

ARTIGO

APLICAÇÃO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA CLASSIFICAÇÃO DE FUGA AO TEMA EM REDAÇÕES¹

CINTIA MARIA DE ARAÚJO PINHO¹

ORCID: <https://orcid.org/0000-0003-0525-5072>
<cintia.pinho01@gmail.com>

MARCOS ANTONIO GASPAR²

ORCID: <https://orcid.org/0000-0002-2422-2455>
<marcos.antonio@uni9.pro.br>

RENATO JOSÉ SASSI²

ORCID: <https://orcid.org/0000-0001-5276-4895>
<sassi@uni9.pro.br>

¹ Centro Estadual de Educação Tecnológica Paula Souza. São Paulo (SP), Brasil.

² Universidade Nove de Julho. São Paulo (SP), Brasil.

RESUMO: O processo de correção manual de redações acarreta algumas dificuldades, dentre as quais se apontam o tempo dispendido para a correção e a 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 tem-se revelado promissora no processo de avaliação automatizada da linguagem escrita. O objetivo desta pesquisa é comparar diferentes técnicas de Inteligência Artificial 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 em 1.320 redações elaboradas em língua portuguesa visando a classificação desses textos para normalizar, identificar padrões e categorizar as redações em 119 temas diferentes. Os resultados indicam que o classificador Rede Neural Convolutacional 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.

¹ Editores participantes do processo de avaliação por pares aberta: Suzana dos Santos Gomes e Eucídio Pimenta Arruda

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 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 Convolutional Neural Network classifier obtained greater gain in relation to the other classifiers analyzed, both in 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 ensayos 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. La evasión del tema es uno de los elementos evaluados en la redacción del Enem que puede anular el ensayo. 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 se ha mostrado prometedor en el proceso de evaluación automatizada del lenguaje escrito. El objetivo de esta investigación es comparar diferentes técnicas de Inteligencia Artificial para la clasificación de evasión del tema en textos e identificar aquella con mejor resultado para habilitar un sistema inteligente de corrección de ensayos. 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 Convolutiva 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 ensayos.

Palabras clave: ensayos, evaluación automática de ensayos, 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, imprescindível para o desenvolvimento da cognição humana. Além de estar inserida nas atividades sociais e no contexto profissional do indivíduo, a escrita integra o seu processo de crescimento pessoal e de ensino-aprendizagem. 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” (Brasil, 2019, p. 5). 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.

Nesse sentido, o professor assume importante papel no desenvolvimento das competências voltadas à leitura e escrita. Lesme (2021) entrevistou uma avaliadora do Enem aplicado no ano de 2021, que corrigiu cerca de cem a 150 redações diárias, levando de três a cinco minutos para corrigir cada redação. Starlles (2022), ao entrevistar outra avaliadora, verificou que ela levou em média um minuto e meio para corrigir cada redação e que somente chegou a esse tempo depois de anos de experiência e execução de trabalho repetitivo.

Em sala de aula, estes tempos podem ser superiores, pois a análise realizada pela Universia, com a participação do Centro de Treinamento e Capacitação para o Exame Nacional do Ensino Médio (Enem), destacou que um professor pode levar de quarenta segundos a dez minutos para corrigir uma redação, de acordo com a qualidade da escrita apresentada no texto (Universia, 2015).

Para exemplificar de forma prática o volume dos dados apresentados nos parágrafos anteriores, caso um professor de língua portuguesa avalie uma redação para cada aluno de um total de 500 estudantes, ele poderá demorar de oito a oitenta e três horas para corrigir as redações, podendo este tempo variar de acordo com sua experiência. Tal estimativa leva em consideração que o profissional gaste, respectivamente, entre um e dez minutos para avaliar cada redação.

O cenário exposto evidencia que os professores passam por dificuldades ao avaliar individualmente os textos de diferentes alunos. Um 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 e que, em alguns casos, após a correção dos textos, os professores comentavam 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), observou-se que mais de 60% dos professores analisados utilizaram menos de 25% de seu tempo para o ensino e devolutivas de redações. Diante desta realidade, é relevante desenvolver soluções que possibilitem otimizar o trabalho do professor ao proporcionar a redução do tempo e dos custos associados ao processo de avaliação de textos. Tais contribuições proporcionariam economia de recursos, que poderiam ser destinados a outras atividades do docente, tais como preparação de aulas, planejamento de conteúdos, atualização de competências e disponibilidade para sanar dúvidas dos alunos, dentre outras possibilidades.

Em relação ao aluno, ressalta-se que os conhecimentos absorvidos durante o ensino fundamental e o médio são avaliados no Enem, uma vez que este exame mensura o desempenho dos alunos ao término do ensino médio. Tal avaliação é uma das principais formas de ingresso ao ensino superior no Brasil. Em 2019, “[...] o ENEM foi responsável por 32% das seleções em processos seletivos de entrada em cursos superiores” (Tokarnia, 2019, n.p.).

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) divulgou como é realizada a avaliação das redações. Os textos são corrigidos por mais de 5 mil avaliadores, que corrigem 150 textos a cada três dias. A cada 50 redações, o avaliador recebe duas delas já avaliadas por outro especialista (Inep, 2022, n.p.). Assim, cada redação é corrigida por dois professores que desconhecem a nota atribuída pelo outro, além de não saberem quem é o candidato avaliado (Inep, 2020). Quando há discrepância nas notas atribuídas pelos dois avaliadores, um terceiro avaliador é convocado.

Este processo ocorre quando há diferença de mais de cem pontos na nota total atribuída ou caso ocorra diferença de 80 pontos por competência (Inep, 2022).

Este processo manual acarreta elevado tempo para a divulgação dos resultados aos candidatos, além de consumir recursos para a manutenção dessa metodologia. Para se ter uma ideia da dimensão desse processo, em 2023 houve 3.476.226 inscrições para o Enem, o que demonstra a envergadura e complexidade do sistema 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, mil pontos como nota atribuída à redação (Inep, 2022).

No entanto, encontram-se resultados insatisfatórios no desempenho dos estudantes nesse tipo de exame. Ao se avaliarem os microdados do Enem 2022 disponibilizados pelo Inep no início de 2023 (Inep, 2023), 2.355.395 candidatos estiveram presentes em todos os dias. Foi identificado que aproximadamente 129 mil estudantes tiraram zero na redação do Enem e somente 32 indivíduos alcançaram a nota máxima em sua produção textual. De acordo com a base analisada, os principais motivos que levaram à atribuição da nota zero são redação em branco (43,9%), fuga ao tema (24,6%) e cópia do texto motivador (20,8%). 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, o que é corroborado por um estudo realizado por Diana (2021), que destaca esse problema na elaboração de redações, o que acarreta ao aluno uma nota muito baixa ou mesmo o cancelamento da prova do Enem.

A fuga ao 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 com a correção da redação.

Nesse contexto, este trabalho trata exclusivamente dos desvios de escrita que tangenciam a competência 2 estabelecida pelo Enem, ou seja, a fuga ao tema. O conjunto de circunstâncias relatado anteriormente ressalta a importância de se estudar uma forma de auxílio aos professores para a 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.

Apesar das possíveis contribuições positivas da aplicação de soluções de Inteligência Artificial nas atividades docentes, faz-se necessário indicar também aspectos que estabelecem pontos de atenção a respeito deste fenômeno. Nesse sentido, Pinto (2005) alerta para a não neutralidade da tecnologia em suas aplicações e, por consequência, seu impacto em indivíduos, profissionais, organizações e sociedade. Um estudo realizado por Silveira e Barros (2021) a respeito das competências do docente de ensino superior indicou que 34,09% das competências analisadas poderão futuramente ser absorvidas pela Inteligência Artificial em algum momento do desenvolvimento das tecnologias a ela associadas. Portanto, tal cenário sinaliza consequências futuras nas atividades do professor que não devem ser desprezadas no contexto da aplicação da Inteligência Artificial na profissão docente.

Não obstante, no contexto específico deste estudo, 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 essa demanda, dentre as quais se destacam como alternativas o Processamento de Linguagem Natural e a Mineração de Textos, além de diferentes técnicas de classificação de textos.

Assim, a presente pesquisa tem como objetivo comparar diferentes técnicas de Inteligência Artificial 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 apoio ao professor. 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.

Além da comparação entre os classificadores, a presente pesquisa apresenta ainda os resultados dos experimentos já aplicados numa plataforma web denominada CRIA (Corretor de

Redações por Inteligência Artificial). O intuito dessa plataforma é proporcionar a análise instantânea de redações simulando as diretrizes e notas do Enem, indicando os desvios cometidos no texto. As instituições de ensino podem disponibilizar a plataforma CRIA para seus alunos a fim estimular a interação do professor com o estudante.

Quando o aluno envia sua redação para a plataforma, ele obtém a nota instantaneamente e depois recebe a correção detalhada com as indicações específicas, inclusive a indicação de fuga ao tema, que sinaliza a probabilidade de ocorrer a fuga e marcação no texto das palavras aderentes à proposta de tema da redação. Ao corrigir os pontos indicados pela plataforma CRIA, o aluno aprenderá com os seus erros. Ao final desse processo, o aluno envia a redação para o professor responsável pelo componente curricular de língua portuguesa. Esse professor analisará as correções sugeridas pela solução de Inteligência Artificial e as alterações realizadas pelo aluno, podendo alterar os desvios apontados e as notas atribuídas, ou ainda acrescentar a indicação de mais erros na redação. A proposta da plataforma desenvolvida volta-se a diminuir a sobrecarga de trabalho e otimizar a gestão das turmas do professor.

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, um 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 com as notas atribuídas em correções feitas por avaliadores humanos. Os autores afirmam ainda que a correção realizada por um programa de computador não sofre determinadas interferências externas, tais como fadiga e alteração de humor, permitindo assim avaliar e analisar sempre de forma equânime. Não obstante, um importante contraponto a ser destacado é que soluções de Inteligência Artificial podem, eventualmente, carregar vieses e preconceitos em sua estrutura. Mesmo 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 Júnior (2017) desenvolveu uma pesquisa para melhorar a qualidade da avaliação automática de textos dissertativos aplicando o Processamento de Linguagem Natural e Redes Neurais Artificiais. 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 zero a mil 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 nesse experimento atribuiu notas para as redações com uma taxa de erro de cem 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 processamento de linguagem natural. 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 soluções para a avaliação automática de redações, abrindo-se, assim, novas possibilidades de pesquisa.

Passero (2018) desenvolveu um projeto para a detecção de fuga ao tema nas redações aplicando técnicas de processamento de linguagem natural 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 12 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 de 4,24%, aquela

que detecta que a redação teve uma fuga ao tema, quando na verdade não teve. Nesse 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 foram com o algoritmo Regressão Logística, que alcançou 75,62% de acurácia.

Riedo (2020) desenvolveu um instrumento embasado em técnicas e procedimentos de Inteligência Artificial para a avaliação qualitativa de produções escritas em educação à distância. Sua solução considerou a relação entre os conceitos elaborados, e, devido às variações na maneira humana de se expressar por escrito, aqueles não inicialmente presentes na base conceitual de avaliação puderam ser “aprendidos” de modo a ampliar a base. Como principal contribuição desse estudo, a solução desenvolvida foi capaz de discriminar qualitativamente as produções escritas.

Bittencourt Júnior (2020) propôs a avaliação automática de redações utilizando redes neurais profundas. 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.

As pesquisas indicadas até aqui trouxeram avanços na aplicação de técnicas da Inteligência Artificial para a correção automatizada de textos. Esta pesquisa, além de aplicar diferentes técnicas de classificação para detecção de fuga ao tema em redações, busca ainda comparar diferentes algoritmos de classificação, além de estabelecer a relação com uma Rede Neural Artificial. Tais comparações foram importantes para a compreensão de que muitas vezes não há a necessidade de se utilizar uma rede neural para resolver todos os problemas de previsão inerentes à classificação de textos. Nesse sentido, o que expressa a efetiva necessidade de utilização de uma rede neural é quando a quantidade de dados já não converge de forma equivalente aos algoritmos mais simples de Aprendizado de Máquina. Assim, nesta pesquisa, pôde ser visto que esses algoritmos convergiram de forma bem parecida, ou seja, demonstrando não ser necessário despender o gasto computacional exigido por uma rede neural, tanto em relação ao treinamento quanto para aplicação dos modelos gerados.

INTELIGÊNCIA ARTIFICIAL APLICADA À EDUCAÇÃO

A Inteligência Artificial 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 Inteligência Artificial 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 Inteligência Artificial 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 Inteligência Artificial na educação são promissoras, ao passo que as máquinas vão se ajustando às necessidades individuais de cada profissional.

A Inteligência Artificial utiliza-se de diferentes técnicas para fornecer informações baseadas em grandes volumes de dados; dentre elas, destacam-se a Mineração de Textos, o Processamento de Linguagem Natural e Aprendizado de Máquina. Este último engloba as técnicas inteligentes de classificação, com destaque para as redes neurais artificiais (Eggers; Schatsky; Viechnicki, 2017; Hariri; Fredericks; Bowers, 2019). Estas técnicas serão tratadas nos próximos tópicos, a fim de compor um breve panorama do conjunto de métodos mencionados.

Mineração de textos (MT)

Morais e Ambrósio (2007) definem MT 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 MT envolve a aplicação de algoritmos computacionais que processam textos e identificam informações úteis e implícitas, que normalmente não poderiam ser recuperadas utilizando-se métodos tradicionais de consulta, pois a informação contida nesses textos não pode ser obtida de forma direta.

Processamento de linguagem natural

O Processamento de Linguagem Natural 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 processamento de linguagem natural são baseadas em modelos de linguagem que definem uma distribuição de probabilidade sobre sequências de palavras, caracteres ou *bytes* em uma linguagem natural (Coneglian, 2018; Goodfellow; Yoshua, 2016).

O processamento de linguagem natural busca padrões e indicativos que auxiliem na compreensão do texto em análise. Assim, os estudos de processamento de linguagem natural 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 (Coneglian, 2018; Goodfellow; Yoshua, 2016).

O processamento de linguagem natural 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 (PNL [...], 2019; Prates, 2019).

Técnicas de inteligência artificial

Atualmente as principais técnicas inteligentes estão inseridas no contexto de Aprendizado de Máquina. Esta área é dedicada ao estudo de algoritmos de previsão e inferência, que buscam simular em computadores o cérebro enquanto máquina de aprendizado. O Aprendizado de Máquina inclui técnicas estatísticas e de Inteligência Artificial 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 (Bianchi, 2020; Muylaert, 2020; Russo, 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. Nesse 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 classificar ou prever ocorrências futuras;
- **Aprendizado Não Supervisionado:** ocorre 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. Esse tipo de aprendizado é, geralmente, aplicado quando o objetivo é gerar agrupamentos;
- **Aprendizado por Reforço:** ocorre 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 no Aprendizado de Máquina 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 (Han; Kamber; Pei, 2011; Ramos *et al.*, 2018).

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, a mineração de textos, a 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 dessas informações para ensinar o sistema a classificar corretamente os documentos.

A Aprendizagem Profunda, do inglês *Deep Learning*, é um ramo do Aprendizado de Máquina 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 (Baberjee, 2020; Bianchi, 2020; Premlatha, 2019). Segundo os autores, a Aprendizagem Profunda é uma das arquiteturas de Aprendizado de Máquina que tem sido aplicada na marcação de classes gramaticais, tradução e classificação de textos.

Classificadores

Com o objetivo de comparar diferentes técnicas de Inteligência Artificial para classificação de fuga ao tema em redações, foram utilizados classificadores de aprendizado supervisionado, explicados anteriormente. Os algoritmos classificadores destacados a seguir são da biblioteca Scikit-Learn (2021a), que utilizam técnicas de Aprendizado de Máquina. Todos os classificadores indicados a seguir foram aplicados nos experimentos realizados neste trabalho.

Multilayer Perceptron (MLP): 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 as classes, 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, recalcular o valor dos pesos do vetor da última camada de neurônios (Leite, 2018; Moreira, 2018).

Árvores de Decisão: o *Decision Tree Classifier* é um algoritmo de aprendizado supervisionado, utilizado 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 treinadas 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 aprendizes fracos com o objetivo de produzir um

melhor modelo preditivo. Um aprendiz fraco (*weak learner*) é definido como um classificador que é ligeiramente correlacionado com a classificação verdadeira. Isso significa que, em um aprendiz fraco, o desempenho em qualquer conjunto de treinamento é ligeiramente melhor que a previsão ao acaso. Na técnica de *Boosting*, cada aprendiz 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. É um método de conjunto que treina e implanta árvores de decisão em série, ou seja, 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. 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 *Support Vector Machines* 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 desse algoritmo são a eficiência e a facilidade de implementação, pois ele 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): conjunto de algoritmos de aprendizado supervisionado usado para classificação, regressão e detecção de *outliers* (um valor atípico ou resultado que foge da média). 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 classificadores de Aprendizado de Máquina destacados anteriormente, também existem os classificadores baseados em Aprendizado Profundo, como a Rede Neural Convolutiva (Scikit-learn, 2021e).

Rede Neural Convolutiva (RNC): algoritmo de Aprendizado Profundo que pode captar uma imagem de entrada, atribuir importância (pesos e vieses que podem ser aprendidos) a vários aspectos e objetos da imagem e ser capaz de diferenciar um do outro. As Redes Neurais Convolutivas 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 Redes Neurais Convolutivas em problemas de Processamento de Linguagem Natural, para os quais resultados promissores têm sido obtidos. Britz (2015), Carneiro (2020) e Rodrigues (2018) destacam que as Redes Neurais Convolutivas se mostram muito eficazes para tarefas de classificação textual.

Métricas de avaliação dos classificadores

Duas etapas principais são realizadas para que uma técnica classifique os dados de uma base. Na primeira etapa, é gerado o modelo que aprende por meio do treinamento dos dados, normalmente utilizando 70% a 80% da base. Para este particionamento, dá-se o nome de *Cross Validation* (CV) ou Validação Cruzada, que é uma técnica muito utilizada para avaliação de desempenho de modelos de aprendizado de máquina. Na segunda etapa, os dados separados são testados, entre 30% e 20% da base, para estimar o desempenho da técnica, mensurando-se, assim, os acertos do modelo (Han; Kamber; Pei, 2011; Ramos *et al.*, 2018). Após a classificação, é necessário avaliar o desempenho do classificador, sendo utilizadas para tanto algumas métricas. Para distinguir entre a classe real e a classe prevista, são utilizados os rótulos “Positivo (P)” e “Negativo (N)”, que são usados para as previsões de classe produzidas por um modelo.

Segundo Ramos *et al.* (2018), dado um classificador e uma instância a classificar, é gerada uma Matriz de Confusão (MC) com quatro resultados possíveis (Quadro 1):

- a) Verdadeiro Positivo (VP), quando o rótulo avaliado é verdadeiro e o modelo trouxe como resultado um valor positivo, indicando acerto do modelo;

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

Quadro 1 – Matriz de confusão

		Prevista	
		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 no Quadro 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), conforme demonstradas no Quadro 2.

Quadro 2 – Métricas de avaliação

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

Fonte: Adaptado de Rodrigues (2019).

As métricas apresentadas podem ser utilizadas em diferentes tipos de classificadores, tais como os que foram demonstrados nos tópicos anteriores: *Multilayer Perceptron* (MLP); *Decision Tree*; *Random Forest*; *Gradiente Boosting*; *Adaboosting*; *Support Vector Machine*; 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. Neste caso, as redações foram corrigidas por um professor, em ambientes de aprendizado voltados aos parâmetros aplicados no Enem. Todas as redações componentes do *corpus* utilizado nos experimentos já tinham sido 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 esta última (fuga) foi responsável pelo processo de classificação, consistindo no atributo-alvo deste estudo.

Quadro 3 – Estrutura da base de dados usada na fase inicial dos experimentos

Redação	Tema	Texto Motivador	Nota comp1	Nota comp2	Nota comp3	Nota comp4	Nota comp5	Total	avaliação	Fuga
Texto com a redação	Tem da redação	Texto fornecido na proposta, falando sobre o assunto da temática.	Notas entre: 0, 40, 80, 120, 160 e 200.	Notas entre: 0, 40, 80, 120, 160 e 200.	Notas entre: 0, 40, 80, 120, 160 e 200.	Notas entre: 0, 40, 80, 120, 160 e 200.	Notas entre: 0, 40, 80, 120, 160 e 200.	Soma total das notas das competências entre de 0 a 1000	Comentário do professor que avaliou a redação	sim ou não

Fonte: Elaborado pelos 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 – Distribuição das notas das redações

Quantidade	Redações Avaliadas	
	Nota	Percentual
266	0 pontos. (Obs.: 230 redações foram classificadas como fuga ao tema).	20,1%
33	20 a 100 pontos	2,5%
172	120 a 300 pontos	13,5%
293	320 a 500 pontos	22,2%
252	520 a 700 pontos	19,2%
192	720 a 900 pontos	14,5%
112	920 a 1.000 pontos	8,5%
Total: 1.320		Total: 100,0%

Fonte: Elaborado pelos autores.

Distribuição dos dados

Para a separação da base de dados, utilizou-se a validação cruzada, com o objetivo de evitar que apenas uma porção de dados de treino e teste pudesse ser muito similar. 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. Dessa forma, os dados foram separados em três conjuntos diferentes, seguindo a ideia demonstrada no Quadro 4, no qual são expostos os três grupos diferentes gerados para a realização dos experimentos.

Esses grupos foram misturados para gerar as três amostras. Assim, foram realizados três experimentos com cada classificador. A proposta foi utilizar os mesmos conjuntos de dados, tanto nos classificadores do Sklearn como na Rede Neural Convolutiva. Dessa forma, foi possível realizar a comparação dos resultados entre as técnicas, já que utilizaram as mesmas distribuições. O processo visualizado no Quadro 4 permite evitar a variância dos dados, além de possibilitar entender se os experimentos realizados trouxeram a mesma média de resultados com diferentes combinações.

Quadro 4 – Demonstração da validação cruzada para o processo de treinamento e testes

1ª amostra 1.320 redações	2ª amostra 1.320 redações	3ª amostra 1.320 redações
Treino	Teste 2	Teste 1
Teste 1	Treino	Teste 2
Teste 2	Teste 1	Treino

Fonte: Elaborado pelos autores.

Para a utilização da validação cruzada nos treinamentos, o *script* foi inserido no processo de treinamento, utilizando-se a biblioteca do *Sklearn cross_val_predict*. Dessa forma, os resultados demonstrados na seção de apresentação e discussão dos resultados já incluem tal distribuição dos dados. No Quadro 5 é apresentado o processo executado com o uso da biblioteca do *Sklearn cross_val_predict*.

Quadro 5 – Exemplo da aplicação da biblioteca do *Sklearn cross_val_predict*

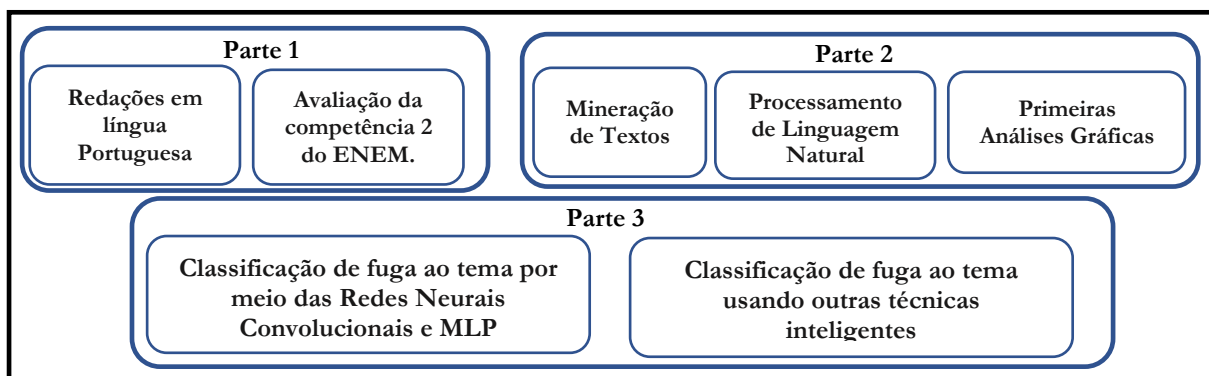
```
>>> from sklearn.model_selection import cross_val_predict
#preciso ter a base de dados importada
#depois separar os dados de entrada em (x) e coluna alvo (y)
#Configuração do Modelo de classificação, com todos os hiperparâmetros
>>>y_pred = cross_val_predict(modelo, x, y, cv=3)
```

Fonte: Elaborado pelos autores.

A aplicação da biblioteca exposta no Quadro 5 ocorre no momento do treinamento do modelo ao se realizar a validação cruzada, a fim de determinar qual conjunto de dados obteve melhor resultado.

Fluxo de atividades para a execução dos experimentos

Além das etapas definidas anteriormente para a pesquisa, também foi gerado fluxo de atividades para a elaboração dos experimentos, conforme apresentado na Figura 1. 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 deste trabalho.

Figura 1 – Fluxo de atividades

Fonte: Elaborado pelos autores.

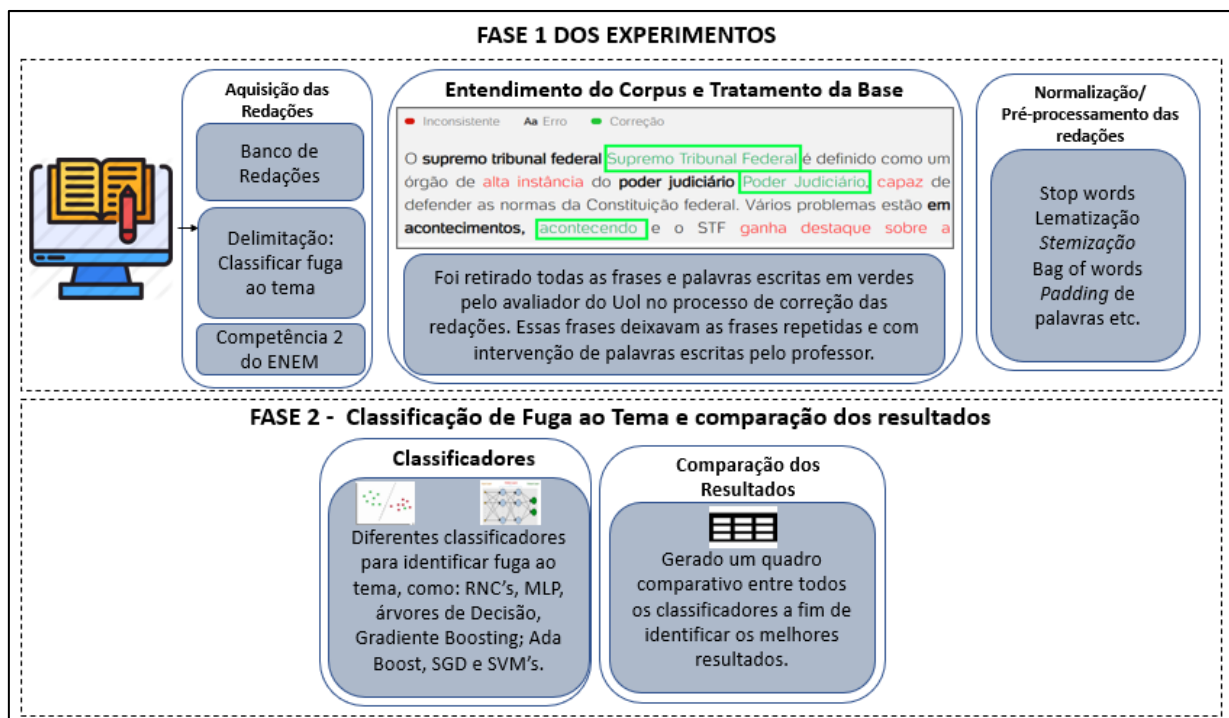
Conforme foi apresentado na Figura 1, destaca-se que foram realizadas as etapas pertinentes à Mineração de Textos, ao Processamento de Linguagem Natural e, por fim, ao treinamento na base, utilizando-se as Redes Neurais Convolucionais e a *Multilayer Perceptron*. Além das redes neurais citadas, o treinamento também foi realizado com outras técnicas de classificação como: *Decision Tree Classifier*, *Random Forest Classifier*, *Gradiente Boosting*, *Ada Boost*, *Stochastic Gradiente Descent* e *Support Vector Machine*. Os resultados esperados após a aplicação do modelo e das técnicas selecionadas buscaram classificar a fuga ao tema como positiva ou negativa, levando-se em consideração as redações sob análise para gerar a futura solução a ser validada.

Detalhamento dos experimentos

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

Dando sequência à análise da Figura 2, após a retirada dos trechos redundantes, gerados na avaliação dos docentes, o próximo passo foi aplicar as primeiras técnicas de Processamento de Linguagem Natural e Mineração de Textos para normalizar os documentos e assim prepará-los para as primeiras análises gráficas, clusterização (agrupamento) e elaboração dos modelos de treinamento.

Figura 2 – Sequência dos experimentos



Fonte: Elaborado pelos autores.

Na fase de normalização dos dados, foi criada uma função em *Python*, que executava toda a tratativa necessária. Nessa 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. Na Figura 3, pode ser visualizado o *script* da função em *Python* desenvolvido para realizar a normalização dos dados.

Figura 3 – Script de normalização dos dados desenvolvido em Python

```
def preprocessamento(text):
    result = []
    pos_tag = ['ADJ', 'NOUN', 'VERB', 'PROPN']
    texto = re.sub(u'^a-zA-Z0-9áéíóúÁÉÍÓÚâêîôÃÊËÔÃÕÇÇ:', '', text)
    doc = pln(texto.lower())
    for token in doc:

        if(token.text in stop_words or token.text in pontuacoes):
            continue

        if(token.pos_ in pos_tag):
            result.append(token.text)

    texto = ' '.join([str(elemento) for elemento in result if not elemento.isdigit()])

    return texto
```

Fonte: Elaborado pelos autores.

Na Figura 4 a seguir, é exposto um exemplo do resultado de um tema de redação normalizada. Observa-se que todo o tema foi transformado para letras minúsculas, 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 a ação de deflexionar uma palavra para determinar o seu lema, como pode ser observado na palavra “fumo”, que se tornou “fumar”, bem como na palavra “combate”, que se tornou “combater”.

Figura 4 – 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: Elaborado pelos autores.

Na última fase dos experimentos, foram aplicadas as técnicas que estabeleceram a classificação das redações sob análise. Em seguida, foi gerado um comparativo entre todas as técnicas, de modo a evidenciar as particularidades da aplicação de cada uma delas nos experimentos, conforme exposto no Quadro 6.

Quadro 6 – Detalhamento dos experimentos

Etapas/bibliotecas/experimentos realizados	
1)	Importação das bibliotecas necessárias para iniciar os experimentos: <i>Spacy</i> , <i>Pandas</i> , <i>Numpy</i> , <i>Matplotlib</i> e <i>Scikit-learn</i> .
2)	Extração da base de dados com 1.320 redações.
3)	Pré-processamento dos dados, realizando-se limpeza na base com o processo de normalização, remoção de <i>stop words</i> , remoção de caracteres especiais, <i>stemização</i> , <i>lematização</i> e <i>padding</i> de palavras.
4)	Antes do processo de treinamento do modelo, procurou-se colher informações a respeito dos textos, gerando-se alguns histogramas para entender a base de dados.
5)	Ainda na fase de entendimento da base de dados e procurando-se formas de auxiliar os professores para saberem quais as temáticas nas quais os alunos possuem maior dificuldade de escrita, foi então realizado o processo de similaridade entre as redações e suas temáticas. Este foi realizado 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. Foram utilizadas 628 redações para o processo de treino e, para o processo de teste, foram utilizadas 209 redações (as redações de teste correspondem a 33% das redações treinadas). No processo de treinamento, foi utilizada validação cruzada. Neste passo, verifica-se como a acurácia do modelo se comporta em diferentes categorias e classes.

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 treinamento, o primeiro passo foi transformar a variável <i>target</i> “fuga” em valores numéricos “0” e “1”. A fuga ao tema passou a ser representada por “1”, e a não fuga ao tema por “0”.
8)	Na sequência, todos os textos precisaram ser vetorizados, passando pelo processo de <i>padding</i> , criando-se um vetor de números que representam cada palavra. Nesta fase, cada redação e proposta do tema foram vetorizadas para ficarem na mesma dimensão. Para a Rede Neural Convolucional, foi utilizada a biblioteca “ <i>tfds</i> ” do “ <i>tensorflow_datasets</i> ”.
9)	Uma vez que toda a preparação dos dados foi realizada, o próximo passo foi a configuração do modelo da rede neural, determinando-se o número de camadas da rede neural, o número de neurônios para cada camada, o número de filtros, a quantidade de documentos avaliada 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 aplicado nas 209 redações separadas para a fase de teste.
11)	Após a aplicação do primeiro teste, foi gerada 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 realizado o processo de normalização das redações. Assim como na Rede Neural Convolucional, foi necessário realizar nova representação dos valores textuais (discretos), uma vez que os classificadores utilizados só entendem números. Assim, foi necessário converter os dados brutos, que estão em formato de texto, para o formato numérico. Isso deve acontecer antes de passar as redações para o classificador. O processo de vetorização para os classificadores do <i>Scikit-learn</i> foi realizado utilizando-se 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-se os hiperparâmetros definidos na própria documentação da biblioteca.
15)	Os resultados dos experimentos serão demonstrados na próxima seção.

Fonte: Elaborado pelos autores.

Os resultados dos experimentos mencionados no Quadro 6 estão demonstrados na próxima seção, a fim de facilitar a visualização de como tais experimentos são apresentados aos usuários finais (professores e alunos). Vale ressaltar que tais resultados são disponibilizados na plataforma CRIA desenvolvida a partir desses experimentos.

APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

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 *post-tagging*, da biblioteca do *Spacy*. Nesse 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 a 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 5, encontra-se exposto um exemplo da análise da correlação entre a redação escrita e a proposta da redação para casos com fuga ao tema.

Figura 5 – 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: Elaborado pelos autores.

Na figura 5 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 se analisar o conteúdo da Figura 5, 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. No entanto, foi possível verificar que o aluno não usou as palavras “drogas” e “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 6 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 6 – Exemplo de correlação entre redação e proposta da redação – Redações sem fuga ao tema

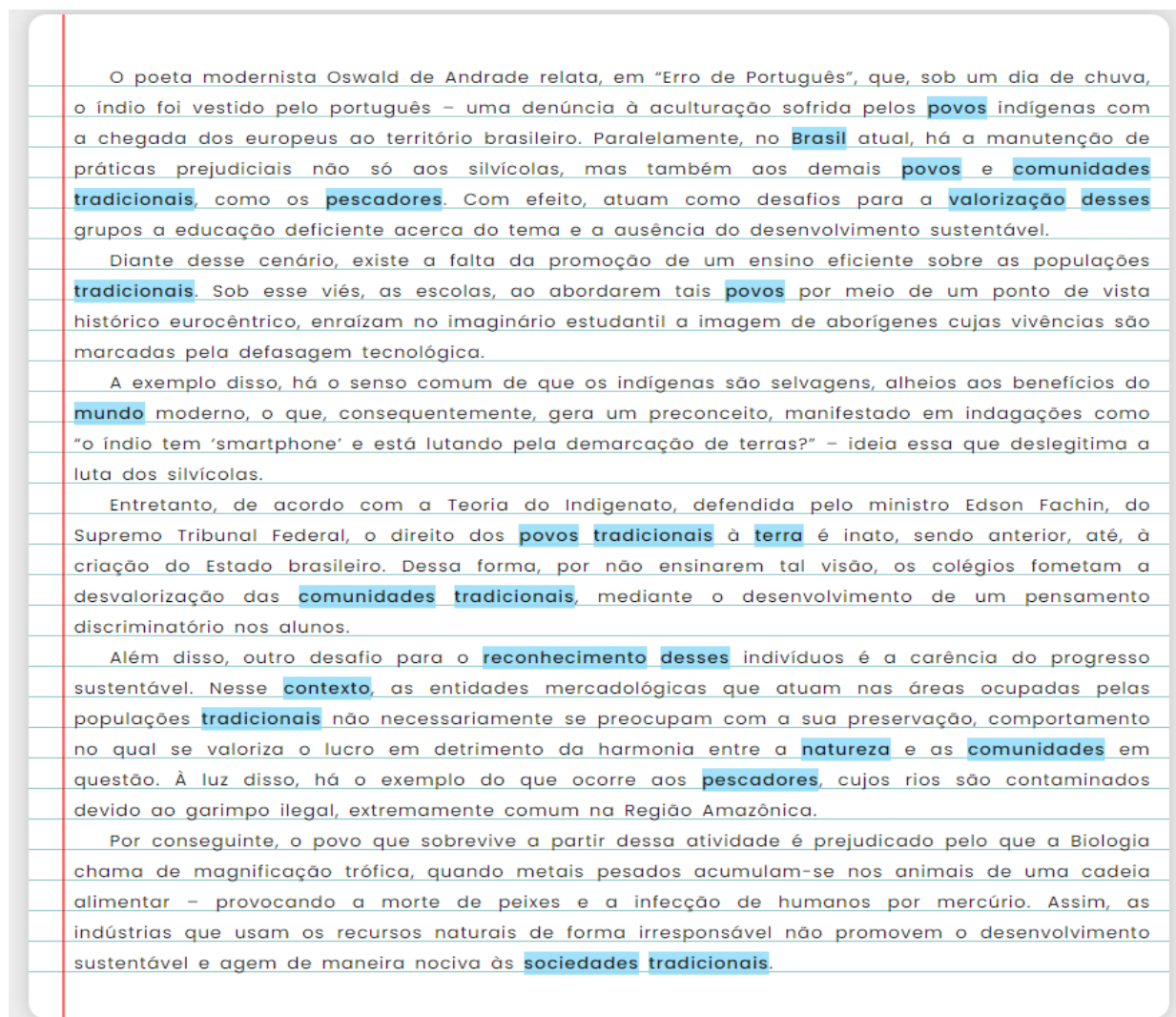
<p>Violência e drogas: o papel do usuário {‘violência’: 3, ‘drogas’: 9, ‘comércio’: 1, ‘ilegais’: 1, ‘lei’: 1, ‘oferta’: 1, ‘procura’: 1, ‘política’: 4, ‘proibicionista’: 2, ‘especial’: 1, ‘país’: 3, ‘descriminalização’: 4, ‘debate’: 2}</p>
<p>Violência e drogas: o papel do usuário {‘violência’: 3, ‘drogas’: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

Fonte: Elaborado pelos autores.

Ao se analisarem os resultados das duas redações expostas na Figura 6, que possuem a mesma temática abordada na Figura 5, foi possível identificar que, além de haver 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. Dessa 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 parte dos experimentos já está disponibilizada para uso na plataforma CRIA desenvolvida. Na Figura 7 é exposto como o aluno e o professor podem identificar as palavras aderentes ao tema proposto no conteúdo de uma redação elaborada.

Figura 7 – Identificação de palavras aderentes em uma redação



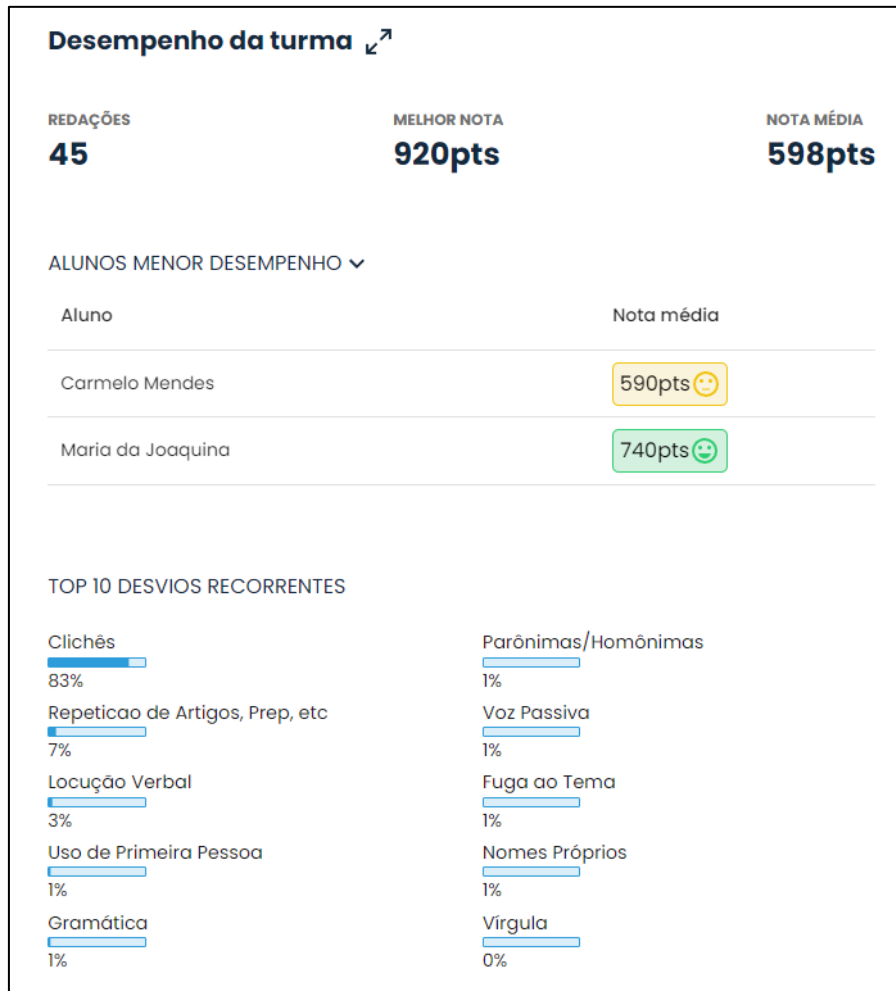
Fonte: Plataforma CRIA (2023).

A plataforma disponibiliza a quantidade e a distribuição dessas palavras por parágrafo. A redação demonstrada na Figura 7 tem como tema: “Desafios para a valorização de comunidades e povos tradicionais no Brasil”. É possível verificar que a redação exposta no exemplo trouxe 25 palavras identificadas pela similaridade ao tema proposto.

Essa primeira fase dos experimentos permitiu gerar informações relevantes ao professor sobre o desempenho de seus alunos e turmas, uma vez que a plataforma disponibiliza os erros recorrentes dos alunos bem como aqueles desvios que são mais cometidos na redação. Dessa forma, o docente identifica aqueles alunos que estão com maiores dificuldades na produção textual, bem como os temas com maior dificuldade de assimilação por parte dos estudantes. Esta solução (CRIA) já está implementada e vem proporcionando benefícios aos alunos, que podem reescrever sua redação antes mesmo de enviá-la ao professor. Este expediente é especialmente válido quando o aluno percebe que pode ter tangenciado o tema proposto para a redação. Além disso, também se observa ganho para o professor, uma vez que este recebe uma redação já previamente analisada pelo aluno e, ainda assim, pode interferir caso não concorde com os apontamentos indicados pela plataforma CRIA.

A solução desenvolvida também proporciona ao professor acesso a um painel de avaliação após a correção realizada pela técnica inteligente. Nela, o professor visualiza o desempenho de cada turma, bem como a exposição dos erros recorrentes nessa turma, conforme apresentado na Figuras 8. Também é disponibilizada a visualização das redações a ele atribuídas para correção. As telas expostas na Figura 8 permitem ao professor avaliar os erros mais recorrentes da turma para trabalhar em sala de aula, bem como validar as correções vindas da correção realizada pela técnica inteligente.

Figura 8 – Painel do Professor após correção da técnica inteligente para acompanhamento do desempenho da turma



Fonte: Plataforma CRIA (2023).

Nota: os nomes dos alunos indicados são fictícios.

Figura 8 – Painel do Professor após correção da técnica inteligente para avaliação da turma (continuação)

Redações				
A CORRIGIR ↙ ↗				
POSTADO EM:	NOME DO ALUNO	TEMA	NOTA	VER
16/06/2023	Carmelo Mendes	O PREPARO DA SOCIE	480pts 😞	
05/04/2023	Carmelo Mendes	Desafios para a valc	440pts 😞	

ATRIBUÍDAS ↙ ↗				
TEMA	INÍCIO	ENTREGA	TURMA	VER
Direito da Criança e dc	25/10/2023	01/11/2023	3C-Informática	
Repercussão do pensc	25/10/2023	01/11/2023	3C-Informática	
A importância dos cui	23/10/2023	06/11/2023	3C-Informática	
A questão da fome err	16/10/2023	23/10/2023	3C-Informática	
Causas e consequênc	16/10/2023	30/10/2023	3C-Informática	

⏪ < 1 2 3 4 > ⏩

Fonte: Plataforma CRIA (2023).

Nota: os nomes dos alunos indicados são fictícios.

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 textos. Assim sendo, aplicou-se o processo de *padding* e vetorização dos textos antes de se iniciar o treinamento do modelo desenvolvido.

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 foi a aplicação dos diferentes modelos de classificação selecionados para esse experimento. Para aplicação dos classificadores, foram gerados os modelos, aplicando-se as duas bases de teste separadas, assim como ocorreu com a Rede Neural Convolutiva. Na sequência, foram aplicadas as métricas de avaliação utilizadas (Acurácia e Matriz de Confusão) geradas para cada modelo em questão.

Classificação de fuga ao tema usando as redes neurais convolucionais

Os resultados da Rede Neural Convolutiva foram dispostos de forma separada dos classificadores posteriores, pois ela 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

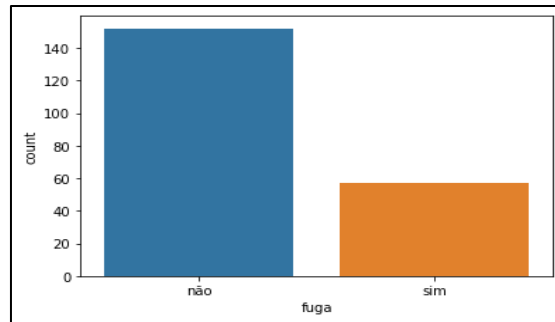
e lematização e *padding* de palavras, conforme exposto no tópico de método e materiais de pesquisa, a base de dados foi finalmente separada entre conjunto de treinamento e conjunto de teste.

Para o treinamento da Rede Neural Convolucional, 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. Esse critério também levou em consideração que essa normalmente é a margem que as universidades utilizam como critério de eliminação dos candidatos (Universidades [...], 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 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: Elaborado pelos autores.

Após a aplicação da base de teste no modelo, uma das métricas utilizadas para avaliar os resultados e comparar os diferentes algoritmos testados foi a Matriz de Confusão. Na Tabela 2, é apresentada a Matriz de Confusão do experimento realizado, com os resultados da Rede Neural Convolucional para o primeiro experimento com a base de teste, demonstrando os erros e acertos do modelo.

Tabela 2 – Matriz de Confusão – Primeira base de teste

0 (Categoria Real)	143	9
1(Categoria Real)	29	28
	0 (Previsão)	1 (Previsão)

Fonte: Elaborado pelos autores.

Ao se realizar a análise dos resultados expostos na Tabela 2, foi identificado que, das 152 redações que não tiveram fuga ao tema, o modelo classificou como fuga nove 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, a técnica classificou como fuga 28 redações, o que equivale a 49,0% de acerto, valor este que representa os verdadeiros positivos, ou seja, a classe fuga ao tema foi determinada como classe positiva. Esta taxa de acertos pode ter-se dado pela limitação da base de dados, que continha poucos exemplos de fuga ao tema. Essa limitação tem sido contornada ao longo do tempo com a extensão da aplicação dos experimentos desta pesquisa na plataforma CRIA, um serviço já implementado junto a escolas e professores. Vale ressaltar que, neste primeiro teste, a acurácia 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 menores ou iguais a 499, ou seja, normalmente aquelas que são desconsideradas por grande parte das universidades para aprovação dos candidatos. Os resultados podem ser visualizados na Tabela 3, que expõe a Matriz de Confusão gerada para as outras 540 redações que não foram utilizadas na fase de treinamento. Na distribuição dessa base, havia 483 redações sem fuga ao tema e 57 redações com fuga.

Tabela 3 – Matriz de Confusão – Segunda base de teste

0 (Categoria Real)	455	28
1(Categoria Real)	29	28
	0 (Previsão)	1 (Previsão)

Fonte: Elaborado pelos 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. Isto porque a taxa de erros em relação aos falsos positivos desta segunda base de teste foi de 5,7%, e de falsos negativos foi de 94,3%. No caso dos verdadeiros negativos, obteve-se o mesmo resultado do experimento anterior.

A Tabela 4 consolida os resultados obtidos no processo de classificação da Rede Neural Convolutacional. 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 ao tema 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 4 – Resultados consolidados da matriz de confusão – Rede Neural Convolutacional

Primeira base de teste com redações de notas superiores a 500 e 57 redações com fuga ao tema					
Classificador RNC	Acurácia	VP	FN	FP	VN
		(Verdadeiro Positivo)	(Falso Negativo)	(Falso Positivo)	(Verdadeiro Negativo)
	81,8%	49,1%	50,9%	5,9%	94,1%
Segunda base de teste com redações de notas inferiores a 500 e 57 redações com fuga ao tema					
Classificador RNC	Acurácia	VP	FN	FP	VN
		(Verdadeiro Positivo)	(Falso Negativo)	(Falso Positivo)	(Verdadeiro Negativo)
	89,4%	49,1%	50,9%	5,7%	94,3%

Fonte: Elaborado pelos autores.

Os resultados da Precisão, *Recall* e *F1-score*, com base nos resultados da aplicação da Rede Neural Convolutacional, são apresentados na Tabela 5, considerando-se “1” para fuga ao tema e “0” para não fuga ao tema.

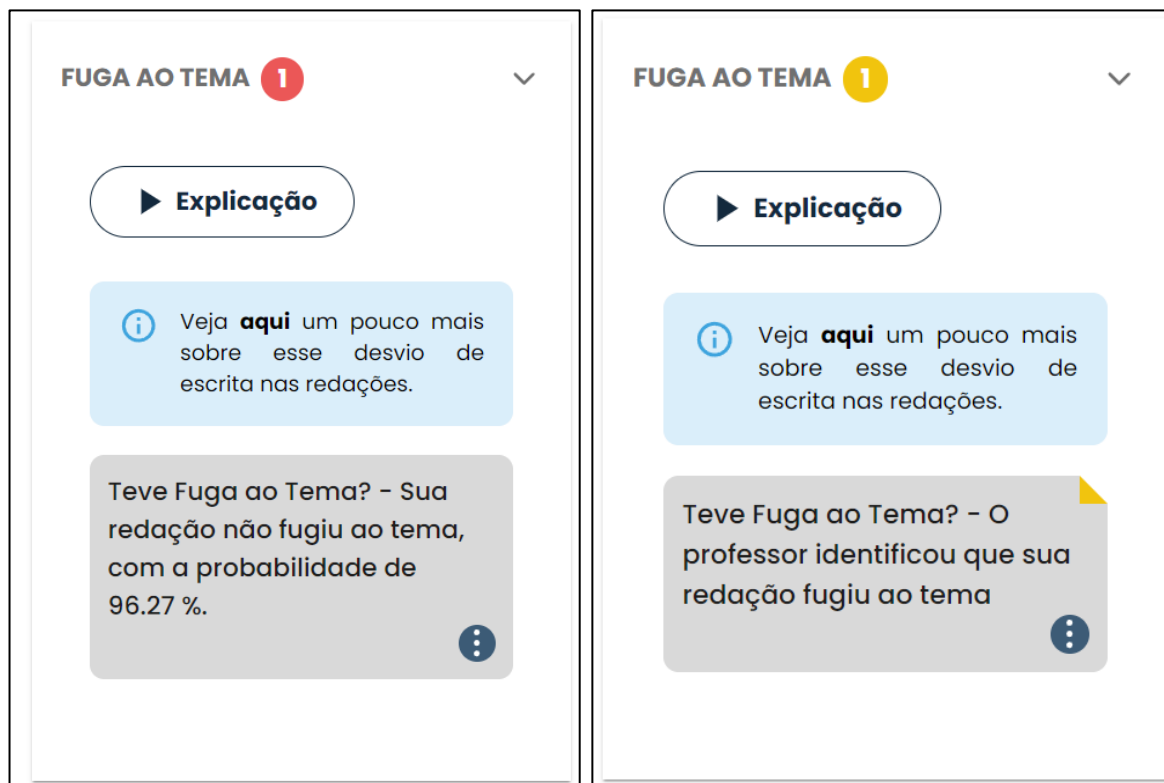
Tabela 5 – Resultados consolidados de precisão, *Recall* e *F1-Score* – Rede Neural Convolutacional

Primeira base de teste com redações de notas superiores a 500 e 57 redações com fuga ao tema				
		Precisão	<i>Recall</i>	<i>F1-Score</i>
Classificador RNC	0	94%	94%	94%
	1	50%	49%	50%
Segunda base de teste com redações de notas inferiores a 500 e 57 redações com fuga ao tema				
		Precisão	<i>Recall</i>	<i>F1-Score</i>
Classificador RNC	0	94%	94%	94%
	1	50%	49%	50%

Fonte: Elaborado pelos autores.

Os resultados apresentados quanto à precisão do modelo indicam todas as classificações de classe positiva que o modelo fez, e quantas estão corretas. Já o *Recall* dentre todas as situações de classe positiva como valor esperado indica quantas estão corretas. 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 10 estão expostas um exemplo prático da apresentação de probabilidade de fuga ao tema, conforme disponibilizado na plataforma CRIA. O aluno e o professor recebem a indicação da probabilidade de fuga ao tema. Caso o professor não concorde, ele pode alterar a indicação da correção feita pela Inteligência Artificial por meio da opção “não concordar”. Dessa forma, o sistema continuará aprendendo com o perfil de avaliação do professor, ao passo que o modelo vai sendo retreinado. Ao lado direito da Figura 10, o aluno consegue identificar se o professor interferiu na avaliação da Inteligência Artificial por meio do texto indicativo.

Figura 10 – Exemplo prático do retorno da probabilidade da fuga

Fonte: Plataforma CRIA (2023).

O percentual da probabilidade de fuga ao tema apresentado pela plataforma ao aluno e ao professor, que no exemplo exposto na Figura 10 foi de 99,61%, é um importante indicador da aderência da redação elaborada ao tema proposto. Assim, quanto menor o percentual apresentado, menor a indicação de ocorrência de fuga ao tema. A partir do percentual apresentado pela plataforma, o professor pode ainda associar esse percentual ao resultado do primeiro experimento desta pesquisa, voltado às palavras aderentes à proposta de tema.

Nesta pesquisa, após o processo de classificação utilizando-se as Redes Neurais Convolucionais, o próximo passo foi testar outros classificadores.

Resultados da classificação de fuga ao tema aplicando outros classificadores selecionados

São apresentados, a seguir, os resultados referentes aos classificadores de Aprendizado de Máquina da biblioteca do *Scikit-Learn*, após nova normalização das redações. Os classificadores do *Scikit-Learn* selecionados foram explicitados no tópico classificadores, sendo estes: *MLPClassifier*, *DecisionTreeClassifier*, *RandomForestClassifier*, *SGDClassifier*, *SVM (SVC)*, *GradientBoostingClassifier*, *AdaBoostClassifier*.

Para todos os classificadores, foi utilizada a mesma base da Rede Neural Convolucional. As redações foram divididas em treino e teste, utilizando-se os mesmos critérios de divisão estipulados para a Rede Neural Convolucional, viabilizando assim a comparação com os resultados obtidos. Foram aplicadas também as mesmas métricas utilizadas nos experimentos realizados com a Rede Neural Convolucional: Matriz de Confusão; a Acurácia; o *Recall* e o *F1-score*.

Na Tabela 6 são expostos os resultados de cada classificador, utilizando-se o mesmo critério aplicado para a Rede Neural Convolucional, ou seja, as notas atribuídas pelo avaliador humano iguais ou superiores a 500 pontos.

Tabela 6 – Classificadores *Scikit Learn* – Consolidado da Matriz de Confusão

Classificador	Acurácia	Primeira base de teste com redações de notas superiores a 500 e 57 redações com fuga ao tema			
		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: Elaborado pelos autores.

Para identificar os melhores resultados da primeira base de teste, levou-se em consideração a acurácia aliada aos Verdadeiros Positivos e Verdadeiros Negativos, além da classificação que apresentou o menor erro dos Falsos Positivos, 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 7 são apresentados os resultados da segunda base de teste, ou seja, aquelas redações que tiveram notas atribuídas pelo avaliador humano abaixo de 500 pontos.

Tabela 7 – Resultados dos Classificadores *Scikit-Learn* – Segunda base de teste

CLASSIFICADOR	ACURÁCIA	VP	FN	FP	VN
		(Verdadeiro Positivo)	(Falso Negativo)	(Falso Positivo)	(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: Elaborado pelos 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 da Tabela 7, apesar de tratar das redações de notas inferiores, obtiveram resultados similares à base de dados anterior. A maior diferença foi verificada nos Falsos Positivos, que, neste caso, obteve como classificadores com menor erro nos Falsos Positivos o *MLPClassifier* e *SGDClassifier*, com respectivamente 3,3% e 8,7%.

Na Tabela 8 são apresentados os resultados das demais métricas de avaliação para auxiliar no entendimento dos resultados do modelo desenvolvido.

Tabela 8 – Métricas Precisão, *Recall* e *F1-Score* – Consolidado de classificadores *Scikit-Learn*

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

Fonte: Elaborado pelos autores.

Ao se avaliarem os resultados da Tabela 8, 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 negativa (não fuga) ficaram acima de 80%, e os acertos da classe positiva (fuga ao tema) mantiveram a média acima de 50%.

O classificador *MLPClassifier* manteve a precisão nas duas classes (positiva e negativa) acima de 70%. Já no item *Recall*, para a classe positiva, obteve um acerto de apenas 33%. O *F1-Score* fez uma média harmônica entre os resultados dos dois anteriores e, para a classe positiva, também se manteve abaixo 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 se a redação está aderente à proposta de tema informada, o que pode trazer importante conhecimento ao avaliador ou docente em relação à evolução dos alunos na produção textual, com marcações parágrafo a parágrafo.

Na avaliação dos resultados obtidos nos classificadores da Rede Neural Convolutiva, foi identificado maior ganho em relação aos classificadores *Scikit-Learn*, tanto em relação à acurácia quanto em relação aos resultados de falsos positivos, métricas de Precisão, *Recall* e *F1-Score*. Apesar de os melhores resultados terem sido obtidos com a Rede Neural Convolutiva, devem-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 não conseguiram atingir o objetivo proposto nesta pesquisa, quais sejam: *Random Forest Classifier* e SVM (SVC).

Na sequência, o modelo que teve a melhor acurácia foi a Rede Neural Convolutiva, 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% de 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 *MLPClassifier*, 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 redações classificadas como fuga ao tema, em especial a Rede Neural Convolutiva. Isto porque, segundo Rodrigues (2018), a Rede Neural Convolutiva obtém melhores resultados com uma grande base de exemplos. Não obstante, a partir dos resultados apresentados nesta pesquisa, constantes experimentos continuaram a ser realizados buscando a melhoria de acurácia e aumento da base de dados. Os estudos realizados nesta pesquisa constituíram a base para colocação em produção da Plataforma CRIA.

Vale ressaltar que, desde outubro de 2022, cinco escolas testaram e validaram a plataforma CRIA antes de seu lançamento oficial. Os professores que a utilizaram relataram economia de tempo, uma vez que o aluno já recebe uma avaliação instantânea. Assim, o discente faz a avaliação crítica de sua própria redação a partir das indicações de desvios de escrita disponibilizadas pela plataforma, incluindo a fuga ao tema, além da indicação de artigos explicativos para leitura. O objetivo é provocar no discente a compreensão “do que” e “por que” ele errou ou por que não deve cometer determinada falha, ajudando-o a compreender as regras do Enem para a elaboração de redação. Já o professor recebe para avaliação uma redação com menos erros, disponibilizando assim mais esforço e tempo para a realização de uma avaliação mais contributiva, com a indicação de pontos de correção que a plataforma não conseguiu detectar de forma automatizada.

O incremento do uso da plataforma CRIA tem possibilitado o aumento da base de dados para novos treinamentos. Isto tem sido alcançado a partir do recebimento de redações de diferentes escolas, faixas etárias e diversidade social dos discentes. A diversidade das redações tem possibilitado a diminuição de vieses na avaliação, bem como a expansão da base de dados com textos já corrigidos por professores.

No contexto de sala de aula, tal aplicação tem proporcionado ganhos em relação ao tempo despendido para correção e menor desgaste do professor na avaliação de textos. Para o aluno, a disponibilização de devolutiva mais rápida tem-se demonstrado positiva, além da garantia de posterior

avaliação do professor da disciplina. Assim, a solução validada nesses experimentos contribui para a redução de tempo e recursos empregados no processo de avaliação de textos produzidos pelos alunos, sem que se descarte o papel do professor como protagonista desse processo.

Nas pesquisas realizadas, não foram encontrados outros autores que se utilizaram dos mesmos métodos e técnicas aplicados 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 CRIA 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.

Dessa 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 professores 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 vem sendo aplicada num sistema para avaliar a evolução dos alunos do decorrer de seus estudos acadêmicos, proporcionando, ao professor, assim, conhecer as dificuldades individuais dos alunos numa turma.

CONCLUSÕES

A aplicação das técnicas mencionadas nos experimentos executados nesta pesquisa buscou indicar quais técnicas de Inteligência Artificial comparadas apresentam melhores resultados para a identificação de fuga ao tema em redações. Assim, entende-se que este estudo alcançou o objetivo de pesquisa estabelecido, 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 Redes Neurais Convolucionais como nos classificadores do *Scikit-Learn*. O modelo que obteve a melhor acurácia foi a Rede Neural Convolutiva, com resultados de até 89,4% de acurácia e taxa de falso positivo (FP) de 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 positiva, 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 *MLPClassifier*, com no máximo 4,6% de erro, taxa de verdadeiros positivos (VP) de 33% e acurácia entre 78% e 90%.

Dessa forma, o objetivo proposto de comparar diferentes técnicas de Inteligência Artificial 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 professores e 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 para o classificador Rede Neural Convolutiva. Tal resultado possibilitou a criação de uma aplicação para fornecer *feedback* automático, como suporte ao professor ou avaliador de textos, o que contribui para diminuir o tempo demandado para correção, além de prestar melhor auxílio às instituições, professores e alunos. Porém, cabe mencionar que outros classificadores podem ser aplicados

paralelamente, de modo a viabilizar um processo de dupla verificação, que contribuirá para sinalizar de forma mais eficiente a indicação de fuga ao tema ao professor/avaliador da redação.

A solução ora desenvolvida visa diminuir a desigualdade nos processos seletivos que utilizam avaliação de redações elaboradas pelos candidatos, oferecendo assim maior oportunidade de aprendizagem independentemente da instituição em que o aluno estude. Isto porque esta solução automatizada poderia ser aplicada também em escolas para fins de treinamento no processo de elaboração e correção de redações. 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 de aspectos básicos nas correções para uma função mais voltada à comprovação da efetividade do aprendizado do aluno na correção do texto. 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 de avaliadores e professores. Em outros contextos, tais contribuições poderiam, inclusive, ser benéficas também aos professores que se dedicam exclusivamente à correção de textos. Isto porque tais profissionais poderiam se valer das indicações providas pela solução e, então, direcionar seus esforços para outras demandas mais complexas da atividade de correção de textos.

As principais contribuições deste estudo buscam permitir ao avaliador, professor ou empresas que aplicam processos seletivos avaliarem 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 ao avaliador, assim, 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 a existência de uma plataforma que possibilite o treino mais frequente e resposta mais rápida poderá contribuir para facilitar não apenas o trabalho do professor como também proporcionar, futuramente, uma escrita mais clara e desenvolvida, possibilitando assim maior maturidade do aluno neste processo.

Uma questão limitadora dos resultados apresentados neste estudo refere-se à base de dados utilizada nos experimentos realizados. Pretende-se adicionar a esta amostra 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 à seleção das técnicas de Inteligência Artificial empregadas nos experimentos executados neste estudo, a cargo dos autores.

Para as pesquisas futuras, recomenda-se a ampliação da base de dados de redações, visando proporcionar mais efetividade aos experimentos com técnicas de Inteligência Artificial. Este fato já está bem próximo da realidade, uma vez que, por meio da plataforma CRIA, se busca aumentar a base de redações e exemplos de fuga ao tema. Além disso, também se indica 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.
- BANERJEE, Dibyendu. Natural language processing (NLP) simplified: A step-by-step guide. *Data Science Foundation*, 2020, publicado em 14/4/2020. 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. *Viceri-Seidor*, 2020, publicado em 27/5/2020. Disponível em: <https://www.viceri.com.br/insights/as-classificacoes-dos-algoritmos-de-machine-learning>. Acesso em: 9 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). Goiânia: Universidade Federal de Goiás, 2020.
- BRASIL. Ministério da Educação e Cultura (MEC). MEC realiza conferência para discutir estratégias de alfabetização no Brasil. *Portal MEC*, 2019, publicado em 22/10/2019. Disponível em: <http://portal.mec.gov.br/component/tags/tag/5?start=60>. Acesso em: 15 abr. 2021.
- BRITZ, Denny. Understanding convolutional neural networks for NLP. *Denny's Blog*, 2015, publicado em 7/11/2015. Disponível em: <http://www.wildml.com/2015/11/understanding-convolutional-neural-networksfor-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.
- 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. *Medium*, 2020, publicado em 7/7/2020. Disponível em: <https://medium.com/data-hackers/redes-neurais-convolucionais-para-processamento-de-linguagem-natural-935488d6901b>. Acesso em: 31 mar. 2021.
- CARVALHO, André C. P. F. de; FAIRHURST, Michael C.; BISSET, David L. An integrated boolean neural network for pattern classification. *Pattern Recognition Letters*, v. 15, p. 807-813, ago./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). São Paulo: Universidade Estadual Paulista – UNESP, 2018.
- COSTA DA SILVA, Josenildo. Aprendendo em uma floresta aleatória. *Medium*, 2018, publicado em 12/3/2018. Disponível em: <https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleat%C3%B3ria-3545f6babdf8>. Acesso em: 31 mar. 2021.
- CRIA – Plataforma CRIA. *Plataforma de Correção automática de Redações*, 2023. Disponível em: <https://web.cria.net.br>. Acesso em: 25 ago. 2023.

DIANA, Daniela B. G. Os 16 maiores erros de redação cometidos pelos estudantes. *Toda Matéria*, 2021, publicado em 8/1/2021. Disponível em: <https://www.todamateria.com.br/erros-de-redacao/>. Acesso em: 29 mar. 2021.

EGGERS William D.; SCHATSKY, David; VIECHNICKI, Peter. AI-augmented government using cognitive technologies to redesign public sector work. *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_Minerao_de_texto_-_Conceitos_e_aplicacoes_praticas. Acesso em: 15 abr. 2021.

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

GRANATYR, Jones. Processamento de Linguagem Natural com Deep Learning. *Expert Academy*, 2020, curso realizado em novembro de 2020. Disponível em: <https://iaexpert.academy/courses/processamento-linguagem-natural-deep-learning-transformer/>. Acesso em: 10 dez. 2020.

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.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *A redação no Enem 2022: cartilha do participante*. Brasília, 2022. Disponível em: https://download.inep.gov.br/download/enem/cartilha_do_participante_enem_2022.pdf. Acesso em: 19 out. 2023.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *Entenda como é calculada a nota do Enem*. 2020. 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.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *Microdados Enem 2022*. 2023. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enem>. Acesso em: 20 ago. 2023.

LEITE, Tiago M. Redes neurais, perceptron multicamadas e o algoritmo backpropagation. *Medium*, 2018, publicado em 10/5/2018. Disponível em: <https://medium.com/ensina-ai/redes-neurais-perceptron-multicamadas-e-o-algoritmo-backpropagation-eaf89778f5b8>. Acesso em: 10 maio 2021.

LESME, Adriano. Enem 2021: corretores podem corrigir até 200 redações por dia. *Brasil Escola – UOL*, 2021, publicado em: 1º/12/2021. Disponível em: <https://vestibular.brasilecola.uol.com.br/enem/enem-2021-corretores-podem-corrigir-ate-200-redacoes-por->

dia/351641.html#:~:text=Cada%20profissional%20ter%C3%A1%20que%20avaliar,de%2015%20a%2020%20dias.&text=Com%20o%20t%C3%A9rmino%20do%20Exame,a%20supervis%C3%A3o%20de%202016%20profissionais. Acesso em: 19 ago. 2023.

LUDERMIR, Teresa B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, v. 35, n. 101, p. 85-94, 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. *Medium*, 2018, publicado em 24/12/2018. Disponível em: <https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9>. Acesso em: 15 abr. 2021.

MÜLLER, Sarah; BERGANDE, Bianca; BRUNE, Philipp. 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. *Anais [...]* Association for Computing Machinery, New York, USA, p. 45-49.

MUYLAERT, Renata. Pandemia do novo coronavírus, Parte 6: inteligência artificial (NLP). *Sobrevivendo na Ciência*, 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). Itajaí: Universidade do Vale do Itajaí, 2018.

PESSANHA, Cíntia. Random Forest: como funciona um dos algoritmos mais populares de ML. *Medium*, 2019, publicado em 20/11/2019. Disponível em: <https://medium.com/cinthiabpessanha/random-forest-como-funciona-um-dos-algoritmos-mais-populares-de-ml-cc1b8a58b3b4>. Acesso em: 27 jul. 2021.

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, 15^o, São Paulo. *Anais [...]* São Paulo: SBGC, 2020, p. 599-615.

PINTO, Álvaro V. *O conceito de tecnologia*. Rio de Janeiro: Contraponto, 2005.

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

PRATES, Wladimir R. Introdução ao processamento de linguagem natural (PLN). *Ciência e Negócios*, 2019, publicado em 1^o/8/2019. Disponível em: <https://cienciaenegocios.com/processamento-de-linguagem-natural-nlp/>. Acesso em: 27 jul. 2021.

PREMLATHA, Karan R. What is AI? In a simple way. *AI Time Journal*, 2019, publicado em 5/2/2019. Disponível em: <https://www.aitimejournal.com/@premlatha.kr/what-is-ai-in-a-simple-way>. Acesso em: 15 abr. 2021.

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). São Carlos: Universidade Federal de 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.

RIEDO, Cassio R. F. *Avaliação qualitativa imediata de produções escritas em EaD*. 266 f. Tese (Doutorado em Educação). Campinas: Universidade Estadual de Campinas, 2020.

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, abr./2010.

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

RODRIGUES, Vitor. Métricas de avaliação – Quais as diferenças? *Medium*, 2019, publicado em 12/4/2019. Disponível em: <https://vitorborbarodrigues.medium.com/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>. Acesso em: 11 maio 2021.

RUSSO, Inês F. D. *O impacto da inteligência artificial na sustentabilidade ambiental: uma agricultura sustentável*. 84 f. Dissertação (Mestrado em Gestão de Sistemas de Informação). Lisboa: Universidade de 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). Maceió: Universidade Federal de Alagoas, 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. *Medium*, 2020, publicado em 22/6/2020. Disponível em: <https://medium.com/equals-lab/uma-breve-introdu%C3%A7%C3%A3o-ao-algoritmo-de-machine-learning-gradient-boosting-utilizando-a-biblioteca-311285783099>. Acesso em: 11 maio 2021.

SILVEIRA, Rita C. B. da; BARROS, Manoel J. F. de. Impacto da inteligência artificial na empregabilidade docente. In: COLÓQUIO INTERNACIONAL DE GESTÃO UNIVERSITÁRIA – GIGU, XX, Florianópolis. *Anais [...]* Florianópolis: UFSC, 2021, p. 1-17.

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. 204-213, 2019.

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

STARLLES, Wender. Confissões de uma corretora de redações do Enem. *Guia do Estudante*, 2022, atualizado em 8/8/2022. Disponível em: <https://guiadoestudante.abril.com.br/enem/confissoes-de-uma-corretora-de-redacoes-do-enem/>. Acesso em: 19 ago. 2023.

TOKARNIA, Mariana. Enem é um dos principais instrumentos de acesso ao ensino superior. *Agência Brasil*, 2019, 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 mar. 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.

UNIVERSIDADE DE SÃO PAULO (USP). USP desenvolve ferramenta de correção automática de redações. *Portal USP São Carlos*, 2021, publicado em 10/3/2021. Disponível em: <http://www.saocarlos.usp.br/usp-desenvolve-ferramenta-de-correcao-automatica-de-redacoes/>. Acesso em: 29 mar. 2021.

UNIVERSIDADES privadas de SP adotam vestibular online e nota do Enem. *Jornal Cruzeiro do Sul*, 2020, publicado em 9/6/2020. Disponível em: <https://www.jornalcruzeiro.com.br/brasil/universidades-privadas-de-sp-adotam-vestibular-online-e-nota-do-enem/>. Acesso em: 29 mar. 2021.

WALTRICK, Camila. Machine learning – O que é, tipos de aprendizagem de máquina, algoritmos e aplicações. *Medium*, 2020, publicado em 7/5/2020. Disponível em: <https://medium.com/camilawaltrick/introducao-machine-learning-o-que-e-tipos-de-aprendizado-de-maquina-445dcfb708f0>. Acesso em: 29 mar. 2021.

Submetido: 18/05/2022

Preprint: 25/04/2022

Aprovado: 13/09/2023

CONTRIBUIÇÃO DOS AUTORES

Autora 1 - Administração do projeto, análise formal, conceituação, curadoria de dados, escrita – primeira versão, investigação, metodologia, recursos, software, supervisão, validação e visualização.

Autor 2 - Administração do projeto, análise formal, conceituação, escrita – primeira versão, escrita – revisão e edição, investigação, metodologia, recursos, supervisão, validação e visualização.

Autor 3 - Conceituação, escrita – revisão e edição, metodologia, recursos, software, supervisão e validação.

DECLARAÇÃO DE CONFLITO DE INTERESSE

Os autores declaram que não há conflito de interesse com o presente artigo.