org/10.11591/ijece.v12i5.pp5226-5235>
- Kamath, S. *et al.* Engagement analysis of students in online learning environments. In: **Machine Learning and Big Data Analytics (Proceedings of International Conference on Machine Learning and Big Data Analytics (ICMLBDA) 2021)**. Springer International Publishing. p. 34-47. 2022. https://doi.org/10.1007/978-3-030-82469-3_4
- Kim, J.; Davis, T.; Hong, L. Augmented Intelligence: Enhancing Human Decision Making. In: **Bridging Human Intelligence and Artificial Intelligence**. Cham:

- Springer International Publishing**. p. 151-170. 2022. https://doi.org/10.1007/978-3-030-84729-6_10
- Mayer, R. E. Multimedia learning (2nd ed.). **Cambridge University Press**. 2009.
- Moraitis, N. *et al.* On the assessment of ensemble models for propagation loss forecasts in rural environments. **IEEE Wireless Communications Letters**, v. 11, n. 5, p. 1097-1101. 2022. <https://doi.org/10.1109/LWC.2022.3157520>
- Muhlbaier, M. D.; Topalis, A.; Polikar, R. Learn, N. C. Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes. **IEEE transactions on neural networks**, v. 20, n. 1, p. 152-168, 2008. <https://doi.org/10.1109/TNN.2008.2008326>
- Nti, I. K.; Adekoya, A. F.; Weyori, B. A. A comprehensive evaluation of ensemble learning for stock-market prediction. **Journal of Big Data**, v. 7, n. 1, p. 1-40. 2020. <https://doi.org/10.1186/s40537-020-00299-5>
- Pereira, A. J.; Gomes, A. S.; Primo, T. T. Especificação de Sistema de Recomendação Educacional de Incentivo as Interações em Plataforma Social de Aprendizagem. **RENOTE**, v. 20, n. 2, p. 1-10. 2022. <https://doi.org/10.22456/1679-1916.129143>
- Pereira, A. J. *et al.* Learning Mediated by Social Network for Education in K-12: Levels of Interaction, Strategies, and Difficulties. **Education Sciences**, v. 13, n. 2, p. 100. 2023. <https://doi.org/10.3390/educsci13020100>
- Pereira, A. J.; Gomes, A. S.; Primo, T. Uma Abordagem de Sistema de Tutoria Inteligente para Cooperação com a Atuação de Tutores Humanos. **RENOTE**, Porto Alegre, v. 21, n. 2, p. 208-219, 2023. <https://seer.ufrgs.br/index.php/renote/article/view/137742>
- Refaeilzadeh, P., Tang, L., Liu, H. Cross-Validation. In: LIU, L., ÖZSU, M.T. (eds) **Encyclopedia of Database Systems**. Springer, Boston, MA. 2009. https://doi.org/10.1007/978-0-387-39940-9_565
- Rokach, L. Ensemble-based classifiers. **Artificial intelligence review**, v. 33, p. 1-39. 2010. <https://doi.org/10.1007/s10462-009-9124-7>
- Srinivasa, KG; Kurni, M.; Saritha, K. Aproveitando o poder da IA para a educação. In: **Métodos de aprendizagem, ensino e avaliação para aprendizes contemporâneos: pedagogia para a geração digital**. Cingapura: Springer Nature Cingapura. p. 311-342. 2022. https://doi.org/10.1007/978-981-19-6734-4_13
- Saleem, F. *et al.* Intelligent decision support system for predicting student's E-learning performance using ensemble machine learning. **Mathematics**, v. 9, n. 17, p. 2078. 2021. <https://doi.org/10.3390/math9172078>
- Saleh, K.; YU, K.; Chen, F. Video-Based Student Engagement Estimation via Time Convolution Neural Networks for Remote Learning. In: **AI 2021: Advances in Artificial Intelligence: 34th Australasian Joint Conference, AI 2021, Sydney, NSW, Australia, February 2-4, 2022, Proceedings**. Cham: Springer International Publishing. p. 658-667. 2022. https://doi.org/10.1007/978-3-030-97546-3_53

- Salta, K. *et al.* Shift from a traditional to a distance learning environment during the COVID-19 pandemic: university students' engagement and interactions. **Science & Education**, v. 31, n. 1, p. 93-122. 2022. <https://doi.org/10.1007/s11191-021-00234-x>
- St-Hilaire, F. *et al.* A New era: Intelligent tutoring systems will transform online learning for millions. **arXiv preprint arXiv:2203.03724**. 2022. <https://doi.org/10.48550/arXiv.2203.03724>
- Sweller, J. Cognitive load theory, learning difficulty, and instructional design. **Learning and Instruction**, v. 4, n. 4, p. 295–312. 1994. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Taser, P. Y. Application of bagging and boosting approaches using decision tree-based algorithms in diabetes risk prediction. In: **Proceedings**. MDPI. p. 6. 2021. <https://doi.org/10.3390/proceedings2021074006>
- Wang, G. *et al.* A comparative assessment of ensemble learning for credit scoring. **Expert systems with applications**, v. 38, n. 1, p. 223-230, 2011. <https://doi.org/10.1016/j.eswa.2010.06.048>
- Werang, B. R.; Leba, S. M. R. Factors Affecting Student Engagement in Online Teaching and Learning: A Qualitative Case Study. **Qualitative Report**, v. 27, n. 2. 2022. <https://doi.org/10.1080/13562517.2017.1319808>
- Zafari, M. *et al.* Artificial intelligence applications in K-12 education: A systematic literature review. **IEEE Access**. 2022. <https://doi.org/10.1109/ACCESS.2022.3179356>