

Estado da publicação: O preprint foi publicado em um periódico como um artigo
DOI do artigo publicado: <https://doi.org/10.1590/0102-469839773>

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES

Cintia Maria de Araújo Pinho, Marcos Antonio Gaspar, Renato José Sassi

<https://doi.org/10.1590/SciELOPreprints.3825>

Submetido em: 2022-03-21

Postado em: 2022-04-25 (versão 1)

(AAAA-MM-DD)

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES

CINTIA MARIA DE ARAÚJO PINHO¹

<https://orcid.org/0000-0003-0525-5072>

MARCOS ANTONIO GASPAR²

<https://orcid.org/0000-0002-2422-2455>

RENATO JOSÉ SASSI³

<https://orcid.org/0000-0001-5276-4895>

RESUMO: O processo de correção manual de redações acarreta algumas dificuldades, dentre as quais apontam-se o tempo dispendido para a correção e devolutiva de resposta ao aluno. Para instituições como escolas de ensino básico e fundamental, universidades e o Exame Nacional do Ensino Médio (ENEM), tal atividade demanda tempo e custo para a avaliação dos textos produzidos. A fuga ao tema é um dos itens avaliados na redação do ENEM que pode anular a redação produzida pelo candidato. Neste contexto, a análise automática de redações com a aplicação de técnicas e métodos de Processamento de Linguagem Natural, Mineração de Textos e outras técnicas de Inteligência Artificial (IA) tem se revelado promissora no processo de avaliação automatizada da linguagem escrita. O objetivo desta pesquisa é comparar diferentes técnicas de IA para classificação de fuga ao tema em textos e identificar aquela com melhor resultado para viabilizar um sistema de correção inteligente de redações. Para tanto, foram executados experimentos computacionais visando a classificação desses textos para normalizar, identificar padrões e classificar as redações em 1.320 redações de língua portuguesa em 119 temas diferentes. Os resultados indicam que o classificador RNC (rede neural convolucional) obteve maior ganho em relação aos demais classificadores analisados, tanto em acurácia, quanto em relação aos resultados de falsos positivos, métricas de precisão, *recall* e *F1-Score*. Como conclusão, a solução validada nesta pesquisa contribui para impactar positivamente o trabalho de professores e instituições de ensino, por meio da redução de tempo e custos associados ao processo de avaliação de redações.

Palavras-chave: redações, avaliação automática de redações, fuga ao tema, inteligência artificial.

APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNIQUES FOR CLASSIFICATION OF ESCAPE FROM THE TOPIC IN ESSAYS

ABSTRACT: The process of manual correction of essays causes some difficulties, among which we point out the time spent for correction and feedback to the student. For institutions such as elementary schools, universities and the National High School Exam in Brazil (ENEM), such activity demands time and cost for the evaluation of the texts produced. Going off-topic is one of the items evaluated in the ENEM essay that can nullify the whole essay produced by the candidate. In this context, the automatic analysis of essays with the application of techniques and methods of Natural Language Processing, Text Mining and other Artificial Intelligence (AI) techniques has shown to be promising in the process of automated evaluation of written language. The goal of this research is to compare different AI techniques for the classification of going off-topic in texts and identify the one with the best result to enable a smart correction system for essays. Therefore, computer experiments were carried out to classify these texts in order to normalize, identify patterns and classify the essays in 1,320 Brazilian Portuguese essays in 119 different topics. The results indicate that the CNN classifier (convolutional neural network) obtained greater gain in relation to the other classifiers analyzed, both in

¹ UNIVERSIDADE NOVE DE JULHO, São Paulo, SP, Brasil. <cintia.pinho01@gmail.com>

² UNIVERSIDADE NOVE DE JULHO, São Paulo, SP, Brasil. <marcos.antonio@uni9.pro.br>

³ UNIVERSIDADE NOVE DE JULHO, São Paulo, SP, Brasil. <sassi@uni9.pro.br>

accuracy and in relation to the results of false positives, precision of metrics, recall and F1-Score. In conclusion, the solution validated in this research contributes to positively impacting the work of teachers and educational institutions, by reducing the time and costs associated with the essay evaluation process.

Keywords: essays, automatic essay evaluation, go off-topic, artificial intelligence.

APLICACIÓN DE TÉCNICAS DE INTELIGENCIA ARTIFICIAL PARA CLASIFICACIÓN DE ESCAPE DE LA SUJECCIÓN EN ENSAYOS

RESUMEN: El proceso de corrección manual de redacciones presenta dificultades como el tiempo dedicado a la corrección y devolución al alumno. Para las escuelas, las universidades y el Examen Nacional de Enseñanza Secundaria en Brasil (ENEM), tal actividad demanda tiempo y costo para la evaluación de los textos producidos. El escape del tema es uno de los elementos evaluados en la redacción del ENEM que puede anular la redacción. El análisis automático de ensayos con la aplicación de técnicas y métodos de Procesamiento del Lenguaje Natural, Minería de Texto y otras técnicas de Inteligencia Artificial (IA) se ha mostrado prometedor en el proceso de evaluación automatizada del lenguaje escrito. El objetivo de esta investigación es comparar diferentes técnicas de IA para la clasificación de la elusión de temas en textos e identificar aquella con mejor resultado para habilitar un sistema inteligente de corrección de redacciones. Por lo tanto, se llevaron a cabo experimentos computacionales para clasificar estos textos con el fin de normalizar, identificar patrones y clasificar los ensayos en 1.320 ensayos en lengua portuguesa en 119 temas diferentes. Los resultados indican que el clasificador red neuronal convolucional obtuvo mayor ganancia con relación a los demás clasificadores analizados, tanto en precisión como en relación con los resultados de falsos positivos, métricas de precisión, recall e F1-Score. La solución validada en esta investigación contribuye a impactar positivamente el trabajo de los docentes y las instituciones educativas, al reducir el tiempo y los costos asociados al proceso de evaluación de redacciones.

Palabras clave: redacciones, evaluación automática de redacciones, escape de tema, inteligencia artificial.

INTRODUÇÃO

A escrita é uma prática de grande importância, seja no mundo acadêmico, corporativo ou até mesmo na vivência social. Segundo Gomes (2020), a escrita é uma competência a ser desenvolvida que é imprescindível para o desenvolvimento da cognição humana. Além de estar inserida nas atividades sociais e no contexto profissional do indivíduo, a escrita integra o processo de crescimento pessoal e de ensino-aprendizagem do indivíduo. A habilidade de se comunicar por meio da escrita continua sendo primordial. Para o estudante que almeja a entrada no ensino superior, a elaboração de uma boa escrita da redação pode facilitar esse processo (SQUARISI; SALVADOR, 2020).

Contudo, o desenvolvimento da escrita nos estudantes ainda é um desafio. Em 2019 foi instituído o decreto nº 9.765 relativo à Política Nacional de Alfabetização (PNA), que ressalta que a educação é uma preocupação central das nações no século XXI. Os resultados obtidos pelo Brasil nas avaliações internacionais e os próprios indicadores nacionais revelam um grave problema no ensino e na aprendizagem da leitura e escrita. Assim sendo, o referido decreto enfatiza ser necessário implementar melhores condições para o ensino e a aprendizagem das habilidades de leitura e de escrita em todo o país (MEC, 2019, s.p.).

Nesse sentido, o professor assume importante papel no desenvolvimento das competências voltadas à leitura e escrita. Análise realizada pela instituição Universia com a participação do Centro de Treinamento e Capacitação para o Enem (Exame Nacional do Ensino Médio), destacou que o professor pode demorar de 40 segundos a 10 minutos para corrigir uma redação, de acordo com a qualidade da escrita apresentada no texto (UNIVERSIA, 2015, s.p.). Para exemplificar de forma prática o volume dos dados apresentados anteriormente, se um professor de língua portuguesa solicitar uma

redação para cada aluno de um montante de 500 estudantes, ele pode demorar de 16 a 41 horas consecutivas para corrigir os textos elaborados pelos discentes. Tal estimativa leva em consideração que o profissional gaste, respectivamente, entre 2 e 5 minutos ao avaliar cada redação.

O contexto atual mostra que os professores passam por dificuldades ao avaliar individualmente os textos de diferentes alunos. Estudo realizado por Riolfi e Igreja (2010) aponta que os professores dedicam apenas 6% do seu tempo em sala de aula para o ensino da redação. Foi identificado ainda que em alguns casos, após a correção dos textos dos alunos, o professor comentava oralmente as redações, ignorando outros problemas textuais individuais em sua exposição coletiva à turma (RIOLFI; IGREJA, 2010, p. 321). Em pesquisa conduzida por Pinho *et al.* (2020), mais de 60% dos professores analisados utilizam menos de 25% de seu tempo para o ensino e devolutivas de redações.

O conhecimento absorvido durante o ensino regular do aluno é avaliado no ENEM, uma vez que este exame mensura o desempenho dos alunos após o término do ensino médio. Tal avaliação é uma das principais formas de ingresso ao ensino superior no Brasil. Só em 2019 o ENEM foi responsável por 32% das seleções em processos seletivos de entrada em cursos superiores (AGÊNCIA BRASIL, 2019, s.p.).

O INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira) divulgou como é executada a avaliação das redações: os textos são corrigidos por mais de 5 mil avaliadores, que corrigem 150 textos a cada 3 dias. A cada 50 redações, o avaliador recebe duas delas já avaliadas por outro especialista (INEP, 2020, s.p.). Assim, cada redação é corrigida por dois professores e eles desconhecem a nota atribuída pelo outro, além de não saberem quem é o candidato avaliado (INEP, 2020, s.p.). Entretanto, esse processo manual acarreta elevado tempo para a divulgação dos resultados aos candidatos, além de altos custos para a manutenção desta metodologia. Para se ter uma ideia da dimensão deste processo, em 2020 houve 6.121.363 inscrições para o ENEM, o que demonstra a envergadura e complexidade do processo de avaliação de redações executado em cada edição do exame.

O único item discursivo da prova do ENEM é a redação, cujo texto produzido pelo aluno é avaliado de acordo com cinco competências. Cada competência avaliada contabiliza até 200 pontos ao candidato autor da redação, de modo que se ele atingir o domínio das cinco competências poderá pontuar, no máximo, 1000 pontos como nota atribuída à redação (INEP, 2020a). Contudo, observa-se resultados insatisfatórios no desempenho dos estudantes neste tipo de exame. Campos (2020) divulga que só em 2019, 143 mil estudantes tiraram nota zero na redação do ENEM e somente 53 indivíduos alcançaram nota máxima em sua produção textual. Segundo o INEP (2019), os maiores motivos que levaram à atribuição de nota zero são redação em branco (39,8%), fuga ao tema (28,4%) e cópia do texto motivador (16,2%). Assim, uma vez que o candidato elabore a redação, verifica-se que o maior motivo de atribuição de nota mínima está relacionado à fuga ao tema proposto. Estudo realizado por Diana (2021, s.p.) destaca que a fuga ao tema é o erro mais comum na elaboração de redações, acarretando ao aluno uma nota muito baixa ou mesmo o cancelamento da prova do ENEM.

A fuga do tema está relacionada à competência 2 estabelecida pelo ENEM e ocorre quando o candidato escreve um texto que não traz nenhuma referência à frase temática indicada na proposição estabelecida para a redação. Assim, se o candidato fugiu à proposta estabelecida, o avaliador não precisa seguir a correção da redação elaborada. Neste contexto, esta pesquisa trata exclusivamente dos desvios de escrita que tangenciam a competência 2 estabelecida pelo ENEM (fuga ao tema). O conjunto de circunstâncias relatado anteriormente ressalta a importância de estudar uma forma de auxílio aos professores para agilização do processo de avaliação de textos discursivos. Para tanto, uma alternativa é dar apoio para a criação de um sistema de correção inteligente de redações. A ideia de tal solução não é substituir totalmente o trabalho do professor na avaliação dos textos, mas tornar mais ágil o processo avaliativo ao fornecer indicações e apontamentos de possíveis falhas na escrita do aluno, relativamente à fuga ao tema proposto.

O uso de ferramentas digitais e inteligentes pode trazer melhorias significativas ao processo de correção de textos dissertativos. Entende-se que a implantação de soluções baseadas em Inteligência Artificial possa contribuir para esta demanda, dentre as quais destacam-se como alternativas o

Processamento de Linguagem Natural (PLN) e a Mineração de Textos (MT), além de diferentes técnicas de classificação.

Assim, a presente pesquisa tem como objetivo comparar diferentes técnicas de IA para classificação de fuga ao tema em textos e identificar aquelas que trazem melhores resultados para viabilizar um sistema de correção inteligente de redações. Para atingir tal objetivo foi realizado um estudo comparativo de diferentes técnicas de inteligência artificial voltadas para o apoio ao processo ensino-aprendizagem, buscando-se assim a execução de experimentos para indicar quais técnicas alcançam melhores resultados.

PESQUISAS SOBRE SOLUÇÕES INTELIGENTES PARA A ANÁLISE DE REDAÇÕES

A aplicação de tecnologia da informação à educação tem suscitado pesquisas em prol de auxiliar professores no processo de correção e identificação de problemas no aprendizado da produção textual de alunos. Especificamente sobre a análise automática de coesão textual em redações, estudo realizado por Nobre e Pellegrino (2010) identificou de forma automática problemas de coesão em 90% dos textos argumentativos e dissertativos analisados no experimento conduzido pelos autores. Os resultados da solução automatizada aplicada no experimento foram compatíveis às notas atribuídas em correções feitas por avaliadores humanos. Os autores afirmam ainda que a correção realizada por um programa de computador não sofre interferências externas, tais como fadiga e alteração de humor, permitindo assim avaliar e analisar sempre de forma equânime. Assim, o processo automatizado diminui a carga de trabalho do avaliador humano e se mostra uma ferramenta para apoio ao processo de correção executado por avaliadores humanos.

Santos Jr. (2017) desenvolveu pesquisa para melhorar a qualidade da avaliação automática de textos dissertativos aplicando o processamento de linguagem natural (PLN) e redes neurais. Em seu experimento o autor procurou tratar dos desvios das redações de forma genérica, sem avaliar especificamente cada competência de avaliação. A rede neural aplicada deveria acertar a pontuação de 0 a 1000 a ser atribuída à redação. Para tanto, foram avaliados 18 temas de redações, com a indicação dos resultados de cada temática tendo sido gerada de forma separada. O melhor resultado alcançado neste experimento atribuiu notas para as redações com uma taxa de erro de 100 pontos para mais ou para menos.

Cândido e Webber (2018) descrevem as possibilidades de se tratar com assertividade a coerência e coesão de redações com uso de ferramentas de PLN. O estudo realizado por eles empregou elementos linguísticos e técnicas computacionais para realizar a avaliação de redações. Os experimentos executados compararam a análise realizada por um software e as avaliações feitas por dois especialistas humanos. Foram encontrados resultados convergentes em 70% dos casos analisados. Considera-se que tais resultados iniciais sejam promissores para o desenvolvimento de solução para a avaliação automática de redações, abrindo-se então novas possibilidades de pesquisa.

Passero (2018) desenvolveu projeto para a detecção de fuga ao tema nas redações aplicando técnicas de PLN e aprendizado de máquina. O autor implementou modelos de detecção de fuga ao tema considerando-se as técnicas de análise textual, empregando para tanto a similaridade semântica textual e algumas técnicas como regressão linear e máquinas de vetores de suporte. Em seus experimentos foram utilizados 2.151 casos de redações sem fuga ao tema, além de doze exemplos de redações com fuga ao tema. Os melhores resultados desse experimento foram obtidos utilizando-se o algoritmo KFF-A, com acurácia média entre 81,13% e 96,76%. O autor relata que sua solução ainda apresenta uma taxa de falsos positivos elevada (4,24%), aquela que detecta que a redação teve uma fuga ao tema, quando na verdade não teve. Neste caso, a presença do avaliador humano ainda seria indispensável.

Ramisch (2020) investigou especificamente a recorrência de desvios de natureza sintática nas redações e as eventuais correlações com determinados atributos linguísticos das sentenças elaboradas. Contudo, em sua pesquisa foram eliminadas as redações anuladas ou com fuga ao tema. Os melhores resultados obtidos pelo *corpus* de teste foi com o algoritmo Regressão Logística, que alcançou 75,62% de acurácia.

Bittencourt Júnior (2020) propôs a avaliação automática de redações utilizando redes neurais profundas para tanto. Para o processo de aprendizado foi utilizado um conjunto de redações

com 18 temas diferentes. O estudo procurou avaliar as cinco competências estipuladas pelo ENEM. Como resultados, aponta-se a proposição de uma nova arquitetura multitema, com base na hipótese de que as características aprendidas pela rede aplicada para a correção de determinado tema poderiam ajudar a aprimorar o desempenho de outros temas, melhorando assim os resultados obtidos a cada avaliação automatizada.

INTELIGÊNCIA ARTIFICIAL APLICADA À EDUCAÇÃO

A Inteligência Artificial (IA) é voltada ao campo de conhecimento associado à linguagem, inteligência, raciocínio, aprendizagem e resolução de problemas. Preuss, Barone e Henriques (2020) argumentam que as técnicas de IA aplicadas a diferentes áreas de estudo trazem inúmeros benefícios. Para Russo (2020) e Ludermir (2021), tais técnicas podem resolver problemas cada vez mais complexos trazendo assim eficiência, significado e agilidade.

Segundo Müller (2018), a aplicação de IA na educação tem sido amplamente discutida, embora ela atenda a um número limitado de cenários de aprendizagem, já que as máquinas inteligentes operam nos limites de seu sistema. Assim, sistemas inteligentes aplicados à educação devem fornecer suporte para os professores e melhorar o seu trabalho. Em complemento, Müller (2018) argumenta que as pesquisas de IA na educação são promissoras, ao passo que as máquinas vão se ajustando às necessidades individuais de cada profissional.

A IA se utiliza de diferentes técnicas para fornecer informações baseadas em grandes volumes de dados, dentre elas destacam-se a Mineração de Textos (MT), o Processamento de Linguagem Natural (PLN) e Aprendizado de Máquina. Este último engloba as técnicas inteligentes de classificação, com destaque para as redes neurais artificiais (EGGERS WILLIAM *et al.*, 2017; HARIRI, 2019). Estas técnicas serão tratadas nos próximos tópicos, a fim de compor um breve panorama do conjunto de métodos mencionados.

Mineração de textos

Moraes e Ambrósio (2007) definem mineração de textos como um processo de descoberta de conhecimento que utiliza técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras. Na visão de Gonçalves (2012) e Souza (2019), a Mineração de Textos envolve a aplicação de algoritmos computacionais que processam textos e identificam informações úteis e implícitas, que normalmente não poderiam ser recuperadas utilizando métodos tradicionais de consulta, pois a informação contida nestes textos não pode ser obtida de forma direta.

Processamento de linguagem natural (PLN)

O Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial que estuda a comunicação humana por métodos computacionais. Assim, busca-se converter a linguagem natural humana em uma representação formal, de forma que se torne mais facilmente manipulável por máquinas. Muitas aplicações de PLN são baseadas em modelos de linguagem que definam uma distribuição de probabilidade sobre sequências de palavras, caracteres ou bytes em uma linguagem natural (GOODFELLOW; YOSHUA, 2016; CONEGLIAN, 2018).

A PLN busca padrões e indicativos que auxiliem na compreensão do texto em análise. Assim, os estudos de PLN e Aprendizado de Máquina convergem cada vez mais devido à grande quantidade de dados que é gerada diariamente, uma vez que é por meio desses dados que o computador aprende (GOODFELLOW; YOSHUA, 2016; CONEGLIAN, 2018).

A PLN tem se voltado ao tratamento e análise de massas de dados não estruturados, especialmente no formato de texto, o que tem proporcionado o surgimento de diferentes áreas de atuação, quais sejam: sistemas de respostas a perguntas de usuários, traduções feitas por máquinas, reconhecimento de voz e diálogos, classificação de documentos, reconhecimento de textos em imagens e análise de sentimentos em textos (PRATES, 2019; STEFANINI, 2019).

Técnicas de inteligência artificial

Atualmente as principais técnicas inteligentes estão inseridas no contexto de Aprendizado de Máquina (AM), esta área é dedicada ao estudo de algoritmos de previsão e inferência, que buscam simular em computadores o cérebro enquanto máquina de aprendizado. O AM inclui técnicas estatísticas para permitir que máquinas aperfeiçoem ao máximo suas tarefas com base nos dados extraídos por experiência. Assim, os algoritmos podem aprender com estes dados, identificar padrões e tomar decisões com pouca intervenção humana (MUYLAERT, 2020; RUSSO, 2020; BIANCHI, 2020).

No contexto de aprendizagem de máquina, existem os seguintes tipos de aprendizado, conforme indicados por Carvalho (1994) e Waltrick (2020):

- **Aprendizado Supervisionado:** o modelo deverá ser, literalmente, ensinado sobre o que deve ser feito. Neste sentido, deverá ser fornecido um conjunto de dados rotulados para o modelo aprender, sendo estes dados particionados entre conjuntos para treinamento e teste. Esse tipo de aprendizado é, geralmente, aplicado quando o objetivo é prever ocorrências futuras;
- **Aprendizado Não Supervisionado (auto-organização):** quando não existe um agente externo indicando a resposta desejada para os padrões de entrada. Diferentemente do aprendizado anterior, aqui é fornecido um conjunto de dados não rotulados e não se ensina ao modelo qual é o objetivo final. Alguns exemplos de técnicas que podem ser aplicadas neste aprendizado são: associação, redução de dimensão e clusterização;
- **Aprendizado por Reforço:** quando um crítico externo avalia a resposta fornecida. É usado nos casos em que o problema não é, basicamente, relacionado a conjunto de dados, mas há um ambiente para lidar, como um cenário de um jogo ou uma cidade onde circulam carros autônomos. Utiliza o método ‘tentativa e erro’, no qual o acerto equivale a uma recompensa, enquanto o erro equivale a uma punição.

Os classificadores empregados na AM buscam organizar objetos entre diversas categorias e, para tanto, o modelo classificador analisa o conjunto de dados fornecidos. Nesse conjunto, cada dado já contém um rótulo indicando a qual categoria ele pertence, a fim de ‘aprender’ como classificar outros novos dados. Na classificação, os algoritmos que implementam esse processo são chamados de classificadores (RAMOS *et al.*, 2018; HAN; KAMBER; PEI, 2011).

Para Affonso *et al.* (2010), a classificação de textos é uma técnica empregada para atribuir automaticamente uma ou mais categorias predefinidas a um *corpus* em análise. As aplicações mais comuns são a indexação de textos, mineração de textos, categorização de mensagens, notícias, resumos e arquivos de publicações periódicas. Nos sistemas computacionais, o processo de classificação envolve técnicas para extração das informações mais relevantes de cada categoria, além da utilização destas informações para ensinar o sistema a classificar corretamente os documentos.

A Aprendizagem Profunda (AP), do inglês ‘*deep learning*’, é um ramo do AM baseado em um conjunto de algoritmos que tenta modelar abstrações de alto nível de dados. Algumas de suas representações são inspiradas na interpretação do processamento de informações e padrões de comunicação em um sistema nervoso (PREMLATHA, 2019; BABERJEE, 2020; BIANCHI, 2020). Segundo os autores, a rede neural é uma das arquiteturas da AM que tem sido aplicada na marcação de classes gramaticais, tradução e classificação de textos.

Classificadores

Com o objetivo de comparar diferentes técnicas de IA para classificação de fuga ao tema em redações foram utilizados classificadores de aprendizado supervisionado explicados anteriormente. Os algoritmos classificadores destacados abaixo são da biblioteca Scikit Learn (2021a) que utilizam técnicas de AM.

MLP (Multilayer Perceptron): o *MLPClassifier* é uma rede neural que possui mais de uma camada de neurônios. Em casos em que não há a possibilidade de uma única reta separar os elementos, a MLP gera um plano de classificações. (AFFONSO *et al.*, 2010; LEITE, 2018). O algoritmo utilizado

para treinamento da MLP é chamado backpropagation ou retropropagação, sendo composto por quatro passos: inicialização, ativação, treinamento dos pesos e iteração. A ideia do algoritmo backpropagation é, com base no cálculo do erro ocorrido na camada de saída da rede neural, recalculando o valor dos pesos do vetor da camada última camada de neurônios (MOREIRA, 2018; LEITE, 2018).

Árvores de Decisão: o *DecisionTreeClassifier* é um algoritmo de aprendizado supervisionado, utilizados em tarefas de classificação e regressão. A árvore de decisão é um método para aproximar funções-alvo com valores discretos, em que a função aprendida é representada por uma árvore de decisão. As decisões são tomadas com base num conjunto de regras do tipo ‘se-então’ (‘if-then’) (MITCHELL, 1997).

As árvores de decisão representam uma das formas mais simplificadas de um sistema de suporte a decisão. A partir de um conjunto de dados o algoritmo cria uma representação do conhecimento ali embutido, em formato de árvore (PESSANHA, 2019).

Florestas Aleatórias (Random Forest): *RandomForestClassifier* é um algoritmo de aprendizagem supervisionada que cria uma floresta de modo aleatório, a ‘floresta’ criada é uma combinação (ensemble) de árvores de decisão, na maioria dos casos treinados com o método de bagging, a ideia principal do método de bagging é que a combinação dos modelos de aprendizado aumenta o resultado geral (COSTA DA SILVA, 2018).

Gradient Boosting: o algoritmo *GradientBoostingClassifier* é uma generalização do boost para funções de perda diferenciáveis arbitrárias. Esse algoritmo é um procedimento preciso e eficaz que pode ser usado para problemas de regressão e classificação em uma variedade de áreas, incluindo a classificação de pesquisa na Web ou ecologia, por exemplo. (SCIKIT LEARN, 2021b).

O algoritmo Gradient Boosting está incluído no grupo de classificadores Ensemble. Esse classificador utiliza uma combinação de resultados de preditores fracos com o objetivo de produzir um melhor modelo preditivo. Na técnica de Boosting, cada classificador fraco é treinado com um conjunto de dados, de forma sequencial e de uma maneira adaptativa, pela qual um modelo base depende dos anteriores e, ao final, são combinados de uma maneira determinística (SILVA, 2020).

AdaBoost: O princípio básico do *AdaBoostClassifier* é ajustar uma sequência de aprendizes fracos (ou seja, modelos que são apenas ligeiramente melhores do que suposições aleatórias, como pequenas árvores de decisão) em versões repetidamente modificadas dos dados (SCIKIT LEARN, 2021b).

O AdaBoost pode ser usado para impulsionar o desempenho de qualquer algoritmo de aprendizado de máquina. É melhor usado com classificadores fracos. Estes são modelos que alcançam precisão logo acima do acaso em um problema de classificação (BROWNLEE, 2016).

Stochastic Gradient Descent (SGD): *SGDClassifier* é um algoritmo simples mas eficiente, empregado para ajustar classificadores lineares e regressores sob funções de perda convexa, como máquinas de vetor de suporte (linear) e regressão logística. (SCIKIT LEARN, 2021c).

O SGD foi aplicado com sucesso a problemas de aprendizado de máquina esparsos e em grande escala, frequentemente encontrados na classificação de texto e no processamento de linguagem natural. As vantagens da aplicação deste algoritmo são a eficiência e facilidade de implementação, pois oferece muitas oportunidades para ajuste de código. A classe *SGDClassifier* implementa uma rotina de aprendizagem de gradiente descendente estocástico simples que suporta diferentes funções de perda e penalidades para a classificação (SCIKIT LEARN, 2021c).

Support Vector Machines (SVM): são um conjunto de algoritmos de aprendizado supervisionado usados para classificação, regressão e detecção de outliers (uma observação que se diferencia tanto das demais observações que levanta suspeitas de que aquela observação foi gerada por um mecanismo distinto). As SVMs são eficazes em espaços dimensionais elevados e quando o número de dimensões é maior do que o número de amostras (SCIKIT LEARN, 2021d).

Além dos classificados de AM destacados anteriormente, também existem os classificadores baseados em Aprendizado Profundo(AP), como a Rede Neural Convolucional (RNC) (SCIKIT LEARN, 2021e).

RNC: é um algoritmo de aprendizado profundo que pode captar uma entrada, atribuir importância (por meio de pesos e vieses que podem ser aprendidos) a vários aspectos, sendo capaz de diferenciar um do outro. As Redes Neurais Convolucionais (RNCs) são responsáveis por avanços na

classificação de imagens, configurando-se no núcleo da maioria dos sistemas de visão por computador atuais, desde a marcação automática de fotos do Facebook até carros autônomos. Mais recentemente, tem-se aplicado RNCs em problemas de Processamento de Linguagem Natural, para os quais vem se obtendo resultados promissores (BRITZ, 2015). Rodrigues (2018) e Carneiro (2020) destacam que a RNC se mostra muito eficaz para tarefas de classificação textual.

Métricas de avaliação dos classificadores

Na tarefa de classificação, duas etapas principais são realizadas. Na primeira etapa é gerado o modelo que aprende por meio do treinamento dos dados, normalmente utilizando 70% a 80% da base de dados. Na segunda etapa, os dados separados são testados, entre 30% e 20% da base para estimar o desempenho, mensurando-se assim os acertos do modelo (RAMOS *et al.*, 2018; HAN; KAMBER; PEI, 2011). Após o processo de classificação é necessário avaliar o desempenho do classificador, sendo utilizadas algumas métricas. Para distinguir entre a classe real e a classe prevista são usados rótulos (P - Positivo, N - Negativo) para as previsões de classe produzidas por um modelo.

Segundo Ramos *et al.* (2018), dado um classificador e uma instância a classificar, é criada uma matriz de confusão com quatro resultados possíveis (Figura 1):

- VP (Verdadeiro Positivo), quando o rótulo avaliado é verdadeiro e modelo trouxe como resultado um valor positivo – indicando acerto do modelo;
- FN (Falso Negativo), quando houve um erro do modelo, previu a classe negativo quando o valor real era positivo;
- VN (Verdadeiro Negativo), quando o rótulo avaliado é negativo e o modelo trouxe como resultado um valor negativo – indicando acerto do modelo;
- FP (Falso Positivo), quando o rótulo avaliado é negativo e o modelo trouxe como resultado um valor positivo, indicando assim erro do modelo.

Figura 1: Matriz de Confusão de Resultados

		Detectada	
		Sim	Não
Real	Sim	Verdadeiro Positivo (VP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (VN)

Fonte: Rodrigues (2019).

A matriz de confusão visualizada na Figura 1 demonstra como as nomenclaturas citadas anteriormente ficam dispostas. Destas, são retiradas métricas de avaliação como a acurácia, precisão, *recall* e *F1-score* (RODRIGUES, 2019), demonstradas no Quadro 2.

Quadro 1: Métricas de Avaliação

Métrica	Descrição	Fórmula
Acurácia	Indica uma performance geral do modelo. Dentre todas as classificações, quantas o modelo classificou corretamente;	$\frac{VP + VN}{VP + FP + FN + VN}$
Precisão	Dentre todas as classificações de classe Positivo que o modelo fez, quantas estão corretas;	$\frac{VP}{VP + FP}$
<i>Recall</i>	Dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas;	$\frac{VP}{VP + FN}$
<i>F1-Score</i>	Média harmônica entre precisão e recall.	$\frac{2 * Precisão * Recall}{Precisão + Recall}$

Fonte: Adaptado - Rodrigues (2019).

As métricas visualizadas podem ser utilizadas em diferentes classificadores, como os que foram demonstrados nos tópicos anteriores: Multilayer Perceptron (MLP); *decision tree*; *random forest*; *gradiente boosting*; *adaboosting*; *Support Vector Machine* (SVM) e redes neurais convolucionais.

MÉTODO E MATERIAIS DE PESQUISA

Esta é uma pesquisa aplicada e experimental, executada por meio da aplicação dos algoritmos e mensuração dos resultados obtidos em experimentos desenvolvidos.

Base de dados e plataforma de experimentos

Para a condução dos experimentos foi utilizada uma amostra com 1.320 redações distribuídas em 119 temas diferentes. Este *corpus* foi extraído da base de dados disponível no repositório do Portal UOL (2019) de fonte aberta e também da plataforma desenvolvida por Pinho *et al.* (2020), com o objetivo de montar um repositório de redações corrigidas por diferentes professores e níveis de alunos. Todas as redações componentes do *corpus* utilizado nos experimentos já foram corrigidas por professores, possuindo, portanto, as respectivas notas e comentários atribuídos pelos docentes ao final do processo de correção.

Quanto à sua estrutura, a base de dados possui doze colunas, conforme exemplificado no Quadro 3. As colunas consideradas para a condução dos primeiros experimentos foram: redação, tema e fuga. Sendo que a última (fuga) foi responsável pelo processo de classificação, consistindo no atributo alvo deste estudo.

Quadro 2: Estrutura da base de dados usada na fase inicial dos experimentos

REDAÇÃO	TEMA	TITULO	TEXTO MOTIVADOR	Nota comp 1	Nota comp 2	Nota comp 3	Nota comp 4	Nota comp 5	Total	Comentário avaliador	Fuga
Texto com a redação	Tema da Redação	Título fornecido pelo aluno	Texto fornecido na proposta falando sobre o assunto da temática	0 a 200	0 a 200	0 a 200	0 a 200	0 a 200	0 a 1000	Comentário do avaliador	Sim ou não

Fonte: Autores.

A distribuição das notas das redações dadas pelos professores para a composição da base de dados pode ser visualizada na Tabela 1, considerando-se que o objetivo inicial desta pesquisa é analisar se a redação fugiu ao tema proposto. Houve 230 redações classificadas como ‘fuga ao tema’, representando 17% da base de dados total considerada para os experimentos desta pesquisa.

Tabela 1: Distribuições das notas das redações

Redações Avaliadas		
Qtde	Nota	Percentual
266	0 pontos.	20,1%
	(Obs.: 230 redações foram classificadas como fuga ao tema).	
33	20 a 100 pontos	2,5%
172	120 a 300 pontos	13,0%
293	320 a 500 pontos	22,2%
252	520 a 700 pontos	19,2%
192	720 a 900 pontos	14,5%
112	920 a 1000 pontos	8,5%
Total:1320		Total: 100%

Fonte: Autores.

Distribuição dos dados

Para a separação da base de dados utilizou o processo de validação cruzada, com o objetivo de evitar que apenas uma porção de dados de treino e teste pudessem ser muito similares. Isto porque, neste caso, quando houvesse testes com novos dados muito diferentes do modelo treinado os resultados seriam insatisfatórios. Trabalhar com diferentes distribuições possibilita diminuir os riscos de vícios de trabalhar com apenas uma amostra. Desta forma, os dados foram separados em três conjuntos diferentes, seguindo a ideia demonstrada na Figura 2, na qual são expostos os três grupos diferentes gerados para a realização dos experimentos.

Estes grupos foram embaralhados para gerar as 3 amostras, assim foram realizados 3 experimentos com cada classificador. A proposta era utilizar os mesmos conjuntos de dados, tanto nos classificadores do Sklearn como na RNC. Desta forma, foi possível realizar a comparação dos resultados entre as técnicas, já que utilizaram as mesmas distribuições. Este processo visualizado na Figura 2 permite evitar a variância, além de possibilitar entender se os experimentos realizados trouxeram a mesma média de resultados com diferentes combinações.

Figura 2: Demonstração da Validação Cruzada para o processo de treinamento e testes

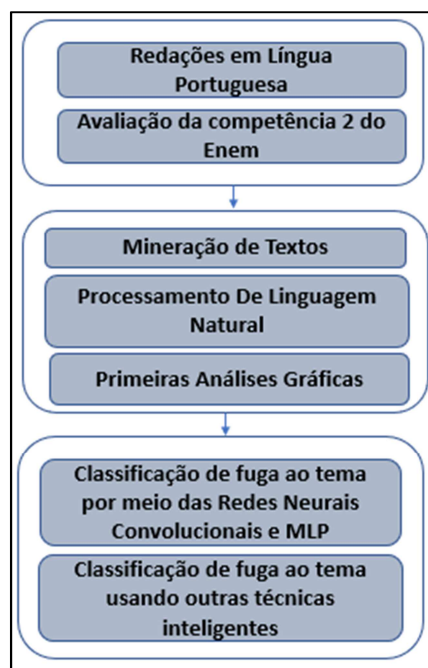
1ª Amostra 1320 redações	2ª Amostra 1320 redações	3ª Amostra 1320 redações
Treino	Teste2	Teste1
Teste1	Treino	Teste2
Teste2	Teste1	Treino

Fonte: Autores.

Fluxo de atividades para a execução dos experimentos

Além das etapas definidas para a pesquisa também foi gerado fluxo de atividades para a elaboração dos experimentos, conforme apresentado na Figura 3. A sequência desenvolvida é constituída inicialmente pelas redações em língua portuguesa para avaliação de desvios estabelecidos pela competência 2 do ENEM (fuga ao tema), foco desta pesquisa.

Figura 3: Fluxo de Atividades



Fonte: Autores.

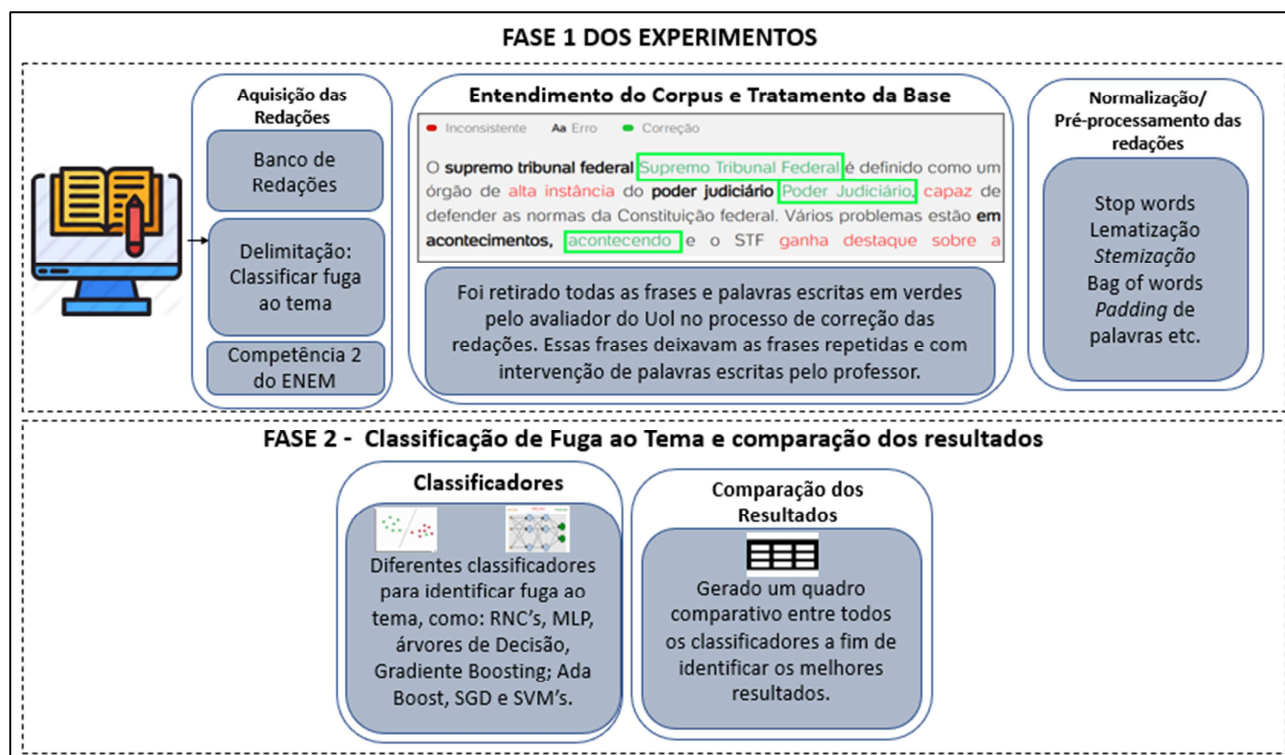
No processo dos experimentos executados nesta pesquisa foram realizados as etapas pertinentes à mineração de textos, processamento de linguagem natural e, por fim, treinamento na base utilizando-se as redes neurais convolucionais Multilayer Perceptron (MLP). Além das redes neurais, o treinamento também foi realizado com outras técnicas e algoritmos de classificação, tais como: Árvore de Decisão, Florestas Aleatórias, Gradiente *Boosting*, *Ada Boost*, *Stochastic Gradiente Descent* (SGD) e *Support Vector Machines* (SVMs). Os resultados esperados após a aplicação do modelo e das técnicas e algoritmos indicados busca a obtenção da classificação da fuga ao tema como positiva ou negativa. Este processo leva em consideração as redações sob análise para gerar a futura solução a ser validada.

Detalhamento dos experimentos

Depois de entender o fluxo dos experimentos, buscou-se detalhar a sequência destes, conforme exposto na Figura 4. Analisando-se a fase 1 do esquema exposto na figura, inicialmente foi utilizada uma base de dados com 1.320 redações para identificar desvios de escrita na competência 2 (fuga ao tema). Em seguida estas redações passaram por uma análise dos pesquisadores, sendo que nesta etapa foi identificado que algumas possuíam palavras e frases repetidas. Estas foram escritas no processo de correção pelo avaliador (pessoa física), sendo que todas as frases e palavras estavam destacadas pela cor verde. Assim tais repetições foram retiradas e os textos voltaram à sua forma original.

Dando sequência a análise da Figura 4, após a retirada dos textos redundantes o próximo passo foi aplicar as primeiras técnicas de PLN e Mineração de Textos para normalizar os documentos e prepará-los para as primeiras análises gráficas, clusterização e elaboração dos modelos de treinamento.

Figura 4: Sequência de experimentos



Fonte: Autores.

Na fase de normalização dos dados foi criada uma função em Python, que executava toda a tratativa necessária. Nesta fase foi utilizada a biblioteca do *Spacy* e criada uma função para deixar todos os caracteres em minúsculo, realizar o processo de tokenização, lematização, remoção de *stop words* e retirada de caracteres especiais das redações do *corpus* em análise.

A Figura 5 expõe exemplos de como ficaram as frases após o processo de normalização anteriormente explicado. Neste exemplo, o tema de redação normalizado fica em minúsculo, sem caractere especial e sem *stop words*, que seriam as palavras 'ao' e 'ou'. Também pode ser visualizado o processo de lematização, no qual há ação de deflexionar uma palavra para determinar o seu lema, como pode ser visto na palavra 'fumo', que passou para 'fumar' e também na palavra 'combate', que se tornou 'combater'.

Figura 5: Visualização dos textos após normalização

Tema redação sem normalizar: Combate ao fumo: autoritarismo ou dever do governo

Tema redação normalizado: combater fumar autoritarismo dever governar

Fonte: Autores.

Na última fase dos experimentos foram aplicadas as técnicas de classificação, que estabeleceram a classificação das redações sob análise. Em seguida, foi gerado um comparativo entre todas as técnicas aplicadas, de modo a evidenciar as particularidades da aplicação de cada uma delas no experimento, conforme detalhamento das etapas, bibliotecas e passos realizados para cada fase dos experimentos executados expostos no Quadro 4.

Quadro 3: Detalhamento dos experimentos

Etapas / bibliotecas / experimentos realizados
1) Importação das bibliotecas necessárias para iniciar os experimentos, como <i>Spacy</i> , <i>Pandas</i> , <i>Numpy</i> , <i>Matplotlib</i> , <i>Scikit-learn</i> etc.
2) Extração da base de dados com 1320 redações

3)	Pré-processamento dos dados, fazendo uma limpeza na base com o processo de normalização, remoção de stop words, remoção de caracteres especiais, <i>stemização</i> e <i>lematização</i> , <i>padding</i> de palavras;
4)	Antes da do processo de treinamento do modelo procurou-se colher informações a respeito dos textos gerando alguns histogramas para entender a base de dados.
5)	Ainda na fase de entendimento da base de dados e procurando formas de auxiliar os professores para saber as temáticas que os alunos possuem maior dificuldade de escrita, foi então realizado o processo de similaridade entre as redações e suas temáticas por meio das palavras relacionadas entre a redação escrita pelo aluno e o texto motivador fornecido pela instituição.
6)	Para o processo de treinamento foram utilizadas as redações com notas superiores a 499 pontos e as redações com fuga ao tema. O total de redação para o processo de treino foram 628 redações e para o processo de teste foram 209 redações. (As redações de teste correspondem a 33% das redações treinadas)
7)	Após divisão da base em treino e teste o primeiro classificador testado foi a rede neural convolucional. Para a entrada da Rede neural no Processo de Treino, o primeiro passo foi transformar a variável target 'fuga' em valores numéricos 0 e 1, na qual a fuga ao tema passou a ser representado por 1 e a não fuga por 0.
8)	Na sequência todos os textos precisaram ser vetorizados passando pelo processo de <i>padding</i> (criado um vetor de números que representam cada palavra). Nesta fase, cada redação e proposta do tema foi vetorizado e trabalhados para ficarem na mesma dimensão. Para a rede Neural Convolucional, este processo ocorreu utilizando a biblioteca ' <i>tfds</i> ' do ' <i>tensorflow_datasets</i> '.
9)	Uma vez que toda preparação dos dados foi realizada o próximo passo foi a configuração do modelo determinando as camadas de entrada, filtros, neurônios, a quantidade de documentos que seriam avaliadas para a atualização dos pesos e a quantidade de épocas para execução do treinamento.
10)	Após o processo de Treinamento o próximo passo foi utilizar o modelo gerado para as 209 redações separadas para a fase de teste.
11)	Após a aplicação do Primeiro Teste, foram geradas a matriz de confusão com os resultados e acurácia do experimento.
12)	Outro teste foi realizado agora com as redações que ainda não tinham sido treinadas, ou seja, aquelas com notas inferiores a 500.
13)	Para os próximos classificadores foi também foi realizado o processo de normalização das redações, assim como na rede neural convolucional foi necessário fazer nova representação dos valores textuais (discretos), os classificadores utilizados "só entendem números". Assim foi necessário converter os dados brutos, que estão em formato de texto, para um formato numérico. Isto deve acontecer antes passar as redações para o classificador, o processo de vetorização para os classificadores do <i>Scikit-learn</i> foi realizado utilizando o <i>TfidfVectorizer</i> .
14)	As redações que já haviam sido divididas em treino e teste foram então configuradas em cada classificador do <i>Scikit-learn</i> utilizando os hiper parâmetros definidos na própria documentação da Biblioteca
15)	Os resultados dos experimentos serão demonstrados no próximo capítulo.

Fonte: Autores.

APRESENTAÇÃO DOS RESULTADOS

Deteção de padrões

O primeiro experimento realizado com o objetivo de detectar padrões nas redações que indicavam fuga ao tema foi a aplicação da técnica de marcação gramatical, também conhecida como *pos tagging* da biblioteca do Spacy. Neste experimento foram extraídas tanto da proposta da redação (texto motivador e tema), como da redação, os verbos, substantivos e adjetivos, considerados os termos que fornecem mais significado à escrita.

Após obtenção dos termos com maior valor semântico de cada texto, estes foram então comparados com o objetivo de encontrar as palavras que o aluno escreveu que tinham aderência à proposta de tema para a elaboração da redação. Na Figura 6 é exposto um exemplo da análise da correlação entre a redação escrita e a proposta da redação para redações com fuga ao tema.

Figura 6: Exemplo de Correlação entre Redação e Proposta da Redação – Redações com fuga ao tema

Violência e drogas: o papel do usuário {'drogas': 2}
Violência e drogas: o papel do usuário {'violência': 3, 'segurança': 1, 'brasil': 2, 'homem': 1, 'pensamento': 1, 'país': 1}

Fonte: Autores.

Na figura anterior são demonstradas duas redações classificadas como fuga ao tema, assim como sua relação entre a redação elaborada pelo aluno e a proposta de tema intitulada “Violência e Drogas: o papel do usuário”. Ao analisar o conteúdo da Figura 6 é possível identificar na primeira redação que apenas a palavra ‘drogas’ apareceu duas vezes em toda a redação, demonstrando que o aluno não usou mais nenhuma palavra relativa à proposta que foi indicada pela instituição aplicadora da redação. Já a segunda redação demonstra que o aluno usou seis palavras que podem ter relação com a temática previamente indicada, palavras estas que se encontravam na proposta da redação. Porém, foi possível verificar que o aluno não usou as palavras ‘drogas’ ou ‘usuário’, bem como as outras palavras que tiveram relação com a proposta, mas que poderiam ter sido empregadas em diferentes contextos, não apenas nessa temática.

Na Figura 7 é exemplificada a correlação entre a redação escrita e a proposta da redação para duas redações classificadas sem fuga ao tema, tendo a primeira recebido nota 700 e a segunda recebido nota 900.

Figura 7: Exemplo de Correlação entre Redação e Proposta da Redação – Redações sem fuga ao tema

Violência e drogas: o papel do usuário {'violência': 3, 'drogas': 9, 'comércio': 1, 'ilegais': 1, 'lei': 1, 'oferta': 1, 'procura': 1, 'política': 4, 'proibicionista': 2, 'especial': 1, 'país': 3, 'descriminalização': 4, 'debate': 2}
Violência e drogas: o papel do usuário {'violência': 3, 'drogas': 3, 'tráfico': 3, 'comércio': 3, 'saúde': 1, 'lei': 2, 'oferta': 1, 'culpa': 2, 'proibição': 3, 'ilegalidade': 1, 'parcela': 1, 'maconha': 1, 'brasil': 1, 'população': 1, 'financia': 1, 'consumo': 3, 'regulamentação': 1, 'problema': 1}

Fonte: Autores.

Ao analisar os resultados das duas redações expostas na Figura 7, que possuem a mesma temática abordada na Figura 6 anterior, foi possível identificar que, além de ter muitas palavras relacionadas ao texto motivador, tais palavras têm maior relação com o tema, a exemplo de: ‘descriminalização’, ‘ilegais’, ‘drogas’, ‘política’ e ‘tráfico’, dentre outras ocorrências.

Esses resultados poderiam ser disponibilizados para o professor em conjunto com a indicação da probabilidade de fuga ao tema, que será exposta nos classificadores dos próximos tópicos. Desta forma, além da indicação do percentual de alinhamento à proposta da redação, o docente teria também a indicação das palavras que mais aderiram à proposta da redação.

Essa primeira fase dos experimentos proporcionou informações relevantes ao professor sobre o desempenho de suas turmas e alunos, identificando aqueles que estão com maiores dificuldades na produção textual, bem como temas com maior dificuldade de assimilação por parte dos estudantes. Esta solução pode ser facilmente implementada num sistema de gerenciamento de turmas, proporcionando assim benefícios tanto para o docente, quanto para a escola.

Preparação da base para os classificadores

Para a aplicação dos classificadores selecionados neste experimento, uma nova etapa foi iniciada. Após o processo de normalização das redações foi necessário ainda adaptar os textos aos classificadores, uma vez que estes não reconhecem valores discretos (textos). Assim sendo, aplicou-se o processo de *padding* e vetorização dos textos antes de iniciar o treinamento do modelo desenvolvido.

Na Figura 8 é demonstrado um exemplo da redação convertida em números. Neste processo cada palavra assume uma codificação. Todos os classificadores considerados nos

experimentos utilizaram esta técnica para treinar seus modelos e, posteriormente, realizar a classificação de fuga ao tema nas redações.

Figura 8: Padding das sentenças

[7	222	4740	365	4	1647	184	9	11214	7707	55	11
1760	7	985	6677	9	781	6784	2818	721	1340	302	1420	
2478	7936	169	3021	5889	133	22546	17160	18149	81	448	2543	
2920	150	12028	380	1565	7	222	3141	63	82	10622	83	
366	81	4500	4332	7	11213	9	770	1075	133	2164	63	
209	21819	49	7691	6	3267	45	2397	1808	74	236	243	
96	36	82	1096	1983	366	14956	6392	2451	7	6620	9	
308	707	239	1096	1983	9	730	2240	7218	306	9	2430	
730	106	276	371	1108	9	7187	1429	365	3116	658	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	

Fonte: Autores.

A dimensão dos vetores criados nesta etapa do experimento foi de 510 palavras, o que significa que após o procedimento de normalização realizado, a maior redação ou texto motivador encontrado possuía 510 palavras. Assim, o algoritmo completa os vetores com zero para deixá-los todos na mesma dimensão antes de começar o treinamento dos próximos modelos.

Uma vez que os dados foram tratados, o próximo passo é aplicar os diferentes modelos de classificação selecionados. Para aplicação no classificadores foram gerados os modelos aplicando-se as duas bases de teste separadas, assim como ocorreu com na RNC. Na sequência as métricas de avaliação utilizadas foram a acurácia e matriz de confusão gerada para cada modelo em questão.

Classificação de fuga ao tema usando as RNCs

Os resultados deste classificador foram dispostos de forma separada dos posteriores, pois ele se utiliza de outras técnicas para sua aplicação, empregando para tanto a biblioteca do Spacy, além de utilizar um método de vetorização diferente dos classificadores do *Scikit Learn*. Na terceira fase dos experimentos, após a realização de todo o processo de pré-processamento das redações por meio das normalizações, remoção de *stop words*, remoção de caracteres especiais, stemização e lematização e *padding* de palavras, conforme exposto no tópico de método e materiais, a base de dados foi finalmente separada entre conjunto de treinamento e conjunto de teste.

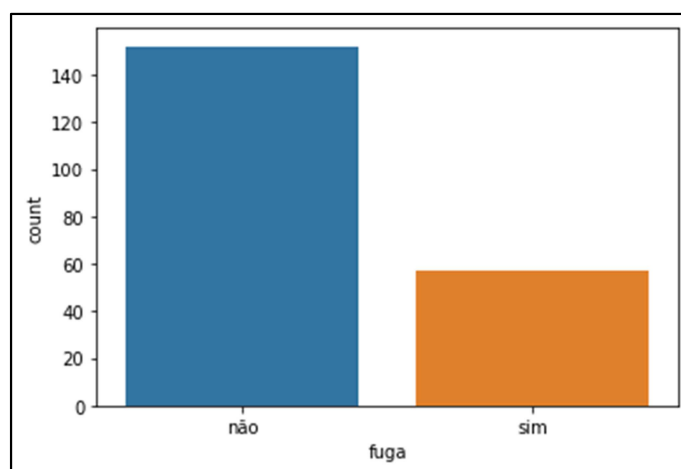
Para o treinamento da RNC foram utilizadas 628 redações, o que equivale a 47% das redações da base em análise. A escolha das redações para treino e teste do modelo levou em consideração a nota real atribuída pelos professores a cada redação. Os melhores resultados foram obtidos quando as redações tinham notas superiores a 499 pontos. Este critério também levou em consideração que essa normalmente é a margem que as universidades utilizam como critério de eliminação dos candidatos (CRUZEIRO DO SUL, 2020).

A configuração do modelo concebido utilizou a biblioteca do *keras* do tensorflow, com os seguintes parâmetros: `emb_dim = 200`; `nb_filters = 700`; `ffn_units = 1000`; `batch_size = 32`; `dropout_rate = 0.2`; `nb_epochs = 40`. Os valores iniciais dos parâmetros foram fornecidos em um curso oferecido por Granatyr (2020) na Plataforma IA Expert. Nos experimentos seguintes os parâmetros foram ajustados de acordo com os resultados apresentados.

Após o processo de treinamento e a geração do modelo, o próximo passo foi o teste com as 209 redações restantes, que equivalem a 33% da base de treinamento. O balanceamento entre redações que fugiram e não fugiram ao tema pode ser visualizado na Figura 9. As redações que fugiram

ao tema equivaleram a 17% do total analisado. Na fase de treinamento também foi considerado o mesmo percentual de redações com fuga ao tema.

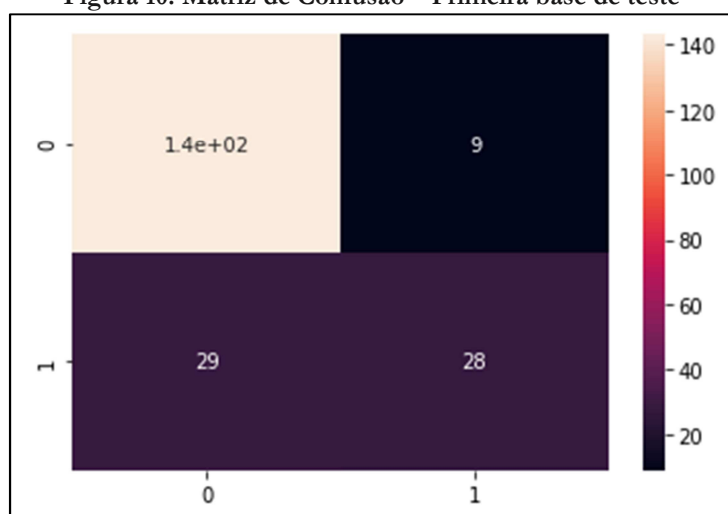
Figura 9: Distribuição da base de teste



Fonte: Autores.

Após a aplicação da base de teste no modelo, os primeiros resultados podem ser visualizados na Figura 10, na qual está representada a Matriz de Confusão criada, com os acertos e erros do modelo.

Figura 10: Matriz de Confusão – Primeira base de teste

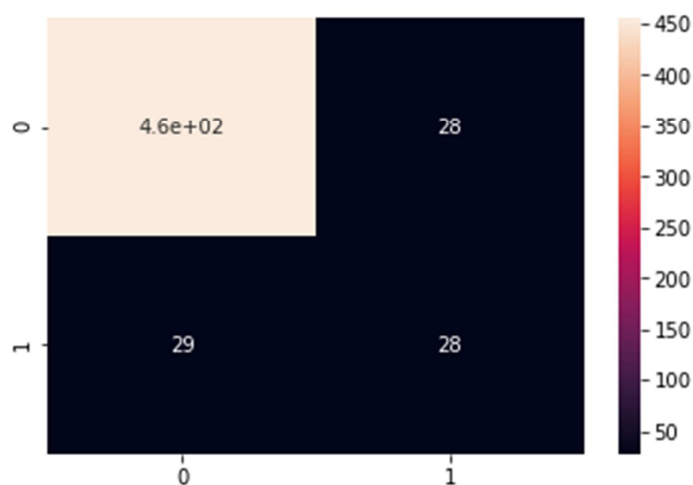


Fonte: Autores.

Ao realizar a análise dos resultados expostos na Figura 10 foi identificado que das 152 redações que não tiveram fuga ao tema, o modelo classificou como fuga 9 redações, indicando assim 5,9% de erro, valor este que representa os falsos positivos. Na segunda linha da matriz de confusão havia 57 redações com fuga ao tema. Destas, o algoritmo classificou como fuga 28 redações, o que equivale a 49% de acerto, este valor representa os verdadeiros positivos. Neste primeiro teste a acurácia (taxa de acertos) ficou em aproximadamente 81,8%, comparando-se os resultados da solução implementada nos experimentos desta pesquisa com os resultados reais oriundos das correções efetuadas pelos professores.

Um segundo teste foi realizado com as redações com notas inferiores a 500, ou seja, normalmente aquelas que são desconsideradas por grande parte das universidades para aprovação dos candidatos. Os resultados podem ser visualizados na Figura 11, que expõe a matriz de confusão gerada para as outras 540 redações que não foram utilizadas na fase de treinamento.

Figura 11: Matriz de Confusão - Segunda base de Teste



Fonte: Autores.

No segundo teste realizado com as redações com notas inferiores obteve-se acurácia de 89,4%. Outra análise foi realizada a partir das previsões para saber se era possível tirar alguma informação dos falsos positivos incorridos, já que o sistema informou que 28 redações foram classificadas como fuga ao tema, mas na verdade não apresentavam fuga ao tema de fato, uma vez que a taxa de erros em relação aos falsos positivos desta segunda base de teste foi de 5,7%.

A Tabela 2 consolida os resultados obtidos no processo de classificação da RNC. Apesar de ter obtido resultados de acurácia superiores a 80% nas duas bases de dados, a maior atenção está em relação aos falsos positivos, ou seja, quando o sistema afirma que a redação teve fuga ao tema de forma equivocada. A taxa de erro ficou entre 5,9% e 5,7% na classificação de fuga para as duas bases, quando na verdade não houve. Ainda assim é importante avaliar se estas redações podem ter fugido parcialmente à temática proposta, o que também irá subtrair nota do estudante.

Tabela 2: Resultados consolidados da Matriz de Confusão - RNC

PRIMEIRA BASE DE TESTE COM REDAÇÕES DE NOTAS SUPERIORES A 500 e 57 REDAÇÕES COM FUGA AO TEMA					
CLASSIFICADOR RNC	ACURÁCIA	VP (Verdadeiro Positivo)	FN (Falso Negativo)	FP (Falso Positivo)	VN (Verdadeiro Negativo)
	81,8%	49,1%	50,9%	5,9%	94,1%
SEGUNDA BASE DE TESTE COM REDAÇÕES DE NOTAS INFERIORES A 500 e 7 REDAÇÕES COM FUGA AO TEMA					
CLASSIFICADOR RNC	ACURÁCIA	VP (Verdadeiro Positivo)	FN (Falso Negativo)	FP (Falso Positivo)	VN (Verdadeiro Negativo)
	89,4%	49,1%	50,9%	5,7%	94,3%

Fonte: Autores.

A Tabela 3 indica os resultados de outras métricas de avaliação, tais como *precision*, *recall* e *F1-score* para cada uma das classes analisadas, sendo '1' para fuga ao tema e '0' para não fuga.

Tabela 3: Resultados consolidados Precisão, Recall e F1-Score - RNC

PRIMEIRA BASE DE TESTE COM REDAÇÕES DE NOTAS SUPERIORES A 500 e 57 REDAÇÕES COM FUGA AO TEMA				
Classificador RNC	Precisão		Recall	F1-Score
	0	94%	94%	94%

	1	50%	49%	50%
SEGUNDA BASE DE TESTE COM REDAÇÕES DE NOTAS INFERIORES A 500 e 7 REDAÇÕES COM FUGA AO TEMA				
		Precisão	Recall	F1-Score
Classificador RNC	0	94%	94%	94%
	1	50%	49%	50%

Fonte: Autores.

Os resultados apresentados quanto à precisão do modelo indicam todas as classificações de classe positivo que o modelo fez, e quantas estão corretas. Já o *Recall* dentre todas as situações de classe Positivo como valor esperado, quantas estão corretas e, por fim, o *F1-Score* faz uma média harmônica entre as outras duas. Essas métricas também apontam um maior grau de acerto para a classe negativa das redações, ou seja, daquelas que não tinham fuga ao tema.

Na Figura 12 é indicado que o algoritmo aplicado também ficou em dúvida. Um padrão verificado foi que os valores inferiores a 80% indicando que havia fuga ao tema, na verdade não deveriam ser classificados dessa forma. Assim, o professor pode utilizar estas previsões para facilitar seu processo de avaliação das redações.

Figura 12: Exemplos de Previsões de falsos positivos

```
-----
{'Fugiu': 0.7833188772201538, 'NaoFugiu': 0.21478599309921265}
Resposta Real: NaoFugiu
Previsão: Fugiu
-----
{'Fugiu': 0.5969784259796143, 'NaoFugiu': 0.4116729497909546}
Resposta Real: NaoFugiu
Previsão: Fugiu
-----
{'Fugiu': 0.7432048916816711, 'NaoFugiu': 0.258165180683136}
Resposta Real: NaoFugiu
Previsão: Fugiu
-----
{'Fugiu': 0.5312862396240234, 'NaoFugiu': 0.47236955165863037}
Resposta Real: NaoFugiu
Previsão: Fugiu
-----
{'Fugiu': 0.6753502488136292, 'NaoFugiu': 0.32637909054756165}
Resposta Real: NaoFugiu
Previsão: Fugiu
-----
```

Fonte: Autores.

A Figura 12 ainda explicita a importância de disponibilizar estes dados ao professor, pois desta forma, mesmo que o sistema classifique a fuga, o avaliador humano conseguirá identificar se algoritmo ficou na dúvida e ainda associar os resultados demonstrados no tópico anterior, quando é possível identificar as palavras que mais aderiram a proposta.

Após o processo de classificação utilizando as redes neurais convolucionais, o próximo passo foi testar outros classificadores, conforme demonstrado nos próximos tópicos.

Classificação de fuga ao tema aplicando outros classificadores selecionados

Nesta subseção são apresentados os resultados referentes aos classificadores de aprendizado de máquina da biblioteca do Scikit Learn. Os classificadores do Scikit Learn selecionados foram explicitados no tópico classificadores, sendo estes: MLP (*MultiLayer Perceptron*), árvores de decisão; Florestas Aleatórias, *Gradiente Boost*, *Ada Boost*, *Stochastic Gradiente Descent (SGD)* e *Support Vector Machines (SVM's)*.

Resultados dos classificadores dos demais classificadores

Neste tópico são demonstrados os resultados apresentados nos classificadores de Aprendizado de Máquina do Scikit Learn, após nova normalização das redações. Para todos os classificadores foram utilizados a mesma base da RNC. As redações foram divididas em treino e teste, utilizando-se os mesmos critérios de divisão estipulados para a RNC, viabilizando assim a comparação com os resultados obtidos.

A primeira métrica utilizada para avaliar os resultados foi a matriz de confusão, avaliando como cômputo a classe real e a classe prevista, com uso dos rótulos ‘P – Positivo’ e ‘N – Negativo’. Outra métrica avaliada foi a acurácia, que mede o percentual de acertos em relação à quantidade de ocorrências. As outras métricas avaliadas foram precisão, *recall* e *f1-score*, assim como no classificador anterior. Na Tabela 4 são expostos os resultados de cada classificador, utilizando-se o mesmo critério na RNC, ou seja, notas atribuídas pelo avaliador humano superiores a 500 pontos.

Tabela 3: Classificadores Scikit Learn – Consolidado da Matriz de Confusão

PRIMEIRA BASE DE TESTE COM REDAÇÕES DE NOTAS SUPERIORES A 500 e 57

REDAÇÕES COM FUGA AO TEMA

Classificador	Acurácia	VP (Verdadeiro Positivo)	FN (Falso Negativo)	FP (Falso Positivo)	VN (Verdadeiro Negativo)
<i>MLPClassifier</i>	78%	33%	67%	4,6%	95,4%
<i>DecisionTreeClassifier</i>	74,6%	14%	86%	2,6%	97,4%
<i>RandomForestClassifier</i>	72,7%	0%	100%	0%	100%
<i>SGDClassifier</i>	78%	47%	53%	11%	89%
<i>SVM (SVC)</i>	72,2%	0%	100%	0%	100%
<i>GradientBoostingClassifier</i>	74,6%	51%	49%	16%	84%
<i>AdaBoostClassifier</i>	77%	44%	56%	11%	89%

Fonte: Autores.

Para identificar os melhores resultados da primeira base de teste levou-se em consideração a acurácia aliada aos VP e VN, além da classificação que apresentou o menor erro do FP, ou seja, aquele resultado que apontou que a redação teve fuga ao tema, quando na verdade não tinha. Isto porque este item é o que anularia a prova do aluno, devendo assumir, portanto, a menor taxa de erro.

Assim, levando-se em consideração tais informações, o classificador que apresentou os melhores resultados foi o *GradientBoostingClassifier*, com maior acerto de fuga ao tema (VP) de 51%, taxa de erro dos FP de 16% e acurácia de 74,6,7%. Outros classificadores que tiveram bons desempenhos foram o *SGDClassifier* e o *MLPClassifier*, com as seguintes taxas de erro: 11% e 4,6%, respectivamente, apresentando a acurácia de 86% para ambos os classificadores.

Na Tabela 5 são exibidos os resultados da segunda base de teste, ou seja, aquelas redações que tiveram notas atribuídas pelo avaliador humano abaixo de 500 pontos.

Tabela 4: Resultados dos Classificadores Scikit Learn – Segunda base de teste

SEGUNDA BASE DE TESTE COM REDAÇÕES DE NOTAS INFERIORES A 500 e 57

REDAÇÕES COM FUGA AO TEMA

CLASSIFICADOR	ACURÁCIA	VP (Verdadeiro Positivo)	FN (Falso Negativo)	FP (Falso Positivo)	VN (Verdadeiro Negativo)
<i>MLPClassifier</i>	91%	33%	67%	3,3%	96,7%
<i>DecisionTreeClassifier</i>	82%	14%	86%	9%	91%
<i>RandomForestClassifier</i>	89,4%	0%	100%	0%	100%
<i>SGDClassifier</i>	86,6%	47%	53%	8,7%	91,3%
<i>SVM (SVC)</i>	89,4%	0%	100%	0%	100%
<i>GradientBoostingClassifier</i>	72,9%	51%	49%	24%	76%
<i>AdaBoostClassifier</i>	71,2%	44%	56%	25%	75%

Fonte: Autores.

Os resultados relativos àquelas redações com notas abaixo de 500 pontos expostos na Tabela 5 expuseram a mesma ordem de hierarquia dos classificadores com melhor desempenho. Os resultados desta tabela, apesar de tratar as redações de notas inferiores, obtiveram resultados similares à base de dados anterior. A maior diferença foi verificada nos FP que, neste caso, obteve como classificadores com menor erro nos Falsos Positivos o MLP e *SGDClassifier*, com respectivamente 3,3% e 8,7%.

Na Tabela 6 são demonstrados os resultados das demais métricas de avaliação para auxiliar no entendimento dos resultados do modelo desenvolvido.

Tabela 5: Métricas Precisão, Recall e F1-Score - Consolidado de Classificadores Scikit Learn

<i>MLPClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	79%	95%	87%
Classe 1 (Fuga ao tema)	73%	33%	46%
<i>DecisionTreeClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	75%	97%	85%
Classe 1 (Fuga ao tema)	67%	14%	23%
<i>RandomForestClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	73%	100%	84%
Classe 1 (Fuga ao tema)	0%	0%	0%
<i>SGDClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	82%	89%	85%
Classe 1 (Fuga ao tema)	61%	47%	53%
<i>SVM (SVC)</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	73%	100%	84%
Classe 1 (Fuga ao tema)	0%	0%	0%
<i>GradientBoostingClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	82%	84%	83%
Classe 1 (Fuga ao tema)	54%	51%	52%
<i>AdaBoostClassifier</i>	Precisão	Recall	F1-Score
Classe 0 (não fuga)	81%	89%	85%
Classe 1 (Fuga ao tema)	61%	44%	51%

Fonte: Autores.

Ao avaliar os resultados da Tabela 6, os classificadores com melhores resultados e que mantiveram valores similares nas métricas de precisão, recall e F1-Score foram o *SGDClassifier* e *GradientBoostingClassifier*. Isto porque os resultados da Classe Negativo (não fuga) ficaram acima de 80% e os acertos da classe Positiva (fuga ao tema) mantiveram a média acima de 50%.

O Classificador MLP manteve a precisão nas duas classes (positivo e negativo) acima de 70%. Já no item *recall*, para a classe positivo obteve um acerto de apenas 33%. O *F1-Score* fez uma média harmônica entre os resultados dos dois anteriores e para a classe positivo também se manteve baixo de 50%.

Avaliação e discussão dos resultados

Levando-se em consideração os experimentos realizados nesta pesquisa, iniciando por aqueles que ainda não utilizaram os classificadores, já foi possível identificar algumas possibilidades para a avaliação da escrita e/ou argumentação dos alunos em redações. Partindo-se deste princípio, entende-se ser possível mensurar o número de redações aderentes à proposta de tema informada, o que pode trazer importante conhecimento ao avaliador ou docente em relação à evolução dos alunos na produção textual. Outro resultado importante ao identificar padrões entre a redação e a proposta do tema foi a possibilidade de identificar as palavras mais aderentes a temática fornecida.

Na avaliação dos resultados obtidos nos classificadores da RNC foi identificado maior ganho em relação aos classificadores do Scikit Learn, tanto em relação à acurácia, quanto em relação aos resultados de falsos positivos, métricas de 'precisão', 'recall' e 'F1-Score'. Apesar dos melhores resultados terem sido obtidos com a RNC, deve-se destacar os bons resultados obtidos com os classificadores Scikit Learn.

Na avaliação para designação do melhor modelo é importante entender como os diferentes modelos seriam aplicados numa situação real. Dessa forma, o percentual de erro do falso positivo (FP) deve ser mínimo, assim como o percentual de verdadeiro positivo (VP) deve ser elevado, o que significa que o modelo identificou mais adequadamente a fuga ao tema, proposta central desta pesquisa.

Dessa forma, para fins de análise dos resultados foram desconsiderados os modelos que não conseguiram identificar a fuga ao tema, ou seja, aqueles que apresentaram a taxa de verdadeiros negativos (VN) de 100%, significando que estes não conseguiram atingir o objetivo proposto nesta pesquisa, quais sejam: *RandomForestClassifier* e SVM (SVC).

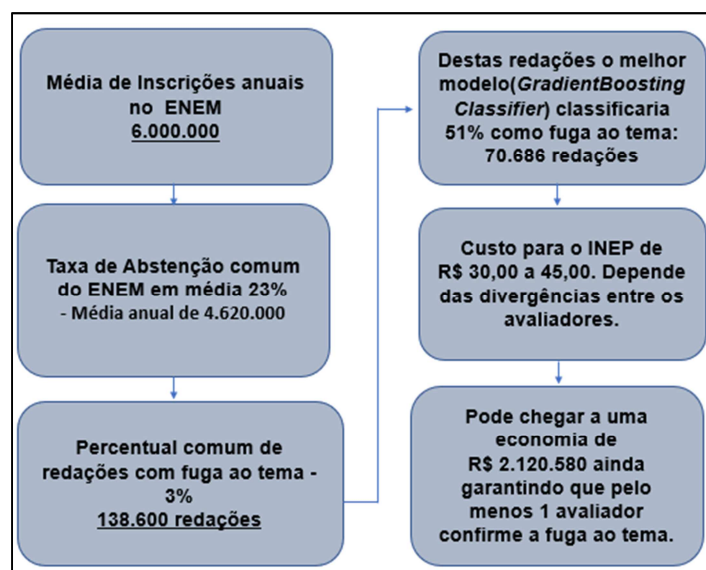
Na sequência, o modelo que teve a melhor acurácia foi a RNC, com resultados de até 89% de acurácia e taxa de falsos positivos (FP) de apenas 5,7%. Contudo, caso se avalie a taxa de verdadeiros positivos (VP), aquela em que o modelo acertou a fuga ao tema, os melhores resultados ocorreram com o *GradientBoostingClassifier*, com 51% e acertos na classe positivo, contudo, a sua taxa de falsos positivos (FP) ficou em 20%, em média. Outro Classificador que obteve resultados melhores em relação aos falsos positivos (FP) foi a MLP, com no máximo 4,6% de erro, taxa de verdadeiros positivos (VP) de 33% e acurácia entre 78% e 90%.

Contudo, tais métricas são passíveis de alcançar melhores resultados quando forem aplicadas em uma base de dados maior e com mais exemplos de fuga ao tema, em especial a RNC. Isto porque, segundo Rodrigues (2018), a RNC obtém melhores resultados com uma grande base de exemplo.

Levando-se em consideração os indicadores alcançados nos modelos testados e a quantidade de redações que são corrigidas anualmente pelo ENEM, em média 4 milhões, de acordo com o portal do INEP (2019), a solução RNC apresentada já poderia indicar de antemão as redações que provavelmente fugiram ao tema, economizando assim o tempo e o custo da correção da redação. Já em sala de aula, a solução delineada nesta pesquisa pode ajudar a agilizar a devolutiva do professor e tornar seu trabalho menos desgastante, por contar com um sistema que indica possíveis falhas de escrita.

A Figura 13 exemplifica uma possível economia de aproximadamente R\$ 2 milhões de reais com a aplicação do modelo RNC validado nesta pesquisa. Ao utilizar o modelo desta pesquisa pressupõe-se que mesmo a redação sendo avaliada por um avaliador humano para garantir que realmente haja fuga ao tema nas redações separadas pelo modelo, a instituição INEP ainda poderia economizar com gastos que seriam dispendidos para até dois avaliadores humanos.

Figura 13: Simulação de economia com modelo aplicado ao ENEM



Fonte: Autores.

Para a sala de aula a aplicação do modelo RNC proporcionaria o ganho em relação ao tempo e menor desgaste do professor na avaliação de textos. Isto seria possível por meio da

disponibilização do modelo validado nesta pesquisa numa ferramenta que indique ao professor a probabilidade de fuga ao tema, indicando ainda as palavras mais aderentes à proposta da redação.

Nas pesquisas realizadas não foram encontrados outros autores que utilizaram da mesma técnica aplicada nesta pesquisa para a avaliação do critério ‘fuga ao tema’ em redações. A pesquisa mais próxima deste trabalho foi o estudo realizado por Passero (2018), que analisou especificamente fuga ao tema. O autor obteve ótimos resultados com uma acurácia de 96,76% e falsos positivos (FP) de 4,24%. Contudo o referido estudo não disponibilizou os resultados da taxa de verdadeiros positivos (VP), ou seja, aqueles resultados que identificaram a fuga ao tema, fator crucial desta pesquisa.

A plataforma CIRA desenvolvida pela Universidade de São Paulo (USP, 2021) também teve o objetivo de corrigir redações automaticamente, contudo sem a verificação de fuga ao tema. Assim, na aplicação que está disponível para uso há uma grande taxa de erros quando avaliadas as redações com notas consideradas baixas. Nos testes realizados, redações que deveriam ter nota 0 (zero) por fuga ao tema, foram avaliadas com notas superiores a 400. Já a pesquisa de Ramisch (2020) também avaliou redações, contudo sua proposta era encontrar problemas de desvios sintáticos, atingindo uma acurácia de 75,6% de acerto.

Desta forma, entende-se que os resultados apresentados nesta pesquisa trazem importantes contribuições à evolução do estudo desta área de pesquisa acadêmica. Assim, a partir dos resultados aqui expostos é possível vislumbrar as primeiras prerrogativas de benefícios da solução ora desenvolvida para auxílio ao trabalho de docentes e avaliadores durante o processo de correção de textos produzidos por alunos ou candidatos. A solução ora validada a partir dos experimentos realizados nesta pesquisa pode também ser aplicada num sistema para avaliar a evolução dos alunos do decorrer de seus estudos acadêmicos, proporcionando assim ao docente conhecer as dificuldades dos alunos numa turma.

CONCLUSÕES

A aplicação das técnicas mencionadas nos experimentos executados nesta pesquisa buscou indicar quais das técnicas de Inteligência Artificial comparadas apresentam melhores resultados para a identificação de fuga ao tema em redações. Assim, entende-se que esta pesquisa respondeu à questão-problema formulada, uma vez que após a aplicação de diferentes classificadores para a formulação de um modelo foi possível indicar aqueles que trazem melhores resultados ao identificar a fuga ao tema nas redações, indicando ainda o percentual de acerto do modelo concebido nesta pesquisa.

Os experimentos trouxeram resultados promissores tanto nas RNCs, como nos classificadores do *Scikit Learn*. O modelo que teve a melhor acurácia foi a RNC, com resultados de até 89,4% de acurácia e taxa de falso positivo (FP) de apenas 5,7%. Contudo, caso se avalie a taxa de verdadeiros positivos (VP), aquela em que o algoritmo acertou a fuga ao tema, os melhores resultados ocorreram com o *GradientBoostingClassifier*, com 51% de acertos na classe Positivo, não obstante sua taxa de Falsos Positivos (FP) tenha sido, em média, de 20%. Outro classificador que obteve resultados melhores em relação aos Falsos Positivos (FP) foi a MLP, com no máximo 4,6% de erro, taxa de verdadeiros positivos (VP) de 33% e acurácia entre 78% e 90%.

Desta forma, o objetivo proposto de comparar diferentes técnicas de IA para classificação de fuga ao tema em textos e identificar aquelas que trouxeram melhores resultados para viabilizar um sistema de correção inteligente de redações foi alcançado. A solução desenvolvida nesta pesquisa possibilita a geração de informações e conhecimentos úteis aos avaliadores de textos educacionais na tarefa de identificação de desvios de escrita e possível fuga ao tema proposto para a elaboração de redação, problemas que, uma vez incorridos, acarretam notas insuficientes aos alunos.

Os resultados dos experimentos aqui realizados demonstram assertividade de 89,4% de acurácia. Tais resultados já possibilitam a criação de uma aplicação para fornecer *feedback* automático como suporte ao professor ou avaliador de textos, o que contribuirá para diminuir o tempo demandado para correção, além de prestar melhor auxílio às instituições, professores e alunos.

A solução ora desenvolvida visa diminuir a desigualdade nos processos seletivos, oferecendo assim maior oportunidade de aprendizagem independente da instituição em que o aluno estude. Em complemento, a solução validada nesta pesquisa, além de proporcionar a possibilidade de

treinar e aperfeiçoar a qualidade na escrita, viabiliza ainda retornos mais rápidos aos envolvidos no processo de ensino-aprendizagem, ou seja, professores e alunos.

Para as instituições de ensino com elevada carga de textos produzidos, a solução provida nesta pesquisa pode viabilizar que o professor passe de uma função de auditoria nas correções, para uma função mais voltada à comprovação da efetividade da correção. Além disso, o fator fadiga do docente seria diminuído, já que atualmente o profissional responsável corrige em torno de 50 redações ao dia, no caso da avaliação de textos produzidos no ENEM. Isto porque, ao adotar a solução validada nesta pesquisa o docente poderia contar com um sistema que indique prováveis erros, o que facilitaria em muito o trabalho dos avaliadores e professores.

A aplicação das primeiras técnicas empregadas nos experimentos realizados já permitiu identificar padrões no conteúdo de cada redação analisada, bem como encontrar relações de similaridade entre as redações sobre um determinado tema em específico, possibilitando assim a classificação da fuga ao tema.

As principais contribuições deste estudo buscam permitir ao avaliador, professor ou empresas que aplicam processos seletivos avaliar as redações com menor esforço, otimizando assim o trabalho e reduzindo o tempo e o custo do processo de avaliação de textos dissertativos. A solução delineada nesta pesquisa pode, portanto, ser primordial na aplicação do ENEM digital, proporcionando assim ao avaliador auxílio na identificação das falhas de escrita e minimizando interferências como fadiga e alteração de humor do avaliador, sintomas estes que podem afetar a correção de um texto dissertativo.

Sob a perspectiva acadêmica, os experimentos realizados e os resultados expostos podem servir de base para estudos que, uma vez alinhados aos conhecimentos dos profissionais de ensino, possam gerar novas abordagens que possibilitem capacitar os alunos a redigirem textos coesos ao tema proposto. Com isso, a desigualdade entre alunos de escolas públicas e privadas pode ser diminuída, uma vez que ter uma plataforma que possibilite o treino mais frequente e resposta mais rápida pode facilitar não apenas o trabalho do professor, como também proporcionar futuramente uma escrita mais clara e desenvolvida, possibilitando assim maior maturidade neste processo.

Uma questão limitadora dos resultados apresentados neste estudo refere-se à base de dados utilizada. Pretende-se adicionar à esta mais redações, notadamente com maior ocorrência de desvios de escrita. Vislumbra-se para tanto, principalmente, aqueles desvios de escrita que acarretaram nota zerada por fuga ao tema, uma vez que no *corpus* deste estudo havia apenas 230 redações com fuga ao tema diagnosticadas pelo avaliador. Outro fator limitador desta pesquisa volta-se a seleção das técnicas de IA empregadas nos experimentos executados neste estudo, a cargo dos autores.

Como sugestões para pesquisas futuras recomenda-se a ampliação da base de dados de redações, visando proporcionar mais efetividade aos experimentos com técnicas de IA. Além disso, também indica-se a possibilidade de adição de outras técnicas de IA, que não aquelas utilizadas neste estudo.

REFERÊNCIAS

AFFONSO, Emmanuel T. F.; SILVA, Alisson M.; SILVA, Michel P.; RODRIGUES, Thiago M. D.; MOITA, Gray F. Uso de redes neurais multilayer perceptron (MLP) em sistema de bloqueio de websites baseado em conteúdo. *Mecânica Computacional*, v. XXIX, n. 93, p. 9075-9090, 2010.

AGÊNCIA BRASIL. **Enem é um dos principais instrumentos de acesso ao ensino superior.** Publicado em 31/10/2019. Disponível em: <https://agenciabrasil.ebc.com.br/educacao/noticia/2019-10/enem-e-um-dos-principais-instrumentos-de-acesso-ao-ensino-superior>. Acesso em: 19/03/2021.

BANERJEE, Dibyendu. *Natural language processing (NLP) simplified : A step-by-step guide*. 2020. Publicado em: 14/04/2020. Data Science Foundation. Disponível em: <https://datascience.foundation/sciencewhitepaper/natural-language-processing-nlp-simplified-a-step-by-step-guide>. Acesso em: 15 abr. 2021.

BIANCHI, Alexandre. *As classificações dos algoritmos de machine learning*. Publicado em 27/05/2020. Disponível em: <https://www.vicere.com.br/insights/as-classificacoes-dos-algoritmos-de-machine-learning>. Acesso em: 09 maio 2021.

BITTENCOURT JÚNIOR, José. A. S. *Avaliação automática de redação em língua portuguesa empregando redes neurais profundas*. 2020. 100 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal de Goiás, Goiânia, 2020.

BRITZ, Denny. *Understanding convolutional neural networks for NLP*. 2015. Publicado em 07/11/2015. Disponível em: <http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/>. Acesso em: 31 mar. 2021.

BROWNLEE, Jason. *Boosting and adaboost for machine learning. Machine learning mastery*. 2016. Disponível em: <https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/>. Acesso em 10 jun. 2021.

CAMPOS, Lorraine V. *Mais de 143 mil participantes tiraram zero na redação do Enem 2019*. 2020. Portal Brasil Escola - UOL. Disponível em: <https://vestibular.brasilecola.uol.com.br/enem/mais-143-mil-participantestiraram-zero-na-redacao-enem-2019/347183.html>. Acesso em 10 jun. 2020.

CÂNDIDO, Thiago G.; WEBBER, Carine G. Avaliação da Coesão Textual: Desafios para Automatizar a Correção de Redações. *RENOTE - Revista Novas Tecnologias na Educação*, v. 16, n. 1, p. 1-10, 2018.

CARNEIRO, Álvaro L. C. *Redes neurais convolucionais para processamento de linguagem natural*. 2020. Publicado em 07/07/2020. Disponível em: <https://medium.com/data-hackers/redes-neurais-convolucionais-para-processamento-de-linguagem-natural-935488d6901b>. Acesso em: 31 mar. 2021.

CARVALHO, A. C. P. F. de.; FAIRHURST, M. C.; BISSET, D. L.. An integrated boolean neural network for pattern classification. *Pattern Recognition Letters*, v. 15, p. 807-813, aug. 1994.

CONEGLIAN, Caio S. *Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na ciência da informação*. 195 f. Tese (Ciência da Informação - FFC) – Universidade Estadual Paulista UNESP, São Paulo, 2018.

COSTA DA SILVA, Josenildo. *Aprendendo em uma floresta aleatória*. 2018. Publicado em: 12/03/2018.

CRUZEIRO DO SUL *Universidades privadas de SP adotam vestibular online e nota do Enem*. 2020. Publicado em: 09/06/2020. Disponível em: <https://www.jornalcruzeiro.com.br/brasil/universidades-privadas-de-sp-adotam-vestibular-online-e-nota-do-enem/>. Acesso em: 29 mar. 2021.

DIANA, Daniela B. G. *Os 16 maiores erros de redação cometidos pelos estudantes*. 2021. Publicado em 08/01/2021. Disponível em: <https://www.todamateria.com.br/erros-de-redacao/>. Acesso em 29 mar. 2021.

EGGERS William, D.; SCHATSKY, David.; VIECHNICKI, p. *AI-augmented government using cognitive technologies to redesign public sector work*. New York: Deloitte, 2017. Disponível em: https://www2.deloitte.com/content/dam/insights/us/articles/3832_AI-augmented-government/DUP_AI-augmented-government.pdf. Acesso em: 30 mar. 2021.

GOMES, Maria. de F. C. A PNA e a unidade dialética afeto-cognição nos atos de ler e escrever. *Revista Brasileira de Alfabetização*, n. 10, edição especial, p. 122-124, 2020. <https://doi.org/10.47249/rba.2019.v1.368>.

GONÇALVES, Eduardo C. Mineração de texto - Conceitos e aplicações práticas. *SQL Magazine*, v. 105, p. 31-44, nov. 2012. Disponível em: https://www.researchgate.net/publication/317912973_Mineracao_de_texto_-_Conceitos_e_aplicacoes_praticas. Acesso em: 15 abr. 2021.

GOODFELLOW, Ian.; YOSHUA Bengio. *Deep learning*. Cambridge: MIT, 2016.

HAN, Jiawei.; KAMBER, Micheline.; PEI, Jian. *Data mining: concepts and techniques: concepts and techniques*. New York: Elsevier, 2011.

HARIRI, Reihaneh. H., FREDERICKS, Erick M.; BOWERS, Kate M. Uncertainty in big data analytics: survey, opportunities, and challenges. *Journal of Big Data*, v; 6, n. 1, p. 1-16, 2019.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *A redação no Enem 2019: cartilha do participante*. Brasília, 2019. Disponível em: http://inep.gov.br/informacao-da-publicacao/-/asset_publisher/6JYIsGMAMkW1/document/id/6736715. Acesso em 20 out. 2019.

INEP. *Entenda como é calculada a nota do Enem*. 2020a. Disponível em: <http://portal.mec.gov.br/ultimas-noticias/418-enem-946573306/84461-entenda-como-e-calculada-a-nota-do-enem>. Acesso em 10 jun. 2020.

INEP. *65% dos inscritos no Enem já concluíram o ensino médio em anos anteriores*. 2020b. Disponível em: http://portal.mec.gov.br/index.php?option=com_content&view=article&id=90701:65-dos-inscritos-no-enem-ja-concluíram-o-ensino-medio-em-anosanteriores&catid=418&Itemid=86. Acesso em 10 jun. 2020.

LEITE, Tiago M. *Redes neurais, perceptron multicamadas e o algoritmo backpropagation*. 2018. Publicado em 10/05/2018. Disponível em: <https://medium.com/ensina-ai/redes-neurais-perceptron-multicamadas-e-o-algoritmo-backpropagation-caf89778f5b8>. Acesso em: 10 maio 2021.

LUDERMIR, Teresa B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, v. 35, n. 101, p. 85-94, Apr. 2021.

MEC – Ministério da Educação. *MEC realiza conferência para discutir estratégias de alfabetização no Brasil*. 2019. Publicado em: 22/10/2019. Disponível em: <http://portal.mec.gov.br/component/tags/tag/5?start=60>. Acesso em: 15 abr. 2021.

MITCHELL, Tom. M. *Machine learning*. New York: McGraw-Hill, 1997.

MORAIS, Edison A. M.; AMBRÓSIO, Ana P. L. *Mineração de textos*. Goiás: UFG, 2007.

MOREIRA, Sandro. *Rede neural perceptron multicamadas*. 2018. Publicado em 24/12/2018.

MÜLLER, Sarah.; BERGANDE, Bianca.; BRUNE, Philipp. (2018). Robot tutoring: on the feasibility of using cognitive systems as tutors in introductory programming education - A teaching experiment. In: EUROPEAN CONFERENCE OF SOFTWARE ENGINEERING EDUCATION (ECSEE'18), 3rd, 2018. *Proceedings...* Association for Computing Machinery, New York, USA, p. 45–49.

MUYLAERT, Renata. *Pandemia do novo coronavírus, Parte 6: inteligência artificial (NLP)*. 2020. Disponível em: <https://marcoarmello.wordpress.com/2020/08/19/coronavirus6/>. Acesso em 27 jul. 2021.

NOBRE, João. C. S.; PELLEGRINO, Sérgio. R. M. *ANAC: um analisador automático de coesão textual em redação*. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION - SBIE, 2010. *Anais...* São Paulo: SBC, 2010, p. 1-12.

PASSERO, Guilherme. *Deteção de fuga ao tema em redações de língua portuguesa*. 145 f. Dissertação (Mestrado em Computação Aplicada) - Universidade do Vale do Itajaí, Santa Catarina, 2018.

PESSANHA, Cinthia. *Random Forest: como funciona um dos algoritmos mais populares de ML*. 2019. Publicado em 20/11/2019.

PINHO, Cíntia M. A., VANIN, Anderson. S.; BELAN; Peterson.; NAPOLITANO, Domingos M. R. Uma ferramenta on-line para ensino de Redação, baseada nos critérios avaliativos do ENEM. In: KMBRASIL 2020 – Congresso Brasileiro de Gestão do Conhecimento, 15º, São Paulo. *Anais...* São Paulo: SBGC, 2020, p. 599-615.

PRATES, Wladimir R. *Introdução ao processamento de linguagem natural (PLN)*. 2019. Publicado em: 01/08/2019.

PREMLATHA, Karan. R. *What is AI? In a simple way*. 2019. Publicado em: 05/02/2019.

PREUSS, Evandro.; BARONE, Dante A. C.; HENRIQUES, Renato V. B. Uso de técnicas de inteligência artificial num sistema de mesa tangível. In: WORKSHOP DE INFORMÁTICA NA ESCOLA, 26º, 2020, *Anais...* Porto Alegre: SBC, 2020, p. 439-448.

RAMISCH, Renata. *Caracterização de desvios sintáticos em redações de estudantes do ensino médio: subsídios para o processamento automático das línguas naturais*. 156 f. Dissertação (Mestrado em Linguística) - Universidade Federal de São Carlos, São Carlos, 2020.

RAMOS, Jorge L. C.; SILVA, João C. S.; PRADO, Leonardo C.; GOMES, Alex S.; RODRIGUES, Rodrigo L. Um estudo comparativo de classificadores na previsão da evasão de alunos em EAD. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION, VII, 2018. *Anais...* São Paulo: SBC, 2018, p. 1463-1472.

RIOLFI, Claudia R.; IGREJA, Suelen G. da. Ensinar a escrever no ensino médio: cadê a dissertação? *Educação e Pesquisa*, v. 36, n. 1, p. 311-324, Apr. 2010.

RODRIGUES, Diego A. R. *Deep learning e redes neurais convolucionais: reconhecimento automático de caracteres em placas de licenciamento*. 37f. Monografia (Ciência da Computação) - Universidade Federal da Paraíba, João Pessoa, 2018.

RODRIGUES, Vitor. *Métricas de avaliação - Quais as diferenças?* 2019. Publicado em 12/04/2019.

RUSSO, Inês F. D. *O impacto da inteligência artificial na sustentabilidade ambiental : uma agricultura sustentável*. 84f. Dissertação (Mestrado em Gestão de Sistemas de Informação) - Universidade de Lisboa, Lisboa, 2020.

SANTOS JÚNIOR, Jário J. dos. *Modelos e técnicas para melhorar a qualidade da avaliação automática para atividades escritas em língua portuguesa brasileira*. 76 f. Dissertação (Mestrado em Informática) – Universidade Federal de Alagoas, Maceió, 2017.

SCIKIT LEARN. *AdaBoost*. 2021a. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html#adaboost>. Acesso em: 11 maio 2021.

SCIKIT LEARN. *Gradient tree boosting*. 2021b. Disponível em: <https://scikit-learn.org/stable/modules/ensemble.html#gradient-boosting>. Acesso em: 11 maio 2021.

SCIKIT LEARN. *Stochastic gradient descent*. 2021c. Disponível em: <https://scikit-learn.org/stable/modules/sgd.html#sgd>. Acesso em: 11 maio 2021.

SCIKIT LEARN. *Support vector machines*. 2021d. Disponível em: <https://scikit-learn.org/stable/modules/svm.html#svm-classification>. Acesso em: 11 maio 2021.

SCIKIT LEARN. *Supervised learning*. 2021e. Disponível em: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. Acesso em: 11 maio 2021.

SILVA, Jonhy. *Uma breve introdução ao algoritmo de machine learning gradient boosting utilizando a biblioteca Scikit-Learn*. 2020. Publicado em: 22/06/2020.

SOUZA, Vanessa F. de.; PERRY, Gabriela T. Mineração de texto em moocs: análise da relevância temática de postagens em fóruns de discussão. *RENOTE - Revista Novas Tecnologias na Educação*, v. 17, n. 3, p. 204-2013, 2019.

SQUARISI, Dad.; SALVADOR, Arlete. *A arte de escrever bem: um guia para jornalistas e profissionais do texto*. 9 ed. São Paulo: Contexto, 2020.

STEFANINI – Group. *PNL: entenda o que é o processamento de linguagem natural*. 2019. Disponível em: <https://stefanini.com/pt-br/trends/artigos/oque-e-processamento-de-linguagem-natural>. Acesso em 20 maio 2021.

UNIVERSIA. *Entrevista com ex-corretor de redação*. 2015. Disponível em: <https://www.universia.net/br/actualidad/orientacion-academica/corretorredaco-do-enem-leva-cerca-2-minutos-prova-diz-professor-1132810.html>. Acesso em: 19 mar. 2021.

USP – Universidade de São Paulo. *USP desenvolve ferramenta de correção automática de redações*. Publicado em 10/03/2021. Disponível em: <http://www.saocarlos.usp.br/usp-desenvolve-ferramenta-de-correcao-automatica-de-redacoes/>. Acesso em 29 mar. 2021.

WALTRICK, Camila. *Machine learning - O que é, tipos de aprendizagem de máquina, algoritmos e aplicações*. 2020. Publicado em: 07/05/2021.

DECLARAÇÃO DE CONTRIBUIÇÃO DOS AUTORES

Nós, Cíntia Maria de Araújo Pinho, Marcos Antonio Gaspar e Renato José Sassi declaramos a seguir a contribuição de cada autor para a elaboração do trabalho intitulado “APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES”.

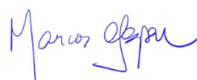
Os autores do trabalho manifestam os respectivos papéis exercidos.

Papéis de autoria	Autor 1 Cíntia Maria de Araújo Pinho	Autor 2 Marcos Antonio Gaspar	Autor 3 Renato José Sassi
Administração do Projeto	X	X	
Análise Formal	X	X	
Conceituação	X	X	X
Curadoria de Dados	X		
Escrita - Primeira versão	X	X	
Escrita - Revisão e Edição		X	X
Investigação	X	X	
Metodologia	X	X	X
Obtenção de Financiamento			
Recursos	X	X	X
Software	X		X
Supervisão	X	X	
Validação	X	X	X
Visualização	X	X	

São Paulo, 04 de março de 2022.



**Cíntia Maria de
Araújo Pinho**



**Marcos Antonio
Gaspar**



Renato José Sassi

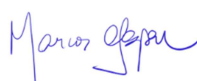
DECLARAÇÃO DE INDICAÇÃO DE SUBMISSÃO DE TRABALHO

Nós, Cíntia Maria de Araújo Pinho, Marcos Antonio Gaspar e Renato José Sassi declaramos a indicação de periódico associado à SciELO para realizar o depósito desta versão *preprint* no repositório SciELO Preprint.

São Paulo, 04 de março de 2022.



**Cíntia Maria de
Araújo Pinho**



**Marcos Antonio
Gaspar**



Renato José Sassi

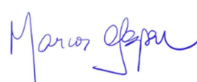
DECLARAÇÃO DE SUBMISSÃO AO SCIELO PREPRINTS

Nós, Cíntia Maria de Araújo Pinho, Marcos Antonio Gaspar e Renato José Sassi declaramos que todos os coautores do trabalho intitulado “APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES” concordam com a sua submissão à plataforma SciELO Preprints.

São Paulo, 04 de março de 2022.



**Cíntia Maria de
Araújo Pinho**



**Marcos Antonio
Gaspar**



Renato José Sassi

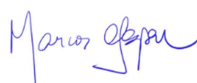
DECLARAÇÃO DE CONFLITO DE INTERESSES

Nós autores, Cíntia Maria de Araújo Pinho, Marcos Antonio Gaspar e Renato José Sassi declaramos não haver conflitos de interesses que inviabilizem a publicação do trabalho intitulado “APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES”.

São Paulo, 04 de março de 2022.



**Cíntia Maria de
Araújo Pinho**



**Marcos Antonio
Gaspar**



Renato José Sassi

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.