

UNIVERSIDADE FEDERAL DO PARANÁ

JULIA INGLÊS MARCOS

CLASSIFICAÇÃO DE NORMAS DO BANCO CENTRAL DO BRASIL APLICADAS
ÀS INSTITUIÇÕES DE PAGAMENTOS COM APRENDIZADO DE MÁQUINA E
PROCESSAMENTO DE LINGUAGEM NATURAL

CURITIBA
2023

JULIA INGLÊS MARCOS

CLASSIFICAÇÃO DE NORMAS DO BANCO CENTRAL DO BRASIL APLICADAS
ÀS INSTITUIÇÕES DE PAGAMENTOS COM APRENDIZADO DE MÁQUINA E
PROCESSAMENTO DE LINGUAGEM NATURAL

Trabalho de Conclusão de Curso apresentado
como requisito parcial à obtenção do grau de
Bacharel em Gestão da Informação no curso de
graduação em Gestão da Informação, Setor de
Ciências Sociais Aplicadas da Universidade
Federal do Paraná.

Orientador: Prof. Dr. Luciano Heitor Gallegos Marin

CURITIBA

2023

RESUMO

O objetivo deste trabalho é utilizar métodos de aprendizado de máquina e processamento de linguagem natural para classificar resoluções e instruções normativas do Banco Central do Brasil em relação às Instituições de Pagamentos. Pretende avaliar a eficácia dos algoritmos utilizados e o risco da sua aplicação. Discorre sobre o contexto das fintechs e instituições de pagamentos, bem como as regulamentações necessárias em conjunto ao Compliance, abordando o cenário regulatório dessas instituições no Brasil. Explica o aprendizado de máquina e o processamento de linguagem natural, destacando suas aplicações na classificação de textos ou documentos. São apresentadas técnicas de limpeza e pré-processamento das resoluções e instruções normativa do Banco Central do Brasil, assim como a descrição detalhada do tratamento dos documentos e a aplicação de algoritmos para a classificação dos mesmos. Verifica o desempenho de diversas técnicas para verificar a mais efetiva para a classificação de resoluções e instruções. Define como os resultados podem auxiliar na aplicação de Aprendizado de Máquina e Processamento de Linguagem Natural nestas resoluções e normas.

Palavras-chave: Instituições de Pagamentos. Aprendizado de Máquina. Processamento de Linguagem Natural. Resoluções e Instruções Normativas. Banco Central do Brasil.

ABSTRACT

The objective of this work is to utilize machine learning and natural language processing methods to classify resolutions and normative instructions from the Central Bank of Brazil in relation to Payment Institutions. It aims to evaluate the effectiveness of the algorithms used and the risk associated with their application. It discusses the context of fintechs and payment institutions, as well as the necessary regulations in conjunction with compliance, addressing the regulatory scenario of these institutions in Brazil. It explains machine learning and natural language processing, highlighting their applications in the classification of texts or documents. Techniques for cleaning and preprocessing the resolutions and normative instructions from the Central Bank of Brazil are presented, along with a detailed description of document treatment and the application of algorithms for their classification. It assesses the performance of various techniques to determine the most effective one for the classification of resolutions and instructions. It defines how the results can assist in the application of machine learning and natural language processing in these resolutions and norms.

Keywords: Payment Institutions. Machine Learning. Natural Language Processing. Resolutions and Normative Instructions. Central Bank of Brazil.

LISTA DE FIGURAS

FIGURA 1 - ÁREAS RELACIONADAS COM IA.....	22
FIGURA 2 - ETAPAS DA ANÁLISE DE LINGUAGEM NATURAL	23
FIGURA 3 - HIERARQUIA APRENDIZADO DE MÁQUINA.....	25
FIGURA 4 - AMOSTRA BASE NORMAS APLICÁVEIS	32
FIGURA 5 - AMOSTRA BASE NORMAS NÃO APLICÁVEIS	33
FIGURA 6 - JUNÇÃO DAS BASES.....	33
FIGURA 7 - FREQUÊNCIA PALAVRAS ANTES DO PRÉ-PROCESSAMENTO.....	34
FIGURA 8 - TEXTO APÓS PRÉ-PROCESSAMENTO	35
FIGURA 9 - FREQUÊNCIA DE PALAVRAS APÓS O PRÉ-PROCESSAMENTO	35
FIGURA 10 - DIVISÃO DA BASE.....	36
FIGURA 11 - APLICAÇÃO MÉTODO TF-IDF VECTORIZER	36
FIGURA 12 - RELATÓRIO DE CLASSIFICAÇÃO RANDOM FOREST	37
FIGURA 13 - RELATÓRIO DE CLASSIFICAÇÃO REGRESSÃO LOGÍSTICA.....	38
FIGURA 14 - RELATÓRIO DE CLASSIFICAÇÃO NAIVE BAYES.....	39
FIGURA 15 - RELATÓRIO DE CLASSIFICAÇÃO ÁRVORE DE DECISÃO	40
FIGURA 16 - RELATÓRIO DE CLASSIFICAÇÃO KNN.....	41
FIGURA 17 - PRÉ-PROCESSAMENTO BASE NOVA.....	42
FIGURA 18 - MÉTODO UTILIZADO PARA CLASSIFICAÇÃO	42
FIGURA 19 - AMOSTRA BASE PÓS CLASSIFICAÇÃO	43

LISTA DE QUADROS

QUADRO 1 - PUBLICAÇÕES RESOLUÇÕES E INSTRUÇÕES BCB POR ANO....	13
QUADRO 2 - MODALIDADE DAS INSTITUIÇÕES DE PAGAMENTO.....	16
QUADRO 3 - DESCRIÇÃO FERRAMENTAS	30
QUADRO 4 - BIBLIOTECAS	31

LISTA DE TABELAS

TABELA 1 - RESULTADO CLASSIFICAÇÃO ALGORTIMOS	43
---	----

LISTA DE ABREVIATURAS OU SIGLAS

BCB	- Banco Central do Brasil
IP	- Instituição de Pagamentos
IA	- Inteligência Artificial
PLN	- Processamento de Linguagem Natural
AM	- Aprendizado de Máquina
ML	- Machine Learning
CMN	- Conselho Monetário Nacional
SFN	- Sistema Financeiro Nacional
KNN	- K-Nearest Neighbors

SUMÁRIO

1 INTRODUÇÃO	10
1.1 PROBLEMATIZAÇÃO	11
1.2 OBJETIVOS	12
1.2.1 Objetivo geral	12
1.2.2 Objetivos específicos.....	12
1.3 JUSTIFICATIVA	13
2 REVISÃO DE LITERATURA	15
2.1 <i>FINTECHS</i>	15
2.2 INSTITUIÇÕES DE PAGAMENTO	16
2.3 BANCO CENTRAL DO BRASIL.....	18
2.4 COMPLIANCE.....	19
2.5 INTELIGÊNCIA ARTIFICIAL	21
2.5.1 Processamento de Linguagem Natural	22
2.5.2 Aprendizado de Máquina.....	25
2.5.3 Classificação de documentos com Aprendizado de Máquina	26
2.5.3.1 TF-IDF	26
2.5.4 Métricas de Avaliação	27
3 METODOLOGIA	29
3.1 MATERIAIS	29
3.1.1 Base de dados	29
3.1.2 Ferramentas	30
3.1.3 Bibliotecas	31
3.2 MÉTODOS	32
3.2.1 Análise inicial da base da base	32
3.2.2 Pré-processamento da base	34
3.2.3 Treinamento da base e escolha do algoritmo.....	35
3.2.3.1 Preparação dos dados.....	35
3.2.3.2 Treinamento da base e algoritmos	36
4 APLICAÇÃO DOS ALGORITMOS E ANÁLISE DOS RESULTADOS.....	42
4.1 ANÁLISE E RISCO DE APLICAÇÃO	44
5 CONSIDERAÇÕES FINAIS	44
5.1 CONTINUIDADE DO ESTUDO.....	45

REFERÊNCIAS.....46

1 INTRODUÇÃO

Desde o seu surgimento, a tecnologia tem desempenhado um papel fundamental na transformação da sociedade e das organizações. O uso de Inteligência Artificial, automação de processos por meio de robôs, Chatbots, telemedicina e uma ampla gama de outras aplicações estão se tornando cada vez mais comuns e influentes em diversos setores. Todos os anos, inovações surgem para resoluções de problemas, otimização de tempo, aumento de produtividade e suprir necessidades de organizações ou de indivíduos no dia a dia.

No âmbito organizacional Schwab (2016) afirma que o mundo caminha para uma nova revolução industrial que vem crescendo de forma acelerada, onde ferramentas de inteligência artificial (IA) e tecnologia serão mais presentes nos setores produtivos no contexto da economia global. A necessidade de aprimorar e melhorar a produtividade, bem como a execução de tarefas repetitivas de forma ágil e dinâmica fez com que as organizações passassem a implementar tecnologias e Inteligência Artificial nos seus processos (CARVALHO, 2021).

No setor financeiro, as instituições financeiras foram impactadas pela influência tecnológica, permitindo que novos modelos de negócio e otimizações de processos surgissem. Tecnologias como IA, computação em nuvem, *blockchain*, pagamentos digitais, contas totalmente digitais, *chatbots* e processos avançados de cibersegurança foram implementados (MCKINSEY, 2016). Além disso, surgiram as *Fintechs*, que é a abreviação de “Financial Technology” ou “Tecnologia Financeira” e refere-se a “empresas que promovem inovações nos mercados financeiros por meio do uso intenso de tecnologia” (BANCO CENTRAL DO BRASIL, 2023). Atualmente no Brasil existem duas categorias de *fintechs* sujeitas à regulamentação do Banco Central do Brasil (BCB), as *fintechs* de crédito e as instituições de pagamentos (IP).

No contexto brasileiro, no ano de 2020, o relatório "*Fintech Report 2020*", apontou que as *fintechs* tiveram um crescimento de 34% chegando em um total de 742 *fintechs* no mercado (DISTRITO, 2020). Já no ano de 2022 o relatório levantou 1289 *fintechs*, representando um aumento de 73% em relação a 2020 (DISTRITO, 2022).

Acompanhando o crescimento e as mudanças do setor financeiro, o Banco Central do Brasil, passou a monitorar, regulamentar e implementar regras para as *fintechs* visando garantir maior segurança e controle das operações para os usuários,

negócios e a própria União. No ano de 2020, o Banco Central do Brasil publicou um total de 126 Resoluções e Instruções Normativas BCB, que são normas regulatórias elaboradas exclusivamente pelo próprio Banco Central. Em contraste, no ano de 2019, nenhuma norma desse tipo foi publicada (BANCO CENTRAL DO BRASIL, 2020). Esse movimento indicou a busca pela regularização e contemplação das instituições financeiras ou instituições de pagamentos às normas de interesse comum.

Dessa forma, para se adequar às exigências regulatórias, as *fintechs* e suas categorias, como as instituições de pagamento passaram a buscar formas de garantir que as normas publicadas fossem cumpridas. Sendo assim, em algumas Instituições de Pagamentos, o setor de Compliance desempenha um papel fundamental ao buscar garantir a conformidade da organização com as leis e normas vigentes. Ele é responsável por captar, analisar e assegurar a aplicabilidade das normas propostas pelos órgãos reguladores, com o objetivo de evitar multas, sanções e assegurar o cumprimento das exigências legais. (SANTOS, 2011).

A fim de auxiliar nos procedimentos e atividades desempenhadas pelos setores como o *Compliance*, a utilização de métodos de IA e Aprendizado de Máquina podem ser explorados e adaptados nesse contexto. Os métodos de IA podem:

[...] diminuir os custos das instituições, trazendo eficiência e agilidade na prestação de serviços, desburocratizando procedimentos e concentrando em uma máquina a possibilidade de executar diversas tarefas que teoricamente demandariam setores e equipes diversas para sua realização, trazendo mais acurácia nas tomadas de decisão, reduzindo riscos e ampliando a conformidade com a regulação (MOTTA e ROSA, 2022, p. 147).

Assim como a IA é utilizada para prevenção de fraudes, levantamento de padrões, a mesma pode auxiliar nos processos de levantamento de normas como consequência aumentar a produtividade e agilidade dos setores responsáveis dentro das *fintechs* e instituições de pagamentos.

1.1 PROBLEMATIZAÇÃO

Dentro da categoria das *fintechs*, uma instituição de pagamentos regulamentada precisa se atentar às Resoluções e Instruções Normativas BCB que lhe cabem. Um processo de captação e análise deve ser realizado nos documentos regulatórios publicados para a verificação da aplicabilidade ou não das regras

presentes no documento no contexto de uma Instituição de Pagamentos. Geralmente os analistas do *Compliance* das instituições ficam responsáveis pelo acesso a plataforma do BCB, levantamento das normas publicadas e posteriormente a realização de uma análise para determinar a aplicabilidade, sendo uma atividade que demanda tempo e é realizada de forma manual.

Diante do cenário apresentado, e das técnicas disponíveis de processamento de linguagem natural e Aprendizado de Máquina, a automação de um processo manual e contínuo pode otimizar tempo dos setores envolvidos, o processo como um todo e posteriormente gerar análises e indicadores das normas.

A proposta de pesquisa do presente estudo visa responder a seguinte questão: **É viável utilizar técnicas de Processamento de Linguagem Natural e Aprendizado de Máquina para classificar resoluções e instruções normativas e determinar sua aplicabilidade em uma Instituição de Pagamentos?**

1.2 OBJETIVOS

Buscando responder à questão levantada para o projeto proposto, os objetivos foram desmembrados em objetivo geral e objetivos específicos

1.2.1 Objetivo geral

O objetivo geral é aplicar técnicas de Aprendizado de Máquina e Processamento de Linguagem Natural para avaliar a viabilidade da classificação de Resoluções e Instruções Normativas publicadas pelo Banco Central do Brasil em relação às instituições de pagamentos.

1.2.2 Objetivos específicos

- I. Compreender a necessidade das Resoluções e Instruções Normativas nas Instituições de Pagamentos;
- II. Aplicar as etapas de processamento de linguagem natural;
- III. Selecionar técnicas de Aprendizado de Máquina para classificação das normas;

- IV. Avaliar o desempenho das técnicas de Aprendizado de Máquina escolhidas e a viabilidade de sua aplicação.

1.3 JUSTIFICATIVA

No contexto de mercado, o volume de normas publicadas pelo Banco Central do Brasil vem crescendo progressivamente ao longo dos anos, sendo resultado da necessidade de adaptação dos órgãos regulatórios diante das mudanças no mercado financeiro. No portal de normas o BCB publica: Resoluções, Circulares, Comunicados, Cartas Circulares e Instruções Normativas.

Dentre os tipos de normas de resoluções e instruções normativas encontrados no portal de normas existem atos normativos do Conselho Monetário Nacional (CMN), Comitê de Regulação e Fiscalização do Mercado Financeiro, de Capitais, de Seguros, de Previdência e Capitalização (COREMEC) e o próprio Banco Central do Brasil, que em seus atos normativos atende a abreviação "BCB".

As "Resoluções BCB" e "Instruções Normativas BCB" são atos normativos publicados do próprio BCB. Já as "Resoluções CMN" e "Instruções Normativas CMN" são atos normativos do Conselho Monetário Nacional (CMN) e as "Resoluções COREMEC" são atos normativos da COREMEC. Como também atos normativos conjuntos denominados "Resolução Conjunta", "Instrução Normativa Conjunta" (BANCO CENTRAL DO BRASIL, 2023). Para a presente pesquisa apenas Resoluções e Instruções Normativas BCB foram contemplados.

Conforme o levantamento feito das Resoluções e Instruções Normativas BCB retirados do portal do Banco Central do Brasil, apresentados no Quadro 1, esse crescimento na quantidade de normas fica evidenciado, durante todo o período de 2013 a 2019 nenhuma norma foi publicada, diferente dos anos seguintes.

QUADRO 1 - PUBLICAÇÕES RESOLUÇÕES E INSTRUÇÕES BCB POR ANO

Ano	Quantidade de normas publicadas
2013-2015	0
2016-2019	0
2020-2021	400
2022-2023 (até maio)	307

Fonte: Elaborado pela autora (2023).

A necessidade de estar em conformidade com as exigências legais demandam das equipes de *Compliance* a necessidade de analisar cada uma das resoluções, sendo uma tarefa repetitiva e que demanda tempo. Dessa forma, soluções para a automatização de atividades passaram a ser consideradas, inclusive com o avanço de Aprendizado de Máquina e a possibilidade da aplicação de seus métodos para análise e classificação de documentos. Dentro do meio jurídico e regulatório, essas inovações e tecnologias estão ganhando cada vez mais espaço. Para Coelho (2019) a IA pode melhorar a identificação, enriquecimento e classificação dos processos jurídicos, e permitir que atividades repetitivas possam ser organizadas e automatizadas. Com a aplicação do processamento de linguagem natural, é possível obter, tratar e categorizar documentos de forma automatizada e inteligente. Isso permite explorar métodos para classificar os documentos de maneira eficiente e precisa (PEIXOTO, 2020).

A contribuição deste estudo para o campo da Gestão da Informação é evidenciada pela abordagem de disciplinas como Análise de Dados, Mineração de Dados e Inteligência Artificial. A aplicabilidade das ferramentas e técnicas utilizadas no projeto foi possibilitada pela incorporação dos temas discutidos em cada disciplina. O campo da gestão da informação oferece a oportunidade de resolver problemas de negócios e otimizar o tempo por meio do uso eficiente e eficaz das informações disponíveis.

No contexto pessoal, a aplicação desta pesquisa permitiu à pesquisadora realizar um estudo aprofundado em sua área de interesse, com o objetivo de buscar soluções para problemas reais por meio do uso de técnicas de aprendizado de máquina.

2 REVISÃO DE LITERATURA

Nesta seção, serão apresentadas as fontes e a fundamentação teórica dos principais conceitos abordados no presente estudo. As informações foram buscadas em livros, dissertações, teses, artigos de periódico e órgãos regulamentadores para embasar a pesquisa dos temas: *Fintechs*, Instituições de Pagamento, *Compliance*, Inteligência Artificial, Aprendizado de Máquina e Processamento de Linguagem Natural.

2.1 FINTECHS

Fintech é um termo originado da junção das palavras "financeiro" e "technology" (tecnologia), e se refere a soluções financeiras impulsionadas por avanços tecnológicos. Segundo o Banco Central do Brasil (2023), as fintechs são empresas que introduzem inovações nos mercados financeiros por meio da aplicação intensiva de tecnologia, oferecendo potencial para desenvolver novos modelos de negócio. Esse novo modelo desconstruiu o antigo modelo bancário, trazendo propostas inovadoras e maior flexibilidade nos produtos (FARIA, 2018). Silva (2020) destaca que as fintechs têm como objetivo tornar os serviços financeiros mais eficientes, proporcionando uma experiência aprimorada aos clientes de forma democrática. Conforme mencionado por Pinto (2018), as *fintechs* apresentam diferenciais em relação às instituições tradicionais, tais como redução de custos, acesso facilitado ao crédito e soluções personalizadas para problemas dos clientes.

De acordo com um estudo especial divulgado pelo BCB, as fintechs têm o potencial de aumentar a concorrência no sistema financeiro ao expandir a oferta de produtos e serviços por meio do uso avançado e especializado de recursos tecnológicos. Além disso, o crescimento das fintechs tem impulsionado as instituições tradicionais a aprimorarem seus processos e formas de funcionamento, adaptando-se ao uso intensivo da tecnologia (BANCO CENTRAL DO BRASIL, 2020).

No Brasil, existem dez categorias de fintechs, de acordo com o Banco Central do Brasil (2023): pagamento, gestão financeira, empréstimo, investimento, financiamento, seguro, negociação de dívidas, criptoativos, Distributed Ledger Technologies (DLTs), câmbio e multisserviços.

Dentre essas categorias, duas estão sujeitas às regulamentações do Banco Central do Brasil: as fintechs de crédito e as instituições de pagamentos. As fintechs de crédito englobam empresas que oferecem serviços relacionados à concessão de crédito, enquanto as instituições de pagamentos são entidades reguladas que fornecem serviços de pagamento (BANCO CENTRAL DO BRASIL, 2023).

2.2 INSTITUIÇÕES DE PAGAMENTO

Uma instituição de pagamentos, de acordo com o Banco Central do Brasil (2020), é uma pessoa jurídica que facilita serviços de compra, venda e movimentação de recursos dentro de um arranjo de pagamento. Ao contrário das instituições financeiras, essas instituições não têm permissão para conceder empréstimos ou financiamentos aos clientes. No entanto, as IPs podem oferecer serviços similares aos bancários, como serviços de pagamento, cartões de crédito e aplicativos de gestão financeira (BANCO CENTRAL DO BRASIL, 2020).

Após a promulgação da Lei 12.865, em 9 de outubro de 2013 (BRASIL, 2013) as primeiras *fintechs* da categoria de pagamentos foram regulamentadas e denominaram-se instituições de pagamento, ficando sujeitas à regulação e supervisão do Banco Central do Brasil (BANCO CENTRAL DO BRASIL, 2020).

De acordo com a Resolução BCB nº 80, publicada em 25 de março de 2021 pelo Banco Central do Brasil, as instituições de pagamentos são classificadas em quatro modalidades (BANCO CENTRAL DO BRASIL, 2021), conforme apresentado no Quadro 2.

QUADRO 2 - MODALIDADE DAS INSTITUIÇÕES DE PAGAMENTO

Modalidade	Atuação
Emissor de moeda eletrônica	Gerencia conta de pagamento de usuário final do tipo pré-paga, disponibiliza transação de pagamento que envolva o ato de pagar ou transferir, na qual os recursos devem ser depositados previamente.
Emissor de instrumento de pagamento pós-pago	Gerencia conta de pagamento do usuário final do tipo pós-paga, na qual os recursos são

	depositados para pagar dívidas previamente assumidas.
Credenciador	Habilita estabelecimentos comerciais para aceitação de instrumento de pagamento, sem gerenciar contas de pagamento de usuários finais.
Iniciador de transação de pagamento	Inicia transação de pagamento ordenada pelo usuário final, porém não gerencia conta de pagamento, nem detém em momento algum os fundos das transações iniciadas

Fonte: Adaptado de BANCO CENTRAL DO BRASIL (2021).

Ainda conforme a Resolução BCB nº 80/2021, uma instituição de pagamentos pode atuar em mais de uma modalidade.

De acordo com o relatório de estudos especiais do BCB (BANCO CENTRAL DO BRASIL, 2020), o modelo de negócios das IPs é projetado para oferecer produtos e serviços que atendam às necessidades dos clientes. Essas instituições se destacam por fornecer contas digitais, agilidade e automação de operações financeiras, além de facilitar pagamentos e transferências. Sua abordagem centralizada no cliente visa a rentabilização e fidelização dos mesmos por meio da diferenciação dos serviços prestados, frequentemente direcionados a nichos específicos.

Inicialmente, IPs não são submetidas às diversas regras e regulamentos que se aplicam às instituições financeiras tradicionais, o que lhes permite oferecer aos clientes interações ágeis e digitais, evitando a necessidade de documentação física e burocracias extensas (MARQUES *et al*, 2022).

Adaptando-se a evolução do mercado e visando garantir a ordem em relação ao setor financeiro, as Resoluções BCB 80 e 81/2021 estabelecem normas e critérios mínimos para a regularização das instituições de pagamentos junto ao Banco Central do Brasil. Esses critérios abrangem diversos aspectos, tais como a modalidade da instituição, tipo de sociedade, denominação social, implementação de políticas de governança que incluem *Compliance*, valor mínimo de movimentações financeiras, definição de capital mínimo, entre outros pontos relevantes (BANCO CENTRAL DO BRASIL, 2021).

Porém, mesmo com o acompanhamento do mercado e a implementação de regulamentações, Klein e Farias (2022) afirmam que a complexidade da regulação nos setores financeiros tem se tornado um desafio crescente. Isso se deve à complexa interação entre diferentes setores e ao rápido avanço da tecnologia, o qual tem um impacto significativo na maneira como os negócios são conduzidos.

Sendo necessários que os órgãos reguladores como o BCB, garantam a prevenção ao risco sistêmico do setor financeiro, determinando que as IPs ou qualquer outra instituição, atuem de forma supervisionada e regularizada garantindo que o sistema financeiro seja protegido e se mantenha íntegro (KLEIN e FARIAS, 2022).

2.3 BANCO CENTRAL DO BRASIL

O Banco Central do Brasil foi constituído pela Lei nº 4.595/1964, onde o mesmo é definido como uma "autarquia federal, caracterizada pela ausência de vinculação a Ministério e que possui autonomia técnica, operacional, administrativa e financeira" (BANCO CENTRAL DO BRASIL, 2023). Sendo assim, conforme sua missão institucional, o BCB tem o objetivo de assegurar a estabilidade de preços, zelar por um sistema financeiro sólido, eficiente e competitivo, suavizar as flutuações do nível de atividade econômica e fomentar o pleno emprego. Guardado (2022) aponta que as funções do Banco Central são: Monopólio de emissão; Banqueiro do Governo; Banco dos Bancos; Executor da política cambial e depositário das reservas internacionais; executor da política monetária; Controlador da Inflação; e principalmente ser o supervisor do sistema financeiro brasileiro.

E no papel de supervisor do sistema financeiro Guardado (2022) afirma que o BCB tem como função:

- Autorizar entidades a funcionar;
- Regular o funcionamento das entidades;
- Monitorar, executar ações e supervisionar as entidades financeiras.

No âmbito do Sistema Financeiro Nacional (SFN), o Banco Central do Brasil desempenha um papel fundamental como um dos supervisores responsáveis pela regulação da moeda, crédito, capitais e câmbio no país. O BCB tem a responsabilidade de assegurar o cumprimento das normas estabelecidas pelo

Conselho Monetário Nacional, além de monitorar, fiscalizar e implementar políticas monetárias, cambiais e de crédito (BANCO CENTRAL DO BRASIL, 2023).

Dentre as entidades supervisionadas pelo BCB temos: Bancos e caixas econômicas; Administradoras de consórcios; Cooperativas de Crédito; Corretoras e distribuidoras; Instituições de Pagamentos; Instituições financeiras; e por fim sete instituições não bancárias, sendo elas agência de fomento, associação de poupança e empréstimo, companhia hipotecárias, sociedade de crédito, financiamento e investimento, sociedade de crédito imobiliário, sociedade de arrendamento mercantil e sociedade de crédito ao microempreendedor.

Todas as instituições mencionadas devem estar sujeitas à supervisão e seguir as regulamentações e normas estabelecidas pelo BCB a fim de assegurar o seu funcionamento adequado. O Banco Central do Brasil disponibiliza um portal que publica resoluções, normas, circulares, comunicados, cartas circulares, instruções normativas e outros atos relevantes. Esses documentos fornecem diretrizes, normas regulatórias e orientações para o setor financeiro, promovendo a transparência e a conformidade com as políticas e exigências estabelecidas.

No portal de normas existem os atos normativos do CMN, COREMEC e as do BCB. Sendo nomeadas com as siglas de cada órgão, como "Resolução BCB", "Resolução CMN" ou "Resolução COREMEC", existe também a possibilidade dos atos normativos recebem a nomenclatura "Conjunta", representando um ato normativo conjunto entre as entidades reguladoras do SFN (BANCO CENTRAL DO BRASIL, 2023).

2.4 COMPLIANCE

O termo Compliance tem origem do verbo "to comply", que traduzindo do inglês, significa "cumprir", "executar", "conformar-se", "satisfazer" algo que lhe foi imposto (LAMBOY, 2018). Sendo assim, dentro de um contexto empresarial Lamboy (2018) afirma que Compliance é o:

[...] dever de cumprir e estar em conformidade com diretrizes estabelecidas na legislação, normas e procedimentos determinados, interna e externamente, para uma empresa, de forma a mitigar riscos relacionados a reputação e a aspectos regulatórios (LAMBOY, 2018, p. 6).

Begnini *et al.* (2022), o *compliance* desempenha um papel fundamental ao garantir que a instituição esteja em conformidade com o código de ética e o cumprimento das normas regulamentadoras internas e externas. Sua implementação permite a existência de monitoramento, controle e fiscalização dos processos, evitando que a instituição corra o risco de sofrer sanções legais ou regulatórias, o que poderia afetar sua reputação e resultar em perdas financeiras significativas.

Considerando os aspectos legislativos relacionados ao *compliance*, é importante destacar que o Decreto Nº 8.420/15 estabelece a responsabilização objetiva de pessoas jurídicas em atos de administração pública, tanto no âmbito nacional quanto estrangeiro. Já a Lei Nº 12.846/13, conhecida como Lei Anticorrupção, onde o programa *compliance* foi chamado de "Programa de Integridade", prevê a implementação do programa (BRASIL, 2013). Essa lei possibilita que as pessoas jurídicas apresentem informações e documentos que comprovem a existência e o funcionamento efetivo do programa de integridade, visando prevenir e combater práticas corruptas e fraudulentas (LAMBOY, 2018).

De acordo com o manual de *compliance* da Federação Brasileira de Bancos (FEBRABAN, 2018), dentro de uma organização a função do *compliance* tem como escopo:

- I. Garantir a disseminação da cultura e dos temas de *Compliance*, fornecendo suporte à Alta Administração na definição de treinamentos e capacitação adequada para todos os colaboradores e prestadores de serviços terceirizados relevantes;
- II. Atuar como área consultiva nos assuntos relacionados a *Compliance*, oferecendo orientação e aconselhamento em questões éticas e de conformidade;
- III. Assegurar a existência de normativos internos atualizados, como políticas, circulares e manuais, além de processos e procedimentos claros que estejam alinhados com os requisitos legais e regulatórios;
- IV. Participar de forma proativa na definição e manutenção de programas relacionados a *Compliance*, como a implementação de controles internos e a realização de monitoramentos periódicos;
- V. Identificar e avaliar a aderência da Instituição ao arcabouço legal e regulatório, às recomendações de órgãos de supervisão e autorregulação,

bem como aos códigos de conduta aplicáveis, levando em consideração os riscos envolvidos;

- VI. Participar na aprovação de produtos, serviços e potenciais parceiros e clientes, garantindo que sejam devidamente avaliados quanto aos aspectos de conformidade;
- VII. Assegurar a existência de processos definidos para o cumprimento de regulamentações específicas aplicáveis à Instituição;
- VIII. Monitorar de forma contínua a exposição aos riscos de *Compliance* e realizar testes de controles para garantir a eficácia das medidas adotadas;
- IX. Relatar de forma sistemática e periódica os resultados das atividades relacionadas a *Compliance* ao Conselho de Administração, à Alta Administração e aos demais níveis organizacionais, fornecendo informações relevantes e análises sobre o desempenho em termos de conformidade;
- X. Manter um relacionamento ético e íntegro com reguladores, autorreguladores, entidades de representação e auditores independentes, garantindo uma comunicação adequada e transparente para atender às demandas regulatórias;

Dentro das Instituições de Pagamentos, o *Compliance* pode captar as resoluções normativas, verificar a aplicabilidade, direcionar aos setores de interesse e garantir que o normativo seja cumprido dentro do prazo e a IP esteja em conformidade com os critérios solicitados pelo órgão regulamentador.

2.5 INTELIGÊNCIA ARTIFICIAL

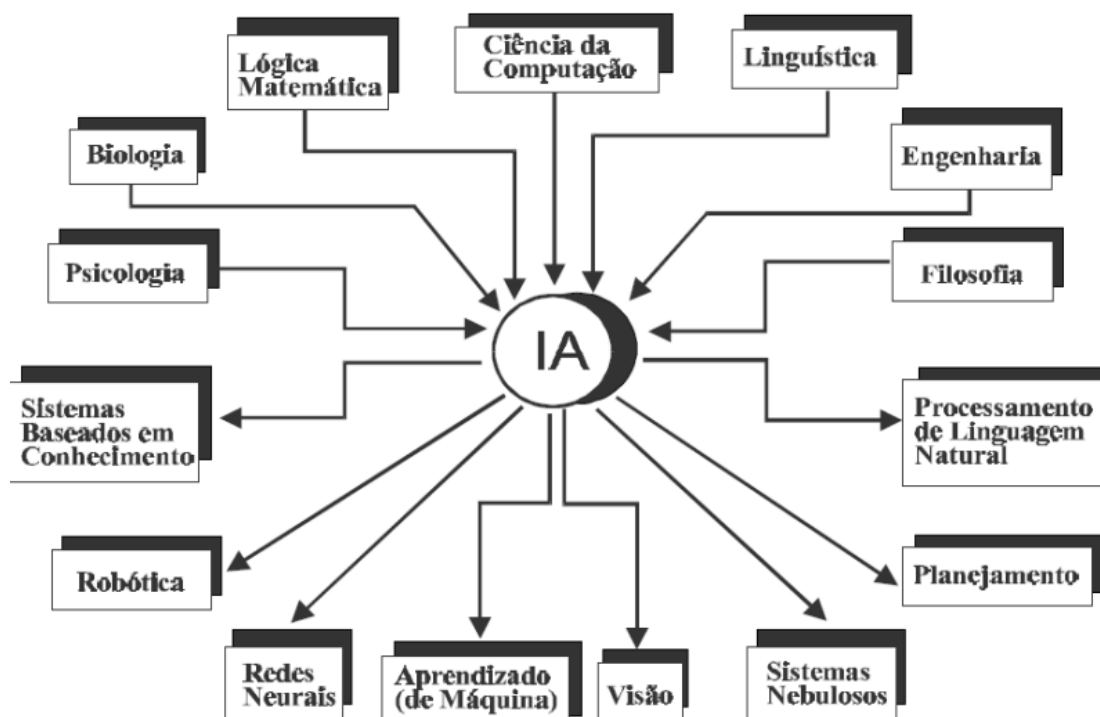
A inteligência artificial (IA) deu seus primeiros passos na década de 1940, impulsionada pelos avanços tecnológicos após a Segunda Guerra Mundial. Um marco importante foi a publicação do estudo "Computing Machinery and Intelligence" por Alan Turing em 1950, considerado o pai da computação. A partir desse momento, uma série de eventos e avanços foram ocorrendo, levando ao desenvolvimento da IA que conhecida hoje.

Para Russell e Norvig (2013), Inteligência Artificial é a arte de criar máquinas que realizam tarefas que exigiriam inteligência se fossem executadas por seres

humanos. A IA utiliza as técnicas de aprendizado de máquina, processamento de linguagem natural, visão computacional e raciocínio lógico para capacitar sistemas automatizados a tomar decisões, aprender com dados e se adaptar a novas situações.

Sendo possível utilizar métodos e estratégias para melhorar performance ou delegar funções repetitivas que podem ser roboticamente praticáveis. Para Monard e Baranauskas (2000) a IA se relaciona com diversas áreas do conhecimento pois tem uma amplitude multidisciplinar, como é apresentado na Figura 1.

FIGURA 1 - ÁREAS RELACIONADAS COM IA



Fonte: MONARD; BARANAUKAS (2000, p. 2).

Ainda de acordo com Monard e Baranauskas (2000) dentre as principais áreas do conhecimento relacionados com o guarda-chuva de IA podemos destacar: Machine Learning (Aprendizado de Máquina) e Processamento de Linguagem Natural.

2.5.1 Processamento de Linguagem Natural

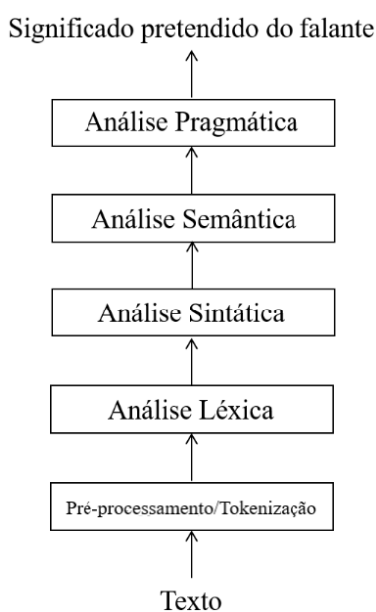
O processamento de linguagem natural (PLN) é uma das subáreas da inteligência artificial (IA) que se dedica ao estudo e desenvolvimento de técnicas para permitir que os computadores compreendam, interpretem e processem a linguagem humana (SANTOS, 2019).

Dentro da computação existe uma linguagem própria, construída com regras rígidas, padronizadas e premeditadas para que o computador consiga compreender de forma rápida e autônoma o comando a ser feito. A linguagem natural geralmente não era compreendida pelas máquinas, o que gerou diversos estudos na área de processamento de linguagem natural para que técnicas fossem aprimoradas e computadores fossem capazes de interpretar e manipular a linguagem natural (Gomes *et al*, 2020).

Conforme Zhang et al. (2023) existe um vasto número de textos em linguagem natural que são gerados todos os dias, seja por rede sociais, *chatbots*, e-mails, reviews de produtos, artigos, livros etc. Essa quantidade de informação estimulou estudos para que as máquinas compreendessem a linguagem humana.

Para Dale (2010) PLN é dividida em etapas, fazendo distinções linguísticas entre sintaxe, semântica e pragmática. Sendo sintaxe a ordem e estrutura; semântica o significado literal; e pragmática se referindo ao significado no contexto do texto. Essa separação, apesar de não ser uma tarefa fácil na aplicação prática de PLN, auxilia na construção de modelos para a análise da linguagem (DALE, 2010). A divisão do processo do processamento de linguagem natural é apresentada na Figura 2.

FIGURA 2 - ETAPAS DA ANÁLISE DE LINGUAGEM NATURAL



Fonte: Adaptado de Dale (2010, p.4)

Santos (2018) destaca aplicações em que o processamento de linguagem natural pode ser utilizado. Algumas das áreas de aplicação incluem:

- Recuperação de documentos: Processo de encontrar documentos relevantes em bases de dados ou repositórios, levando em consideração critérios específicos ou consultas do usuário;
- Classificação de documentos: Atribuição de categorização ou a aplicação de rótulos em documentos com base no seu conteúdo por meio de critérios aplicados;
- Extração de Informação: Identificação de informações relevantes e recuperação das informações em documentos ou textos;
- Sumarização de documentos: Criação de resumos de documentos ou textos;
- Assistentes Virtuais e *chatbots*: Sistemas que interagem com usuários utilizando a linguagem humana, são construídos para compreender a linguagem e interagir com usuários.
- Análise de sentimentos: Técnicas para analisar e determinar a emoção de textos, sendo positivos, negativos ou neutros.
- Tradução de textos: Permite a leitura e tradução de textos;
- Correção de textos: Tem a capacidade de ler, interpretar e corrigir erros ortográficos ou de coesão na língua determinada pelo usuário.

Dentre os recursos utilizados na PNL existem alguns que são amplamente usados, conforme Anchiêta *et al.* (2021):

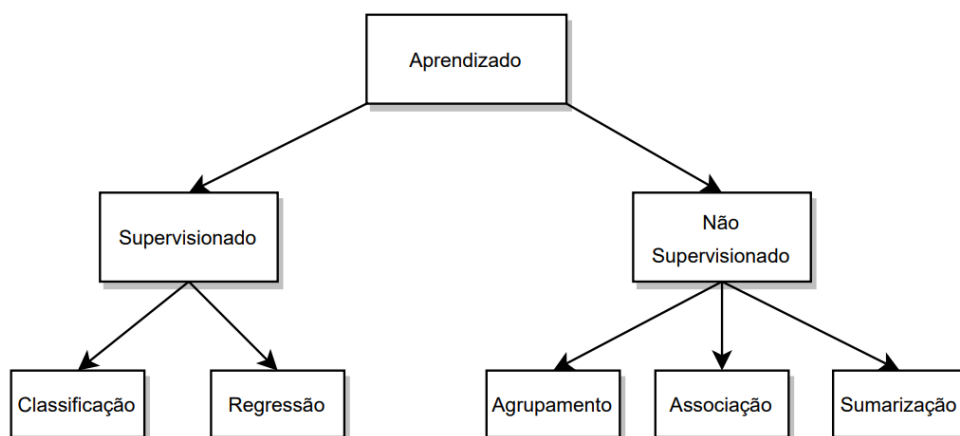
- Tokenização: Processo de separar o texto em tokens, sendo em números, palavras ou símbolos;
- Normalização: Processo de padronizar os dados processados, convertendo para letras maiúsculas ou minúsculas, corrigir erros gramaticais, remover imagens, remover números, pontuações, letras, palavras repetidas e as *stopwords*;
- *Stopwords*: Palavras que são consideradas irrelevantes para a análise do texto, possuem pouco conteúdo lexical, como artigos, pronomes e preposições;
- Lemmatização/*Lemmatizer*: Técnica para buscar o lema da palavra, mantendo o significado junto ao radical.

2.5.2 Aprendizado de Máquina

Para Peixoto (2010, *apud* Murphy, 2012) Machine Learning ou Aprendizado de Máquina é "um conjunto de métodos para detectar padrões em dados de forma automática e, utilizando esses padrões, realizar uma projeção ou encaminhar uma ação".

O processo de AM permite que modelos sejam construídos para que a partir de uma entrada de dados e análise estatística, seja possível realizar automação de processos de tomada de decisão (GASPARETTO, 2021). Segundo Ferreira (2019), o aprendizado de máquina busca ensinar computadores como aprender sem que exista a necessidade de ser programado para realizar tarefas específicas. Gasparetto (2021) explica que existem dois tipos de aprendizagem dentro do Aprendizado de Máquina: Aprendizagem Supervisionada e Aprendizagem não Supervisionada. A divisão por categoria define quais tarefas cada aprendizado tem capacidade de abordar, como preditivas e descritivas. Na Figura 3, Faceli *et al* (2011) apresenta a hierarquia das categorias de AM.

FIGURA 3 - HIERARQUIA APRENDIZADO DE MÁQUINA



Fonte: Adaptado de Faceli *et al* (2011, p. 6).

No aprendizado supervisionado, os elementos de entrada de um determinado conjunto de dados são classificados com rótulos conhecidos que correspondem às saídas desejadas. Isso permite que o algoritmo de aprendizado seja treinado para encontrar padrões e fazer previsões com base nos exemplos fornecidos. O computador poderia buscar em dados históricos disponíveis e identificar possíveis

padrões para classificar determinado evento (GASPARETTO, 2021). Dentro do aprendizado supervisionado existe a classificação, onde a tarefa é identificar, por semelhança, objetos entre diversas categorias pré-definidas (MADEIROS; SILVA, 2020).

2.5.3 Classificação de documentos com Aprendizado de Máquina

Segundo Mastella (2020), o Brasil ainda está nos passos iniciais da aplicação de automação de processos e classificação de textos com PLN e AM, porém já existem análises e estudos sendo desenvolvidos. O Instituto de Economia Aplicada (IPEA), já realizou estudos quanto a aplicabilidade da IA e PLN em processos judiciais presentes em juizados federais e regionais (Gomes *et al*, 2020).

No Tribunal de Justiça de Pernambuco, a IA denominada Elis, já agilizou tarefas repetitivas e triagem de diversos processos, classificando e identificando processos (SILVA; MAIRINK, 2019).

Inúmeras abordagens para a classificação de documentos estão disponíveis. Algoritmos como *Naive Bayes*, *Árvore de decisão*, *Random Forest*, *Regressão logística* e *Redes Neurais* possuem uma abordagem mais clássica, porém ainda funcionais. Atualmente algumas outras abordagens como *Deep Learning*, *Recurrent Neural Networks*, *Long Short-Term Memory* estão ganhando espaço (MASTELLA, 2020).

2.5.3.1 TF-IDF

TF-IDF, sigla de "Term Frequency-Inverse Document Frequency", é uma técnica de quantificação de palavras em conjunto de dados. Nessa técnica as palavras recebem peso de acordo com sua importância no conteúdo do texto ou documento (MOSER, 2021). Sua base de cálculo é:

$$TF_IDF = TF * IDF$$

Onde TF (Term Frequency) é a frequência de uma palavra "p" presente em um documento "d", dado pela equação:

$$TF(p,d) = \log_{10}(\text{Count}(p,d) + 1)$$

Já o IDF (Inverse Document Frequency) é a inversa da frequência de um documento, balanceado o TF, isso faz com que palavras com maior frequência nos documentos se tornem menos importantes. No cálculo D representa o conjunto de documentos, “df” a frequência do termo “t” nos documentos (MOSER,2021):

$$IDF(t) = \log_{10}(\text{Count}(D) / df(t))$$

2.5.4 Métricas de Avaliação

Para verificar a performance e efetividade dos modelos utilizados em aprendizado de máquina, existem alguns indicadores que avaliam o desempenho, como: precisão, recall, acurácia e f-score. Conforme Santos (2019) as métricas podem ser definidas como:

- Acurácia: Avalia o percentual de acertos, verificando o total de acertos dividido pelo total de entradas. Sendo sua fórmula:

$$\text{Total de Acertos} / \text{Total de Itens}$$

- Precisão: Avalia a quantidade de verdadeiros positivos sobre o total de valores positivos reais. Sendo sua fórmula:

$$\text{Positivos} / (\text{Positivos} + \text{Falsos Positivos})$$

- F-Score: Média harmônica calculada pela combinação de precisão e recall. O valor de F leva em conta a precisão e o recall. Sendo sua fórmula:

$$f1 = 2 * (\text{precisão} * \text{sensibilidade} / (\text{precisão} + \text{sensibilidade}))$$

- Recall: Avalia a capacidade da técnica em detectar com sucesso os resultados dados como positivos na classificação. Sendo sua fórmula:

$$\text{Positivos} / (\text{Positivos} + \text{Falsos Negativos})$$

- Curva ROC: ROC é a abreviação de “*Receiver Operating Characteristics*” e apresenta a efetividade de um modelo de classificação binária.
- Curva AOC: Já a curva AOC, abreviação de “Area Under the Curve”, representa a área abaixo da curva ROC. Permite uma medida agregada de desempenho nos limites de classificação possíveis, seus valores variam no intervalo de 0 e 1, onde quanto mais próximo do 1, melhor o desempenho da classificação.

3 METODOLOGIA

Nesta seção será apresentado o ambiente da pesquisa, materiais, ferramentas, construção e descrição da base e os métodos utilizados para a realização do estudo. A pesquisa se caracteriza como descritiva, tendo caráter quantitativo com base nos métodos utilizados.

3.1 MATERIAIS

3.1.1 Base de dados

As normas foram obtidas por meio do portal "Busca Normas" do Banco Central do Brasil, que disponibiliza os atos normativos, como resoluções e instruções normativas. A seleção e extração dos documentos foram realizadas por um advogado especialista em *Compliance*, com experiência em Instituições de Pagamentos, que também classificou as normas. Essa classificação foi baseada na análise individual de cada norma dentro do portal do BCB por parte do especialista, determinando sua aplicabilidade ou não às instituições de pagamentos, o advogado encaminhou dois arquivos contendo as normas classificadas para treinamento e teste.

A extração levou em conta o período de 2020 até maio de 2023 e somente Resoluções e Instruções BCB foram consideradas. No total, foram levantadas 110 normas, divididas entre aplicáveis e não aplicáveis a uma IP e armazenadas em uma planilha da ferramenta Microsoft Excel, uma base com 30 normas aleatórias foi construída também para fim de testes após o treinamento dos algoritmos. A organização dos dados se deu pela criação de três planilhas, sendo elas:

- I. Normas aplicáveis: Foram selecionadas 55 normas aplicáveis em uma instituição de pagamentos;
- II. Normas não aplicáveis: Foram selecionadas 55 normas não aplicáveis em uma instituição de pagamentos;
- III. Base de testes: Foram selecionadas 30 normas aleatórias divididas entre normas aplicáveis ou não aplicáveis para testes.

3.1.2 Ferramentas

Para o desenvolvimento do projeto, foram utilizadas quatro ferramentas principais: Microsoft Excel, Jupyter Notebook, Python e Anaconda. No Quadro 3, são apresentadas as descrições de cada ferramenta, juntamente com suas respectivas versões.

QUADRO 3 - DESCRIÇÃO FERRAMENTAS

Ferramenta	Descrição	Licença
Microsoft Excel – Office 365	Excel é um software de planilhas eletrônicas da Microsoft, possui recursos de criação de tabelas, controles, funções, fórmulas, estruturação e armazenamento de dados.	Paga
Jupyter Notebook 6.5.4	Aplicativo de código aberto que permite o desenvolvimento e a execução de códigos dentro de notebooks. Ele oferece suporte para várias linguagens de programação, incluindo Python, R, Julia, entre outras.	Gratuita
Anaconda 4.12.0	Anaconda é uma plataforma de distribuição das linguagens de programação Python e R, voltada para a área de computação científica. Seu principal objetivo é simplificar o gerenciamento voltado para a área de ciência de dados	Gratuita/Paga
Python 3.10.7	Python é uma linguagem de programação de alto nível. Suporta programação orientada a objetos, programação imperativo e funcional. Amplamente utilizada em análise de dados	Gratuita

Fonte: A autora (2023).

3.1.3 Bibliotecas

Para realizar o tratamento, pré-processamento dos dados, análise exploratória e execução dos algoritmos, foram utilizadas nove bibliotecas de Python, conforme o quadro 4:

QUADRO 4 - BIBLIOTECAS

Biblioteca	Descrição	Finalidade no projeto
Pandas – Versão 1.4.2	Biblioteca que permite a análise e manipulação de dados de forma rápida e flexível;	Manipulação da base
Seaborn – Versão 0.11.2	Biblioteca para visualização de dados baseada no Matplotlib;	Visualização
Matplotlib – Versão 3.5.1	Biblioteca para criação de visualizações;	Visualização
Re – Versão 3.10.7	Biblioteca que permite trabalhar com expressões regulares	Manipulação do texto e palavras
String – Versão 3.10.7	Biblioteca que permite a manipulação e formatação dos termos	Manipulação do texto e palavras
Unicode – Versão 3.10.7	Codificação de caracteres padrão, garante que o texto seja o mesmo em diferentes plataformas. Permite trabalhar com textos e alterar a representação dos caracteres	Manipulação do texto e palavras
SpaCy – Versão 3.5.0	Biblioteca para processamento de linguagem natural	Processamento de linguagem natural
NLTK – Versão 3.7	Abreviação de "Natural Language Toolkit", em português conjunto de ferramentas para processamento de linguagem natural. É uma biblioteca utilizada para o PLN	Processamento de linguagem natural

Sklearn – Versão 1.2.2.	Biblioteca que possui diversas ferramentas para análise de dados, aprendizado de máquina e estatística	Algoritmos de classificação e análise estatística
----------------------------	--	---

Fonte: A autora (2023).

3.2 MÉTODOS

Com o objetivo de classificar resoluções e instruções normativas, são apresentadas as etapas de pré-processamento, treinamento da base, escolha do modelo e teste, todas realizadas por meio de um código Python¹.

3.2.1 Análise inicial da base da base

No processo de análise da base de dados, foram utilizadas duas planilhas: "Normas Aplicáveis" e "Normas Não Aplicáveis". Ambas as planilhas possuem quatro colunas: "Norma", "Tipo", "Texto" e "CLASSE" e 55 linhas cada uma. A coluna "Norma" contém o número identificador de cada normativo, a coluna "Tipo" indica se é uma "Resolução BCB" ou "Instrução Normativa BCB", e a coluna "Texto" contém o texto completo das normas retirado do portal do BCB.

Para distinguir as normas em cada planilha, foi atribuída a coluna "CLASSE", onde a classe 0 representa as normas aplicáveis e a classe 1 representa as normas não aplicáveis. A Figura 4 apresenta uma amostra da base de normas aplicáveis, enquanto a Figura 5 mostra uma amostra da base de normas não aplicáveis.

FIGURA 4 - AMOSTRA BASE NORMAS APLICÁVEIS

Norma	Tipo	Texto	CLASSE
262	Resolução BCB	RESOLUÇÃO BCB Nº 262, DE 22 DE NOVEMBRO DE 202...	0
319	Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 319, DE 4 DE NOVEMB...	0
318	Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 318, DE 4 DE NOVEMB...	0
321	Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 321, DE 10 DE NOVEM...	0
325	Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 325, DE 21 DE NOVEM...	0

Fonte: A autora (2023).

¹ Código disponível na plataforma github: <https://github.com/julingles/TCC.git>

FIGURA 5 - AMOSTRA BASE NORMAS NÃO APLICÁVEIS

Norma	Tipo	Texto	CLASSE
204	Resolução BCB	RESOLUÇÃO BCB Nº 204, DE 22 DE MARÇO DE 2022\...	1
206	Resolução BCB	RESOLUÇÃO BCB Nº 206, DE 22 DE MARÇO DE 2022\...	1
207	Resolução BCB	RESOLUÇÃO BCB Nº 207, DE 22 DE MARÇO DE 2022\...	1
210	Resolução BCB	RESOLUÇÃO BCB Nº 210, DE 22 DE MARÇO DE 2022\...	1
211	Resolução BCB	RESOLUÇÃO BCB Nº 211, DE 22 DE MARÇO DE 2022\...	1

Fonte: A autora (2023).

Após a análise inicial das duas bases, é aplicado o método "Shuffle" da biblioteca Sklearn para combinar as duas bases de dados em uma única, conforme mostrado na Figura 6.

FIGURA 6 - JUNÇÃO DAS BASES

Norma	Tipo	Texto	CLASSE
38	270 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 270, DE 1º DE ABRIL...	0
32	271 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 271, DE 1º DE ABRIL...	0
43	255 Resolução BCB	RESOLUÇÃO BCB Nº 255, DE 1º DE NOVEMBRO DE 202...	0
102	384 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 384, DE 18 de maio ...	1
58	213 Resolução BCB	RESOLUÇÃO BCB Nº 213, DE 30 DE MARÇO DE 2022\...	1
...	...	...	...
34	273 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 273, DE 1º DE ABRIL...	0
87	298 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 298, DE 23 DE AGOST...	1
90	323 Instrução Normativa BCB	INSTRUÇÃO NORMATIVA BCB Nº 323, DE 16 DE NOVEM...	1
27	219 Resolução BCB	RESOLUÇÃO BCB Nº 219, DE 30 DE MARÇO DE 2022\...	0
3	28 Resolução BCB	RESOLUÇÃO BCB Nº 28, DE 23 DE OUTUBRO DE 2020\...	0

Fonte: A autora (2023).

Com a fusão das bases de dados, foi realizada uma análise preliminar do número total de palavras, juntamente com a frequência de ocorrência de cada palavra, conforme apresentado na Figura 7.

FIGURA 7 - FREQUÊNCIA PALAVRAS ANTES DO PRÉ-PROCESSAMENTO

	Palavra	Frequência
18	de	13093
35	e	4273
13	do	4174
38	a	3589
121	-	3015
50	da	2387
55	o	2385
48	no	2287
73	com	2007
131	em	1986
52	que	1744
312	ou	1620
5	DE	1206
30	para	1178
101	Art.	1089
10	as	1029
221	atributos	1012
210	dos	997
68	nº	940
142	§	925

Fonte: A autora (2023).

3.2.2 Pré-processamento da base

Para a etapa de pré-processamento dos dados foi realizado o processo de normalização, remoção das *stopwords* e lematização.

Com o suporte das bibliotecas *re*, *string* e *unicode*, o processo de normalização removeu caracteres especiais, números, acentos, pontuações e espaços em branco desnecessários dos documentos.

Para realizar a remoção das *stopwords*, utilizamos a biblioteca "*NLTK*" em conjunto com o pacote "*portuguese*", que contém os recursos específicos da língua portuguesa, para remoção de preposições e termos considerados irrelevantes para a análise do texto. Além das *stopwords* padrão fornecidas pela biblioteca, foram adicionados novos termos específicos para o processamento dos documentos em questão. Esses termos incluem: 'oh', 'ah', 'ó', 'n', 'bcb', 'resolucao', 'b', 'art', 'arts', 'lei', 'i', 'iii', 'ii', 'iv', 'viii', 'v', 'vii', 'vi', 'ix', 'nao', 'janeiro', 'fevereiro', 'marco', 'abril', 'maio', 'junho', 'julho', 'agosto', 'setembro', 'outubro', 'novembro' e 'dezembro'.

Por fim, no processo de lematização, foi utilizada a biblioteca *SpaCy* em conjunto com o pacote "*pt_core_news_sm*", que possui recursos exclusivos para a língua portuguesa.

Na Figura 8, é apresentada uma amostra da base, onde é possível visualizar o texto original antes do pré-processamento no campo denominado "Texto". Após a

aplicação do pré-processamento, o resultado é exibido no campo denominado "texto_limpo".

FIGURA 8 - TEXTO APÓS PRÉ-PROCESSAMENTO

	Texto	texto_limpo
62	RESOLUÇÃO BCB Nº 221, DE 30 DE MARÇO DE 2022 Dispõe sobre a prestação de informações pelas cooperativas singulares de crédito a respeito de seus cooperados e dos municípios depositantes, incluindo seus órgãos ou e...	dispoe sobre prestacao informacao cooperativo singular credito respeito cooperar municipio depositante incluir orgaos entidade empres controlador diretoria colegiado banco central Brasil sessao realizar base compleme...
8	RESOLUÇÃO BCB Nº 93, DE 6 DE MAIO DE 2021 Dispõe sobre a atividade de auditoria interna nas administradoras de consórcio e nas instituições de pagamento. A Diretoria Colegiada do Banco Central do Brasil, em ses...	dispoe sobre atividade auditorio interno administradora consorcio instituicao pagamento diretoria colegiado banco central Brasil sessao realizar base inciso aline r s l capitulo objeto ambito aplicacao regulamento a...
2	RESOLUÇÃO BCB Nº 7, DE 12 DE AGOSTO DE 2020 Dispõe sobre os critérios e os procedimentos para reconhecimento contábil e mensuração dos componentes do ativo intangível e veda o registro de ativo diferido pelas admi...	dispoe sobre criterio procedimento reconhecimento contabil mensuracao componente ativo intangivel veda registro ativo diferir administradora consorcio instituicao pagamento diretoria colegiado banco central Brasil se...

Fonte: A autora (2023).

Na Figura 9, é exibido o levantamento da frequência de cada palavra após a conclusão do pré-processamento. Esse levantamento permite identificar a mudança na distribuição das palavras e a sua frequência no documento.

FIGURA 9 - FREQUÊNCIA DE PALAVRAS APÓS O PRÉ-PROCESSAMENTO

Palavra	Frequencia
instituicao	1583
atributo	1028
dever	952
pagamento	921
central	898
valor	872
Brasil	835
banco	812
financeiro	785
registrar	679
titulo	677
funcao	626
cujo	615
operacao	521
tratar	504
ativo	467
outro	451
credito	448
servico	426
informacao	421

Fonte: A autora (2023).

3.2.3 Treinamento da base e escolha do algoritmo

3.2.3.1 Preparação dos dados

Considerando que o modelo utilizado para aplicação dos algoritmos é de aprendizagem supervisionada, a base de dados foi dividida em conjuntos de treino e teste. Para essa divisão, 80% (88 normas) dos dados foram destinados ao

treinamento e 20% (22 normas) aos testes das 110 normas aplicáveis e não aplicáveis, conforme a Figura 10.

FIGURA 10 - DIVISÃO DA BASE

```
X_train, X_val, y_train, y_val = train_test_split(df["texto_limpo"],
                                                df["CLASSE"],
                                                test_size=0.2,
                                                shuffle=True) #Divisão da Base
```

Fonte: A autora (2023).

Após separar a base de dados em conjuntos de teste e treinamento, é aplicado o método "TfidfVectorizer" da biblioteca Sklearn. Esse método é utilizado para converter o texto dos documentos em uma matriz de recursos, permitindo calcular a métrica estatística TF-IDF.

Em conjunto com o "TfidfVectorizer", é utilizado o método "fit_transform" para aplicar ao conjunto de dados de treinamento o ajuste vetorizado aos dados e em seguida a transformação dos textos em uma matriz que contém o resultado do cálculo TF-IDF. A Figura 10 evidencia o uso do método.

FIGURA 11 - APLICAÇÃO MÉTODO TF-IDF VECTORIZER

```
tfidf_vectorizer = TfidfVectorizer(use_idf=True)
X_train_vectors_tfidf = tfidf_vectorizer.fit_transform(X_train)
X_val_vectors_tfidf = tfidf_vectorizer.transform(X_val)
```

Fonte: A autora (2023).

3.2.3.2 Treinamento da base e algoritmos

Após a preparação dos dados, foram selecionados cinco métodos de classificação para serem aplicadas à base de dados: *Naive Bayes (MultinomialNB)*, *Random Forest (RandomForestClassifier)*, *Árvore de decisão (Decision Tree)*, KNN e Regressão Logística (*LogisticRegression*). Todas essas técnicas estão disponíveis na biblioteca *Sklearn*.

Cada método foi avaliado para determinar sua melhor performance no conjunto de dados, a fim de selecionar o método com os resultados mais satisfatórios. Posteriormente, os métodos foram aplicados à uma nova base, com dados não

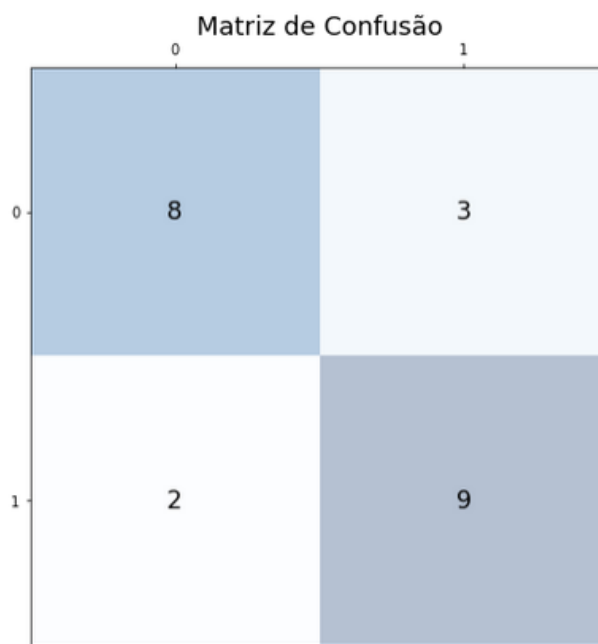
treinados. Com o relatório de classificação e a matriz de confusão disponibilizado pela biblioteca *Sklearn* foi possível avaliar a performance dos cinco métodos utilizados.

O primeiro método utilizado foi o *Random Forest* (*RandomForestClassifier*). O modelo apresentou uma acurácia de 77%. A precisão para normas aplicáveis foi de 80%, enquanto para normas não aplicáveis foi de 75%, representados por 0 e 1, respectivamente. O recall para normas aplicáveis foi de 73% e para normas não aplicáveis foi de 82%, conforme ilustrado na Figura 12.

FIGURA 12 - RELATÓRIO DE CLASSIFICAÇÃO RANDOM FOREST

	precision	recall	f1-score	support
0	0.80	0.73	0.76	11
1	0.75	0.82	0.78	11
accuracy			0.77	22
macro avg	0.78	0.77	0.77	22
weighted avg	0.78	0.77	0.77	22

AUC: 0.756198347107438



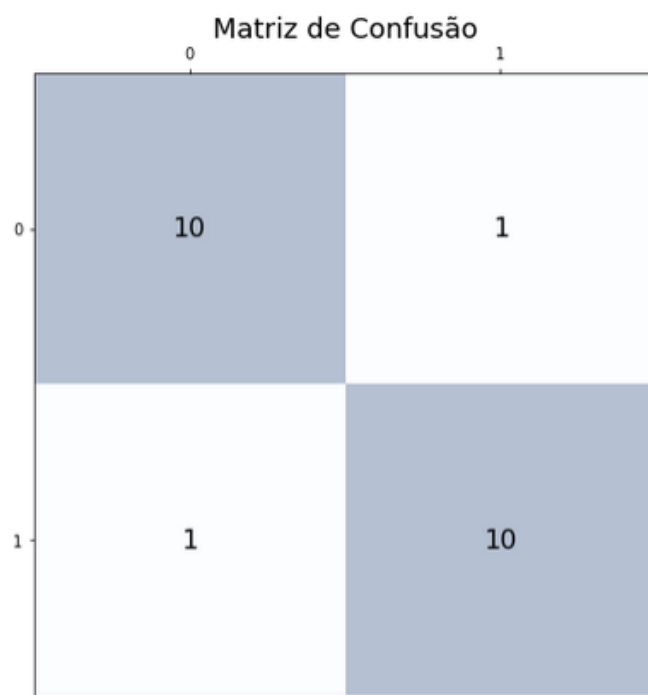
Fonte: A autora (2023)

O segundo método utilizado foi a Regressão Logística (*LogisticRegression*), ilustrado na Figura 13, apresentando uma acurácia de 91%. A precisão de normas aplicáveis e não aplicáveis ficou em 91% bem como o recall.

FIGURA 13 - RELATÓRIO DE CLASSIFICAÇÃO REGRESSÃO LOGÍSTICA

	precision	recall	f1-score	support
0	0.91	0.91	0.91	11
1	0.91	0.91	0.91	11
accuracy			0.91	22
macro avg	0.91	0.91	0.91	22
weighted avg	0.91	0.91	0.91	22

AUC: 0.9338842975206612



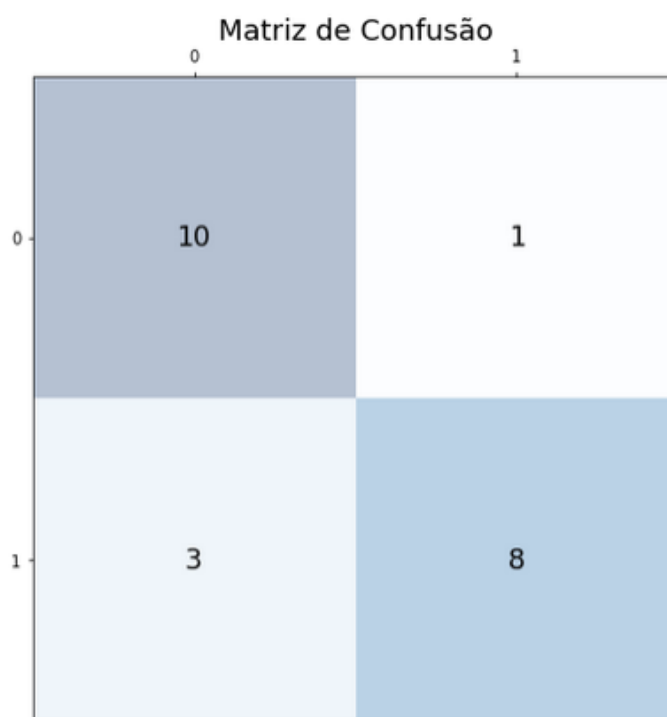
Fonte: A autora (2023)

O terceiro método utilizado foi o *Naïve Bayes*, apresentando uma acurácia de 82%. A precisão para normas aplicáveis foi de 77%, enquanto para normas não aplicáveis foi de 89%. O recall para normas aplicáveis ficou em 91%, enquanto para normas não aplicáveis foi de 73%, conforme ilustrado na Figura 14.

FIGURA 14 - RELATÓRIO DE CLASSIFICAÇÃO NAIVE BAYES

	precision	recall	f1-score	support
0	0.77	0.91	0.83	11
1	0.89	0.73	0.80	11
accuracy			0.82	22
macro avg	0.83	0.82	0.82	22
weighted avg	0.83	0.82	0.82	22

AUC: 0.8925619834710743



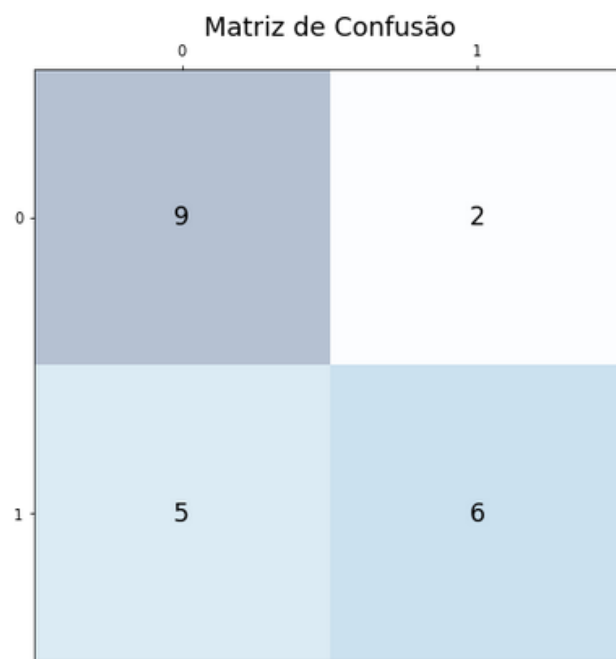
Fonte: A autora (2023)

No modelo de Árvore de Decisão (*Decision Tree*), a acurácia foi de 68%, representando a menor acurácia entre todos os modelos treinados, conforme a Figura 15. A precisão para normas aplicáveis ficou em 64%, enquanto para normas não aplicáveis foi de 75%. O recall para normas aplicáveis foi de 82%, enquanto para normas não aplicáveis ficou em 55%.

FIGURA 15 - RELATÓRIO DE CLASSIFICAÇÃO ÁRVORE DE DECISÃO

	precision	recall	f1-score	support
0	0.64	0.82	0.72	11
1	0.75	0.55	0.63	11
accuracy			0.68	22
macro avg	0.70	0.68	0.68	22
weighted avg	0.70	0.68	0.68	22

AUC: 0.68181818181817



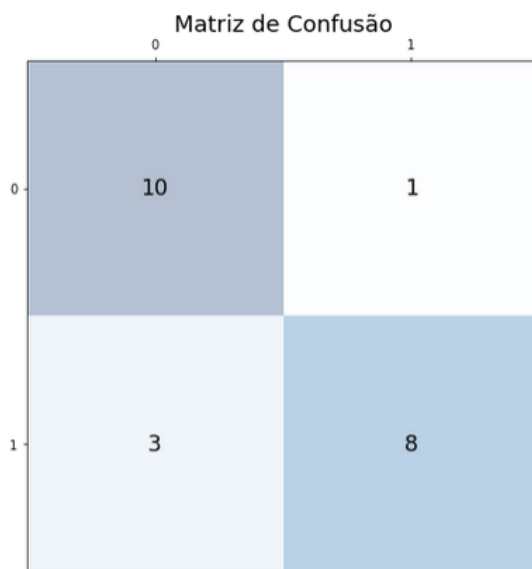
Fonte: A autora (2023).

O último modelo testado foi o *K-Nearest Neighbors* ou KNN (*KNeighborsClassifier*), que obteve uma acurácia de 82%, apresentado na Figura 16. A precisão para normas aplicáveis foi de 77%, enquanto para normas não aplicáveis foi de 89%. O recall para normas aplicáveis ficou em 91%, enquanto para normas não aplicáveis foi de 73%.

FIGURA 16 - RELATÓRIO DE CLASSIFICAÇÃO KNN

	precision	recall	f1-score	support
0	0.77	0.91	0.83	11
1	0.89	0.73	0.80	11
accuracy			0.82	22
macro avg	0.83	0.82	0.82	22
weighted avg	0.83	0.82	0.82	22

AUC: 0.8636363636363636



Fonte: A autora (2023).

Durante a etapa de treinamento e teste, o modelo de Regressão Logística apresentou a maior acurácia, alcançando 91%. Os modelos *Naive Bayes* e KNN obtiveram uma acurácia de 82%, enquanto o modelo de Árvore de Decisão teve a pior performance, com 68% de acurácia. Para a etapa da aplicação dos algoritmos na base de testes apenas os métodos Regressão Logística, Naive Bayes, *Random Forest* e KNN foram considerados, visto que o método de árvore de decisão apresentou uma eficácia abaixo de 70%.

4 APLICAÇÃO DOS ALGORITMOS E ANÁLISE DOS RESULTADOS

Após a conclusão da etapa de treinamento, testes e seleção dos métodos de classificação, os algoritmos foram aplicados a uma nova base não treinada contendo 30 resoluções e instruções normativas BCB. Essas normas foram selecionadas aleatoriamente e misturadas entre normas aplicáveis e não aplicáveis. A base de dados também possui uma coluna chamada "CLASSE", na qual as normas aplicáveis são representadas pelo valor 0 e as não aplicáveis pelo valor 1.

Para a nova base o mesmo método de pré-processamento é realizado antes da aplicação dos modelos treinados, a Figura 17 apresenta uma amostra da base antes do pré-processamento e depois do pré-processamento.

FIGURA 17 - PRÉ-PROCESSAMENTO BASE NOVA

Norma	Texto	texto_limpo	CLASSE
0	274 RESOLUÇÃO BCB Nº 274, DE 13 DE DEZEMBRO DE 2022\nAltera a Resolução BCB nº 131, de 20 de agosto de 2021, que consolida as normas sobre o rito do processo administrativo sancionador, a aplicação de penalidades, o te...	alter consolir norma sobre rito processo administrativo sancionador aplicacao penalidade termo compromisso medida acautelatorio multa cominatorio acordo administrativo processo supervisao prever parametro aplicacao p...	0
1	282 RESOLUÇÃO BCB Nº 282, DE 31 DE DEZEMBRO DE 2022\nAltera a Circular nº 3.978, de 23 de janeiro de 2020, que dispõe sobre a política, os procedimentos e os controles internos a serem adotados pelas instituições autor...	alter circular dispoe sobre politica procedimento controles interno ser adotar instituicao autorizar funcionar banco central Brasil visar prevencao utilizacao sistema financeiro praticar crime lavagemn ocultacao bem ...	0

Fonte: A autora (2023).

Concluindo a etapa de pré-processamento, para o processo de classificação das resoluções e instruções normativas BCB, o texto tratado passa por uma transformação de dados utilizando a vetorização TF-IDF (tfidf_vectorizer) e em seguida passa por cada algoritmo treinado para gerar as previsões. Na Figura 18, é apresentado uma demonstração do código de classificação da base utilizando o método de Regressão Logística.

FIGURA 18 - MÉTODO UTILIZADO PARA CLASSIFICAÇÃO

```
#Testando modelo com Regressão Logística
df_test = pd.read_excel('Base normas teste.xlsx') # carregando nova base
df_test['texto_limpo'] = df_test['Texto'].apply(lambda x: finalprocessamento(x)) #pré-processamento
X_test=df_test['texto_limpo']
X_vector=tfidf_vectorizer.transform(X_test) #convertendo X_test para vetor
y_predict = lr_tfidf.predict(X_vector) #usando o modelo treinado no X_vector
y_prob = lr_tfidf.predict_proba(X_vector) #probabilidade associada do modelo
df_test['target']= y_predict #Cria uma nova coluna com a predição do modelo
```

Fonte: A autora (2023).

No conjunto de dados, uma coluna denominada "*target*" é criada para a classificação das normas pelo algoritmo. Na Figura 19, é possível observar a distribuição das classes nessa coluna, onde as normas aplicáveis são representadas por 0 e as não aplicáveis por 1.

FIGURA 19 - AMOSTRA BASE PÓS CLASSIFICAÇÃO

	Norma	Texto	texto_limpo	CLASSE	target
0	274	RESOLUÇÃO BCB Nº 274, DE 13 DE DEZEMBRO DE 202...	alter consolir norma sobre rito processo admin...	0	0
1	282	RESOLUÇÃO BCB Nº 282, DE 31 DE DEZEMBRO DE 202...	alter circular dispoe sobre politica procedime...	0	1
2	296	RESOLUÇÃO BCB Nº 296, DE 23 DE FEVEREIRO DE 20...	alter disciplina constituicao funcionamento in...	0	0
3	309	RESOLUÇÃO BCB Nº 309, DE 28 DE MARÇO DE 2023\n...	estabelecer procedimento contabel sobre defini...	0	0
4	310	RESOLUÇÃO BCB Nº 310, DE 12 DE ABRIL DE 2023\n...	alter consolida criterio geral elaboracao divu...	0	0
5	311	RESOLUÇÃO BCB Nº 311, DE 12 DE ABRIL DE 2023\n...	alterar dispoe sobre criterio geral elaboracao...	0	0
6	266	RESOLUÇÃO BCB Nº 266, DE 25 DE NOVEMBRO DE 202...	alter circular estabelecer procedimento calcul...	0	1

Fonte: A autora (2023).

Após realizar o pré-processamento e executar todos os métodos de classificação, podemos visualizar os resultados na Tabela 1, que exhibe a classificação feita pelos algoritmos.

TABELA 1 - RESULTADO CLASSIFICAÇÃO ALGORTIMOS

Método	Classificação Correta	% Acertos
Random Forest	24	80,00%
Regressão Logística	23	76,67%
Naive Bayes	22	73,33%
KNN	19	63,33%

Fonte: A autora (2023).

Observa-se que ao aplicar os algoritmos treinados a uma nova base contendo 30 normas, o método *Random Forest* obteve o melhor desempenho, com uma taxa de acerto de 80%. Em seguida, a regressão logística e o *Naive Bayes* apresentaram taxas de acerto de 76,67% e 73,33%, respectivamente. Por outro lado, o método KNN obteve a pior performance, com uma taxa de acerto de apenas 68,33%.

4.1 ANÁLISE E RISCO DE APLICAÇÃO

Embora o resultado obtido com o algoritmo Random Forest tenha alcançado aproximadamente 80% de taxa de acertos, o que pode ser considerado satisfatório para o tema e estudo subjetivo, no contexto de regulamentações que podem resultar em multas e sanções dos órgãos regulatórios, é desejável um nível de acurácia acima de 98% para reduzir o risco de classificação incorreta das normas para apenas 2%. Isso significa que, com a classificação atual, ainda existe um risco de 20% de classificação incorreta. Embora seja um resultado razoável para a situação atual, mostra que a classificação de normas é possível, mas há espaço para melhoria.

5 CONSIDERAÇÕES FINAIS

O número de normas, incluindo as Resoluções e Instruções Normativas BCB, tem crescido ao longo dos anos. No período de 2022 até maio de 2023, essas normas já representaram 76% do total de resoluções e instruções normativas publicadas entre 2020 e 2021.

Esse aumento significativo de normas tem exigido que as instituições de pagamentos se preparem e se adequem para cumprir com essas exigências regulatórias. Trata-se de uma tarefa que consome tempo e demanda esforços contínuos, visto que cada norma publicada pelo Banco Central do Brasil pode ou não se encaixar no contexto da instituição, sendo ela de pagamentos ou não. É preciso que exista a constante atualização e acompanhamento das normas pelas equipes de governança e *Compliance*. Diante do cenário das necessidades regulatórias, o uso de processamento de linguagem natural em conjunto com técnicas de aprendizado de máquina se apresenta como uma alternativa viável para a classificação de resoluções e instruções normativas do Banco Central do Brasil (BCB).

O objetivo geral deste trabalho foi investigar a viabilidade da aplicação de técnicas de Processamento de Linguagem Natural (PNL) e Aprendizado de Máquina na análise e verificação da aplicabilidade de normas do Banco Central do Brasil em instituições de pagamento, testando as principais técnicas de AM e avaliando as melhores performances. O foco do estudo foi avaliar essa possibilidade no contexto geral de qualquer IP, sem considerar suas especificidades individuais.

Após a implementação dos métodos, foi constatada a viabilidade da aplicação das técnicas propostas, porém com possíveis riscos na sua aplicação. Este trabalho demonstrou a possibilidade de classificar as normas atingindo uma acurácia de 80% com o método *Random Forest*, entretanto, por se tratar de n regulatórios o número de acertos deve ser pelo menos 98%, diminuindo o impacto de risco em menos de 2% em caso de normas classificadas erroneamente, que podem acarretar multas por parte do Banco Central do Brasil. Os resultados obtidos na classificação foram validados por dois profissionais da área de *Compliance*, responsáveis pela análise e estudo das normas do BCB.

Os resultados obtidos são considerados satisfatórios dentro do contexto geral, cumprindo o objetivo proposto. No entanto, abre-se espaço para futuros estudos visando melhorar os métodos utilizados, ampliar a base de dados e aumentar a acurácia dos resultados. A contribuição desse estudo pode servir como ponto de partida para projetos futuros e aprofundamento do uso de aprendizado de máquina em análises mais subjetivas relacionadas às regulamentações dos órgãos brasileiros.

5.1 CONTINUIDADE DO ESTUDO

Para futuros trabalhos, recomenda-se a ampliação do escopo para incluir outras categorias normativas, como resoluções e instruções do CNM, COREMEC, Circulares e Cartas Circulares. Essa expansão proporcionará uma análise mais abrangente e completa das normas. Ao ampliar o escopo, será possível aumentar o volume de dados, o que terá um impacto positivo no aprimoramento da performance das técnicas utilizadas, proporcionando resultados mais robustos e precisos.

Além disso, ao aumentar a especificidade em relação aos atributos considerados na avaliação, seria possível realizar uma classificação mais precisa e específica para cada contexto das instituições de pagamento. Isso significa que o modelo poderia levar em conta características particulares de cada instituição, levando a uma classificação mais adequada e personalizada. Tudo isso visando alcançar o menor risco possível na classificação das normas, para evitar sanções e multas por parte dos órgãos reguladores.

REFERÊNCIAS

ANCHIÊTA *et al.* PNL: Das técnicas tradicionais aos modelos de deep learning. In: BEZERRA *et al.* **Minicursos da ERCEMAPI 2021**. Porto Alegre: Ercemapi, 2021. p. 1-85. Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/79/344/603-1>. Acesso em: 10 jun. 2023.

BANCO CENTRAL DO BRASIL. **Instituições de pagamento e seus modelos de negócio**. 2020. Disponível em: https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE088_Instituicoes_de_pagamento_e_seus_modelos_de_negocio.pdf. Acesso em: 07 fev. 2023.

BANCO CENTRAL DO BRASIL. **O que é o Banco Central?**. 2023. Disponível em: <https://www.bcb.gov.br/meubc/faqs/p/o-que-e-o-banco-central>. Acesso em: 07 fev. 2023.

BANCO CENTRAL DO BRASIL. **O que são fintechs?**. 2023. Disponível em: <https://www.bcb.gov.br/meubc/faqs/p/o-que-sao-fintechs>. Acesso em: 07 fev. 2023.

BANCO CENTRAL DO BRASIL. **Resolução BCB nº 81, 25 de março de 2021**. Disciplina os processos de autorização relacionados ao funcionamento das instituições de pagamento e à prestação de serviços de pagamento por parte de outras instituições autorizadas a funcionar pelo Banco Central do Brasil. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=81>. Acesso em 01 jun. 2022.

BANCO CENTRAL DO BRASIL. **Resolução BCB nº 80, de 25 de março de 2021**. Disciplina a constituição e o funcionamento das instituições de pagamento, estabelece os parâmetros para ingressar com pedidos de autorização de funcionamento por parte dessas instituições e dispõe sobre a prestação de serviços de pagamento por outras instituições autorizadas a funcionar pelo Banco Central do Brasil. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/exibenormativo?tipo=Resolu%C3%A7%C3%A3o%20BCB&numero=80>. Acesso em 01 jun. 2022.

BEGNINI, Débora de Lima et al. Relação entre valor de mercado e *Compliance* anticorrupção. **Pensamento Contemporâneo em Administração**, Rio de Janeiro, v. 16, n. 2, p. 134-149, ago. 2022. Disponível em: <https://periodicos.uff.br/pca/article/view/53077/32696>. Acesso em: 10 fev. 2023.

BRASIL. **Lei nº 12.865, de 09 de outubro**
COELHO, Alexandre Zavaglia. **A ciência de dados e a inteligência artificial no Direito em 2018**: parte I. Parte I. 2019. Disponível em: <https://www.conjur.com.br/2019-jan-01/zavaglia-ciencia-dados-inteligencia-artificial-direito>. Acesso em: 10 fev. 2023.

CONGRESSO NACIONAL. **Termo:** resolução. Resolução. 2023. Disponível em: <https://www.congressonacional.leg.br/legislacao-e-publicacoes/glossario-legislativo/-/legislativo/termo/resolucao>. Acesso em: 08 fev. 2023.

DALE, R. Classical approaches to natural language processing. In: **Handbook of natural language processing**. 2. ed. Boca Raton: Chapman & Hall/CRC, 2010. **de 2013**. Dispõe sobre os arranjos de pagamento e as instituições de pagamento integrantes do sistema de pagamentos brasileiro (spb). Disponível em: <https://legislacao.presidencia.gov.br/atos/?tipo=LEI&numero=12865&ano=2013&ato=af1MzYU50MVpWT376>. Acesso em: 13 fev. 2023

DISTRITO. **Fintech Report**. 2022. Disponível em: <https://distrito.me/materiais-ricos/>. Acesso em: 05 jun. 2023.

FACELI, Katti *et al.* **Inteligência Artificial:** uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

FARIA, Emerson. **Fintechs de crédito e intermediários financeiros:** uma análise comparativa de eficiência. 2018. 123 f. Dissertação (Mestrado) - Curso de Administração, Universidade de São Paulo, São Paulo, 2018. Disponível em: <https://www.teses.usp.br/teses/disponiveis/12/12142/tde-07012019-112337/publico/CorrigidoEmerson.pdf>. Acesso em: 06 jun. 2023.

FEBRABAN (Federação Brasileira de Bancos). **BOAS PRÁTICAS DE COMPLIANCE**. 2018. Disponível em: https://cmsarquivos.febraban.org.br/Arquivos/documentos/PDF/febraban_manual_compliance_2018_2web.pdf. Acesso em: 07 jun. 2023.

FERREIRA, Hugo Honda. **Processamento de Linguagem Natural e Classificação de textos em Sistemas Modulares**. 2019. 71 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade de Brasília, Brasília, 2019. Disponível em: https://bdm.unb.br/bitstream/10483/25114/1/2019_HugoHondaFerreira_tcc.pdf. Acesso em: 13 fev. 2023.

FINTECHLAB. **8ª edição do Radar Fintechlab registra mais de 600 iniciativas**. 2019. Disponível em: <https://fintechlab.com.br/index.php/2019/06/12/8a-edicao-do-radar-fintechlab-registra-mais-de-600-iniciativas/>. Acesso em: 11 fev. 2023.

GASPARETTO, William Virgilio. **Machine Learning para identificação de NFS-e**. 2021. 24 f. Monografia (Especialização) - Curso de Inteligência Artificial, Universidade Federal do Paraná, Curitiba, 2021. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/76250/R%20-%20E%20-%20WILLIAM%20VIRGILIO%20GASPARETTO.pdf?sequence=1&isAllowed=y>. Acesso em: 14 fev. 2023.

GOMES, Lucas Moreira *et al.* **Línguas naturais e máquinas artificiais:** Aplicação de técnicas de mineração de texto para a classificação de sentenças judiciais brasileiras. 2020. IPEA. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/10327/1/td_2612.pdf. Acesso em: 14 fev. 2023.

GUARDADO, Fernanda. O papel do Banco Central do Brasil na Economia. 2022. 19 slides. Disponível em: https://www.bcb.gov.br/conteudo/home-ptbr/TextosApresentacoes/FG_Aula%20Magna%20PUC%202022_25.3.22.pdf. Acesso em: 14 fev. 2023.

PEIXOTO, Fabiano Hartmann. Direito e Inteligência Artificial. **Coleção Inteligência Artificial e Jurisdição**. Volume 2. Brasília, 2020. Disponível em: <https://tinyurl.com/mrynjj82>. Acesso: 10 fev. 2023.

KLEIN, Vinicius; FARIAS, Rodolfo Rodrigues da Costa. **Instituições de pagamento e requisitos mínimos de patrimônio**: considerações sobre seu risco sistêmico. Revista Eletrônica Direito e Política, Programa de Pós-Graduação Stricto Sensu em Ciência Jurídica da UNIVALI, vº 17, nº 3, 3º quadrimestre de 2022. Disponível em: <https://periodicos.univali.br/index.php/rdp> -ISSN 1980-7791. Acesso em: 05 jun. 2023.

LAMBOY, Christian K. de. **Manual de Compliance**. 2018. Disponível em: <https://viaetica.com/images/Manual-de-Compliance-Amostra.pdf>. Acesso em: 10 jun. 2023.

MARQUES, Frank Borges *et al.* Cadê o Banco que estava aqui? O Impacto dos Bancos Digitais no Mercado Brasileiro. **Journal Of Information Systems And Technology Management**, Uberlândia, v. 19, n. 202219002, p. 1-16, mar. 2022. Disponível em: <https://www.scielo.br/j/jistm/a/cwsXS5GScbCvYjYsCDzrQ3b/?format=pdf&lang=pt>. Acesso em: 07 jun. 2023.

MASTELLA, Juliana Obino. **UMA METODOLOGIA USANDO AMBIENTES PARALELOS PARA OTIMIZAÇÃO DA CLASSIFICAÇÃO DE TEXTOS APLICADA A DOCUMENTOS JURÍDICOS**. 2020. 68 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2020. Disponível em: https://tede2.pucrs.br/tede2/bitstream/tede/9908/2/JULIANA%20OBINO%20MASTELLA_DIS.pdf. Acesso em: 13 fev. 2023.

MCKINSEY. **Digital finance for all**: powering inclusive growth in emerging economies. POWERING INCLUSIVE GROWTH IN EMERGING ECONOMIES. 2016. Disponível em: <https://www.mckinsey.com/~media/mckinsey/featured%20insights/Employment%20and%20Growth/How%20digital%20finance%20could%20boost%20growth%20in%20emerging%20economies/MGI-Digital-Finance-For-All-Executive-summary-September-2016.ashx>. Acesso em: 05 jun. 2023.

MONARD, M. C.; BARANAUKAS, J. A. **Aplicações de Inteligência Artificial: uma visão geral**. São Carlos: Instituto de Ciências Matemáticas e de Computação de São Carlos, 2000. Disponível em: <https://dcm.ffclrp.usp.br/~augusto/publications/2000-laptec.pdf>. Acesso em: 11 fev. 2023.

MONTAÑO, R. A. N. **Aplicação de Técnicas de Aprendizado de Máquina na Mensuração Florestal**. 2016. 102 f. Tese (Doutorado) - Curso de Ciência da Computação, Universidade Federal do Paraná, Curitiba, 2016. Disponível em: <https://acervodigital.ufpr.br/bitstream/handle/1884/45346/R%20-%20T%20-%20RAZER%20ANTHOM%20NIZER%20ROJAS%20MONTANO.pdf?sequence=1&isAllowed=y>. Acesso em: 15 fev. 2023.

MOSER, Gustavo Vicente Barroso. **ANÁLISE DE SIMILARIDADE ENTRE TF-IDF E MODELOS CONTEXTUALIZADOS DE LINGUAGEM BASEADOS EM TOKENS**. 2021. 98 f. Tese (Doutorado) - Curso de Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2018. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/243553/monografia.pdf?sequence=1&isAllowed=y>. Acesso em: 10 jun. 2023.

PINTO, Andréia Aparecida Bressani. **Fintechs: o futuro dos serviços financeiros no Brasil**. 2018. 33 f. Monografia (Especialização) - Curso de Gestão Financeira, Universidade Tecnológica Federal do Paraná, Curitiba, 2018. Disponível em: https://repositorio.utfpr.edu.br/jspui/bitstream/1/19580/1/CT_GESFIN_III_2018_02.pdf. Acesso em: 08 jun. 2023.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial**. 3. ed. Rio de Janeiro: Elsevier, 2013. 1324 p. Disponível em: <https://www.cin.ufpe.br/~qtsa/Periodo/PDF/4P/SI.pdf>. Acesso em: 11 fev. 2023.

SANTOS, Keila Barbosa Costa dos. **Categorização de Textos por Aprendizagem de Máquina**. 2018. 86 f. Dissertação (Mestrado) - Curso de Modelagem Computacional, Universidade Federal de Alagoas, Maceió, 2019. Disponível em: <https://keilabcs.github.io/publications/dissertacao.pdf>. Acesso em: 10 jun. 2023.

SANTOS, Renato Almeida dos. Compliance como ferramenta de mitigação e prevenção da fraude organizacional. In: CONCURSO DE MONOGRAFIAS DA CGU - 2011, 6., **2011, Brasília. Prevenção e combate à corrupção no Brasil**. Brasília: Esaf, 2011. p. 161-229. Disponível em: <https://repositorio.enap.gov.br/handle/1/5687>. Acesso em: 07 jun. 2023.

SCHWAB, Klaus. **A quarta revolução industrial**. São Paulo: Edipro, 2016.

SILVA, J. A. S.; MAIRINK, C. H. P. Inteligência artificial: aliada ou inimiga. **LIBERTAS: Rev. Ciências Sociais Aplicadas**, Belo Horizonte, v. 9, n. 2, p. 64-85, ago./dez. 2019. Disponível em: <https://famigvirtual.com.br/famig-libertas/index.php/libertas/article/view/247>. Acesso em: 10 fev. 2023.

SILVA, Lucas Santana. **O papel da Fintech no sistema financeiro nacional**. 2020. 71 f. TCC (Graduação) - Curso de Ciências Econômicas, Universidade Federal da Bahia, Salvador, 2020. Disponível em: <https://repositorio.ufba.br/bitstream/ri/33540/1/Tcc%20Lucas%20Santana%20Silva.pdf>. Acesso em: 07 jun. 2023.

Zhang, Aston *et al.* **Dive into Deep Learning**. 2023. Disponível em: <https://d2l.ai/index.html>. Acesso em: 12 fev. 2023.