



**Universidade Federal da Paraíba**  
**Centro de Ciências, Humanas, Sociais e Agrárias**  
**Repositório Institucional - RI**

**Termo de Autorização para Publicação Eletrônica no Repositório Institucional da UFPB**

Autorizo e estou de acordo, na qualidade de titular dos direitos de autor da publicação auto-depositada, conforme Lei no 9610/98, os seguintes termos:

**Da Distribuição não exclusiva**

*O autor declara que:*

- a) O documento entregue é seu trabalho original, e que detém o direito de conceder os direitos contidos neste de qualquer outra pessoa ou entidade. Declara também que a entrega do documento não infringe, tanto quanto lhe é possível saber, os direitos de qualquer outra pessoa ou entidade.
- b) Se o documento entregue contém material do qual não detém os direitos de autor, declara que obteve autorização do detentor dos direitos de autor para conceder a Universidade Federal da Paraíba os direitos requeridos por este termo, e que esse material cujos direitos são de terceiros está claramente identificado e reconhecido no texto ou conteúdo do trabalho entregue.
- c) Se o documento entregue é baseado em trabalho financiado ou apoiado por outra instituição que não a Universidade Federal da Paraíba declara que cumpriu quaisquer obrigações exigidas pelo respectivo contrato ou acordo.
- d) Com a apresentação desta licença, você (o autor (es) ou o titular dos direitos de autor) concede ao Repositório Institucional da UFPB o direito de reproduzir, traduzir, e/ou distribuir a sua publicação (incluindo o resumo) por todo o mundo no formato impresso e eletrônico e em qualquer meio, incluindo os formatos áudio ou vídeo.
- e) Você concorda que o Repositório Institucional da UFPB pode, sem alterar o conteúdo, transpor a sua publicação para qualquer meio ou formato para fins de preservação.
- f) Você concorda que o Repositório Institucional da UFPB pode manter mais de uma cópia de sua publicação para fins de segurança, backup e preservação.

**Da Licença de Uso**

O autor autoriza a disponibilização, de acordo com a Licença Creative Commons, o conteúdo completo da obra de sua autoria, através do Repositório Institucional da UFPB no endereço: <https://repositorio.ufpb.br>. Os direitos da licença Creative Commons são das **atribuições sem derivações 3.0 Brasil (CC BY-ND 3.0 BR)**, disponível em <https://creativecommons.org/licenses/by-nd/3.0/br/>

**Dos Embargos e Restrições de Acesso**

O embargo poderá ser mantido por até 1 (um) ano, podendo ser prorrogado por igual período, com a necessidade de anexar documentos comprobatório. O resumo e os metadados descritivos serão disponibilizados no Repositório Institucional da UFPB. O depósito do trabalho é obrigatório, independente do embargo. Quando embargado, o trabalho permanecerá indisponível enquanto vigorar as restrições. Passado o período do embargo, o trabalho será automaticamente disponibilizado no Repositório Institucional da UFPB.

**1. Tipo de produção intelectual (ex: Relatório Técnico; monografia; dissertação; tese; outro)**

TCC (Trabalho de conclusão de curso)

**2. Agência de Fomento (ex: Capes; UFPB), caso não possua, favor deixar o espaço em branco.**

UFPB

### 3. Informação de acesso ao documento

Esse trabalho é confidencial? ( ) sim; (X) não.

Esse trabalho ocasionará registro de patente? ( ) sim; (X) não.

Qual é a amplitude da liberação da publicação? (X) total; ( ) parcial; ( ) não pode ser publicada, exceto, o sumário.

Em caso de publicação parcial, indicar restrições: \_\_\_\_\_

### 4. Identificação da Dissertação

#### Autor

<b>Nome:</b>	Eduardo Santos da Costa
<b>Matricula:</b>	
<b>Identidade:</b>	
<b>CPF:</b>	
<b>Telefone</b>	
<b>E-mail:</b>	

#### Publicação

<b>Título do trabalho</b>	A INFLUÊNCIA DAS NOTÍCIAS VEICULADAS PELO G1 E DO ÍNDICE ESG NA EXPLICAÇÃO DO COMPORTAMENTO DO MERCADO ACIONÁRIO BRASILEIRO
<b>Data de defesa</b>	14 de Novembro de 2023
<b>Orientador(es)</b>	Gustavo Correia Xavier
<b>CPF dos orientadores</b>	

#### ANUÊNCIA DO AUTOR


Assinatura do autor
<u>Bananeiras</u>
Local
<u>07 / 12 / 2023</u>
Data

#### ANUÊNCIA DO ORIENTADOR


Assinatura do orientador
<u>Bananeiras</u>
Local
<u>07 / 12 / 2023</u>
Data

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

C838i Costa, Eduardo Santos da.

A influência das notícias veiculadas pelo G1 e do índice ESG na explicação do comportamento do mercado acionário brasileiro / Eduardo Santos da Costa. - Bananeiras, 2023.

28 f. : il.

Orientação: Gustavo XAVIER.  
TCC (Graduação) - UFPB/cchsa.

1. Sentimento textual. 2. ESG. 3. Mercado brasileiro. I. XAVIER, Gustavo. II. Título.

UFPB/CCHSA-BANANEIRAS

CDU 658 (042)

**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS HUMANAS, SOCIAIS E AGRÁRIAS  
DEPARTAMENTO DE CIÊNCIAS SOCIAIS APLICADAS  
CURSO DE BACHARELADO EM ADMINISTRAÇÃO**

**EDUARDO SANTOS DA COSTA**

**A INFLUÊNCIA DAS NOTÍCIAS VEICULADAS PELO G1 E DO ÍNDICE ESG NA  
EXPLICAÇÃO DO COMPORTAMENTO DO MERCADO ACIONÁRIO  
BRASILEIRO**

**Bananeiras - PB**

**2023**

**Resumo.** Este estudo tem como objetivo central analisar como um sentimento textual baseado na relação entre notícias do G1 e o índice ESG ajuda a explicar o comportamento dos preços do mercado acionário brasileiro. Para realizar essa análise, foram coletados dados de notícias relacionadas à economia e à política do site G1, abrangendo o período de dezembro de 2020 a maio de 2023, resultando em uma amostra de 17.999 matérias. A metodologia empregada envolveu pré-processamento do texto com o método TF-IDF e previsão de retorno acionário com o uso de modelos de *machine learning*. O estudo revelou que, em média, as notícias analisadas demonstraram uma inclinação predominantemente negativa, embora tenham ocorrido variações notáveis ao longo do período analisado. Além da análise do sentimento das notícias, a pesquisa também explorou diferentes modelos de previsão de preços, incluindo *LASSO* e *XGBoost*. Essas abordagens demonstraram a capacidade de prever tendências no mercado financeiro com base nas informações extraídas do sentimento textual das notícias. Em resumo, este estudo investigou a relação entre o mercado financeiro brasileiro e um sentimento textual baseado em notícias e no índice ESG, destacando o papel crucial do sentimento textual no entendimento das dinâmicas do mercado.

Palavras-chave: Sentimento textual, ESG, Mercado brasileiro.

## Sumário

<b>1.INTRODUÇÃO</b> .....	4
<b>2. FUNDAMENTAÇÃO TEÓRICA</b> .....	5
2.1 - HME .....	5
2.2 Sentimento textual.....	7
2.3. ESG .....	8
2.4. Índices de sustentabilidade .....	10
2.5.1 TF-IDF.....	11
2.5.2 Previsão do preço .....	12
<b>3. METODOLOGIA</b> .....	12
3.1 Pré-processamento do texto com aplicação do TF-IDF .....	13
3.2 Medidas de Sentimento Textual .....	14
3.3 Modelos de Previsão com o uso de <i>Machine Learning</i> .....	14
3.4 Avaliação dos Modelos de Previsão .....	15
<b>4. ANÁLISE DOS RESULTADOS</b> .....	15
4.1 Sentimento textual das notícias (SentNews) .....	16
4.2 Sentimento textual condicional ao ESG (SentNewsESG).....	19
4.3 Análise de previsão do retorno acionário baseado no índice ESG .....	21
<b>5. CONCLUSÃO</b> .....	24
<b>REFERÊNCIAS</b> .....	27

## 1.INTRODUÇÃO

A flutuação dos preços dos ativos é um tema recorrente e desafiador no mercado financeiro, que tem sido objeto de estudo há décadas. Apesar disso, compreender o comportamento do mercado diante de novos eventos e informações continua sendo uma tarefa difícil. O preço de uma ação é influenciado pelas expectativas dos investidores em relação à capacidade da empresa em gerar valor. Além disso, a volatilidade dos preços e os fenômenos de valorização e desvalorização também desempenham um papel crucial nesse contexto.

Nesse sentido, uma das linhas de pensamento que busca explicar o funcionamento do mercado é a Hipótese de Mercado Eficiente (HME), desenvolvida por Eugene Fama na década de 1960. Segundo essa teoria, um mercado eficiente é aquele em que os preços refletem plenamente as informações disponíveis, impossibilitando obter lucros anormais com base nessas informações (FAMA, 1960).

No entanto, a HME não é a única abordagem para entender o comportamento do mercado. Uma área de estudo em ascensão é o sentimento textual, que se refere à análise das emoções, opiniões e atitudes expressas em textos. Através da análise do sentimento textual em notícias, relatórios e redes sociais, sendo possível identificar o sentimento do mercado e antecipar tendências (LOUGHRAN; MCDONALD, 2020).

Além disso, nos últimos anos, as questões ambientais, sociais e de governança corporativa (ESG) ganharam destaque no mercado financeiro. O conceito do tripé da sustentabilidade, passou a considerar além dos resultados financeiros os resultados socioambientais das empresas (BOFINGER; HEYDEN; ROCK, 2022). Nesse sentido, esse trabalho agrega à literatura ao estudar como as notícias do G1 ajudam a explicar o comportamento da bolsa brasileira. A adicionalmente inova ao propor uma medida de sentimento textual baseada no índice ESG.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1 - HME

Uma das linhas de pensamentos advém das Finanças Modernas, fundamentada em uma abordagem microeconômica, tendo como paradigma central a premissa de que os agentes econômicos tomam decisões financeiras com base na racionalidade plena, e o preço de um ativo reflete diretamente as informações disponíveis. Os indivíduos ordenam de forma lógica e racional suas preferências, buscando maximizar a utilidade de suas escolhas (LUCENA ET AL, 2014).

Fundamentando a base da Hipótese de Eficiência de Mercado (HEM). Essa hipótese foi formulada na década de 1960 pelo economista americano Eugene Fama em sua dissertação de mestrado. Seu trabalho de maior relevância sobre o assunto denomina-se *Efficient capital markets: a review of theory and empirical work*, publicado em 1970.

Com base na HME, é possível afirmar que um mercado se torna eficiente devido às informações, impossibilitando obter lucros anormais com base nessas informações, pois os preços refletem as próprias informações. A afirmação de um mercado eficiente, os preços "refletem plenamente" as informações disponíveis (FAMA, 1970).

De acordo com Silva e Oliveira (2012, p. 2), “a Hipótese de Mercados Eficientes (HME) defende que todos os preços refletem de forma precisa as informações públicas em todos os momentos. Em outras palavras, os ativos financeiros são sempre adequados considerando aquilo que é conhecido publicamente a todo instante”.

Ross (2002) (apud RABELO; IKETA, 2004, p. 98), afirma que, a teoria de mercados eficientes diz que todos os títulos sempre são corretamente precificados, o que significa que o mercado como um todo é realmente sábio.

O impacto das informações sobre o preço das ações pode variar em relação ao tempo. Diante disso, Fama (1970) propôs três tipos de eficiência de mercado: fraca, semiforte e forte.

Um mercado de capitais é considerado eficiente de forma fraca se ele incorpora completamente as informações dos preços passados.

Um mercado é semiforte se os preços refletem toda a informação disponível publicamente, incluindo informações como demonstrações contábeis publicadas pela empresa, assim como informações históricas de preços (ROSS; WESTERFIELD; JAFFE, 2015, p.465).

Um mercado é considerado eficiente na forma forte se seus preços refletem todas as informações disponíveis, sejam públicas ou privadas. Ou seja, as chamadas informações privilegiadas. Até mesmo um *insider*, ao saber de um fato relevante sobre uma empresa, não conseguiria transformar essa informação em uma fonte de lucro extraordinário, já que o mercado perceberia sua intenção e o preço da ação em questão seria ajustado à nova informação (RABELO; IKETA, 2004, p. 100).

Fama (1970) propôs três condições para verificar a eficiência dos mercados: inexistência de custos de transação; toda a informação está disponível a custo zero, a todos os participantes do mercado; e todos concordam quanto aos efeitos das informações nos preços atuais dos ativos, assim como em suas distribuições futuras. Apesar dessas condições serem suficientes para a existência de um mercado eficiente, elas não são estritamente necessárias. Mesmo na ausência dessas condições, ainda é possível que exista um mercado eficiente.

É válido ressaltar que, embora a HME determine a existência de eficiência através de certas condições, seus princípios de preço justo e a força das informações no mercado, não garantem que os preços dos ativos refletem sempre seu valor justo ou que os investidores não possam obter retornos abaixo ou acima da média no mercado, por meio de habilidades ou informações privilegiadas, como pode ocorrer em um mercado com hipótese de mercado forte.

## 2.2 Sentimento textual

O sentimento textual refere-se à expressão de emoções, opiniões ou atitudes por meio de textos escritos. Pode ser considerada como um conjunto de áreas dividido em vários métodos como análise de sentimentos, modelagem de tópicos, medidas de similaridade de documentos ou legibilidade (LOUGHRAN; MCDONALD, 2016).

Por meio do sentimento textual procura-se identificar e categorizar as emoções ou sentimentos presentes em um texto, como positivo, negativo ou neutro. Dessa forma, permite entender e antecipar as tendências do mercado. Através da análise de dados qualitativos, em notícias, relatórios, redes sociais, entre outros tipos de texto, com isso é possível identificar o sentimento do mercado. O estilo ou o tom linguístico desses textos podem conter informações úteis sobre o desempenho futuro da empresa (KEARNEY; LIU, 2014).

Um das primeiras obras a investigar a interação entre sentimento textual de notícias e o mercado financeiro, foi o estudo de Tetlock (2007), onde foi analisada a interação entre as notícias do *Wall Street Journal* e o mercado de ações, entre 1984 e 1999. No trabalho, o autor encontrou evidências de que valores elevados de pessimismo na mídia resultam em pressão sobre os preços de mercado. Também foi observado que quando baixa nível de notícias pessimistas na mídia, há um aumento nas negociações no mercado. Por outro lado, quando os retornos do mercado estão baixos, o pessimismo na mídia tende a aumentar.

García (2013) analisou o impacto do sentimento textual nos preços dos ativos no mercado de ações dos Estados Unidos, no período de 1905 a 2005. Utilizando colunas de notícias financeiras do *New York Times*, para o sentimento textual e verificando o índice DJIA para verificar o mercado. Foram identificados padrões de previsibilidade nos retornos dos ativos financeiros relacionados ao conteúdo das notícias, especialmente durante períodos de recessão.

Aguiar (2012) procurou extrair a opinião de relatórios da administração divulgados por firmas brasileiras e sua relação com o mercado durante 1997 a 2009. Nesse trabalho, o autor obteve resultados como a relação entre a divulgação de notícias positivas e o aumento do retorno no mercado de ações, notícias positivas estão relacionadas a uma diminuição de volatilidade no mercado de ações, e que não há relação entre o volume de negócios e os relatórios da administração.

O estudo de Faustino (2013) teve como objetivo pesquisar as notícias e seus efeitos no Ibovespa. Os dados foram extraídos do *LexisNexis Academic*, entre 2010 e 2012, e a análise dos sentimentos das notícias foi baseada no dicionário de palavras de Loughran e McDonald (2011). Como um dos resultados de sua pesquisa, foi observado que quando sentimento negativo aumenta a incerteza com relação ao investimento, isso faz com que estes investidores negociem mais e, portanto, aumenta a volatilidade (FAUSTINO, 2013).

### **2.3. ESG**

A partir do início do século XXI, surgiu a agenda de sustentabilidade, onde reconhecia certos problemas que não eram abordados pelas empresas. Problemas ambientais como: mudanças climáticas, poluição das águas, resíduos, emissões de gases, dentre outros. Além de questões sociais, incluindo: diversidade, inclusão, impacto das operações na comunidade ao redor, greves etc.

Dessa forma, surgiu uma ampla discussão sobre o conceito do tripé da sustentabilidade, conhecido originalmente como "*Triple Bottom Line*", desenvolvido por John Elkington em 2001. Esse conceito busca romper com o paradigma da forma como as empresas avaliam seu desempenho, considerando apenas os resultados financeiros, introduzindo resultados não-financeiros, introduzindo aspectos ambientais e sociais.

O termo ESG foi mencionado pela primeira vez no relatório "*Who Cares Wins*" ("Quem se importa, ganha"), publicado em 2004 pelo Programa das Nações Unidas para o Meio Ambiente (PNUMA). O relatório em questão destacou a importância dos critérios de sustentabilidade de uma organização e o impacto de suas atividades nas dimensões ambiental, social e governança, na sociedade evidenciando como esses fatores podem impactar socialmente. A iniciativa teve como propósito elaborar parâmetros e recomendações para a integração dos fatores ambientais, sociais e de governança no mercado e na tomada de decisão de investimento.

Posteriormente, as organizações têm buscado adotar melhores práticas ambientais, sociais e de governança corporativa, com a finalidade de alinhar as novas métricas e indicadores, com o objetivo de geração de valor para acionistas e obter vantagem competitiva

no mercado. As questões pertinentes à temática ESG foram alçadas ao topo das agendas corporativas, impulsionadas por instituições financeiras e investidores que buscam oportunidades de investimento em organizações socialmente responsáveis (BOFINGER; HEYDEN; ROCK, 2022; BLUEPRINT, 2020).

As métricas e indicadores de ESG para Costa e Ferezin (2021, p. 84) estão “cada vez mais valorizados e discutidos no ambiente corporativo, chegando até obter importância nas Bolsas de Valores mundiais, fazendo as ações de organizações sustentáveis serem mais rentáveis que de organizações não tão preocupadas com esta questão.”

### **2.3.1. ESG no Brasil**

Com a adoção de novos parâmetros de sustentabilidade pelas organizações, alcançando nível significativo de relevância no mercado, as ações das empresas sustentáveis tornaram-se mais rentáveis em comparação com aquelas que não priorizam as questões socioambientais. O termo ESG se ajustou como pauta obrigatória nas discussões do topo da agenda empresarial das organizações, em especial para as empresas avaliadas pelo mercado financeiro, identificando questões cruciais não somente para as organizações, mas para toda a sociedade (NIEMOLLER, 2021).

No recente estudo intitulado "*The ESG ultimatum: Profit or perish*", conduzido pelo *IBM Institute for Business Value* (IBV), foi revelado que 76% dos executivos atribuem alta prioridade ao ESG, considerando-o um ponto fundamental para suas estratégias de negócios. Além disso, os resultados da mesma pesquisa indicam que 72% dos executivos encaram o ESG como um facilitador de receitas, enquanto 45% esperam que os esforços dedicados ao ESG resultem em maior lucratividade. “Todas as empresas, devem aproveitar a oportunidade para se associar a investidores dispostos a recompensá-los por criar valor a longo prazo para a sociedade como um todo” (ECCLES; KLIMENKO, 2019).

Em decorrência da tendência perante a sustentabilidade, surgiram os Investimentos Socialmente Responsáveis, do original *Socially Responsible Investing* (SRI). Neves (2009, p. 30) descreve como investimentos que consideram o desempenho financeiro das empresas, além de considerar as práticas socialmente responsáveis e sua sustentabilidade objetivando o longo prazo. Além de estarem mais preparadas para enfrentar riscos econômicos, sociais e ambientais.

## 2.4. Índices de sustentabilidade

Com a crescente conscientização dos investidores e o aumento de fundos dedicados a investimentos sustentáveis, tornou-se necessário a idealização de um índice voltado para atender a demanda crescente por investimentos socialmente responsáveis e sustentáveis. Dessa forma, no ano de 1999, em Nova Iorque, foi criado o primeiro índice de sustentabilidade em bolsa de valores, o *Dow Jones Sustainability Index* (DJSI).

A criação do primeiro índice de sustentabilidade foi um importante marco para promover a importância do ESG, além de incentivar as empresas a melhorarem seus desempenhos nas áreas socioambientais. Com isso, o índice proporcionou aos investidores uma ferramenta para tomar decisões mais informadas, considerando além do desempenho financeiro, o desempenho socioambiental.

O segundo foi o FTSE4Good, de Londres, criado em 2001; e o terceiro, lançado em 2003, foi o JSE, de Joanesburgo, África do Sul. Em 1º de dezembro de 2005, o Brasil entrou para esse grupo apresentando o Índice de Sustentabilidade Empresarial BM & FBOVESPA (ISE) (MARCONDES; BACARJI, 2010, p.18).

O Índice de Sustentabilidade Empresarial (ISE) é um indicador criado pela B3 (Bolsa de Valores brasileira) com o objetivo de mensurar o desempenho das empresas listadas em termos de sustentabilidade e transparência. Com o objetivo de mostrar o desempenho de mercado de uma carteira formada por empresas que adotam os princípios de gestão sustentável (MARCONDES; BACARJI, 2010, p.18).

O objetivo do ISE B3 é ser o indicador do desempenho médio das cotações dos ativos de empresas selecionadas pelo seu reconhecido comprometimento com a sustentabilidade empresarial. Apoiando os investidores na tomada de decisão de investimento e induzindo as empresas a adotarem as melhores práticas de sustentabilidade (B3, 2023).

A B3 também é responsável pela criação do Índice Carbono Eficiente (ICO2 B3), criado no ano de 2010. Segundo a B3 (2021), esse índice tem como objetivo servir como instrumento para efetivar o combate às mudanças climáticas. Englobando empresas comprometidas com práticas transparentes em relação às suas emissões de gases de efeito estufa.

Em setembro de 2020, a B3 juntamente com a *S & P Dow Jones*, anunciaram o lançamento do índice S & P/B3 Brasil ESG. Tendo como objetivo do Índice S & P/B3 Brasil ESG é servir como um índice amplamente representativo do mercado brasileiro de valores com um perfil ESG melhorado, enquanto oferece características de risco/retorno (KITCHENER; RAPOPORT, 2020).

### 2.5.1 TF-IDF

Com o crescente volume de conteúdo textual gerado diariamente na *Web* torna-se um desafio considerável no que diz respeito à análise e organização. Dentro desse panorama, a adoção do método TF-IDF torna-se imprescindível para processar e compreender a vasta quantidade de informações disponíveis, viabilizando a identificação de padrões, tópicos pertinentes e discernimentos valiosos em meio a esse profundo oceano de dados. Para o TF-IDF cada documento é uma amostra e, cada palavra, um atributo. O resultado será uma matriz de termos-frequências, com dimensões: quantidade de documentos e quantidade de palavras (MILANI, 2020).

O procedimento do TF-IDF assume um papel crucial na realização da classificação dos textos adquiridos na etapa de extração de informações. O TF-IDF também é utilizado para a mineração de texto e sua principal utilidade é a descoberta de palavras importantes em um texto não estruturado ou semiestruturado (LIMA; CASTRO, 2012), atribuindo um peso a cada termo, considerando sua frequência no texto, e em todos os documentos do conjunto de dados indicando sua importância (MILANI, 2020).

Com a extração dos textos concluída, em seguida, os dados passam pelo método de filtragem TF-IDF, uma técnica que utiliza a frequência dos termos nos documentos. Para Salton, Yang (1973) TF-IDF é uma abordagem estatística usada para avaliar o grau de importância de uma palavra em um conjunto de documentos. A sigla se origina dos seus componentes: *term frequency* (TF) e *inverse document frequency* (IDF). O TF mede a importância de um termo para o documento, enquanto, o IDF mede a raridade de um termo em todo o conjunto de documentos. Segundo Matos, Sousa, Ruivo (2019, p. 8) quanto mais uma palavra aparecer nos documentos, mais próximo de 0 será o seu valor IDF, enquanto seu valor se aproxima de 1 caso seja mais raro. Um exemplo para este caso seria o termo “Jesus” ao analisar o Novo Testamento (CHOI et al., 2013).

### 2.5.2 Previsão do preço

A previsão desempenha um papel fundamental em diversas áreas, incluindo economia e finanças. Ela envolve o uso de modelos estatísticos e análise de dados para antecipar eventos futuros com base em informações disponíveis no presente. A precisão dessas previsões é crucial para tomadores de decisão, como analistas financeiros e investidores, que dependem delas para orientar suas estratégias de investimento.

Garcia (2013) foi examinado como o sentimento influenciou os preços dos ativos ao longo do século XX. O autor utilizou textos do *The New York Times* para desenvolver indicadores de sentimento por meio de análise de texto. Os resultados destacam que a habilidade de antecipar os retornos das ações com base no conteúdo das notícias é particularmente notável em momentos de recessão.

O método LASSO, conhecido como "*Least Absolute Shrinkage and Selection Operator*" (Operador de Redução Absoluta Mínima e Seleção), é como uma ferramenta da área de *Machine Learning* e análise de dados. O LASSO funciona como um "organizador" de informações que simplifica conjuntos de dados complicados, tornando mais fácil identificar o que realmente importa.

No estudo conduzido por Barbosa e Guimarães (2015), o método LASSO desempenha um papel fundamental na seleção de variáveis instrumentais eficazes para prever os retornos da educação em diferentes regiões do Brasil. O uso do LASSO ocorreu devido à carência de dados confiáveis, como uma opção viável e promissora para melhorar a estimação de equações relacionadas aos retornos educacionais.

## 3. METODOLOGIA

Este estudo tem como propósito a investigação da relação entre as notícias financeiras e o mercado financeiro nacional. Para a elaboração das variáveis dessa pesquisa, utilizou-se como base de dados as notícias dos cadernos de política e economia do G1. Tendo como principal objetivo compreender o comportamento do índice de ESG na bolsa brasileira.

Os dados das notícias financeiras foram obtidos por meio de *scraping* no site G1. *Scraping*, segundo Mitchell (2019) é coletar dados de maneiras diferentes da interação direta de pessoa com um navegador *web*. Ou seja, através de um programa automatizado que consulta os *sites* e então realiza a extração dos dados estruturados e não estruturados.

*Scraping*, ou raspagem de dados, é uma técnica utilizada para extrair informações de páginas da *web* de forma automatizada. Realizada por meio do Python, através da biblioteca *BeautifulSoup*, que percorre o código HTML coletando os dados especificados e formando o banco de dados necessários para a pesquisa. O objetivo desta fase é a extração do objeto de interesse, que consiste nos títulos encontrados nas páginas do portal G1.

Após a formação do banco de dados, a pesquisa avança para a etapa seguinte, que consiste em compreender o comportamento textual utilizando o *software* estatístico R Studio. Essa ferramenta, por meio do processamento de linguagem, tem como objetivo identificar a intensidade do tom otimista ou pessimista nas obras textuais.

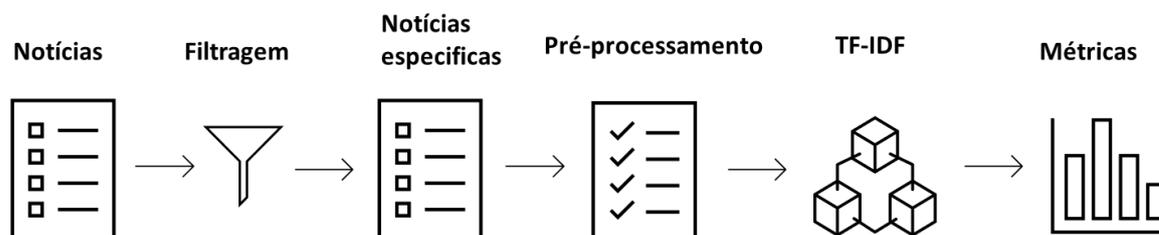
Para isso foi utilizado técnicas estatísticas no ambiente de programação R Studio, juntamente com o dicionário sentiLex\_PT02 de Silva et al. (2010), por meio do pacote LexinconPT do R, com o objetivo de analisar o tom textual das notícias.

### 3.1 Pré-processamento do texto com aplicação do TF-IDF

A análise de conteúdo foi conduzida por meio da aplicação do TF-IDF para ponderação das palavras nas notícias extraídas do G1. Esse procedimento envolveu a avaliação da relevância de cada termo, considerando o conjunto completo de termos observados. Esse processo visa atribuir pesos individuais a cada termo, permitindo uma compreensão mais precisa sobre quais termos se destacam em relação aos documentos de texto, com base em sua importância relativa.

O cálculo do TF-IDF é realizado multiplicando a frequência do termo (TF) pela frequência inversa do documento (IDF) para cada termo em um documento específico. A figura 1 demonstra os processos para a extração dos resultados:

**Figura 1:** Exploração e Aplicações do Processo TF-IDF na Análise de Texto



Fonte: Adaptação, Matos, Souza e Ruivo (2019).

### 3.2 Medidas de Sentimento Textual

Para fazer a previsão baseada em texto foi estimada duas medidas de sentimento textual, o sentimento textual do dia (SentNews) e o sentimento textual do dia, baseado no índice ESG (SentNewsESG). Para tanto seguiu-se os seguintes passos:

- 1) **Estimativa do *score* de sentimento para cada palavra.** Essa estimativa é obtida através da multiplicação do peso da palavra (contida no dicionário) pelo índice TF-IDF da palavra. Dessa forma, o sentimento textual da palavra tem relação com a polaridade e com a frequência relativa da palavra em cada documento.
- 2) **Criação da medida de sentimento para cada dia** baseada na média do dia do *score* sentimento de cada palavra.
- 3) **Criação de um Sentimento Textual ESG.** Essa medida é feita ao estimar um modelo de *machine learning* (*XGBoost*) no qual:
  - a) O retorno acionário do índice ESG é explicado pela matriz esparsa de palavras (1088 variáveis).
  - b) Após a estimativa do modelo é avaliado as palavras que têm mais importância na explicação do retorno do índice ESG.
  - c) E por fim, a partir do modelo estimado é obtido a série de Sentimento Textual ESG a partir do valor previsto pelo modelo *XGBoost* mencionado no item (a).

### 3.3 Modelos de Previsão com o uso de *Machine Learning*

Para conduzir a análise, foi adotada uma abordagem abrangente. Inicialmente, foi estimado vários modelos de *machine learning* por meio de um método de AutoML, utilizando a biblioteca PyCaret. Essa etapa nos permitiu explorar diversas opções e estratégias, automatizando o processo de seleção e ajuste de modelos.

Após a análise inicial, pode ser identificado a necessidade de otimizar ainda mais o desempenho do modelo. Nesse sentido, realizamos a estimativa de um modelo *XGBoost*, um algoritmo amplamente reconhecido e eficaz, com hiperparâmetros finamente ajustados. Essa etapa permitiu que o modelo atingisse um nível superior de precisão e capacidade de generalização, tornando-o ainda mais adaptado aos requisitos específicos de nossa tarefa analítica.

### **3.4 Avaliação dos Modelos de Previsão**

Na condução da análise, foi realizado o cálculo de diversas métricas a fim de comparar o desempenho dos modelos em questão. É importante destacar que, como referência, foi utilizado um modelo base, o qual se baseia exclusivamente na média histórica para realizar previsões dos retornos.

A métrica principal para avaliar a qualidade de previsão dos modelos é o RMSE (*Root Mean Square Error*), fornecendo uma medida do erro médio das previsões em relação aos valores reais. Através do RMSE, busca identificar qual dos modelos apresenta a menor discrepância entre as previsões e os resultados reais, o que é um indicativo da precisão do modelo.

Além disso, também foi realizada uma análise da importância das variáveis. Neste contexto, é avaliado a relevância das variáveis Média Histórica, SentNews e SentNewsESG. Essa análise permite identificar qual das variáveis desempenhou um papel mais significativo na qualidade das previsões dos modelos.

Assim, com base nas métricas calculadas e na análise de importância das variáveis, pode-se avaliar e comparar o desempenho dos modelos e determinar qual deles se destacou na previsão de retornos, bem como a importância relativa das variáveis em nossa análise.

## **4. ANÁLISE DOS RESULTADOS**

O objetivo deste estudo foi investigar a interação entre o conteúdo textual das notícias e as variações nos preços das ações do mercado de ações brasileiro. Para isso, empregamos dados relacionados ao índice ESG e avaliamos o sentimento associado às notícias que abordam assuntos econômicos e políticos nacionais.

As informações sobre notícias financeiras em formato de texto foram coletadas por meio de uma pesquisa nas reportagens publicadas no *site* G1, que estavam disponíveis nos seus cadernos de economia e política. Conduzimos essa coleta de dados no período de dezembro de 2020 a maio de 2023, o que resultou em uma amostra composta por 17.999 matérias.

A escolha de utilizar matérias econômicas e políticas do G1 como fonte de dados se justifica pela sua popularidade, seu sólido histórico de excelência na cobertura econômica em

âmbito regional e nacional, além de sua posição de destaque como um dos dez maiores portais de notícias do Brasil.

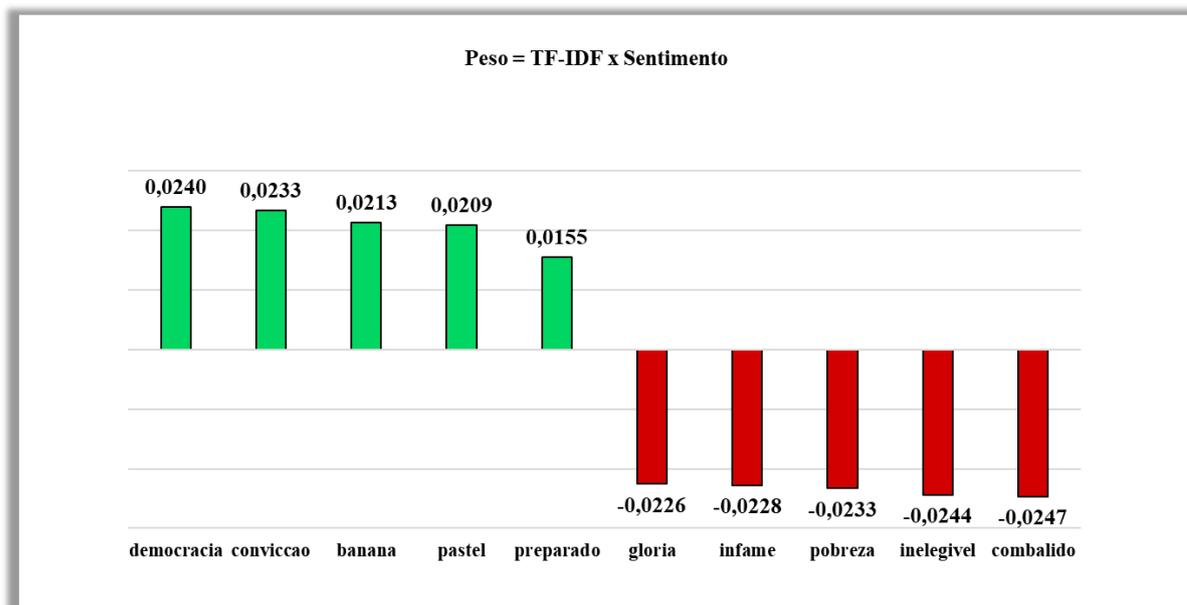
#### **4.1 Sentimento textual das notícias (SentNews)**

As notícias foram submetidas a uma análise de conteúdo, a fim de examinar a frequência e a ponderação das palavras usadas nos textos. Para isso, foi contabilizada a ocorrência de cada palavra em cada texto, de forma isolada. Em seguida, aplica-se o filtro TF-IDF, que avalia a importância relativa de cada termo com base em todos os dados coletados. Esse filtro atribui pesos aos termos, permitindo compreender quais termos são mais relevantes em relação aos documentos de texto.

Em seguida, é necessário determinar a polaridade das palavras com base em suas classificações de acordo com o dicionário Lexicon SentLex de Silva et. al. (2010). Por fim, calcula-se a estimativa do sentimento diário ao calcular a média diária do produto entre o peso TF-IDF e a polaridade de cada palavra avaliada. Permitindo classificar os sentimentos de todas as palavras, destacando as pontuações positivas e negativas das notícias.

No Gráfico 1 exibe as palavras que se destacaram com maior peso na amostra coletada. Observa-se uma diversidade nas palavras que se destacaram com maior peso, abrangendo tanto conotações positivas quanto negativas.

**Gráfico 1 - Análise de relevância das palavras**



Fonte: Dados da Pesquisa, 2023

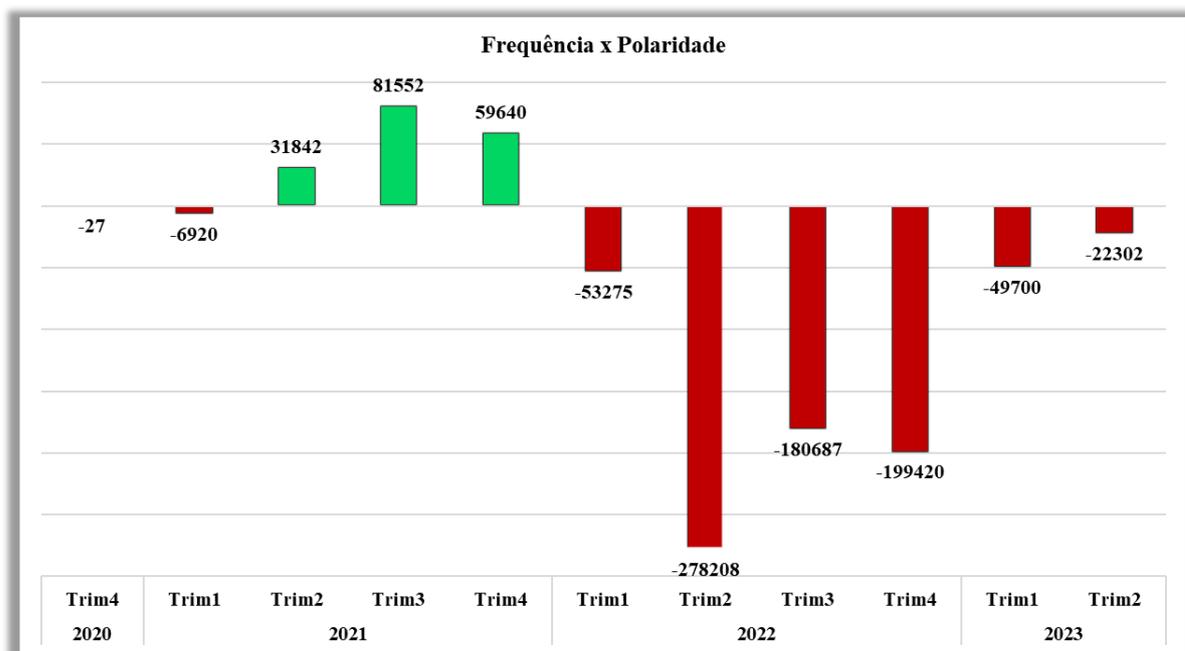
Com o intuito de fazê-lo, foi realizado um exame do significado e da conotação das palavras, resultando em uma lista de 5.130 palavras, verificando a presença de 2201 termos positivos, 2372 termos negativos e 557 termos neutros. Os termos que remetem a positividade que se destacam são: “democracia”, “convicção”, “banana”, “pastel” e “preparado”.

A elaboração da lista de termos negativos seguiu os mesmos critérios adotados na análise das palavras positivas e incluiu palavras associadas à negatividade, tais como: “glória”, “infame”, “pobreza”, “inelegível” e “combalido”. Além das palavras que expressam negatividade e positividade, também foram incluídos termos que remeteram à incerteza, como: “analista”, “jovem”, “agradecer”, “votar”, “permanente” dentre outras.

Também observamos palavras que parecem neutras, como "pastel" e "banana", no entanto, o dicionário utilizado atribui uma conotação positiva a esses termos. O aumento na classificação pode ser explicado pela frequência de uso dessas palavras.

O gráfico 2 exhibe o resultado da amostra das variáveis do estudo. Os dados foram agrupados de acordo com o sentimento textual da mídia e foram divididos por trimestre. Ao analisar o sentimento textual, é evidente que a mídia, de modo geral, demonstra uma tendência ao pessimismo em suas matérias, uma vez que -0,005277284 foi o sentimento médio.

**Gráfico 2** - Correlação entre Frequência e Polaridade na Análise de Textos



Fonte: Dados da Pesquisa, 2023

Com base na amostra, nota-se que, em média, as notícias analisadas na pesquisa apresentam uma predominância de sentimentos negativos. Através da análise, é evidente uma marcante alteração no sentimento textual das notícias ao longo do período de 2020 a 2021. Começando com um tom negativo no final de 2020 e no primeiro trimestre de 2021, houve uma virada significativa no segundo trimestre, que se acentuou ainda mais no terceiro trimestre, indicando uma tendência de aumento no otimismo nas notícias. O quarto trimestre ainda manteve um tom positivo, embora com uma leve queda em relação ao trimestre anterior.

Em resumo, o ano de 2022 começou com um tom negativo nas notícias no primeiro trimestre, com um agravamento significativo no trimestre seguinte. O terceiro trimestre mostrou uma melhora relativa, embora o sentimento negativo ainda prevalecesse, e o quarto trimestre continuou mantendo um tom desfavorável, embora com uma ligeira redução do pessimismo em comparação com o trimestre anterior.

O ano de 2022 foi marcado por um clima de negatividade nas notícias devido à persistência da pandemia da COVID-19 e às crises econômicas que afetaram diversos países. As flutuações na pandemia e os impactos econômicos geraram incertezas e ansiedade, refletidas

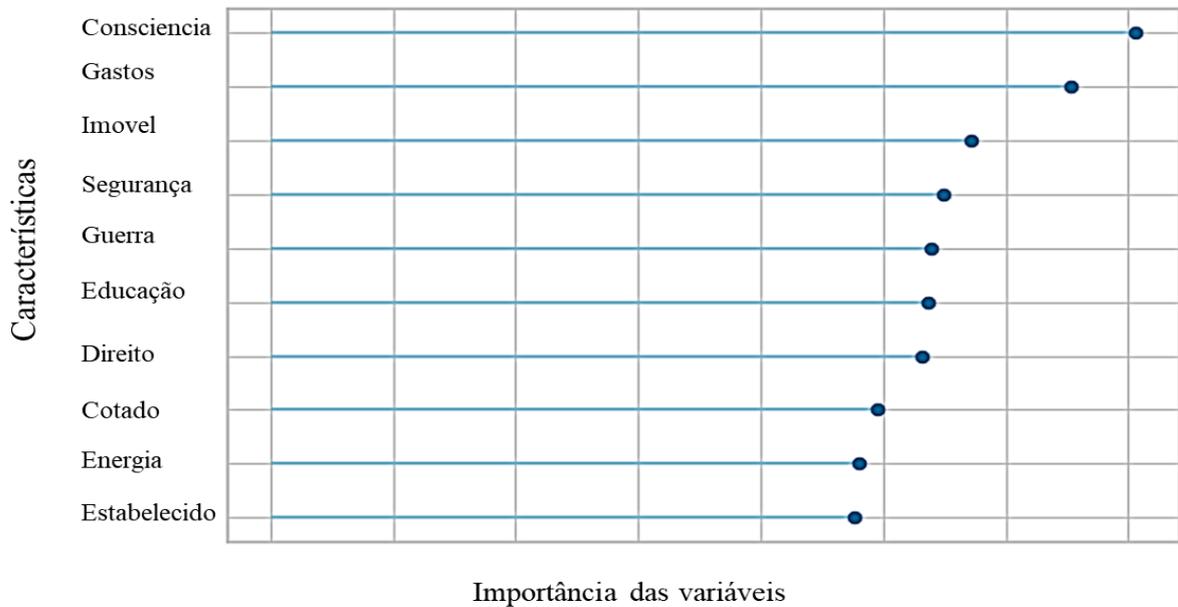
nas manchetes negativas e na cobertura intensa. Esses fatores interagiram, amplificando as preocupações e tornando o ano desafiador para a sociedade global.

O período de transição de 2022 para 2023 manteve um tom negativo nas notícias, com uma melhoria relativa no segundo trimestre de 2023 em relação aos trimestres anteriores, embora o sentimento negativo ainda seja predominante. A melhoria relativa no segundo trimestre de 2023 sugere que pode haver sinais de mudança do tom das notícias, mas ainda é cedo para determinar uma mudança completa no sentimento geral das notícias.

#### **4.2 Sentimento textual condicional ao ESG (SentNewsESG)**

Nesta subseção, é realizada uma análise do Sentimento Textual das Notícias, condicionada ao Índice ESG (SentNewsESG). Resumidamente, analisa uma medida de sentimento que utiliza técnicas de *machine learning* para, ao longo do tempo, identificar quais palavras são mais relevantes na explicação do retorno das ações. Foi estimado um modelo XGBoost, que é uma variante de modelo baseado em árvores de decisão, com esse propósito. Nessa etapa foram utilizados apenas os dados textuais para tentar explicar o retorno, mas sem agregação, ou seja, cada palavra foi colocada como uma variável preditora formando um modelo com mais de 1000 preditores. Na Figura 2 é possível analisar as palavras mais relevantes na explicação do retorno.

**Figura 2** – Análise da Importância das Variáveis em Relação às Características



Fonte: Dados da Pesquisa, 2023

Na Figura 4, é possível realizar uma análise mais minuciosa das palavras de maior significado na previsão do retorno do índice ESG. A palavra "consciência" se destaca como a de maior relevância, sugerindo que é um fator crucial na compreensão desse retorno. Em segundo lugar, encontramos a palavra "gastos". Seguindo essas palavras mais relevantes, observamos um grupo de termos, "imóvel", "segurança", "guerra", "educação" e "direito", com níveis de relevância bastante similares. Em um patamar ligeiramente inferior, temos "cotado", "energia" e "estabelecimento", mais uma vez com importância semelhante.

Ao analisar a importância de certas palavras na previsão do retorno deste índice, podemos identificar conexões significativas. A palavra "consciência" destaca a importância da conscientização sobre questões ESG. Os "gastos" também se destacam, pois representam o compromisso financeiro. Em seguida, palavras como "imóvel", "segurança", "guerra", "educação" e "direito", refletem a diversidade de critérios ESG, abrangendo desde questões ambientais, como eficiência energética, até questões sociais, como a segurança dos trabalhadores e o acesso à justiça.

Em um nível ligeiramente inferior de importância, temos "cotado", "energia" e "estabelecimento", que indicam a influência do mercado financeiro e da infraestrutura nas

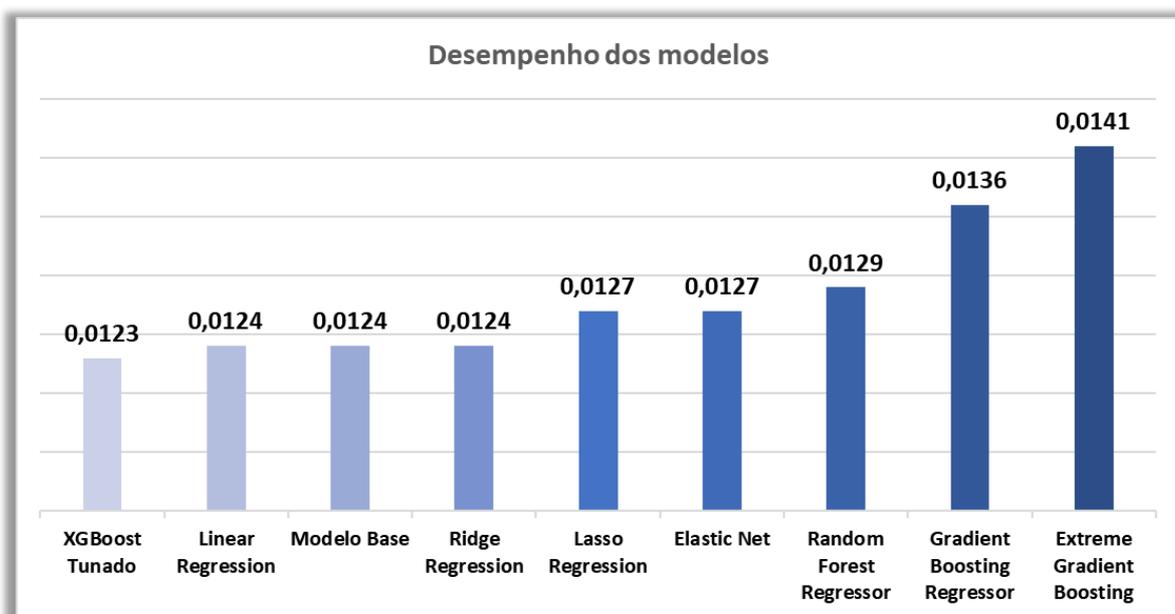
pontuações do índice ESG. Empresas que utilizam fontes de energia sustentável e mantêm práticas de gestão de instalações eficientes podem obter vantagens.

De maneira geral, as palavras que o modelo selecionou como importante parecem realmente ter uma relação com o tema da sustentabilidade. Este resultado é um indicativo de que o modelo realmente está captando essa relação e que o sentimento textual (SentNewsESG) obtido a partir do valor previsto do modelo pode realmente conter as palavras do noticiário relacionadas ao tema da sustentabilidade.

### 4.3 Análise de previsão do retorno acionário baseado no índice ESG

Após a estimação do sentimento textual das notícias (SentNews) e do sentimento textual condicional ao índice ESG (SentNewsESG), foi feita uma análise de vários modelos preditivos onde as variáveis Média Histórica, SentNews e SentNewsESG, que foram utilizados como preditores do retorno acionário do IBOVESPA no dia seguinte. Nesse sentido, o gráfico 3 apresenta métricas de desempenho de diferentes modelos de regressão em um contexto de previsão, com o foco principal na métrica de erro RMSE.

Gráfico 3 - Avaliação de Desempenho de Modelos



Fonte: Dados da Pesquisa, 2023

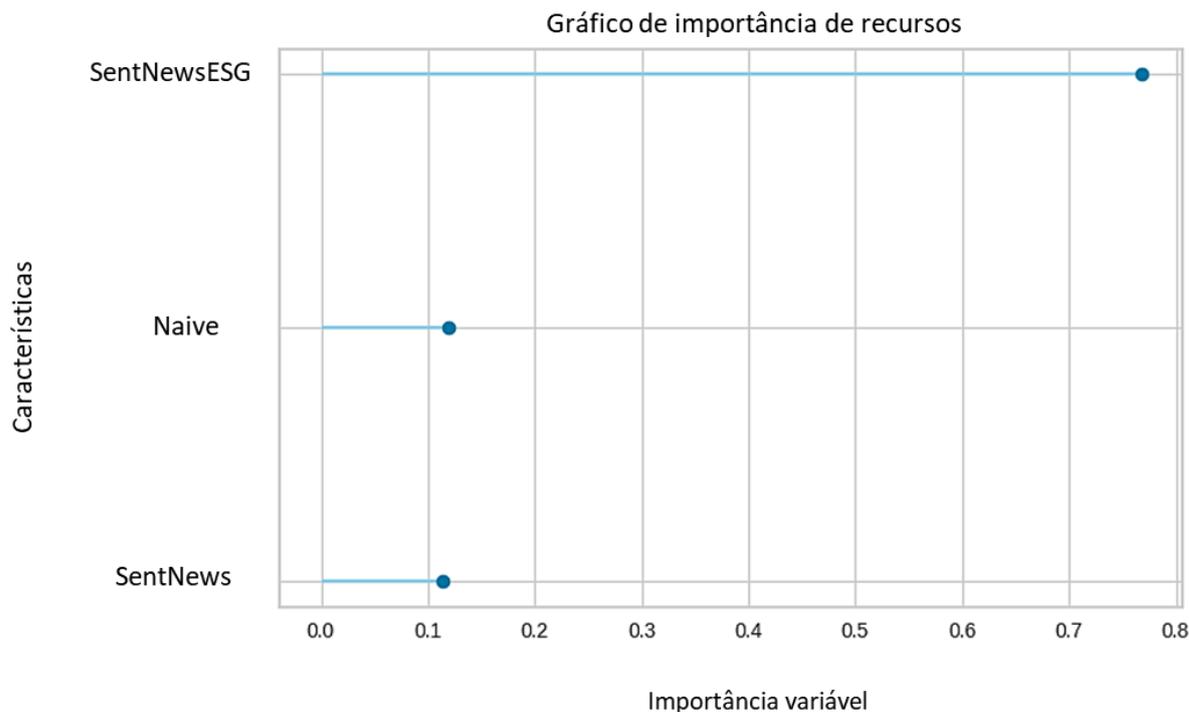
Com base nas métricas apresentadas, o modelo XGBoost Tunado se apresenta como uma escolha promissora para a tarefa de regressão. Com o menor valor de RMSE, uma métrica amplamente aceita na avaliação de modelos de regressão, desempenhando melhor em comparação aos demais modelos, minimizando os erros médios entre as previsões e os valores reais. Isso é um indicativo sólido de que o XGBoost Tunado tem um desempenho levemente superior em relação às outras abordagens avaliadas.

O XGBoost (*Extreme Gradient Boosting*) é uma abordagem sofisticada no campo de *machine learning* que se baseia na construção de árvores de decisão. O que torna o XGBoost particularmente notável é sua capacidade de empregar uma técnica chamada "*boosting*". Nesse processo, diversas árvores de decisão, individualmente menos poderosas, são combinadas de forma inteligente para criar um modelo global substancialmente mais robusto e preciso.

Já o XGBoost Tunado é o modelo que ajusta seus hiperparâmetros para otimizar seu desempenho. A intuição por trás disso é encontrar a combinação ideal de configurações que permita ao modelo aprender melhor a partir dos dados, melhorando assim sua capacidade de fazer previsões ou tomar decisões. É como ajustar as configurações de um carro para obter o melhor desempenho possível.

A Figura 3 exibe os resultados da comparação entre os modelos: o modelo simples, o modelo com a inclusão de análise de sentimento e o modelo com a utilização das palavras.

**Figura 3** - Análise Comparativa dos Modelos Preditivos



Fonte: Dados da Pesquisa, 2023

A figura 5 apresenta o comparativo entre os modelos simples, modelo com palavras e o modelo com sentimento. O modelo com palavras é representado pelo XGBoost Tunado, deixando evidente que o modelo melhorado por meio das melhorias utilizando as características "palavras" supera consideravelmente o modelo base. Dessa forma, o XGBoost Tunado se destaca como uma opção mais avançada e potente em comparação aos demais.

O modelo aprimorado com a inclusão das características "palavras" destaca-se como particularmente adequado para lidar com desafios de natureza complexa e conjuntos de dados de alta dimensionalidade. Em situações em que a simplicidade do modelo base pode se mostrar insuficiente para capturar as relações intrínsecas nos dados, o modelo enriquecido com informações das "palavras" revela sua capacidade de fornecer *insights* mais detalhados e precisos. Essa característica o transforma em uma opção de grande valor em situações em que uma análise minuciosa e uma compreensão profunda das relações nos dados são cruciais para o sucesso da atividade em pauta.

## 5. CONCLUSÃO

Neste trabalho, foi analisada a relação entre o sentimento textual das notícias de economia e política do G1 e as variações nos preços das ações no mercado financeiro brasileiro. Para tanto, foi estimada uma medida de sentimento textual das notícias baseada na relação entre as palavras contidas no texto e a movimentação do índice ESG. Ao explorar essa relação, buscou-se uma forma de compreender como o movimento do mercado acionário pode ser explicado por informações contidas no texto das notícias que têm alguma relação com as relevantes questões relacionadas ao meio ambiente, sustentabilidade e governança.

Na análise do sentimento textual, observou-se a prevalência de um sentimento predominantemente negativo nas notícias, embora tenham ocorrido variações significativas ao longo desse período. Essas flutuações no sentimento textual refletiram os eventos econômicos, políticos e sociais que moldaram o contexto financeiro brasileiro nesse período. A análise do sentimento textual se mostrou uma ferramenta valiosa para compreender como as notícias afetam a percepção e o comportamento dos investidores no mercado financeiro.

Além disso, o estudo explorou diferentes modelos de previsão de preços, incluindo um modelo simples baseado na média histórica (*naive*) e alguns modelos de *machine learning* como LASSO e XGBoost. O modelo base, ou modelo *naive*, serviu de *benchmarking* de forma que fosse possível avaliar como os modelos propostos aqui podem contribuir no poder informacional. De maneira geral, os modelos de *machine learning*, como o LASSO e o XGBoost, mostraram resultados promissores, indicando que a análise do sentimento textual pode ser uma abordagem eficaz para prever movimentos de preços no mercado financeiro.

Os resultados obtidos realçam a importância da seleção cuidadosa do modelo e do método de análise de sentimento textual em aplicações financeiras. O XGBoost Tunado, em particular, se destaca como a escolha preferencial para tarefas de regressão, ao apresentar o menor RMSE, demonstrando um desempenho superior em comparação ao modelo base, oferecendo uma previsão mais precisa e consistente em comparação com outros modelos considerados.

Dessa forma, abre-se um leque de oportunidades para pesquisadores do mercado que buscam *insights* confiáveis e acurados para embasar suas decisões financeiras. É importante destacar que, dada a flexibilidade desses modelos e a possibilidade de ajustar diferentes configurações, outros profissionais e pesquisadores podem explorar abordagens e parâmetros alternativos para obter novos resultados e perspectivas, contribuindo para o constante avanço na compreensão e previsão dos mercados financeiros. Essa adaptabilidade oferece um campo fértil para o desenvolvimento contínuo de estratégias financeiras baseadas em análise de sentimento textual.

No contexto da análise métrica realizada, é relevante ressaltar a importância da aplicação de técnicas de *machine learning* na extração de informações das notícias veiculadas pelo G1 e do índice ESG de forma que contribuíssem na explicação do comportamento do mercado acionário brasileiro. A utilização dessas técnicas possibilitou uma análise mais abrangente, permitindo a identificação de padrões, tendências e a comparação de modelos. Evidenciando como a integração de *machine learning* com a análise de conteúdo e a avaliação de sentimentos em textos podem proporcionar informações preciosas, melhorando a tomada de decisões embasadas e enriquecendo a compreensão do comportamento do mercado de ações.

É válido ressaltar que esta pesquisa possui limitações em termos do curto período da amostra, tendo em vista que o código de *Web Scraping* permitiu alcançar até dezembro de 2020. Para uma compreensão mais completa e robusta do comportamento do mercado, futuras pesquisas podem considerar a ampliação do período de coleta de dados e a inclusão de informações mais atualizadas, a fim de abranger um espectro mais amplo de cenários e tendências financeiras.

Portanto, este estudo destaca a importância do sentimento textual nas notícias financeiras como um fator que influencia a dinâmica do mercado financeiro brasileiro, reforçando a definição de Fama (1970) sobre um mercado eficiente, onde os preços refletem todas as informações disponíveis. A aplicação de técnicas de *machine learning* nesse contexto torna-se um aliado valioso, possibilitando a análise de grandes volumes de dados de texto de forma eficiente e a extração de padrões subjacentes que podem ser usados na previsão de movimentos de preços das ações. Assim, a pesquisa ilustra o potencial transformador da inteligência artificial e do *machine learning* na análise e previsão do mercado financeiro,

reforçando a relevância do sentimento textual como uma fonte valiosa de informações para a tomada de decisões nos mercados globais.

## REFERÊNCIAS

B3. **Índice Carbono Eficiente (ICO2 B3)**, 2021 Disponível em: [https://www.b3.com.br/pt\\_br/noticias/ico2-b3.htm#:~:text=Criado%20em%202010%2C%20o%20ICO2,ações%20mais%20líquidas%20na%20B3](https://www.b3.com.br/pt_br/noticias/ico2-b3.htm#:~:text=Criado%20em%202010%2C%20o%20ICO2,ações%20mais%20líquidas%20na%20B3). Acesso em junho de 2023.

B3. **Índice de Sustentabilidade Empresarial (ISE) B3**, 2023 Disponível em: [https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-de-sustentabilidade/indice-de-sustentabilidade-empresarial-ise-b3.htm#:~:text=O%20objetivo%20do%20ISE%20B3,comprometimento%20com%20a%20sustentabilidade%20empresarial](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-de-sustentabilidade/indice-de-sustentabilidade-empresarial-ise-b3.htm#:~:text=O%20objetivo%20do%20ISE%20B3,comprometimento%20com%20a%20sustentabilidade%20empresarial). Acesso em junho de 2023.

Barbosa, R. B., & Guimarães, D. B. **Estimando retornos da educação no Ceará: uma aplicação do método do LASSO para seleção de variáveis instrumentais fracas**. In: ENCONTRO ECONOMIA DO CEARÁ EM DEBATE, 11., 2015, Fortaleza. Anais... Fortaleza: IPECE, p. 1-17, 2015

Blueprint. **Environmental, Social, and Governance (ESG) Investing**. *Environmental, Social, and Governance (ESG) Investing*, v. 07030, n. 201, p. 9930, 2020

BOFINGER, Yannik; HEYDEN, Kim J.; ROCK, Björn. **Corporate social responsibility and market efficiency: Evidence from ESG and misvaluation measures**. *Journal of Banking & Finance*, v. 134, p. 106322, 2022.

Cavalcanti, A. P. **Uma medida de similaridade textual para identificação de plágio em fóruns educacionais**. Recife, 2018

Choi, D., Ko, B., Kim, H., & Kim, P. **Text analysis for detecting terrorism-related articles on the web**. *Journal of Network and Computer Applications*, ISSN 1084-8045, 2013

Eccles, G. R., & Klimenko, S. **The Investor Revolution**. *Harvard Business Review*, 2019 Disponível em: <https://hbr.org/2019/05/the-investor-revolution>. Acesso em maio de 2023.

Garcia, D. **Sentiment during recessions**. *The Journal of Finance*, 68(3), 1267-1300, 2013.

Junior, T. S. R., & Ikeda, H. R. **Mercados eficientes e arbitragem: um estudo sob o enfoque das finanças comportamentais**. *Revista Contabilidade & Finanças - USP*, vol. 15, núm. 34, 2004.

Kitchener, S., & Rapoport, I. **Capturando ESG no Brasil: O Índice S&P/B3 Brasil ESG**. 2020.

LOUGHRAN, Tim; MCDONALD, Bill. **Textual analysis in finance**. *Annual Review of Financial Economics*, v. 12, p. 357-375, 2020.

LOUGHRAN, Tim; MCDONALD, Bill. **Textual analysis in accounting and finance: A survey**. Journal of Accounting Research, v. 54, n. 4, p. 1187-1230, 2016.

Lucena, W. G. L., Santos, J. M. A., Assis, J. T., & Santos, M. C. **Fatores que influenciam o endividamento e a inadimplência no setor imobiliário da cidade de Toritama-PE à luz das finanças comportamentais**, p. 4, 2014.

Marcondes, A. W., & Bacarji, C. D. **ISE Sustentabilidade no mercado de capitais**. 1ª ed. São Paulo: Report, 2010.

Matos, B. M., Sousa, M. de O., & Ruivo, E. L. P. **Aplicação do TF-IDF para Classificação de Notícias ESG Relativas a Empresas**. Universidade Presbiteriana Mackenzie (UPM). Endereço: Rua da Consolação, 930, Consolação, São Paulo - SP, 01302-907, Brasil, 2019.

Milani, T. G. **Aprendizado de máquina para detecção de spam: Um estudo comparativo de algoritmos de mineração de texto e classificadores**. São Paulo, 2020.

Mitchell, R. **Web Scraping com Python: Coletando mais dados da web moderna**. Novatec Editora. ISBN 9788575227343, 2019.

Neves, B. B. **Fundos de Investimento Socialmente Responsáveis: Uma análise empírica da rentabilidade**, Rio de Janeiro. 2009.

Niemoller, J. **ESG and Sustainability – What’s the difference?** PeaSoup Cloud, 2021. Disponível em: <https://peasoup.cloud/2021/06/14/esg-and-sustainability-hats-the-difference/>. Acesso em: maio de 2023.

Ross, S. A., Westerfield, R. W., & Jaffe, J. F. **Administração financeira: corporate finance**. 10ª ed. São Paulo: McGraw-Hill, 2014.

Salton, G., Yang, C. S. **On the specification of term values in automatic indexing**. Journal of documentation, 29(4), 351-372, 1973.

Silva, C. A. T., Oliveira, J. D. V. **O que move o mercado acionário brasileiro?**. REACE, Unoesc, v. 10, n. 2, p. 295-312, jul./dez, 2011.

SILVA, Mário J. et al. **Automatic expansion of a social judgment lexicon for sentiment analysis**, 2010.