

DOI: 10.5748/19CONTECSI/PSE/DSC/7139

PROCESSO PARA AVALIAÇÃO DE PERFORMANCE DE INVESTIMENTOS FINANCEIROS ESG BRASILEIROS COM SCRAPY E MACHINE LEARNING

Claudene Oliveira Gonçalves ; <https://orcid.org/0000-0002-3733-2784>
IPT - Instituto de Pesquisas Tecnológicas

Vagner Luiz Gava ; <https://orcid.org/0000-0001-5965-957X>
IPT - Instituto de Pesquisas Tecnológicas



PROCESS FOR PERFORMANCE EVALUATION OF BRAZILIAN ESG FINANCIAL INVESTMENTS WITH SCRAPY AND MACHINE LEARNING

ABSTRACT

The adoption and growth of ESG Investments in the financial markets of different countries in recent years demonstrate a strong tendency of companies and investors to act with investments that aim at both profit and long-term sustainability. Developing countries like Brazil currently have timid investments compared to European and North American countries, but have the potential to benefit given the social challenges and economic growth with sustainability that affect them.

Such demand highlights the benefits and challenges that involve the machine learning pipeline to support the growth and development of ESG investments in a plural and global way, especially in countries with a natural language other than English, involving data tracking, creation of metrics and the prediction of long-term return on investment. The objective of this article is to propose a process that covers the collection, organization, training and evaluation phases of ESG investments focused on Brazil using machine learning models based on a systematic literature review. The results obtained showed that the systematic review presents relevant solutions to mitigate the complexity of ESG financial investments and shows that the set of neural network models such as LSTM, which is also part of the machine learning universe, are the most suitable for both sentiment analysis, classification and prediction of ESG investments.

Keywords: ESG, sustainable investing, machine learning, deep learning, sentiment analysis

PROCESSO PARA AVALIAÇÃO DE PERFORMANCE DE INVESTIMENTOS FINANCEIROS ESG BRASILEIROS COM SCRAPY E MACHINE LEARNING

RESUMO

A adoção e crescimento dos Investimentos **ESG** no mercado financeiros de diferentes países nos últimos anos demonstram um forte tendência das empresas e investidores em atuar com investimentos que visam tanto o lucro quanto sustentabilidade a longo prazo. Países em desenvolvimento como o Brasil, atualmente possuem investimento tímidos comparados aos países Europeus e da América do Norte, porém apresentam potencial de se beneficiarem dado os desafios sociais e de crescimento econômico com sustentabilidade que os afetam.

Tal demanda evidencia o benefício e desafios que envolvem o *pipeline* de *machine learning* para suportar o crescimento e desenvolvimento dos investimentos ESG de maneira plural e global, especialmente em países de língua natural diferente do inglês, envolvendo o rastreamento dos dados, criação de métricas e a predição do retorno dos investimentos a longo prazo. O objetivo deste artigo é propor um processo que cubra a fases de coleta, organização, treinamento e avaliação dos investimentos ESG focados no Brasil utilizando modelos de *machine learning* com base na revisão sistemática da literatura. Os resultados obtidos mostraram que a revisão sistemática apresenta soluções relevantes para mitigar a complexidade dos investimentos financeiros ESG e evidencia que o conjunto de modelos de redes neurais como LSTM, que também faz parte do universo de *machine learning* são os mais indicados tanto para análise de sentimentos, classificação e predição dos investimentos ESG.

Palavras-chave: ESG, investimentos sustentáveis, machine learning, deep learning, análise de sentimentos

1. INTRODUÇÃO

1.1 APRESENTAÇÃO E CONTEXTUALIZAÇÃO

Investimentos ESG (E - ambiental do inglês *environment*, S - social, e G - governança) são investimentos financeiros sustentáveis e com responsabilidade social, que também envolve outras denominações como *Corporate Social responsibility* (CSR). ESG é um critério utilizado para medir e avaliar práticas e riscos ambientais, sociais e de governança relacionadas às empresas e setores da indústria (MAURICIO AMORMINO JUNIOR, 2021).

Na Figura 1, as estratégias de investimentos ESG são apresentadas com relação à transparência, medição e relatórios, comprometimento, impacto e intencionalidade. A primeira estratégia é aplicada na formação de investimentos a partir do rastreamento negativo ou exclusão de certos setores ou empresas, com os níveis de medição, comprometimento, impacto e intencionalidade mais baixos. Na sequência, a segunda estratégia visa ao rastreamento positivo, com foco nas empresas ou setores de melhor desempenho ESG. A estratégia com base na integração completa de ESG foca em empresas onde as questões ambientais, sociais e de governança são integradas como parte essencial da análise do investimento. Os investimentos temáticos e de impacto não giram em torno do impacto externo que as empresas produzem em diversos setores da sociedade.

De acordo com relatório Global Sustainable Investments Review publicada em 2020 pela Global Sustainable Alliance (HAND et al., 2020), os investimentos sustentáveis no mundo, alcançaram a marca de US\$ 35.3 trilhões no início de 2020, um crescimento de 15% em relação aos dois últimos anos (2018-2020) com Estados Unidos e Europa representando mais de 80% destes investimentos. O relatório também informa que ESG integrada é a abordagem com maior valor alocado, estimado em US\$ 25 trilhões. Por sua vez, o relatório Annual Impact Investor Survey 2020 realizado pela GIIN - Global Impact Investing Network (HAND et al., 2020), analisou que os investimentos de impacto atingiram US\$ 715 bilhões.

No Brasil, a Bolsa de Valores B3 (B3 S.A., [s.d.]) declarou que ao final de 2020, os títulos temáticos ESG totalizaram R\$ 6 bilhões. Para os investimentos de impacto, o Brasil aparece como o líder da América Latina neste seguimento, com US\$ 785 milhões alocados entre 2018 e 2019, de acordo com o relatório realizado com países da América Latina pela ANDE - Aspen Network of Development Entrepreneurs, elaborado por Hume et al., 2020. Cerca de 28% destes investimentos estão direcionados a empresas do setor de manufatura, seguido por agricultura e alimentos (13%) e saúde (11%).

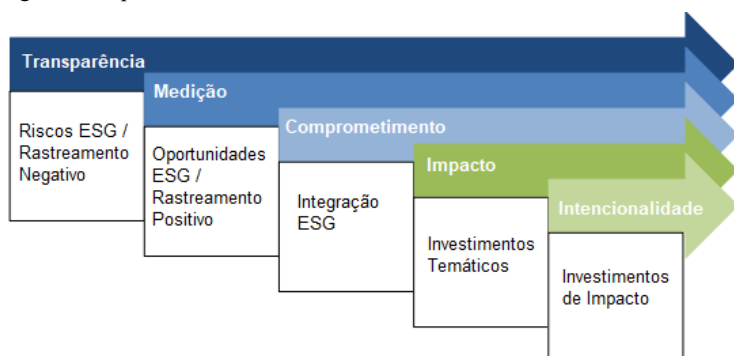
Os números do Brasil são tímidos comparados aos dos Estados Unidos e Europa. Entretanto, dado os diversos problemas socioambientais. No entanto, os atores do mercado financeiro acreditam ser este um excelente celeiro de oportunidades para os investimentos ESG. Esta análise pode ser verificada nos estudos de caso presente no GUIA ASG elaborado pela ANBIMA - Associação Brasileira das Entidades dos Mercados Financeiro e de Capitais (ANBIMA, 2020)

A tendência dos investimentos de impacto e ESG é que ambos cresçam ainda mais nos próximos anos, impulsionados pela população das gerações milênios (67%) e X (44%), que investem de acordo com os valores sociais, políticos e ambientais e acreditam que investimentos de impacto socioambientais podem ter retornos financeiros de acordo com o mercado. Além disso, este perfil de investidor entende que este tipo de investimento pode influenciar mudanças climáticas e combater a pobreza (INDERST; STEWART, 2018).

Conforme a demanda de investimentos ESG cresce, aumenta a necessidade de soluções tecnológicas envolvendo coleta de dados e *machine learning*. Tais técnicas

combinadas podem garantir a rastreabilidade, consistência, confiabilidade e transparência dos dados, informações sobre os principais riscos e tendências, permitindo aos investidores e corporações mitigar seus riscos e, ao mesmo tempo, otimizar o desempenho em nível de portfólio e entidade (KALYANI INAMPUDI, 2020).

Figura 1 - Tipos de Investimentos ESG



Fonte: Kirppu (2019) adaptado pela autora

Ao aplicar *machine learning* como uma tecnologia de suporte a estes investimentos, observa-se desafios para capturar dados de investimentos financeiros em ambientes diversos como observado no trabalho de Manuel et al., (2018), para rastrear e validar se os investimentos estão de fato alinhados com seu propósito, assim como combinar os dados qualitativos e quantitativos para prever o retorno e a performance desses investimentos aliados aos seus riscos.

No framework de *machine learning* (Figura 2) exposto por (Fischer et al., (2021), o fluxo de trabalho passa pela coleta de dados via *web scrappy*, processamento em linguagem natural dos dados para análise de sentimentos e *machine learning* para modelagem das soluções preditivas.

Entretanto, ao analisar os investimentos de impacto e ESG, dos três fundos de investimentos do Brasil no Quadro 1 extraído de Naiara, Bertão (2021), observa-se que os negócios são diversos. A captura dos dados estruturados e não estruturados é uma atividade extensa e desafiadora devido à complexidade e variedade de fontes a serem consideradas (MANUEL et al., 2018), como, por exemplo, buscar dados das empresas signatárias aos “Princípios de Empoderamento das Mulheres”, buscar dados dos negócios do mercado de água ou verificar as iniciativas de impacto social da HubSocial.

No contexto da análise de sentimentos dos dados de investimentos financeiros dos países em desenvolvimento por meio de processamento de linguagem natural (NLP), observa-se que os modelos foram amplamente desenvolvidos para o inglês, porém pouco estudadas para textos em outras línguas como o português (ROCHA; CARDOSO, 2018), sendo este um ponto a ser tratado para dados financeiros que possuem linguagem natural diferente do inglês.

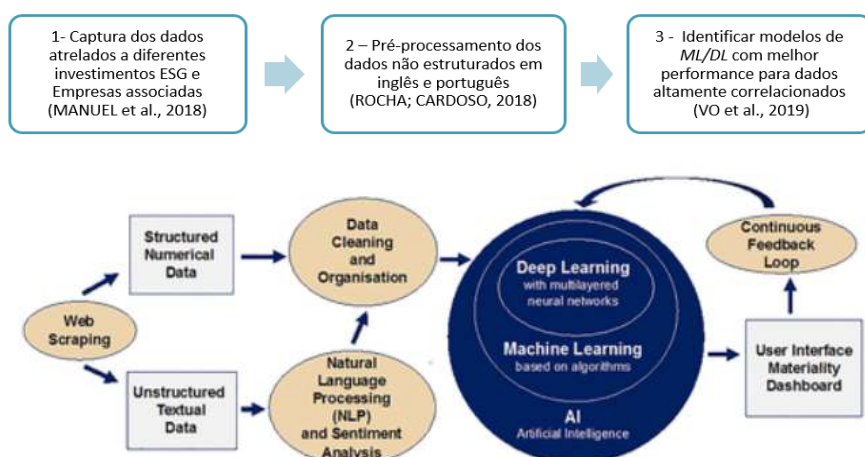
Além disso, o trabalho de VO et al. (2019) analisa que séries históricas do mercado financeiro possuem, por natureza, dados altamente correlacionados, impactando em problemas de ruídos em algoritmos de redes neurais e/ou, adicionalmente, impactando a performance dos algoritmos de *machine learning* tradicionais, como *Random Forest Regressor*

Quadro 1 – Fundos de Investimentos de Impacto e ESG

Fundos	Estratégia
BB Ações Equidade (*)	Ações de empresas brasileiras e algumas estrangeiras que sejam signatárias aos "Princípios de Empoderamento das Mulheres"
BNP Paribas Blackrock Estratégias ESG FIA IE	Investir de forma ativa e passiva em empresas no mundo que tenham seus negócios no mercado de água, especialmente em três frentes: infraestrutura, tratamento e serviços
NAU Capital Impacto FIQ de FIA	Fundo de fundos de ações, cuja taxa de administração é revertida à Hub Social, instituição que dá suporte às iniciativas de impacto social

Fonte: Naiara Bertão, (2021) adaptado pela autora

Tabela formatada

Figura 2- Framework de Projeto com *Machine Learning*

Fonte: Fischer I. et al (2021) adaptada pela autora

O objetivo deste trabalho é realizar uma revisão sistemática da literatura que permita identificar técnicas de coleta, organização, treinamento e análise de investimentos financeiros ESG relacionados a empresas brasileiras utilizando *machine learning* e com isso, propor um processo que minimize os desafios apontados.

Com este objetivo, foi formulada uma pergunta de pesquisa principal e duas secundárias, que norteiam este artigo:

- i. *Como obter um processo baseado nas melhores técnicas de scrapy e machine learning mapear dados e prever o retorno dos investimentos ESG do Brasil?*
 - a. *Quais as melhores técnicas de machine learning para mapear os dados e prever o retorno dos investimentos ESG brasileiros? e*
 - b. *Como o resolver os desafios de processamento de linguagem natural (NLP) para informações em linguagens diferentes do Inglês?*

As seções seguintes estão organizadas da seguinte maneira: Na seção 2 é apresentada a fundamentação teórica que servirá de base para o entendimento do processo proposto. Em seguida, a seção 3 apresenta o método de pesquisa utilizado, cujos resultados serão analisados na seção 4. Por fim, na seção 5 são apresentadas as conclusões deste trabalho, bem como suas limitações e sugestões para estudos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Scrapy

Scrapy ou *web scrapy* é o termo utilizado para definir uma estrutura de extração dos dados que via processá-los e salvá-los em arquivo ou base de dados. A Figura 3 exemplifica o processo de scrapy, no qual é possível por meio de API ou bibliotecas de linguagens como Python que encapsulam os métodos HTTP Request e Response, acessar os dados que estão disponíveis online e posteriormente salvá-los em uma base interna (GANEGEDARA, 2019).

Figura 3 - Fluxo de Web Scrapy



Fonte: Ganegedara (2019)

2.1 Análise de Sentimentos e Natural Language Process (NLP)

Análise de sentimentos: utiliza processamento de linguagem natural (NLP), para avaliar o sentimento de uma frase ou texto, com base em cálculos matemáticos que definem a polaridade em positivo ou negativo ou mesmo se o texto é objetivo ou subjetivo. NLP também é utilizado para classificação e agrupamento de textos. A Figura 4 demonstra o fluxo de aplicação de NLP clássico ou utilizando *Deep Learning*.

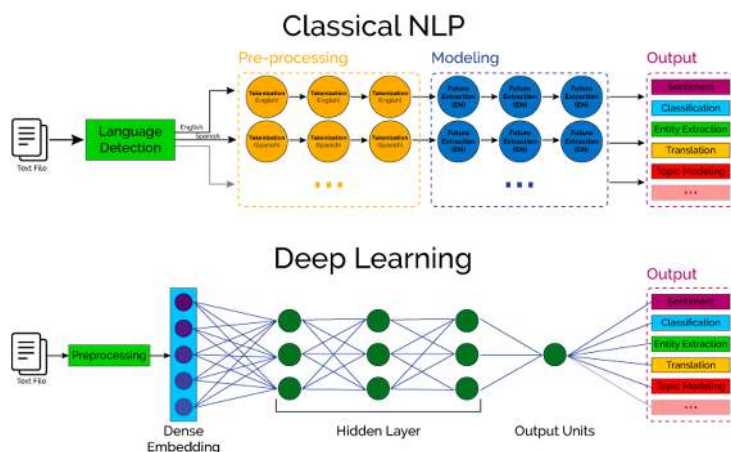
Tokenizer x Embeddings: são algoritmos NLP que transformam texto para números. *Tokenizer* converte um texto em uma lista de inteiros e *Embeddings* converte a lista de inteiros em uma lista de vetores. Exemplos de *Embeddings*:

- Word2Vec e GloVe: modelos de incorporação preditiva
- FastText: permite algoritmos de aprendizado supervisionado e não supervisionado
- BERT - Modelo baseado em redes neurais que analisa o contexto à esquerda e à direita da palavra.
- BERTIMBAU: Análise de sentimentos em português
- FINBERT: Análise de sentimentos em Financeiro

2.1 Redes Neurais

Redes Neurais é uma estrutura computacional que funciona de maneira semelhante à dos neurônios biológicos (Figura 5). Possui a capacidade de aprender os padrões subjacentes dos dados. Exemplos de Modelos de Redes Neurais:

Figura 4 - Fluxo de aplicação de NLP

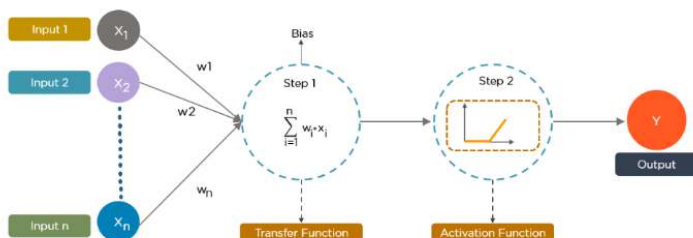


Fonte: Kaur, (2022)

- RNN: cria ligações entre as unidades para formar um ciclo direcionado
- MLP: Perceptron multicamadas
- CNN: Rede Neural Convolutacional
- LSMT: Memória de Curto e Longo Prazo, que ajudam a rede a aprender dados temporais
- BI-LSMT: Similiar a LMST, porém está ao passado ao futuro
- BPN: calcula os valores de entrada dos retornos na camada oculta

Figura 5 - Estrutura de Redes Neurais

- The input layer
- The hidden layer(s)
- The output layer



Fonte: Biswal, (2022)

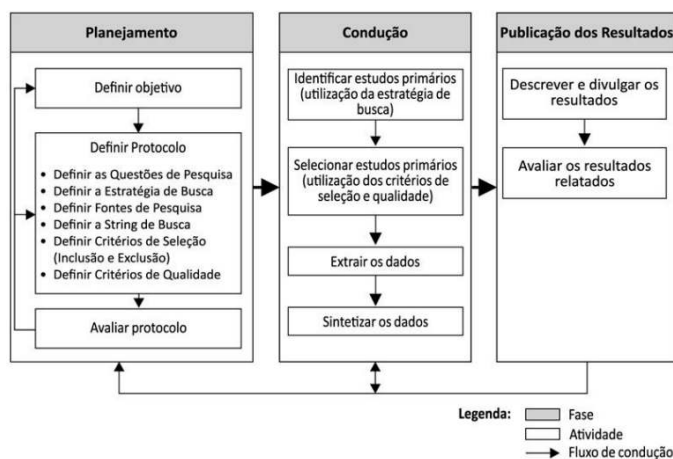
3. MÉTODO

3.1 REVISÃO SISTEMÁTICA

Segundo (Felizardo et al., (2017), a revisão sistemática é um processo que visa analisar e interpretar todas as pesquisas relevantes à um determinado estudo com o objetivo de apresentar uma avaliação justa utilizando uma metodologia confiável, procura reduzir o viés de uma revisão informal. A revisão sistemática envolve as fases de planejamento, condução e análise dos dados para publicação dos resultados.

Na fase de planejamento é identificado a necessidade da revisão da literatura e a elaboração do protocolo da pesquisa. Na fase de condução é a realizado a seleção e avaliação dos estudos. Por fim, na fase de publicação dos resultados ocorre a seleção dos estudos finais, a apresentação dos resultados e elaboração da conclusão (Figura 6)

Figura 6 Fases do processo da Revisão Sistemática



Fonte: Felizardo et al. (2017)

3.2 PROTOCOLO DA REVISÃO SISTEMÁTICA

A fase de planejamento visa identificar a real necessidade para executar uma RS e definir o protocolo da revisão com a especificação da pergunta de pesquisa, estratégia, fontes de busca e critérios de seleção dos trabalhos (FELIZARDO et al., 2017)

O protocolo adotado nesta pesquisa está detalhado abaixo e segue o modelo descrito em Felizardo *et al.* (2017):

Objetivo: Verificar quais as melhores técnicas de *machine learning* e *scrapy* para: classificar os sentimentos de notícias (em português) das empresas associadas aos investimentos ESG do Brasil; prever tais investimentos, com base na análise de sentimentos e nas melhores práticas de *machine learning* tradicionais e/ou redes neurais identificadas.

Questão de pesquisa: *Como obter um processo baseado nas melhores técnicas de scrapy e machine learning para mapear dados e prever o retorno dos investimentos ESG do Brasil?"*

A revisão da literatura também leve em conta as questões secundárias da pesquisa:

- a. *Quais as melhores técnicas de machine learning para mapear os dados e prever o retorno dos investimentos ESG brasileiros? e*
- b. *Como o resolver os desafios de processamento de linguagem natural (NLP) para informações em linguagens diferentes do Inglês?*

A pergunta de pesquisa foi formulada com base nos critérios PICO (*Population, Intervention, Comparison, Outcome*) detalhado em Felizardo *et al.* (2017), O Quadro 1 apresenta os critérios PICO aplicados a esta pesquisa:

Quadro 1 - Critérios PICO aplicados à pesquisa

Critério PICO	Descrição do critério	Aplicação neste trabalho
Population (P)	Papel ou grupo que será observado	Investimentos financeiros ESG relacionados as empresas brasileiras
Intervention (I)	Objeto da investigação	Processo baseado em técnicas de coleta de dados e <i>machine learning</i> para prever a pontuação e o retorno dos investimentos financeiros ESG
Comparison (C)	Objeto ou parâmetro de comparação	Artigos levantados e avaliados durante a revisão sistemática;
Outcome (O)	Resultados ou efeitos observados	Identificação dos melhores modelos e técnicas de coleta de dados e <i>machine learning</i> que permitam desenvolver o processo proposto.

Fonte: Elaborado pela autora

Palavras-chave: Machine learning, deep learning, reinforcement learning, genetic algorithm, artificial intelligence, heterogeneous information network, multivariate analytics, NLP, natural language, natural language processing, sentiment analysis, text analytics, predictive models, prediction, forecast, scrapy, crawler, data mining, ESG, ESG integration, sustainable investing, responsible investing, corporate social responsibility, socially responsible investing, social and governance investing, environment social and governance investing, emerging countries, Brazil, Brazilian, Portuguese

Idioma: Inglês

String de busca I: (("machine learning" OR "deep learning" OR "reinforcement learning" OR "genetic algorithm" OR "artificial intelligence" OR "Heterogeneous information network" OR "Multivariate analytics" OR "NLP" OR "natural language" OR "natural language processing" OR "Sentiment Analysis" OR "text analytics" OR "predictive models" OR "prediction" OR "predicting" OR "forecasting" OR "forecast" OR "scraping" OR "crawler" OR "crawling" OR "data mining") AND ("ESG" OR "ESG integration" OR "sustainable investing" OR "responsible investing" OR "corporate social responsibility" OR "socially responsible investing" OR "social and governance investing" OR "environment social and governance investing"))

A *string* de busca acima, não retornou trabalhos relacionados a investimentos financeiros ESG que utilizam principalmente linguagem de processamento natural para investimentos brasileiros. Portanto, foi necessário desenvolver uma segunda *string* de busca, contemplando investimentos tradicionais além dos investimentos ESG, com foco no Brasil.

String de busca II: ("NLP" OR "natural language" OR "natural language processing" OR "Sentiment Analysis") AND ("Brazil" OR "Brazilian" OR "Portuguese") AND ("ESG"

OR "ESG integration" OR "sustainable investing" OR "responsible investing" OR "corporate social responsibility" OR "socially responsible investing" OR "social and governance investing" OR "environment social and governance investing" OR "impact investing", "financial" OR "stock", "market"))

Critérios para seleção de fontes: as fontes utilizadas neste trabalho são bibliotecas digitais indexadas de conteúdos científicos, que permitem obter os textos na íntegra, adaptação da *string* de busca e exportação dos resultados

Lista das fontes de busca:

- Scopus (<https://www.elsevier.com/pt-br/solutions/scopus>)
- ACM Digital Library (<https://dl.acm.org>)
- Web of Science (<https://clarivate.com/products/web-of-Science/databases/>)
- IEEEExplore (<http://ieeexplore.ieee.org/Xplore/>)

Estratégia de busca nas fontes:

As buscas automáticas por meio das *strings* de busca (I e II) nas bases eletrônicas, no período entre 2018 e 2022, ou seja, abrangendo o resultado da busca no período dos cinco últimos anos de publicações.

Critérios para inclusão de estudos:

- 1) O estudo aborda ML/DL para predição dos investimentos financeiros/ESG
- 2) O estudo aborda *scrapy*/rastreamento de dados para investimentos financeiros/ESG
- 3) O estudo apresenta informações de bases ESG
- 4) O estudo aborda investimentos ESG nos países emergentes/em desenvolvimento
- 5) O estudo cobre tópicos de NLP/análise de sentimentos de modo geral
- 6) O estudo cobre tópicos de NLP para conteúdo diferente de inglês
- 7) O estudo apresenta resultados de ML/DL/NLP/*scrapy* para investimentos financeiros/ESG
- 8) O estudo apresenta análise e predição de investimentos ESG
- 9) O estudo pode ser reproduzido por meio das técnicas e/ou códigos compartilhados

Critérios para exclusão de estudos:

- A) O estudo aborda ML/DL/NLP/*scrapy* para um contexto diferente de investimentos financeiros/ESG
- B) O foco do estudo não está investimentos financeiros/ESG com ML/DL/NLP/*scrapy*
- C) O estudo não aborda os resultados alcançados
- D) Estudo duplicado ou encontrado em outras bases de dados
- E) O estudo não está disponível para acesso por meio digital
- F) O idioma do estudo está diferente dos definidos (inglês e português)

Extração de dados

Os estudos selecionados na fase final, foram extraídos utilizando as mesmas perguntas dos critérios de inclusão conforme Figura 7.

Foi também para esta fase, um critério de pontuação para auxiliar na análise quantitativa e qualitativa dos dados extraídos, de tal forma que para cada pergunta analisada é possível atribuir uma nota ao dado extraído, onde a categoria “Altamente” equivale a 5 pontos e significa que o artigo possui o dado investigado com alta relevância para esta pesquisa. É possível que os dados extraídos em cada pergunta, também seja categorizado

como “Médio” equivalente a 3 pontos ou “Baixo” equivalente a 1 ponto. A categoria “Não” equivalente a 0 pontos, significa que o artigo não possui o dado investigado.

Figura 7 - Formulário de Extração de dados

Questions	
1) O estudo aborda ML/DL para predição dos investimentos financeiros/ESG	edit remove
2) O estudo aborda scraping/rastreamento de dados para investimentos financeiros/ESG	edit remove
3) O estudo apresenta informações de bases ESG	edit remove
4) O estudo aborda investimentos ESG nos países emergentes/em desenvolvimento	edit remove
5) O estudo cobre tópicos de NLP/análise de sentimentos de modo geral	edit remove
6) O estudo cobre tópicos de NLP para conteúdo diferente de inglês	edit remove
7) O estudo apresenta resultados de ML/DL/NLP/scraping para investimentos financeiros/ESG	edit remove
8) O estudo apresenta análise e predição de investimentos ESG	edit remove
9) O estudo pode ser reproduzido através das técnicas e/ou códigos compartilhados	edit remove

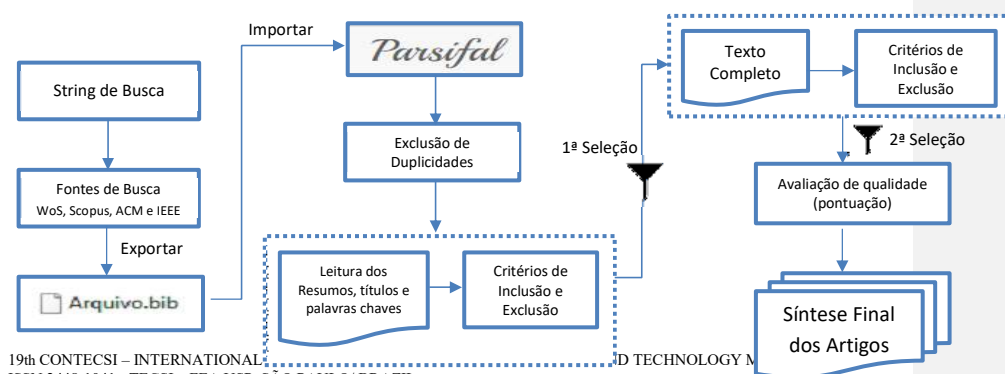
Fonte: Elaborado pela autora

Estratégia para seleção dos estudos:

A Figura 8 exibe o processo de revisão sistemática utilizada neste artigo. Inicialmente, é aplicado a pesquisa automática nas fontes bibliográficas, utilizando as *strings* de busca documentadas no protocolo. Os resultados são exportados direto das fontes em arquivos no formato BibTex. Estes arquivos são importados na ferramenta online Parsif.al (<https://parsif.al/about/>). Esta ferramenta foi projetada para apoiar pesquisadores na realização de revisões sistemáticas da literatura e possibilita a identificação e remoção automática de arquivos duplicados, a configuração e atribuição dos critérios de inclusão e exclusão, a criação do formulário de extração e categorias com pontuações para avaliar a qualidade dos artigos.

Na primeira seleção dos arquivos é aplicado os critérios de inclusão e exclusão para cada um deles, com base na leitura dos resumos, títulos e palavras chaves. Na sequência, temos a leitura do texto completo e a aplicação dos critérios de seleção. Ao final do processo é extraído a avaliação qualitativa com a pontuação dos artigos identificados na segunda seleção e a suas respectivas sínteses.

Figura 8 - Processo utilizado na revisão sistemática



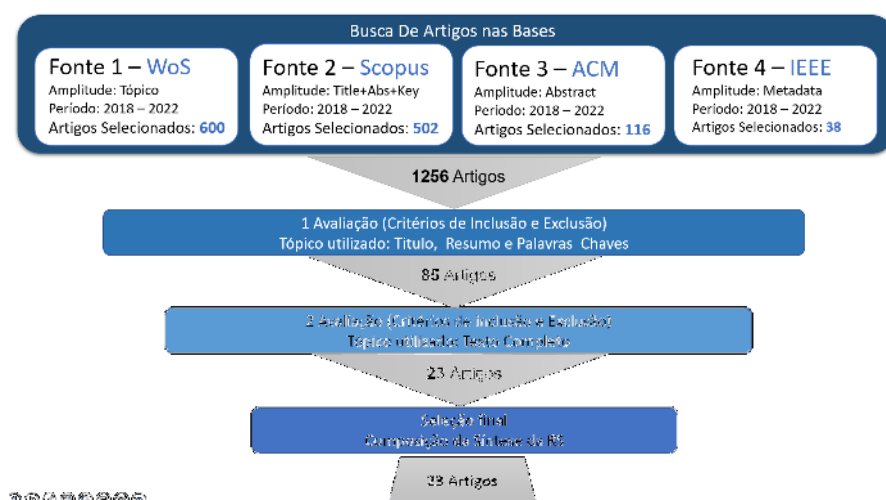
Fonte: Elaborado pela autora

4. RESULTADOS

4.1 CONDUÇÃO DA REVISÃO SISTEMÁTICA

A revisão sistemática foi conduzida em outubro de 2022. A Figura 9 representa o resumo da condução revisão sistemática.

Figura 9 - Resultado da Condução da Revisão Sistemática



Na segunda avaliação, foi realizada a leitura completa 85 artigos selecionados na etapa anterior e novamente os critérios de inclusão e exclusão foram considerados para cada artigo. Durante a leitura, foi considerado os dados previamente levantados no mapeamento do formulário de extração. Após a leitura completa dos artigos, foram selecionados 26 artigos, aos quais foram aplicados a análise qualitativa explicada na sessão 3.

O Quadro 2 apresenta os artigos selecionados na revisão sistemática na ordem da pontuação extraída na análise qualitativa.

Quadro 2 - Artigos selecionados na RSL

Id	Título	Autor	Ano	Fonte	Pontos
1	Application of grey relational analysis and artificial neural networks on corporate social responsibility (CSR) indices	Diaz & Nguyen	2021	Web of Science	23
2	Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning	De Oliveira Carosia et al.	2021	Scopus	23

19th CONTECSI

Id	Título	Autor	Ano	Fonte	Pontos
3	The Role of ESG in Sustainable Development: An Analysis Through the Lens of Machine Learning	Gupta et al	2021	IEEE	23
4	Sentiment Analysis Applied to News from the Brazilian Stock Market	Januário et al.	2022	IEEE	23
5	Fundamental ratios as predictors of ESG scores: a machine learning approach	D'Amato et al.	2021	Scopus	22
6	Using big data to evaluate corporate social responsibility and sustainable development practices	Barbeito-Caamaño & Chalmeta	2020	Scopus	22
7	Quantifying ESG alpha using scholar big data: An automated machine learning approach	Chen & Liu	2020	Scopus	22
8	Detecting Environmental, Social and Governance (ESG) Topics Using Domain-Specific Language Models and Data Augmentation	Nugent et al.	2020	Web of Science	22
9	Towards the Categorization of Brazilian Financial Market Headlines	Schmitz, Immich, et al.	2022	Web of Science	22
10	Proposing an Integrated Approach to Analyzing ESG Data via Machine Learning and Deep Learning Algorithms	Lee et al.	2022	Web of Science	21
11	Weak Supervision and Black-Litterman for Automated ESG Portfolio Construction	Sokolov et al.	2021	Scopus	20
12	Socially Responsible Investment Portfolio Construction with a Double-Screening Mechanism considering Machine Learning Prediction	Zhang & Chen	2021	Scopus	19
13	Deep learning for decision making and the optimization of socially responsible investments and portfolio	Vo et al.	2019	Scopus	19
14	Identifying Corporate Sustainability Issues by Analyzing ShareholderResolutions: A Machine-Learning Text Analytics Approach	Raghupathi et al	2020	Web of Science	19
15	Data Analysis of ESG Stocks in the Chinese Stock Market Based on Machine Learning	Yu et al.	2022	Scopus	18
16	The effect of green energy, global environmental indexes, and stockmarkets in predicting oil price crashes: Evidence from explainablemachine learning	Ben Jabeur et al.	2021	Web of Science	17
17	Prediction of ESG compliance using a heterogeneous information network	Hisano et al.	2020	Web of Science	15
18	ESG score prediction through random forest algorithm	D'Amato et al.	2022	Web of Science	15
19	The Influence of Tweets and News on the Brazilian Stock Market Through Sentiment Analysis	De Carosia et al.	2019	ACM	15

Id	Título	Autor	Ano	Fonte	Pontos
20	Determinants and Prediction Accuracy of Price Multiples for South East Asia: Conventional and Machine Learning Analysis	Joshi & Chauhan	2020	Web of Science	14
21	Modeling of the Relationship between Corporate Social Responsibility and Stock Price with Artificial Neural Network	Sukthomya & Laosiritaworn	2018	IEEE	13
22	Deep Learning Model for Stock Excess Return Prediction Based on Nonlinear Random Matrix and Esg Factor	Meng et al	2022	Scopus	12
23	Rho AI – Leveraging artificial intelligence to address climate change: Financing, implementation, and ethics	Fischer et al.	2021	Scopus	10

Fonte: Elaborado pela autora

A seção seguinte apresenta a análise dos resultados extraídos destes estudos.

4.2 ANÁLISE DOS RESULTADOS

O trabalho de Fischer et al (2021) aborda uma solução denominada Rho IA, que prevê a integração de tecnologia automatizada envolvendo scrapy, processamento de linguagem natural e aprendizado de máquina, para permitir que os investidores avaliem o impacto climático das empresas e usem essa avaliação como base para a realização de investimentos nas empresas. O trabalho não chegou a implementar a solução proposta, mas sugeriu isso como trabalho futuro.

O trabalho de Gupta et al (2021) propõe a criação de um conjunto de dados financeiro e ESG de empresas de capital aberto de todo o mundo (incluindo países de média e baixa renda) coletados na base Yahoo Finance por meio de web scrapers em Python e a aprendizado de máquina para avaliar a importância dos parâmetros ESG (coletados na base Sustainalytics) no desempenho financeiro das empresas. O autor demonstrou, que empresas com boa classificação ESG também se saem bem na frente financeira. Quando as médias estatísticas das empresas foram calculadas com suas classificações ESG, o retorno sobre o patrimônio líquido (ROE) foi 14% maior em comparação com empresas com classificações ESG não tão boas. O trabalho também identificou que variáveis como “margem de lucro” e “retorno sobre o ativo” podem ser previstas com melhor precisão quando os dados ESG são usados nos parâmetros de entrada dos modelos de regressão.

Sukthomya e Laosiritaworn (2018) investigaram o impacto da prática de CSR nos preços das ações e conclui que empresas que se engajam em comportamentos socialmente responsáveis, se beneficiam do aumento da competitividade e do melhor desempenho das ações por aplicação de redes neurais. Já o trabalho de Barbeito-Caamaño e Chalmers (2020) utilizou de técnicas de BigData para analisar informações sobre as práticas de empresas CSR nas redes sociais por meio de modelos de *machine learning* para a avaliação dos comentários do Twitter em positivos e negativos feitos sobre uma empresa considerada CSR.

Schmitz et al. (2022) desenvolveram *crawlers* (scrapy automatizado com agendamentos de execuções) em Python para coletar dados de sites notícias financeiras tradicionais em português, tais como InfoMoney, MoneyTimes e Suno Reaserch. O trabalho desenvolveu uma solução denominada GOOSE para a categorização de títulos de notícias utilizando vetores semânticos treinados em domínio específico. A solução proposta foi

composta com o modelo de GloVe para associar as palavras em um vetor n-dimensional e o modelo de redes neurais Bi-LSTM para classificação dos textos. Já o trabalho de Raghupathi et al. (2020) aplicou NLTK para o reconhecimento das palavras e o algoritmo de *machine learning* K-Neighbors para identificar e analisar os principais problemas de sustentabilidade nas resoluções propostas pelos acionistas de investimento de ações e ETF (fundos de ações).

Hisano et al. (2020) por sua vez desenvolveram um processo de *scrapy* para coletar dados estruturados e não estruturados para criar métricas ESG, com o interesse em prever se uma empresa entrará na lista de exclusão de empresas inapropriadas para investimentos. Para isso, faz uma varredura das notícias negativas em bases de dados que dispõem de informações relevantes de empresas localizadas em diversos países, assim como uso de BigData para gerenciamento dos dados e de *machine learning* por meio do algoritmo *Random Forest* para classificação das notícias. Porém não aprofundou em seu trabalho as técnicas utilizadas.

Os trabalhos de Carosia et al. (2020), De Oliveira Carosia et al. (2021) e Januario et al., (2022) apresentam similares. Os três extraíram e aplicaram análise de sentimentos de notícias e comentários pertencentes ao domínio financeiro brasileiro publicadas online, contribuindo assim, para análise das técnicas de NLP com textos em português utilizando técnicas de *word embedding* (NLTK e FastText) e modelos de redes neurais. Além da análise de sentimentos. De Oliveira Carosia et al. (2021) também utilizaram o teste de causalidade de Granger para analisar a correlação entre os sentimentos das notícias e o mercado financeiro, como conclusão o trabalho sugere que para mitigar os desafios de previsão do mercado de ações brasileiro, se faz necessário investigar a correlação entre o sentimento predominante das notícias e as mudanças no mercado em diferentes períodos. Importante notar que os três trabalhos são direcionados aos investimentos financeiros tradicionais, ou seja, não mencionam os investimentos ESG especificamente.

Ainda no contexto de NLP, Nugent et al (2020) propõem uma solução para mitigar os problemas de adaptação de domínio e aumento de dados em conjunto de dados de controvérsias ESG, por meio de técnicas de *deep learning*. O trabalho aplica retraduações de textos entre francês e inglês. Apesar do estudo não trazer comparações com outras linguagens além das duas mencionadas, o autor consegue demonstrar as possibilidades de atingir uma boa performance no pré-processamento de dados de diferentes linguagens naturais por meio dos algoritmos de BERT e Tensor2Tensor.

Joshi e Chauhan (2020) sugerem diversas abordagens para analisar dados ESG utilizando diferentes algoritmos de *machine learning* tradicionais e redes neurais para classificar se investidores realizaram investimentos excelentes ou ruins e a previsão das próprias pontuações ESG baseada na classificação (E, S ou G) de notícias das empresas. Na mesma linha investigativa, o trabalho de Sokolov et al. (2020) focou em desenvolver uma solução para pontuação ESG com base em artigos de notícias, por meio do método de Black-Litterman (direcionado para a diversificação de investimentos) para agregação de dados ESG e utilização do modelo pré-treinado BERT para identificação de categorias ESG.

Na predição dos valores dos investimentos ESG, os trabalhos de D'Amato et al. (2021) e D'Amato et al. (2022) demonstram que é possível extrair valor (alfa) de investimentos ESG utilizando redes neurais e que isso tem impacto no desempenho financeiro das empresas, assim como a utilização das informações estruturais das empresas (análise fundamentalista) para explicar seus respectivos indicadores ESG por meio do algoritmo de classificação *Random Forest*. Chen e Liu (2020) por sua vez, utilizaram uma base de dados ESG acadêmica como dados alternativos para uma abordagem de aprendizado de máquina para quantificar o prêmio ESG de uma empresa e capturar o alfa ESG.

Meng et al. (2022) sugerem a aplicação de uma matriz aleatória para realizar o mapeamento dos fatores ESG e com isso, prever o excesso do retorno de ações utilizando *deep learning*. Já o trabalho de Zhang e Chen (2021) utilizou o modelo de média-variância (MV) de Markowitz para análise de portfólio predição do retorno dos investimentos socialmente responsáveis. Yu et al., (2022) demonstraram a performance dos investimentos ESG utilizando *machine learning* não somente em termos dos retornos elevados, mas também em termos de risco baixo.

O estudo de Vo et al., (2019) por sua vez, realizou a predição do retorno dos investimentos responsáveis e de impacto, por meio de um novo modelo híbrido (denominado DRIP) que combina diversas técnicas de predição com base em algoritmos de *deep learning* e *reinforce learning*. A solução demonstrou assertividade nas predições periódicas para tais investimentos. O estudo não evidencia o comportamento do modelo com dados heterógenos como os realizado nos estudos de Hisano et al (2020) e Nugent et al (2021), sendo este um ponto que pode ser desenvolvido em trabalhos futuros.

O trabalho de Diaz e NguYen (2021) aplicou modelos de redes neurais para prever os retornos do índice CSR com base nos índices de ações S&P500, de volatilidade, do dólar americano, comercial, de commodities (CRB) e do petróleo bruto (Brent). Na comparação dos modelos, BPN foi o que se saiu melhor segundo os autores. Além disso, foi utilizado o método *grey relational analysis* (GRA) para determinar quais variáveis, ou seja, os seis índices utilizados na previsão, possui o maior efeito sobre os retornos do índice CSR por meio de seus graus relacionais. O trabalho concluiu que na pontuação GRA, os índices de dólar e o S&P 500 apresentam maior potencial de predição do índice CSR.

Ben Jabeur et al (2021) buscaram prever os preços do petróleo durante a pandemia da COVID-19 analisando recursos de energia verde, índices ambientais globais (ESG) e de mercados de ações. O estudo emprega técnicas de *machine learning*, para capturar a tendência das mudanças nos preços do petróleo e reduzir o impacto da pandemia de COVID-19 nesses preços.

Analisando todos os trabalhos detalhados acima, é possível identificar que uma boa parte utilizou modelos de redes neurais tanto para as técnicas de processamento de linguagem natural como análise de sentimentos e textual, quanto para as predições dos retornos dos valores dos investimentos ESG. Além disso, uma investigação nas bases mencionadas pelos 23 trabalhos, levou esta pesquisa à conclusão da ausência de fontes com dados ESG de acesso aberto como Yahoo Finance para empresas brasileiras. Portanto, o mapeamento da pontuação ESG relatada em alguns dos trabalhos selecionados como o de Joshi e Chauhan (2020) é uma alternativa para construção de uma base com indicadores e pontuação ESG própria por meio da classificação de notícias. Outro fator relevante, é ausência de pesquisas direcionadas aos investimentos ESG brasileiros.

O Quadro 3 compara as bases de dados, o método de análise e os modelos utilizado no treinamento de cada trabalho selecionado.

Quadro 3 - Comparação dos artigos selecionados

Autor	Coleta e Bases Dados	Método de Análise	Modelo de Treinamento
Diaz & Nguyen, 2021	Google Finance, Forexpros, Quote, Jefferies e Theice	Grey relational analysis (GRA) e Análise de sentimentos	BPN, RNN e RBFNN
De Oliveira Carosia et al, 2021	G1, Estado, Folha de São Paulo	Teste de causalidade de Granger e Análise de sentimentos	FastText Word Embedding, MLP, LSTM e CNN

Autor	Coleta e Bases Dados	Método de Análise	Modelo de Treinamento
Gupta et al, 2021	Scrapy: Bases: Yahoo Finance e Sustainalytics	Média do ROI das melhores e piores empresas	RandomForest Regressor
Januário et al., 2022	Folha de Sao Paulo, Estadão, InfoMoney, MoneyTimes, Portal Exame, Último Segundo, UOL, Yahoo e G1	Análise de sentimentos	NLTK, Naive Bayes e MLP
D'Amato et al, 2021	MSCI World, Thomson Reuters e Sustainalytics3	Análise Fundamentalista	Machine Learning
Barbeito-Caamaño & Chalmers, 2020	Scrapy - Twitter	Análise de sentimentos	Sentimentr & plyr
Zhang & Chen, 2021	Microsoft Academic Graph	Retorno cumulativo, Sharpe e indicadores financeiros	LSTM, Random Forest, SVR, Linear Regression, Lasso and Ridge
Nugent et al., 2020	BookCorpus e Reuters News	Análise textual	BERT
Schmitz, Immich, et al., 2022	Crawlers - InfoMoney, MoneyTimes e Suno Reaserch	Análise textual	Glove Word Embedding e Bi-LSTM
Lee et al., 2022	Kaggle (400k-nyse-randominvestments-financial-ratios; fossil; fossil-free-funds) Korean website "Naver"	Análise Fundamentalista	DecisionTree, RandomForest, ExtraTree, LGBM, KNN, XGB LSTM, GRU
Sokolov et al., 2021	Scrapy - New York Times, Compustat	Black-Litterman Análise textual	BERT
Zhang & Chen, 2021	Thomson Reuters ASSET 4	O modelo de média-variância (MV) de Markowitz para análise de portfólio de investimentos socialmente responsáveis	ELM, SVR, DNN e LSTM.
Vo et al., 2019	–	–	BERT & LSTM, BI-LSTM e Reinforce Learning
Raghupathi et al, 2020	Scrapy - CERES	Análise textual	NLTK e K-Neighbors
Yu et al., 2022	Yahoo Finance	Análise do Sharp	KNN, SVM e AdaBoost
Ben Jabeur et al., 2021	Yahoo Finance	Indicadores de investimento	LGBM, CatBoost, XGB e RandomForest, ANN

Autor	Coleta e Bases Dados	Método de Análise	Modelo de Treinamento
Hisano et al., 2020	Dow Jones, Dow Jones Watchlist, Capital IQ Company Screening Report, FactSet, FactShip, Reuters, Panama papers, Dbpedia	Análise textual	Random Forest Classification
D'Amato et al., 2022	Thomson Reuters Refinitiv	Análise Fundamentalista	RandomForest
De Carosia et al., 2019	Estadão, Folha de São Paulo e G1.	Análise de Sentimentos	NLTK, MLP
Joshi & Chauhan, 2020	Kaggle dataset (400k-nyse-randominvestments-financial-ratios; fossil; fossil-free-funds) Korean website "Naver"	Análise de texto	DecisionTree, RandomForest, ExtraTree, LGBM, KNN, XGB LSTM e GRU
Sukthomya & Laosiritaworn, 2018	Market for Alternative Investment Bloomberg Data History.	—	ANN
Meng et al, 2022	—	Análise do Excesso de Retorno	Deep Learning
Fischer et al., 2021	Scrapy - Dados Estruturados e Não estruturados	—	Machine Learning Deep Learning

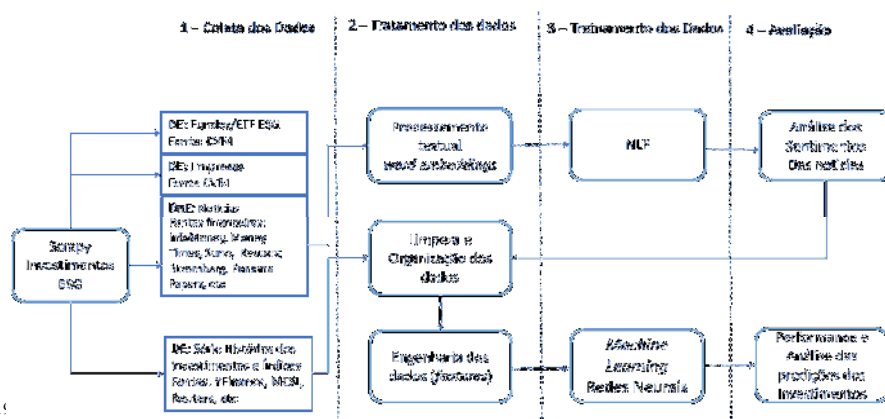
Fonte: Elaborado pela autora

4.3 PROCESSO PROPOSTO

O objetivo desta seção é construir um processo para mitigar os desafios de *machine learning* para os investimentos ESG brasileiros.

Segue abaixo a descrição dos passos do processo proposto (Figura 10)

Figura 10 - Processo Proposto



Fonte: Elaborado pela autora

1. Coletar os dados: Primeira etapa do processo

- a. Inicialmente é preciso realizar um *web scrappy* em alguma base dos investimentos ESG. Nesta pesquisa adotou-se como exemplo a fonte Comissão de Valores Imobiliários (CVM) que possui a lista de fundos de investimentos brasileiros ESG,
- b. Em seguida é necessário identificar as empresas que compõem os fundos de investimentos. A lista de empresas brasileiras também está disponível na CVM (<https://www.gov.br/cvm/pt-br>),
- c. Extrair com técnicas de *scrappy* notícias das empresas relacionadas aos investimentos ESG (identificadas no passo 1.b) em diferentes fontes como sites de notícias comuns ou especializados da área financeira e
- d. Extrair a série histórica dos valores dos investimentos ESG e dos índices em bases como Yahoo Finance, MCSI e Reuters.

2. Tratamento dos dados

- a. Para os dados não estruturados (DnE) aplicar algumas das técnicas de *word embeddings* como GloVe ou FastText para posteriormente avaliados com algum modelo já pré-treinado, ou ainda serem aplicados em algum modelo de treinamento do zero na fase 3,
- b. Limpeza e Organização dos dados estruturados e
- c. Aplicação das técnicas de engenharia de dados por meio de algum método de análise financeira como os observados no Quadro 3.

3. Treinamento dos Dados

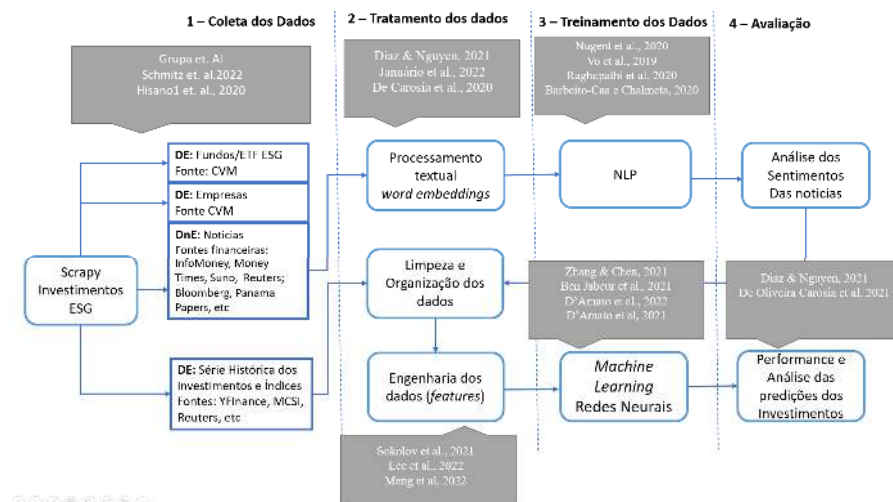
- a. Com os dados das notícias já estruturados na fase 2.a é possível aplicar algum modelo já pré-treinado como BERT, ou utilizar modelos de *machine learning/deep learning* para realizar o treinamento do zero e
- b. Por meio das *features* identificadas no passo 2.c, as quais se juntam também a classificação das notícias realizados no passo 4.a, o processo segue com o treinamento dos dados para previsão da performance dos retornos, risco ou volatilidade dos investimentos.

4. Avaliação

- a. Após o treinamento dos dados das notícias, é importante avaliar a classificação dos sentimentos. O resultado desta classificação será agregado aos dados da série histórica dos investimentos, com a retomada do passo 2.b (Figura 10) e
- b. Finalmente, é realizada a avaliação da performance e a análise da predição dos investimentos ESG.

A Figura 11 exibe os trabalhos que contribuíram para o processo proposto.

Figura 11 – Artigos de suporte ao processo proposto



Fonte: Elaborado pelo autor

5. CONCLUSÃO

Este artigo responde à pergunta de pesquisa “Como obter um processo baseado nas melhores técnicas de scrapy e machine learning para mapear dados e prever o retorno dos investimentos ESG do Brasil?” por meio de uma revisão sistemática da literatura, mostrando os passos para coleta dos dados, tratamento dos dados não estruturado em português, extração de *features* preditoras utilizando métodos de análise financeiras e utilização de diferentes modelos de *machine learning* tradicionais ou de redes neurais

Para as perguntas secundárias: “a. Quais as melhores técnicas de machine learning para mapear os dados e prever o retorno dos investimentos ESG brasileiros?” e “b. Como o resolver os desafios de algoritmos de machine learning ao lidar com informações em linguagens diferentes do Inglês?” é possível apontar redes neurais (que fazem parte do universo de *machine learning*) como o melhor conjunto de algoritmos para atuar tanto no pré-processamento de dados em português, quanto para predição dos retornos dos investimentos ESG, combinados com outras técnicas de *machine learning* e/ou *reinforce learning* para atingir modelos que consigam gerenciar a complexidade e diversidade dos investimentos mencionados.

A revisão sistemática efetuada e o processo proposto são contribuições para revisão das práticas mais recomendadas da literatura, ampliação dos estudos relacionados aos investimentos financeiros ESG e discussão dos desafios existentes de *machine learning* relacionados à captura de dados heterogêneos e predições com análise de sentimentos de texto em português.

Este artigo não identificou na revisão sistemática trabalhos relacionados aos investimentos ESG brasileiros utilizando *machine learning* e de base de dados aberta com indicadores ESG para as empresas brasileiras.

Recomendam-se como trabalhos futuros a implementação do processo proposto neste artigo..

6. REFERÊNCIAS

ANBIMA. **Incorporação dos aspectos ASG nas análises de investimento GUIA ASG**. [s.l.: s.n.].

Disponível em:

<<https://www.anbima.com.br/data/files/1A/50/EE/31/BFDEF610CA9C4DF69B2BA2A8/ANBIMA-Guia-ASG-2019.pdf>>. Acesso em: 30 out. 2021.

B3 S.A. **Produtos e Serviços ESG**. Disponível em:

<https://www.b3.com.br/pt_br/b3/sustentabilidade/produtos-e-servicos-esg/green-bonds/>.

Acesso em: 9 nov. 2021.

BARBEITO-CAAMAÑO, A.; CHALMETA, R. Using big data to evaluate corporate social responsibility and sustainable development practices. **Corporate Social Responsibility and Environmental Management**, v. 27, n. 6, p. 2831–2848, 1 nov. 2020.

BEN JABEUR, S.; KHALFAOUI, R.; BEN ARFI, W. The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: Evidence from explainable machine learning. **Journal of Environmental Management**, v. 298, 15 nov. 2021.

BISWAL, A. **Top 10 Deep Learning Algorithms You Should Know in 2023**. Disponível em:

<<https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm>>. Acesso em: 18 nov. 2022.

CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. **Applied Artificial Intelligence**, v. 34, n. 1, p. 1–19, 2 jan. 2020.

CHEN, L. Stock Price Prediction using Adaptive Time Series Forecasting and Machine Learning Algorithms. [s.d.].

CHEN, Q.; LIU, X.-Y. **Quantifying ESG alpha using scholar big data**. Proceedings of the First ACM International Conference on AI in Finance. **Anais...**New York, NY, USA: ACM, 15 out. 2020.

Disponível em: <<https://dl.acm.org/doi/10.1145/3383455.3422529>>. Acesso em: 20 out. 2021

D'AMATO, V.; D'ECCLESIA, R.; LEVANTESI, S. Fundamental ratios as predictors of ESG scores: a machine learning approach. **DECISIONS IN ECONOMICS AND FINANCE**, v. 44, n. 2, p. 1087–1110, dez. 2021.

D'AMATO, V.; D'ECCLESIA, R.; LEVANTESI, S. ESG score prediction through random forest algorithm. **Computational Management Science**, v. 19, n. 2, p. 347–373, 1 jun. 2022.

DE OLIVEIRA CAROSIA, A. E.; COELHO, G. P.; DA SILVA, A. E. A. Investment strategies applied to the Brazilian stock market: A methodology based on Sentiment Analysis with deep learning. **Expert Systems with Applications**, v. 184, p. 115470, dez. 2021.

19th CONTECSI

DIAZ, J. F.; NGUYEN, T. T. Application of grey relational analysis and artificial neural networks on corporate social responsibility (CSR) indices. **Journal of Sustainable Finance and Investment**, 2021.

FELIZARDO, K. et al. **Revisão sistemática da literatura e engenharia de software**. 1. ed. Rio de Janeiro: Elsevier, 2017.

FISCHER, I.; BESWICK, C.; NEWELL, S. Rho AI – Leveraging artificial intelligence to address climate change: Financing, implementation and ethics. **Journal of Information Technology Teaching Cases**, v. 11, n. 2, p. 110–116, 2021.

GANEGEDARA, T. **Stock Market Predictions with LSTM in Python**. Disponível em: <<https://www.datacamp.com/tutorial/lstm-python-stock-market>>. Acesso em: 18 nov. 2022.

GUPTA, A.; SHARMA, U.; GUPTA, S. K. **The Role of ESG in Sustainable Development: An Analysis Through the Lens of Machine Learning**. 2021 IEEE International Humanitarian Technology Conference, IHTC 2021. **Anais...**Institute of Electrical and Electronics Engineers Inc., 2021.

HAND, D. et al. **IMPACT Authored by the GIIN Research Team About the Global Impact Investing Network (GIIN)**. [s.l: s.n.]. Disponível em: <<http://www.thegiin.org/>>. Acesso em: 27 out. 2021.

HISANO, R.; SORNETTE, D.; MIZUNO, T. Prediction of ESG compliance using a heterogeneous information network. **Journal of Big Data**, v. 7, n. 1, 1 dez. 2020.

HUME, V.; DAVIDSON, A.; GUTTENTA, M. **Investimentos de Impacto na América Latina**. [s.l: s.n.]. Disponível em: <www.andeglobal.org>.

INDERST, G.; STEWART, F. **Incorporating ENVIRONMENTAL, SOCIAL and GOVERNANCE (ESG) Factors into FIXED INCOME INVESTMENT**. [s.l: s.n.]. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175830>. Acesso em: 4 nov. 2021.

JANUARIO, B. A. et al. Sentiment Analysis Applied to News from the Brazilian Stock Market. **IEEE LATIN AMERICA TRANSACTIONS**, v. 20, n. 3, p. 512–518, mar. 2022.

KALYANI INAMPUDI, B. **The Impact of AI on Environmental, Social and Governance (ESG) Investing: Implications for the Investment Value Chain**. [s.l: s.n.]. Disponível em: <<https://doi.org/10.1002/9781119551966.ch35>>. Acesso em: 12 nov. 2021.

KAUR, J. **Applications of Natural Language Processing For Businesses**. Disponível em: <<https://www.xenonstack.com/blog/natural-language-processing-nlp>>. Acesso em: 18 nov. 2022.

KIRPPU, A. **Impact investing in private equity**. [s.l: s.n.]. Disponível em: <https://aaltodoc.aalto.fi/bitstream/handle/123456789/39847/master_Kirppu_Annemari_2019.pdf?sequence=1&isAllowed=y>. Acesso em: 3 nov. 2021.

MANUEL, J. et al. **Real Estate Market Data Scraping and Analysis for Financial Investments**. [s.l: s.n.]. Disponível em: <<https://repositorio-aberto.up.pt/bitstream/10216/116510/2/296684.pdf>>. Acesso em: 10 nov. 2021.

MAURICIO AMORMINO JUNIOR. **VISÕES DE FUTURO PARA A AGENDA DE IMPACTO NO BRASIL**. [s.l: s.n.]. Disponível em: <<http://aliancapeloimpacto.org.br/wp-content/uploads/2021/03/alianca-interativo.pdf>>. Acesso em: 30 out. 2021.

19th CONTECSI

MENG, T.; YAHYA, M. H.; CHAI, J. Deep Learning Model for Stock Excess Return Prediction Based on Nonlinear Random Matrix and Esg Factor. **Mathematical Problems in Engineering**, v. 2022, p. 1–10, 14 set. 2022.

NAIARA BERTÃO. **Conheça os fundos de investimentos ESG ou de impacto disponíveis no Brasil - ValorInvest**. Disponível em:

<<https://valorinveste.globo.com/google/amp/produtos/fundos/noticia/2021/03/07/conheca-os-fundos-de-investimentos-esg-ou-de-impacto-disponiveis-no-brasil.ghtml>>. Acesso em: 4 nov. 2021.

NUGENT, T.; STELEA, N.; LEIDNER, J. L. Detecting ESG topics using domain-specific language models and data augmentation approaches. 16 out. 2020.

RAGHUPATHI, V.; REN, J.; RAGHUPATHI, W. Identifying corporate sustainability issues by analyzing shareholder resolutions: A machine-learning text analytics approach. **Sustainability (Switzerland)**, v. 12, n. 11, 1 jun. 2020.

ROCHA, G.; CARDOSO, H. L. Recognizing textual entailment: Challenges in the Portuguese language. **Information (Switzerland)**, v. 9, n. 4, 29 mar. 2018.

SCHMITZ, M. et al. **Towards the Categorization of Brazilian Financial Market Headlines**. [s.l.: s.n.].

SOKOLOV, A. et al. **Building Machine Learning Systems to Automate ESG Index Construction**. [s.l.: s.n.]. Disponível em: <https://www.seco.risklab.ca/wp-content/uploads/2020/11/ESG__paper_PMR_ESG.pdf>. Acesso em: 20 out. 2021.

SUKTHOMYA, D.; LAOSIRITAWORN, W. **Modeling of the relationship between corporate social responsibility and stock price with artificial neural network**. 2018 7th International Conference on Industrial Technology and Management, ICITM 2018. **Anais...**Institute of Electrical and Electronics Engineers Inc., 6 abr. 2018.

VO, N. N. Y. et al. Deep learning for decision making and the optimization of socially responsible investments and portfolio. **Decision Support Systems**, v. 124, 1 set. 2019.

YU, G. et al. **Data Analysis of ESG Stocks in the Chinese Stock Market Based on Machine Learning**. 2022 2nd International Conference on Consumer Electronics and Computer Engineering, ICCECE 2022. **Anais...**Institute of Electrical and Electronics Engineers Inc., 2022.

ZHANG, J.; CHEN, X. Socially Responsible Investment Portfolio Construction with a Double-Screening Mechanism considering Machine Learning Prediction. **Discrete Dynamics in Nature and Society**, v. 2021, 2021.