

Classificação do Diálogo Freireano em Mensagens de Fóruns de Discussão: Uma Análise de Desempenho do TF-IDF e o BERT para Sentenças

Francisco Romes da Silva Filho¹, Gabriel Antoine Louis Paillard²,
Rafael Augusto Ferreira do Carmo², Ernesto Trajano de Lima²,
Michel Sales Bonfim¹

¹ Programa de Pós-Graduação em Computação (PCOMP)
Universidade Federal do Ceará (UFC) – Quixadá – CE – Brasil

²Instituto Universidade Virtual (IUVI)
Universidade Federal do Ceará (UFC) – Fortaleza – CE – Brasil

romesfilho_cc@alu.ufc.br, {gabriel, carmorafael, ernesto}@virtual.ufc.br, michelsb@ufc.br,

Abstract. *Freire's dialogue concepts have been applied and organized in recent years to be able to delimit forum message characteristics with theory. This work proposed a binary text classifier for the presence of Valorization of Autonomy in forum messages and also performed a comparison of the performance of two text encoding techniques. The results indicated with statistical significance that the Sentence-BERT was superior to the TF-IDF method as a encoding method.*

Resumo. *Os conceitos do diálogo freireano foram aplicados e organizados nos últimos anos de forma a ser possível delimitar características de mensagens de fóruns com a teoria. Esse trabalho propôs um classificador de texto binário para a presença da Valorização da Autonomia em mensagens de fóruns e ainda realizou uma comparação do desempenho de duas técnicas de codificação de texto. Os resultados indicaram com significância estatística que o Sentence-BERT foi superior ao método TF-IDF como método de codificação.*

1. Introdução

Ambientes Virtuais de Aprendizagem (AVA) são programas de computadores comumente utilizados no ensino a distância. Eles incorporam uma série de ferramentas para o auxílio às atividades de ensino. Dentre tais ferramentas, destacamos os fóruns, que constituem espaços de discussão assíncrona mais presentes em AVA e que permitem que os participantes do curso, docentes e discentes, mantenham discussões relativas às temáticas abordadas, o que pode ser feito a qualquer tempo ou lugar [Wu et al. 2020].

Tendo em vista isso, percebe-se a necessidade de estruturar o entendimento e avaliar diferentes elementos da comunicação mediada por computador, especificamente em fóruns de AVAs [Barbosa et al. 2021, Hlosta et al. 2017, Gomedede et al. 2020]. Pois é, por exemplo, que através do conteúdo produzido em fóruns de discussão de AVAs, modelos de aprendizado de máquina (AM) – assim como outras soluções computacionais – buscam identificar possibilidades de evasão, resultado em uma disciplina, estilo de aprendizagem, dentre outros [Al-Shabandar et al. 2018, Gomedede et al. 2020].

Seguindo essa abordagem de explorar o potencial da experiência educacional em um fórum, tendo a vivência do estudante como um ponto a ser bastante valorizado. As ideias desenvolvidas pelo Paulo Freire defendem a participação autônoma dos estudantes na construção do seu processo de aprendizagem e são antagônicas a ideia de que a educação é uma simples transferência de conhecimentos [David 2010]. As relações sociais estabelecidas entre os atores de um ambiente estudantil, por exemplo, podem ser tratadas como o diálogo e, segundo o Paulo Freire, são cinco os pilares do diálogos: o amor, a humildade, a fé nos homens, a esperança e o pensar crítico.

A automatização da categorização dos tipos de diálogos existentes entre usuários de fóruns de discussão adicionaria um instrumento extremamente importante e valioso para os mantenedores dos fóruns de discussão [David et al. 2015, Singh and Lal 2013]. [David et al. 2015] afirma “que o conhecimento empírico utilizado no processo manual de identificação das categorias do diálogo poderia ser melhor compreendido se fosse formalizado de alguma maneira”.

Em [David et al. 2015], optou-se por desenvolver uma ontologia para o domínio do Diálogo. Contudo, tendo vista que o AM é bastante utilizado para identificação automática de categorias (classificação) [Singh and Lal 2013]. Assim como outros estudos que realizaram automações com AM em fóruns, o presente estudo escolheu desenvolver um classificador binário com AM. Da mesma forma que outros estudos, esse estudo requer atenção para o processo engenharia de atributos que pode contribuir e facilitar o processo de predição, com a redução de dimensionalidade – problema que pode afetar até técnicas avançadas e não lineares - e interpretabilidade [Oliveira et al. 2019, Yoo et al. 2022].

Portanto, o objetivo do presente trabalho é propor através de um classificador de conteúdo de mensagens de fóruns de discussão de AVA, com base no Diálogo Freireano. Especificamente, para o domínio desse estudo, classificação de texto, o presente trabalho apresenta comparação do desempenho de dois métodos de vetorização de texto – técnica de engenharia de atributos – são eles o Frequência do Termo – Frequência Inversa dos Documentos (*TF-IDF*, do inglês *Term Frequency - Inverse Document Frequency*) e *Sentence BERT*.

Este artigo está organizado da seguinte forma: A Seção 2 descreve conceitos importantes para este estudo. A Seção 3 apresenta alguns trabalhos relacionados à proposta deste trabalho. A Seção 4 apresenta a metodologia e suas etapas definidas para guiar o presente trabalho. Os resultados e análises obtidos são apresentados, discutidos e comentados na Seção 5. Por fim, a Seção 6 apresenta as considerações finais.

2. Fundamentação teórica

Esta seção contém a descrição dos conceitos mais importantes que orientam esse trabalho.

2.1. Diálogo Freireano

O autor Paulo Freire manifesta apoio ao conceito da educação dialógica. Expondo que o diálogo é uma forma de preparar as pessoas, e a habilitarem a construir algo novo, coletivamente. Dessa forma, o diálogo teria 5 pilares: amor, humildade, fé nos homens, esperança e um pensar crítico [David et al. 2015]. Ao longo dos anos, houve a expansão desses conceitos, introdução na educação à distância e a busca por organizar esses pilares

em fóruns de discussão de ambientes virtuais de aprendizagem (AVA) – forma explícita de estabelecer um diálogo em ferramentas de ensino à distância.

Uma organização possível para esses pilares, resgatada por [David et al. 2015] seria que: “(i) amor, seria traduzido em Afetividade, e identificada pela presença de respeito mútuo e proximidade; (ii) humildade, traduzida em Simetria Discursiva, e identificada pela igualdade de papéis entre os participantes (alunos e professor-tutor); (iii) fé nos homens, traduzida em Valorização da autonomia e identificada nas mensagens do professor-tutor pelo incentivo à livre expressão por parte dos alunos; (iv) esperança, traduzida em Exercício da Autonomia e identificada pelo aprofundamento dos conceitos pelos alunos; (v) pensar crítico, traduzido em Reflexividade crítica e identificado na demonstração de reflexão quanto ao próprio processo de aprendizagem e/ou dos demais interlocutores”.

2.2. Aprendizado de Máquina

Aprendizado de Máquina (AM) é um artefato computacional muito importante na análise de dados. A aplicação do AM permite realiza automatizações, como, por exemplo, a categorizações através da tarefa de classificação do aprendizado supervisionado. Dessa forma, com o uso de AM seria possível automatizar a classificação do Diálogo Freireano. [L’Heureux et al. 2017]. Por isso, o conceito é introduzido aqui.

Projetos de AM, via de regra e em essência, são constituídos de duas fases: treinamento e teste (avaliação). Muitas vezes, quando se fala em classificação, em particular a classificação de texto, as seguintes etapas são frequentemente executadas: (i) identificação de um atributo alvo nos dados (rótulo/classe) e os valores de classe para realização do treinamento; (ii) identificação de um subconjunto dos atributos necessários para a realização da classificação (ou seja, redução de dimensionalidade); (iii) vetorização de texto para transformar a linguagem natural entendível para a máquina; (iv) fazer o modelo aprender usando os dados de treinamento como entrada; e (v) utilizar o modelo treinado para classificar os dados desconhecidos [Burkov and Lutz 2019].

2.2.1. Vetorização de Texto

Inicialmente, para que um modelo de AM aprenda com os dados de entrada é necessário ser feita a transformação desse dado a uma representação interpretável pela máquina. Por isso, apresenta-se aqui o conceito de vetorização de texto ou codificação de texto que é uma técnica de extração de atributos considerada um passo primordial para pré-processar e transformar texto (linguagem natural) em uma estrutura de atributos em que seja possível aplicar modelagens matemáticas dos classificadores [Chistol 2020].

Uma abordagem comum para a aplicação desse conceito é a técnica de *Bag of Words* que através da frequência das palavras em um texto gera uma representação. Por exemplo, dado o seguinte dicionário de palavras {“Um”, “Livro”, “Isso”, “Boa”, “Sara”, “Texto”, “É”, “Troca”} a frase “Isso é um texto” teria a seguinte representação baseada na ordem e frequência das palavras: (“1”, “0”, “1”, “0”, “0”, “1”, “1”, “0”) [Chistol 2020]. Neste trabalho, são abordadas duas técnicas de naturezas diferentes para o processo de codificação de texto, Frequência do Termo – Frequência Inversa dos Documentos (*TF-IDF*, do inglês *Term Frequency - Inverse Document Frequency*) e *Sentence BERT*.

O TF-IDF é um método estatístico projetado para representar a importância do vocabulário para o conjunto de dados da família de técnicas *bag of words*. O valor do TF-IDF é proporcional ao número de vezes que a palavra aparece no documento, contudo é geralmente compensado pela frequência da palavra nos dados, ajudando a evitar que algumas palavras geralmente estejam presentes nos dados e sejam dados maior importância [Yang 2017]. TF-IDF é o produto da frequência do termo (TF) pela frequência inversa do documento (IDF), TF é o número de vezes que o termo t aparece em uma amostra e o IDF é igual a $\log \frac{N}{N_t}$. Considere que a variável t representa uma palavra e N é o número total de documentos e N_t é o número de documentos com o termo t [Mishra and Vishwakarma 2015].

Uma alternativa a abordagem de *bag of words* apresentada anteriormente para o processo de codificação de texto como vetores de números é o processo conhecido como *sentence embedding*. Esse processo tem como característica o mapeamento das sentenças – não somente palavras e sim um conjunto maior de palavras – para vetores de número reais. Esse processo pode ser considerado como uma extensão de outro método bastante conhecido neste domínio que é o *word embedding*, nele existe um mapeamento com inteiros, mas observa-se com que frequência uma palavra aparece em cada contexto [Reimers and Gurevych 2019a].

[Reimers and Gurevych 2019a] apresenta o Sentence-BERT (SBERT), o BERT para Sentenças, “como uma modificação do uso das redes siamesas e trigêmeas BERT que consegue derivar incorporações de sentenças semanticamente significativas”, o uso desse tipo de redes está geralmente relacionado a computação da similaridade entre duas entradas.

Essa solução é baseada em um importante modelo presente no estado da arte que é o BERT (do inglês *Bidirectional Encoder Representations from Transformers*), uma rede neural do tipo *transformer* pré-treinada com vários métodos do estado da arte do processamento de linguagem natural (PLN ou NLP, do inglês *Natural Language Processing*).

2.3. Algoritmos de Classificação

Por intermédio da aplicação do passo detalhado anteriormente, os dados já se tornam interpretáveis por modelos de AM. Dessa forma, é possível selecionar algoritmos para classificação binária, tarefa desse estudo, para realizar a predição da classe alvo. Dois bastante conhecidos e presentes nessa pesquisa são a Regressão Logística e o *Gradient Boosting*.

A Regressão Logística é um método que utiliza-se de conceitos de probabilidade e através de uma função logística que dada uma entrada e um limiar comumente definido em 0.5, devolve um valor no intervalo entre 0 e 1. Então, o valor da classe é definido na posição desse valor retornado em relação ao limiar definido [Raschka and Mirjalili 2019]. Já o *Gradient Boosting* é um algoritmo pertencente a classe de *ensemble learning* em que agrupam múltiplos modelos preditivos para realizar a predição da classe alvo. A composição do *Gradient Boosting* é feita a partir de várias unidades de árvores de decisão para a predição final.

2.3.1. Métricas

Por fim, para interpretar e avaliar o desempenho dos modelos construídos, é importante avaliar com métricas utilizadas e discutidas pelo estado da arte. As métricas normalmente presentes em projetos de AM para a tarefa de classificação binária são acurácias, precisão, revocação e *f1-score*. A acurácia representa quantos exemplos foram de fato classificados corretamente. A precisão é uma métrica definida pela razão entre a quantidade de exemplos classificados como positivos corretamente e o total de exemplos classificados como positivos. A revocação enfatiza os erros por falso negativo. O *F1 Score* é definido pela média harmônica entre a precisão e a revocação, dessa forma, ela consegue mensurar melhor o acerto para cada classe do conjunto de dados [Sammut and Webb 2011].

Nesta seção foram discutidos temas centrais para o estudo, tendo em vista que o presente trabalho usa um conjunto de dados, onde o processo de anotação foi baseado no Diálogo Freireano e o objetivo é construir um modelo computacional que usa o AM. Além disso, há técnicas auxiliares e derivadas de AM – como vetorização de texto, para simplificar e melhorar o processo de classificação – e, ao final, foram discutidas métricas mais comuns aos problemas de AM do tipo classificação (aprendizado supervisionado).

3. Trabalhos Relacionados

Nessa seção apresentaremos trabalhos relacionados a pesquisa desenvolvida pelo presente trabalho.

[Gomede et al. 2020] reforça a ideia que é possível classificar os estudantes a partir da forma que eles reconhecem, processam e armazenam informações. Os autores desse trabalho propõem um classificador multiclasse para classificar os estilos de aprendizagem com base na teoria de Felder-Silverman. Como resultados, o estudo apresenta uma rede neural artificial – com Multi Layer Perceptron (MLP) – para identificação automática de estilos de aprendizagem a partir da teoria educacional e um conjunto de atributos mais importantes para essa tarefa, oferecendo maior embasamento para tomada de decisões e intervenções nos ambientes virtuais de aprendizagem. O melhor modelo apresentado pelo trabalho obteve 85%, 76%, 75% e 80% de acurácia em cada categoria presente nos dados, “active_reflective”, “sensing_intuitive”, “visual_verbal” e “sequential_global”, respectivamente.

[David et al. 2015] concebe um estudo baseado na teoria educacional de Paulo Freire argumentando em favor da educação dialógica. Com base nessa teoria, os autores apresentam um modelo de uma ontologia. É apresentado que o diálogo é composto de cinco parâmetros: afetividade, simetria discursiva, valorização da Autonomia, exercício da autonomia e reflexividade crítica. Dessa forma, o estudo analisa o conteúdo de mensagens de fóruns de discussão de uma Ambiente Virtual de Aprendizagem (AVA) para inferir a forma que esses parâmetros se apresentam – totalmente, parcial, ou não se apresenta (ausência). Através da análise do conteúdo das mensagens, conceberam-se as seguintes categorias para identificação automática: afetividade, valorização da autonomia e reflexividade crítica.

[Ferreira et al. 2020] apresenta um classificador automático para o conteúdo de mensagens em fóruns de discussão online de acordo com categorias de presença social. Essas categorias derivam de uma teoria educacional conhecida como Comunidade

de Inquirição (CoI, do inglês *Community of Inquiry*). Portanto, as mensagens foram anotadas por especialistas como sendo das categorias *Affective*, *Interactive* e *Cohesive*, com cada mensagem podendo ter zero ou mais categorias presentes. Os autores empregaram um modelo preditivo construído com o algoritmo floresta aleatória com um conjunto de dados, dividido em 86% para treino e 14% para teste. Portanto, ao final o trabalho apresentou 3 classificadores binários com floresta aleatória para cada categoria. O desempenho dos modelos teve médias de 89% e 73% em termos de acurácia e *kappa*, respectivamente.

Em resumo, diversos estudos aliam conceitos e práticas educacionais e computacionais visando apresentar conhecimento e propor intervenções em ambientes educacionais. Igualmente, a presente pesquisa planeja aliar o diálogo freireano segundo a abordagem de [David et al. 2015] e realizar a classificação de mensagens de fóruns de discussão com AM, empregando o aprendizado supervisionado com classificação binária.

4. Metodologia

Nessa seção os passos metodológicos para a execução desse experimento são descritos.

4.1. Conjunto de Dados

O conjunto de dados utilizados por esse estudo é o mesmo utilizado por [David et al. 2015] que projeta ontologias para a representação do diálogo freireano. Os dados apresentam 295 amostras de mensagens de fóruns de discussão de uma turma de ensino superior de uma universidade pública brasileira, anotados com base na teoria do diálogo freireano. As amostras foram anotadas com base na presença de Valorização da Autonomia (fé nos homens). A anotação feita em [David et al. 2015] indica a presença e a ausência de valorização da autonomia nas amostras.

4.2. Pré-processamento

Como passo auxiliar a metodologia, são executados transformações de dados na anotação dos dados, para enquadrar o conjunto de dados para construir classificadores binários. O presente trabalho considerou que as amostras que tiveram indicações da presença de Valorização da Autonomia serão considerados da classe positiva (1) e as que indiquem ausência ou indefinição de Valorização da Autonomia serão consideradas negativas (0).

Antes do processo de codificação das mensagens para serem inseridos como entradas para os modelos. Foram feitos procedimentos utilizando a biblioteca spaCy¹, como a retirada de pontuação, remoção de dígitos, remoção de *stop words* – palavras que podem ser consideradas irrelevantes, como: as, e, os, de, para, com, sem, foi – e, por último, foi aplicado a técnica de lematização – técnica para reduzir a palavra à sua raiz, retirando todas as inflexões e chegando ao *lemma*. Porém, essa redução sempre resultará em uma palavra que realmente existe na gramática. Finalmente, aplica-se a vetorização de texto. O presente trabalho escolheu duas abordagens, o uso do *TF-IDF* e *Sentence BERT* – com a implementação fornecida por [Reimers and Gurevych 2019b].

¹spaCy – <https://spacy.io>

4.3. Treinamento e Avaliação dos modelos

Nesse passo, a linguagem *Python* e as bibliotecas *scikit-learn*², *XGBoost*³ e *mlxtend*⁴ são utilizadas para construir, treinar e avaliar os modelos. Para selecionar o melhor modelo com base em seu desempenho é utilizado o teste *T* pareado *5x2cv*. Esse procedimento foi proposto por [Dietterich 1998], como uma alternativa para tratar problemas de outros métodos (como o teste *t k-fold cross validation*). Esse método considera dois estimadores para uma tarefa supervisionada. Normalmente, utiliza-se a divisão dos dados do tipo *hold-out* (treino e teste). No teste *t* pareado *5x2cv* é realizado também uma divisão em duas partes, porém a divisão – 50% para o conjunto de treino e 50% para o conjunto de teste – é repetida 5 vezes [Raschka 2018, Dietterich 1998].

A cada uma das 5 interações ajusta-se os dois modelos comparados ao conjunto de treino e avalia-se seu desempenho com o conjunto de teste. O próximo passo é inverter os conjuntos de treino e teste e calcular o desempenho novamente, o que resulta em 2 medidas de diferença de desempenho. Então, estima-se a média e variância das diferenças. A variância é calculada por 5 interações e usada para calcular a estatística *t* [Raschka 2018, Dietterich 1998].

Na estatística *t*, assume-se que segue aproximadamente uma distribuição *t* com 5 graus de liberdade, sob a hipótese nula de que os dois modelos comparados têm desempenho igual. Se o valor de *p* for menor que $\alpha = 0.05$, rejeitamos a hipótese nula e aceitamos haver uma diferença significativa nos dois modelos [Raschka 2018, Dietterich 1998].

Por fim, o desempenho dos modelos é apresentado em termos de *f1 score*. Essa métrica foi escolhida, por conta da maior capacidade dela de mensurar melhor o desempenho de modelos com dados desbalanceados.

5. Resultados

Essa seção apresenta os resultados alcançados dos experimentos realizados pelo presente trabalho.

5.1. Análise Exploratória dos Dados

O conjunto de dados têm 181 mensagens com a presença da categoria valorização da autonomia e 114 mensagens não apresentam a categoria. Dessa forma, o conjunto de dados têm 61% de amostras da classe positiva e 39% da classe negativa. Com base nas palavras do conjunto de dados, uma nuvem de palavras foi gerada para a ilustração dos termos mais presentes, como pode ser observada na Figura 1.

5.2. Comparação entre TF-IDF e SBERT

O intuito inicial desses experimentos é avaliar o desempenho dos codificadores e identificar qual seria o melhor para a construção do modelo final. Para isso, foram escolhidos a Regressão Logística, um modelo linear clássico da biblioteca *scikit-learn*, e o *XGBoost*, modelo de uma biblioteca otimizada de *Gradient Boosting*, ele é bem-sucedido nas

²*scikit-learn* – <https://scikit-learn.org/>

³*XGBoost* – <https://xgboost.readthedocs.io/en/stable/>

⁴*mlxtend* – <http://rasbt.github.io/mlxtend/>

Figura 1. Nuvem de palavras do conjunto de dados



competições do Kaggle⁵. Outro ponto é que as principais soluções em KDD-CUP15⁶ usaram o *XGBoost* [Hlosta et al. 2017].

Observa-se que os codificadores produziram dados com dimensionalidades diferentes e o *TF-IDF* produziu com uma alta dimensionalidade – já que sempre apresentou uma matriz com mais de 2000 colunas. De forma contrária, a codificação do *Sentence BERT* apresentou sempre 512 colunas. Somente a codificação *TF-IDF* teve a sua entrada com um pré-processamento prévio, para diminuir a dimensionalidade da entrada.

As hipóteses do nosso experimento são: H_0 - Não há diferença estatisticamente significativa no desempenho dos modelos e H_1 - Há diferença estatisticamente significativa no desempenho dos modelos. A H_0 é a nossa hipótese nula, ou seja, quando o *p-value* for maior ou igual ao *alpha* de 0.05 mantemos afirmação de H_0 . Caso contrário, rejeita-se H_0 e aceita H_1 .

A Tabela 1 apresenta os resultados do experimento em que foram comparadas dois estimadores do tipo Regressão Logística, em que cada um foi construído com uma codificação (*TF-IDF* e *Sentence BERT*). Observando os valores, pode-se verificar que é possível rejeitar H_0 , então, existe uma diferença significativa entre utilizar diferentes codificações para a Regressão Logística.

Tabela 1. Teste T pareado: *TF-IDF* e *Sentence BERT*

Estatística T	Alpha	p-value
-7.0345	0.05	0.0009

Já a Tabela 2 apresenta os resultados do experimento em que foram comparadas dois estimadores do tipo *XGBoost*, em que cada um foi construído com uma codificação (*TF-IDF* e *Sentence BERT*). Observando os valores, pode-se verificar que é possível rejeitar H_0 , então, existe uma diferença significativa entre utilizar diferentes codificações para o *XGBoost* também.

Tabela 2. Teste T pareado: *TF-IDF* e *Sentence BERT*

Estatística T	Alpha	p-value
-2.7113	0.05	0.0422

⁵<https://www.kaggle.com>

⁶<http://kddcup2015.com>

É possível observar com base nos resultados que com significância estatística, pode-se afirmar que a codificação *Sentence BERT* tem um desempenho superior ao *TF-IDF*. Mesmo ao utilizar um algoritmo capaz de lidar bem com grande dimensionalidade, como *XGBoost*.

5.3. TF-IDF – Comparação entre Regressão Logística e XGBoost

A Tabela 3 apresenta os valores do Teste T pareado para comparação de dois algoritmos de AM diferentes, utilizando a mesma codificação de mensagens, o *TF-IDF*. Os valores indicam que é possível rejeitar H_0 . Portanto, existe uma diferença significativa entre os algoritmos. Ambos algoritmos receberam como entrada dados com grande dimensionalidade, porém, *XGBoost* teve um melhor desempenho, por conta de ser um algoritmo mais robusto, por utilizar de uma técnica de agrupamento (*ensemble*) de vários modelos mais simples. Veja a Tabela 7 para conferir o desempenho de todos os modelos.

Tabela 3. Teste T pareado: Regressão Logística e XGBoost

Estatística T	Alpha	p-value
-3.6416	0.05	0.0149

A Tabela 4 apresenta os coeficientes do modelo de regressão logística. Através dos coeficientes do modelo de regressão logística espera-se encontrar um maior nível de interpretabilidade do modelo. Os coeficientes indicam o aumento ou redução da probabilidade da amostra ser classificada como presente ou ausente. Então, dessa forma, é possível através dos coeficientes do modelo avaliar palavras ditas como atributos terão mais influência na classificação. De forma similar a tabela de coeficientes, a Tabela 5 apresenta a importância dos atributos (palavras presentes nas mensagens) para a classificação com *XGBoost*. As palavras com mais influência de ambos modelos não aparentam isoladamente se enquadrar nas representações mais comuns da categoria valorização da autonomia e nem aparecem nas palavras mais frequentes do dicionário do conjunto de dados [David et al. 2015].

Tabela 4. Tabela de coeficientes do modelo de regressão logística com TF-IDF

Positivos		Negativos	
Peso	Atributo	Peso	Atributo
0.871	concordo	-0.440	papel
0.772	peçoal	-0.444	criar
0.688	comentário	-0.458	ideal
0.661	tiago	-0.470	fórum
0.600	silveira	-0.488	promover
0.534	atribuição	-0.554	pedagógico
0.505	<BIAS>	-0.589	participar
0.504	colocar	-0.597	trabalho
0.502	wesley	-0.665	processo
0.501	principalmente	-0.773	professor

Tabela 5. Tabela de importância dos atributos no modelo XGBoost com TF-IDF

Positivos		Negativos	
Peso	Atributo	Peso	Atributo
0.0579	tutoria	0.0234	complementar
0.0274	principal	0.0220	concordo
0.0257	muito	0.0216	participação
0.0253	específico	0.0214	desenvolvimento
0.0245	promover	0.0199	resolução

5.4. SBERT – Comparação entre Regressão Logística e XGBoost

A Tabela 6 apresenta os valores do Teste T pareado para comparação de dois algoritmos de AM diferente, utilizando a mesma codificação de mensagens, o *Sentence BERT*. Os valores indicam que não é possível rejeitar H_0 . Então, não existe uma diferença significativa entre os algoritmos. Ambos algoritmos receberam como entrada dados com menor dimensionalidade, o XGBoost teve uma ligeira vantagem no desempenho. Veja a Tabela 7 para constatar o desempenho de todos os modelos.

Tabela 6. Teste T pareado: Regressão Logística e XGBoost

Estatística T	Alpha	p-value
-0.0944	0.05	0.9285

Como a Tabela 7 nos apresenta por meio do *F1 Score* médio e a avaliação estatística realizada com o Teste T pareado nos revela, a codificação do tipo *Sentence BERT* apresenta vantagem em relação ao *TF-IDF*. O XGBoost demonstrou uma vantagem em relação a Regressão Logística, até mesmo quando utiliza-se a codificação *TF-IDF* que como os resultados apresentaram produziu uma grande dimensionalidade, e considerando o algoritmo possivelmente temos uma matriz esparsa também servindo de entrada para os modelos, prejudicando, então, os resultados.

Tabela 7. Sumário dos resultados por *F1 Score* médio

	Sentence BERT	TF-IDF
XGBoost	0.69	0.57
Regressão Logística	0.63	0.47

6. Conclusão

O presente trabalho apresentou um conjunto de experimentos visando construir modelos de AM para classificar mensagens com base na teoria do diálogo freireano. Na construção, desses modelos foi empenhado uma análise de desempenho aplicado a metodologia teste t pareado 5x2cv, para seleção de algoritmos e, nesse experimento, a seleção da melhor técnica de codificação de texto. Os resultados indicaram que os modelos quando treinados com a codificação *Sentence BERT* tem desempenho superior e através do teste t pareado é possível afirmar que há diferença estatisticamente significativa em comparação ao *TF-IDF*. Dessa forma, o melhor modelo alcançado pelos experimentos foi o XGBoost que utilizou a codificação *Sentence BERT* que obteve 0.69 de *F1 Score* médio.

Como trabalhos futuros, projeta-se a evolução do presente estudo ao lidar com as limitações da versão atual do estudo. Portanto, considera-se a aplicação do uso de aprendizado profundo (novas arquiteturas como *transformers*, redes neurais convolucionais e redes neurais recorrentes) e outro tipos de codificações baseadas em aprendizado profundo. E ainda levando em consideração a pequena quantidade de amostras presentes e o desbalanceamento do conjunto de dados, é possível estipular o uso de técnicas de balanceamento (sobreamostragem ou geração de dados sintéticos). Por fim, após a otimização do modelo, espera-se utilizar o classificador em um AVA, por exemplo, com a classificação automática das categorias do diálogo freireano logo após a postagem de uma nova mensagem no fórum.

Referências

- Al-Shabandar, R., Hussain, A. J., Liatsis, P., and Keight, R. (2018). Analyzing learners behavior in moocs: An examination of performance and motivation using a data-driven approach. *IEEE Access*, 6:73669–73685.
- Barbosa, A., Ferreira, M., Ferreira Mello, R., Dueire Lins, R., and Gasevic, D. (2021). The impact of automatic text translation on classification of online discussions for social and cognitive presences. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, LAK21, page 77–87, New York, NY, USA. Association for Computing Machinery.
- Burkov, A. and Lutz, M. (2019). *The Hundred-Page Machine Learning Book en français*.
- Chistol, M. (2020). A comparative study of parametric versus non-parametric text classification algorithms. In *2020 International Conference on Development and Application Systems (DAS)*, pages 208–213.
- David, P., Lima, E., and Mendes, F. (2015). Uma ontologia de domínio para a análise do diálogo freireano em fóruns de discussão da educação a distância. *Braz. Symp. on Computers in Education (SBIE)*, 26(1):1082.
- David, P. B. (2010). *Interações contingentes em ambientes virtuais de aprendizagem*. PhD thesis, Universidade Federal do Ceará, Faculdade de Educação, Programa de Pós-graduação em Educação Brasileira.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923.
- Ferreira, M., Rolim, V., Mello, R. F., Lins, R. D., Chen, G., and Gašević, D. (2020). Towards automatic content analysis of social presence in transcripts of online discussions. LAK '20, New York, NY, USA. Association for Computing Machinery.
- Gomede, E., Miranda de Barros, R., and de Souza Mendes, L. (2020). Use of deep multi-target prediction to identify learning styles. *Applied Sciences*, 10(5).
- Hlosta, M., Zdrahal, Z., and Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. LAK '17, New York, NY, USA. Association for Computing Machinery.
- L’Heureux, A., Grolinger, K., Elyamany, H. F., and Capretz, M. A. M. (2017). Machine learning with big data: Challenges and approaches. *IEEE Access*, 5:7776–7797.
- Mishra, A. and Vishwakarma, S. (2015). Analysis of tf-idf model and its variant for document retrieval. In *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, pages 772–776.
- Oliveira, M. M. d., Barwaldt, R., Pias, M. R., and Espíndola, D. B. (2019). Understanding the student dropout in distance learning. In *2019 IEEE Frontiers in Education Conference (FIE)*, pages 1–7.
- Raschka, S. (2018). Mlxtend: Providing machine learning and data science utilities and extensions to python’s scientific computing stack. *The Journal of Open Source Software*, 3(24).

- Raschka, S. and Mirjalili, V. (2019). *Python Machine Learning*. Packt Publishing, Birmingham, UK, 3 edition.
- Reimers, N. and Gurevych, I. (2019a). Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.
- Reimers, N. and Gurevych, I. (2019b). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proc. of the 2019 Conf. on Empirical Methods in Nat. Lang. Processing*. Association for Computational Linguistics.
- Sammut, C. and Webb, G. I., editors (2011). *Encyclopedia of Machine Learning*. Springer Reference. Springer, New York.
- Singh, S. and Lal, S. P. (2013). Using feature selection and association rule mining to evaluate digital courseware. In *2013 Eleventh International Conference on ICT and Knowledge Engineering*, pages 1–7.
- Wu, J.-Y., Hsiao, Y.-C., and Nian, M.-W. (2020). Using supervised machine learning on large-scale online forums to classify course-related facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments*, 28(1):65–80.
- Yang, Y. (2017). Research and realization of internet public opinion analysis based on improved tf - idf algorithm. In *2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES)*, pages 80–83.
- Yoo, J. E., Rho, M., and Lee, Y. (2022). Online students' learning behaviors and academic success: An analysis of lms log data from flipped classrooms via regularization. *IEEE Access*, 10:10740–10753.