

FinBERT-PT-BR: Análise de Sentimentos de Textos em Português do Mercado Financeiro

FinBERT-PT-BR: Sentiment Analysis of Texts in Portuguese from the Financial Market

Lucas L. Santos¹, Reinaldo A. C. Bianchi², Anna H. R. Costa¹

¹Escola Politécnica – Universidade de São Paulo (USP)

²Centro Universitário da FEI

lucaslssantos99@usp.br, rbianchi@fei.edu.br, anna.reali@usp.br

Abstract. *This article contributes a sentiment analysis model for financial news in Portuguese using the BERT neural network architecture. The model was trained in two stages: language modeling and sentiment modeling, with 1.4 million texts and 500 labeled texts, respectively. The model showed better performance than current state-of-the-art models across various metrics and can be used to build sentiment indices, investment strategies, and analyze macroeconomic data. The study demonstrates the potential of natural language processing and transformers for quantitative finance.*

Resumo. *Este artigo contribui com um modelo de análise de sentimento para notícias financeiras em língua portuguesa usando a arquitetura de rede neural BERT. O modelo foi treinado em duas etapas: modelagem de linguagem e modelagem de sentimentos, com 1,4 milhão de textos e 500 textos rotulados, respectivamente. O modelo apresentou melhor desempenho do que os modelos atuais do estado da arte em diversas métricas e pode ser usado para construir índices de sentimento, estratégias de investimento e analisar dados macroeconômicos. O estudo demonstra o potencial do processamento de linguagem natural e transformers para finanças quantitativas.*

1. Introdução

O mercado de capitais é vivo e dinâmico, assim como dita a hipótese de mercados adaptativos [Lo 2004], em que os preços dos ativos são adaptativos e refletem informações e emoções da população. Os agentes que atuam nessa área devem, portanto, ter uma excelente capacidade de processamento de informações, a fim de tomar as melhores decisões, evitando vieses comportamentais. Um sistema que capture as emoções refletidas nos textos divulgados a respeito do mercado financeiro podem auxiliar nesta tomada de decisão.

Atualmente, os algoritmos no estado da arte em processamento de linguagem natural (PLN) são baseados em redes neurais profundas com mecanismos de atenção [Vaswani et al. 2017] e atingem alto desempenho em diversas tarefas [Devlin et al. 2018]. Entretanto, a maioria desses modelos são treinados na língua inglesa e são modelos de propósito geral, os quais podem não apresentar resultados tão bons por conta do vocabulário específico do mercado financeiro. Portanto, para atingir resultados no estado da arte em finanças, o ideal é treinar os modelos com textos específicos do domínio e língua desejados [Araci 2019].

Assim, o objetivo deste artigo é apresentar um modelo de linguagem do estado da arte no contexto do mercado financeiro em português do Brasil, FinBERT-PT-BR, além de efetuar um ajuste fino para a tarefa de análise de sentimentos, contribuindo com o modelo SentFinBERT-PT-BR. Por fim, mostramos que o SentFinBERT-PT-BR possibilita a construção de sinais para análise e construção de estratégias de investimentos. Para atingir este objetivo, primeiramente fez-se a coleta de dados textuais extraídos de notícias de finanças de veículos de comunicação. Parte destes dados foi então rotulada manualmente para o treinamento do modelo SentFinBERT-PT-BR, avaliando o sentimento expresso nas notícias em três classes: positivo, negativo e neutro. O modelo treinado foi então avaliado e comparado com outros modelos, mostrando desempenho superior.

O restante do artigo está organizado da seguinte forma: a Seção 2 apresenta os conceitos básicos referentes à PLN e análise de sentimentos; a Seção 3 descreve o desenvolvimento dos modelos FinBERT-PT-BR e SentFinBERT-PT-BR. A Seção 4 mostra nossos experimentos e finalmente as conclusões e alguns direcionamentos futuros são apresentados na Seção 5.

2. Conceitos Básicos

O processamento de linguagem natural (PLN) [Manning and Schütze 1999] é uma área da inteligência artificial que se concentra em permitir que computadores compreendam e processem a linguagem humana. Ele envolve a análise de texto escrito ou falado para que os computadores possam entender a intenção, o significado e a estrutura das informações.

A análise de sentimentos [Pang and Lee 2004, Liu 2012] é uma técnica de PLN utilizada para identificar e extrair informações sobre as emoções expressas em um texto, como positividade, negatividade ou neutralidade. O objetivo é entender como as pessoas se sentem em relação a um determinado assunto ou produto, por exemplo. Para realizar a análise de sentimentos, são utilizados algoritmos de PLN que examinam o texto em busca de palavras ou expressões que indiquem emoções. Esses algoritmos utilizam técnicas de aprendizado de máquina e mineração de dados para identificar padrões e classificar o texto em categorias de sentimentos. A análise de sentimentos pode ser aplicada em várias áreas, como marketing, atendimento ao cliente, política e ciências sociais.

Especificamente para a língua portuguesa, a análise de sentimentos tem sido utilizada por diversos autores nas mais diversas tarefas: [Silva 2018] analisa as percepções da população sobre corrupção durante as eleições presidenciais no Brasil em 2018 utilizando mensagens postadas no Twitter; [Pereira 2019] também realizou uma análise de sentimentos da população brasileira em relação a eleição presidencial de 2018 utilizando Twitter; [Xavier et al. 2020] apresentam resultados de estudos demonstrando como a análise de dados de redes sociais pode ser utilizada nas atividades de vigilância, tendo como estudo de caso a pandemia da Covid-19.

[Souza et al. 2020] propuseram a criação e avaliação de modelos de linguagem pré-treinados baseados na arquitetura BERT para a língua portuguesa do Brasil, utilizando uma grande quantidade de dados textuais em português brasileiro para treinar os modelos de linguagem. Os autores também avaliaram os modelos em uma variedade de tarefas de PLN, incluindo análise de sentimentos, classificação de textos e previsão de sentenças seguintes. Os resultados mostraram que os modelos pré-treinados BERTimbau superaram modelos treinados a partir do zero e outros modelos pré-treinados disponíveis

publicamente. Entre as conclusões dos autores, eles indicam que o BERTimbau devem ser útil para uma variedade de aplicações, incluindo análise de sentimentos no mercado financeiro, o que motivou o seu uso no trabalho aqui proposto.

A análise de sentimentos para tarefas relativas ao mercado financeiro [Man et al. 2019], como predição de tendências e de valores de ações e outras commodities, surgiu na década passada, e é uma área bastante pesquisada na interseção entre IA e Mercados Financeiros. Um dos artigos mais importantes da área, escrito por Bollen, Mao, e Zeng, demonstra que o sentimento coletivo expresso no Twitter pode prever a direção do mercado de ações dos EUA com uma precisão de até 87,6 % [Bollen et al. 2011]. Diversos outros trabalhos seguiram esta linha de análise, utilizando a relação entre o sentimento expresso no Twitter e o comportamento do mercado de ações [Pagolu et al. 2016, Kordonis et al. 2016], o preço de criptomoedas [Otabek and Choi 2022, Kraaijeveld and De Smedt 2020], o preço do ouro [Junjie and Mengoni 2020] e a taxa de câmbio Euro-Dólar [Ardia et al. 2015]. Finalmente, uma revisão mais recente pode ser encontrada em [Tan et al. 2023].

Embora existam poucos estudos específicos que examinem a relação entre análise de sentimento e o mercado financeiro brasileiro, podemos citar algumas referências de artigos recentes que utilizam PLN para análise de sentimento no mercado financeiro em língua portuguesa: [Medeiros and Borges 2019] realizam uma análise do uso de sentimento em tweets de mercado financeiro em Português, de forma similar a [de Souza et al. 2021] e [Januário et al. 2022], que além de tweets utilizam também outras fontes de notícias.

3. Desenvolvimento dos Modelos

Neste artigo propomos dois modelos, FinBERT-PT-BR e SentFinBERT-PT-BR. O modelo FinBERT-PT-BR é treinado de modo semi-supervisionado, no contexto de mercado financeiro, a partir da sintonia fina (*fine-tuning*) de um modelo de escopo geral treinado para a língua portuguesa, o BERTimbau [Souza et al. 2020]. O modelo FinBERT-PT-BR é então treinado com uma base de dados anotada de notícias do mercado financeiro para gerar o modelo SentFinBERT-PT-BR, que é um classificador de sentimentos das sentenças de notícias da área financeira. A Figura 1 ilustra esta metodologia.

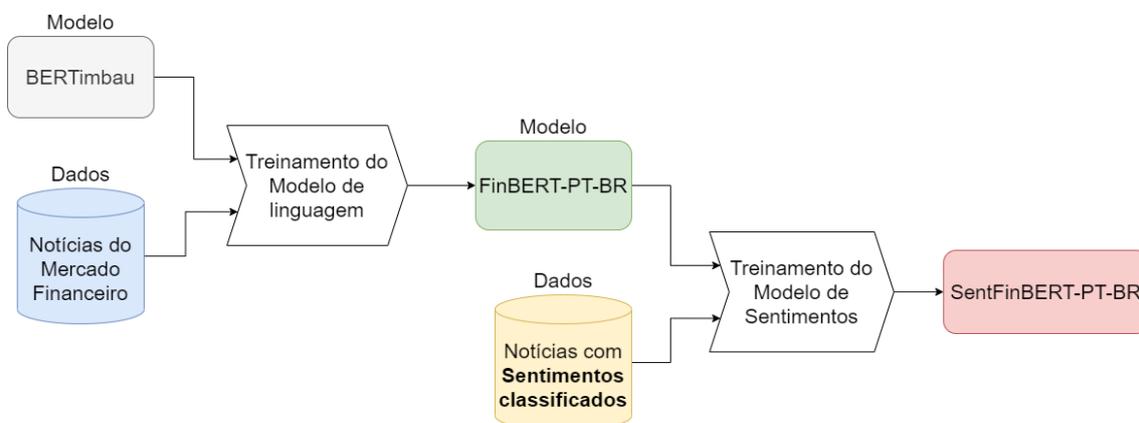


Figura 1. Arquitetura de treinamento do FinBERT-PT-BR e do SentFinBERT-PT-BR.

3.1. Desenvolvimento do Modelo FinBERT-PT-BR

O modelo FinBERT-PT-BR foi gerado a partir do *fine-tuning* do BERTimbau [Souza et al. 2020], com notícias coletadas de canais de notícias financeiras das plataformas *Valor Econômico*, *Exame* e *Infomoney*.

O processo de aquisição foi automatizado por meio da ferramenta *Scrapy*, utilizada para percorrer os sítios de notícias financeiras e coletar os textos e metadados das notícias, como título, subtítulo, data de publicação, data de atualização da publicação, nome e página do autor e *links*. No total foram coletadas 2,7 milhões de sentenças de notícias do mercado financeiro entre 2006 e 2022. A distribuição dos textos coletados por plataforma é apresentada na Tabela 1.

Tabela 1. Número total de 2,7 milhões de textos financeiros coletados, englobando 130 milhões de palavras.

Valor Econômico	1,23 milhões
Exame	1,01 milhões
Infomoney	0,46 milhões

Após a coleta dos textos foi necessário realizar uma limpeza dos dados utilizando expressões regulares que identificaram padrão de textos que não eram relevantes para o modelo, como, por exemplo, textos com má formação que possuíam caracteres especiais e códigos fonte, resultando em uma base final com 1,6 milhão de sentenças.

FinBERT-PT-BR foi treinado utilizando *PyTorch* e *Hugging Face*, tendo como pesos iniciais os pesos do modelo BERTimbau [Souza et al. 2020]. Como o modelo BERT suporta até 512 *tokens* na sua entrada, foi necessário realizar uma filtragem dos textos para que fossem utilizados apenas os textos com até 512 *tokens*, resultando em 1.428.867 sentenças de notícias.

O treinamento foi realizado em uma máquina do *Kaggle* com 30GB de RAM e 30GB de GPU (2x Nvidia T4), o que limitou o tamanho do lote (*batch size*) e o carregamento de todos os textos em memória. O *batch size* foi definido como 16, por limitação de alocação do modelo e dos textos na GPU simultaneamente. Devido à abundância de textos foi necessário implementar o processo de treinamento com alocação dinâmica de memória, ou seja, ao longo do treinamento os textos são carregados do disco para a memória e depois para a GPU. Apesar das limitações, foi possível realizar o treinamento de duas épocas em 11 horas.

Para o treinamento do modelo de linguagem, os dois principais parâmetros que não dependem diretamente de recursos computacionais são: taxa de aprendizado (*learning rate*) e probabilidade de máscara. A taxa de aprendizado foi definida como $2e-5$, o valor recomendado pela literatura [Sun et al. 2019] para o *fine tuning* de modelos de linguagem. A probabilidade de máscara é a probabilidade da ocultação de *tokens* e foi definida como 15%, sendo a mesma utilizada no treinamento do modelo BERT original [Devlin et al. 2018].

Para avaliar o modelo de linguagem foi utilizada a **métrica de perplexidade** [Chen et al. 1998], que indica o quão bem o modelo consegue prever uma palavra dado o contexto. Essa métrica foi calculada numa amostra de 100 mil sentenças, que não foram

utilizadas no treinamento do modelo. FinBERT-PT-BR apresentou uma perplexidade de 1,24, sendo uma métrica satisfatória, visto que o modelo BERTimbau original apresenta uma perplexidade de 1,51.

3.2. Desenvolvimento do Modelo SentFinBERT-PT-BR

O modelo SentFinBERT-PT-BR recebe um texto como entrada e classifica em uma das três categorias: positivo, negativo ou neutro. Para isso, SentFinBERT-PT-BR foi treinado com uma base de dados anotada manualmente por três pessoas, sendo duas com formação em engenharia e uma com formação em linguística. Cada um dos textos foi anotado por pelo menos duas pessoas.

No total foram anotados 1000 textos, sendo 497 textos descartados, pois foram classificados como “Não se aplica” ou não houve concordância entre os anotadores. Após a remoção dos textos descartados, a base de treinamento ficou com 503 textos, sendo 160 textos positivos, 203 textos negativos e 140 textos neutros.

A concordância entre os anotadores foi verificada utilizando a porcentagem de concordância e o *Krippendorff's alpha* [Artstein and Poesio 2008].

A porcentagem de concordância varia entre 0% e 100%, sendo 100% a concordância perfeita e, para N textos, é dada por:

$$\% \text{Concordância} = \frac{1}{N} \sum_{i=1}^N agr_i, \text{ com} \quad (1)$$

$$agr_i = \begin{cases} 1, & \text{Se os avaliadores classificaram o texto na mesma categoria} \\ 0, & \text{Se os avaliadores não classificaram o texto na mesma categoria} \end{cases} \quad (2)$$

Por conta da sua simplicidade, a porcentagem de concordância possui algumas desvantagens como: viés para bases de dados desbalanceadas e baixa robustez para mais de 2 avaliadores. A métrica obtida na base final de anotação foi de 90,4%.

Já o *Krippendorff's alpha* [Krippendorff 2018] varia entre -1 e 1 , sendo 1 a concordância perfeita, -1 a discordância perfeita e 0 representa anotação aleatória e é baseada em duas principais variáveis, D_o que é a variância amostral das respostas dos anotadores e D_e que é a variância esperada dos anotadores computada a partir das frequências das classes presentes na base dados:

$$\alpha = 1 - \frac{D_o}{D_e}. \quad (3)$$

Quando todos os avaliadores concordam, a variância amostral será $D_o = 0$, resultando assim num $\alpha = 1$, indicando uma alta confiabilidade na base de anotações. Um $\alpha = 0$ ocorre quando não possível distinguir as anotações de dados puramente aleatórios, pois a variância amostral das classificações será alta, indicando assim uma baixa confiabilidade. A métrica obtida na base final de anotação foi de $\alpha = 0,88$, indicando que a anotação foi realizada consistentemente entre os anotadores. Os resultados das duas métricas estão mostrados na Tabela 2.

O modelo de sentimentos SentFinBERT-PT-BR foi treinado utilizando o modelo de linguagem FinBERT-PT-BR pré-treinado com os textos do mercado financeiro, porém

Tabela 2. Métricas de concordância entre os anotadores.

Métrica	Valor
Porcentagem de concordância	90,4%
<i>Krippendorff's alpha</i> (α)	0,88

com a adição de uma camada de classificação de sentimentos na arquitetura da rede neural. Essa camada foi adicionada na primeira dimensão de saída do BERT, seguindo a recomendação da literatura [Devlin et al. 2018], conforme ilustra a Figura 2.

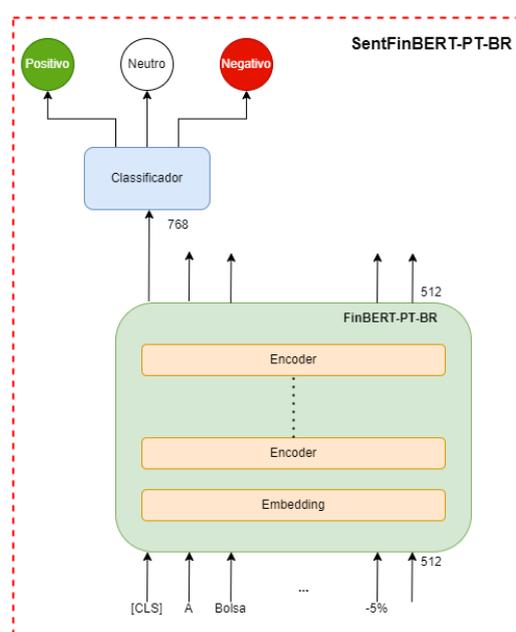


Figura 2. Arquitetura do modelo de classificação de sentimento.

O processo de utilizar o modelo de linguagem FinBERT-PT-BR para a tarefa de análise de sentimento é caracterizado como um *transfer learning*. Um problema comum durante a aplicação de *transfer learning* em modelos de linguagem é o esquecimento de informações durante o treinamento. Para evitar o esquecimento foi aplicada a técnica de *Gradual Unfreezing* [Devlin et al. 2018] durante o treinamento do modelo de sentimentos SentFinBERT-PT-BR. Essa técnica consiste em desbloquear a atualização dos pesos da rede neural gradativamente durante o treinamento. Dessa forma, a rede neural consegue aprender a tarefa de classificação de sentimento sem esquecer as informações aprendidas durante o treinamento do modelo de linguagem. Visto que o BERT possui 11 camadas de *Encoder*, a técnica de *Gradual Unfreezing* foi aplicada para desbloquear as camadas de *Encoder* gradativamente. A taxa de aprendizado adotada foi de $5e-6$ e o número de épocas foi 11.

Foi também utilizada a técnica de validação cruzada com 70% da base com 5 divisões, isto é, a base de treinamento foi dividida em 5 partes, sendo 4 partes para treinamento e 1 parte para validação. O processo de validação cruzada foi realizado 5 vezes, sendo que em cada iteração uma parte da base foi utilizada para validação e as outras 4 partes para treinamento. O modelo de sentimento com a melhor convergência da

validação (dada pela mínima função de custo) foi utilizado para realizar a classificação dos sentimentos dos textos da base de teste, que corresponde a 30% da base classificada.

Para comparação do desempenho do nosso sistema proposto, SentFinBERT-PT-BR, foi treinado o modelo Sent-BERTimbau, no qual foi também realizado um *transfer learning* do BERTimbau original [Souza et al. 2020] para a tarefa de classificação de sentimentos. SentFinBERT-PT-BR obteve o melhor resultado na base de teste, com uma acurácia de 0,76 e um F1-Score de 0,73 para predição das 3 categorias de sentimento, conforme mostra a Tabela 3. Sent-BERTimbau apresentou uma acurácia de 0,67 e um F1-Score de 0,63.

Tabela 3. Resultados dos modelos de classificação de texto

Nome do Modelo	Acurácia	F1-Score
Sent-BERTimbau	0,67	0,63
SentFinBERT-PT-BR	0,76	0,73

4. Aplicações do SentFinBERT-PT-BR

Visando avaliar o SentFinBERT-PT-BR em tarefas relacionadas a finanças, um índice de sentimentos de mercado foi construído e sua relação com os fatos ocorridos no Brasil e no mundo foi avaliada. Além disso, também é proposta uma estratégia de investimento que toma como base o índice de sentimentos.

4.1. Índice de Sentimentos e sua Relação com o Mercado

O índice de sentimento é uma série temporal construída a partir das classificações das notícias ao longo do tempo [Hiew et al. 2019]. Cada ponto desta série temporal é dado pelo índice de um intervalo entre t e $t - k$,

$$\text{Índice}_{t-k,t} = \frac{\text{Pos}_{t-k,t} - \text{Neg}_{t-k,t}}{\text{Pos}_{t-k,t} + \text{Neu}_{t-k,t} + \text{Neg}_{t-k,t}}, \quad (4)$$

onde $\text{Pos}_{t-k,t}$, $\text{Neg}_{t-k,t}$, $\text{Neu}_{t-k,t}$ representam o número de notícias positivas, negativas e neutras compreendido entre o intervalo de tempo $t - k$ e t , respectivamente.

Como o índice de sentimentos reflete a maior parte de notícias coletadas sobre o mercado no período, é possível realizar uma análise de correlação para entender momentos em que ele se apresenta mais otimista ou pessimista. O índice aqui foi calculado a partir de uma amostra das notícias gerais coletadas e pode ser visto na Figura 3, relacionando-o a alguns fatos relevantes de quebra de expectativa do mercado de capitais, que ocasionaram grande volatilidade na economia.

O primeiro fato relevante destacado no índice de sentimentos foram as manifestações de 2013, que ocorreram no Brasil por conta da alta do preço dos transportes públicos, marcadas por protestos e vandalismo. Avaliando o índice de sentimentos, é possível perceber que o sentimento do mercado teve uma grande variação negativa no mês de junho de 2013, quando ocorreu o primeiro protesto. Logo após a finalização das manifestações, o sentimento do mercado apresentou outra grande variação, dessa vez positiva, no mês de agosto.

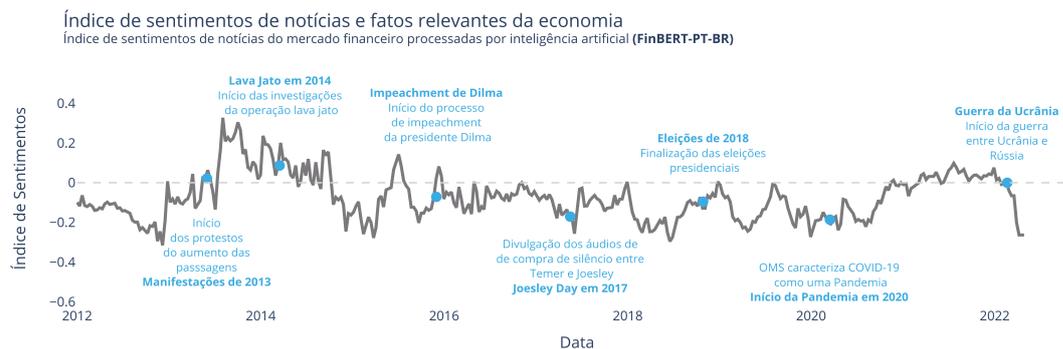


Figura 3. Índice de sentimentos e fatos relevantes da economia.

O segundo fato relevante destacado no índice de sentimentos foi o início da operação Lava Jato, que ocorreu em março de 2014. Esse período foi marcado por uma sequência de meses com revelações de escândalos de corrupção envolvendo diversos políticos e empresas. Nota-se que no médio prazo houve uma tendência de queda no sentimento do mercado, que deve ter sido causado pelo acúmulo de notícias negativas sobre corrupção. O valor mínimo do índice de sentimentos foi atingido em março de 2015, quando grandes executivos foram presos.

O final de 2015 foi marcado pelo início do processo de *impeachment* da presidente Dilma Rousseff, finalizado em agosto de 2016. Logo após o início do processo de *impeachment*, o sentimento do mercado apresentou uma grande variação positiva seguida de uma queda no sentimento do mercado. Visto que o país estava em recessão, o sentimento do mercado tendeu a um valor positivo por uma expectativa de resolução da crise econômica, porém foi apenas uma pequena janela de euforia.

Após o *impeachment*, já no governo de Michel Temer, o sentimento do mercado apresentou uma grande variação negativa no denominado “Joesley Day”. Esse dia, 17 de maio de 2017, foi marcado pela divulgação de uma gravação de uma conversa entre o empresário Joesley Batista e o presidente Michel Temer, em que o assunto da conversa era a compra do silêncio do ex-presidente da Câmara dos Deputados Eduardo Cunha, que estava preso por corrupção. Por envolver o presidente da República, houve muita incerteza sobre o futuro do governo, e o sentimento do mercado apresentou uma grande variação negativa. Visto que o presidente não foi destituído, o sentimento de mercado foi amenizado e voltou a um valor positivo.

Em 2018, o sentimento de mercado apresentou um leve aumento após a conclusão das eleições presidenciais. Tal aumento pode ser explicado pelo fato de que o mercado estava otimista com a vitória de Jair Bolsonaro, que era visto como um candidato favorável ao mercado. Porém, após a posse de Bolsonaro, o sentimento de mercado retornou a um valor negativo, o que pode ser explicado pela incerteza sobre o futuro do governo. Depois, ao longo do primeiro ano de governo, o sentimento apresentou algumas sequências de alta que podem ser atribuídas a notícias positivas, como a aprovação da reforma da previdência. O principal destaque de 2020 foi a pandemia da COVID-19, que começou em janeiro, classificada como pandemia pela OMS em março. A pandemia trouxe ao mundo muitas incertezas, e o sentimento de mercado não apresentou grandes variações

justamente por conta das incertezas. Enquanto o vírus ainda não havia se espalhado pelo Brasil, alguns veículos comunicavam que no Brasil a crise sanitária não seria tão grave por conta do clima tropical, já outros veículos noticiavam que o Brasil estava despreparado para lidar com a pandemia. Portanto, devido a essa incerteza, o sentimento de mercado não apresentou grandes variações.

Mais para o final de 2020, o sentimento de mercado apresentou uma grande variação positiva, que pode ser explicada pelo desenvolvimento de vacinas e estímulos econômicos dos governos. O valor de pico do sentimento de mercado foi atingido no final do ano, quando alguns países começaram a vacinar a população. Outro fato que pode ter contribuído para o aumento do índice de sentimento do mercado foram os estímulos econômicos do governo, implementados para tentar minimizar os efeitos da pandemia, e que acarretaram grandes altas nos mercados.

Por fim, o último fato relevante destacado no índice de sentimentos foi início da invasão da Ucrânia pela Rússia. De longe, esse foi o fato que mais impactou o sentimento de mercado. Com o início da guerra, vários países impuseram sanções econômicas à Rússia, e o sentimento de mercado apresentou uma grande variação negativa. Nesse período o mundo ainda estava se recuperando da pandemia, havia sinais inflacionários por conta dos estímulos econômicos e a guerra trouxe mais pressão inflacionária ao mundo. Por conta disso, o sentimento de mercado apresentou uma grande variação negativa.

4.2. Apostando contra o sentimento

Nesta seção é proposta uma nova estratégia de investimento baseada no índice de sentimento construído, denominada *apostando contra o sentimento*. Essa estratégia consiste em apostar contra o sentimento do mercado, ou seja, investir em ações que possuem uma alta relação negativa de seu preço com o índice de sentimentos. O raciocínio para essa estratégia é que investidores pessimistas ou ansiosos tendem a vender ações influenciados pelas notícias negativas, mas os fundamentos das empresas não são necessariamente afetados pelas notícias negativas de todo o mercado, mas sim são influenciados por fatores macroeconômicos e intrínsecos da empresa. Então, quando investidores pessimistas vendem tais ações, há um aumento no prêmio de risco dessas ações. Portanto, investir em ações com alta correlação negativa com o índice de sentimentos é uma estratégia de investimento que pode gerar retornos positivos, visto que o prêmio de risco dessas ações tende a ser maior que as demais ações.

Para testar essa estratégia, foi utilizado o índice de sentimentos e uma base de preços de ativos da bolsa de valores brasileira. A base de preços de ativos foi obtida do site *Yahoo Finance* e contém os preços de fechamento das ações negociadas na bolsa de valores brasileira. A base de preços de ativos contém os preços de fechamento das ações de 2014 a 2022. Para cada ação, foi calculada a correlação entre o índice de sentimentos e os retornos históricos de cada ação, e as ações com alta correlação negativa com o índice de sentimentos foram selecionadas a cada mês.

O resultado da simulação da estratégia de investimento *apostando contra o sentimento* é mostrado na Figura 4. A estratégia proposta apresentou um retorno acumulado de 683% ao longo de 8 anos, enquanto o índice Bovespa apresentou um retorno acumulado de 254% no mesmo período. O retorno acumulado da estratégia *apostando contra o sentimento* foi 2,7 vezes maior que o retorno acumulado do índice Bovespa.



Figura 4. Simulação da estratégia de investimento *apostando contra o sentimento* e do índice Bovespa.

5. Conclusão

O presente trabalho apresentou um modelo de análise de sentimentos de textos em português referentes ao mercado financeiro, SentFinBERT-PT-BR, a partir de um modelo de linguagem sintonizado para a língua portuguesa do Brasil e para o contexto de textos da área financeira, denominado FinBERT-PT-BR.

FinBERT-PT-BR foi gerado a partir do treinamento do BERTimbau, um modelo BERT geral para a língua portuguesa, com 1,4 milhão de textos no contexto do mercado financeiro. Já SentFinBERT-PT-BR é um classificador treinado com 503 notícias com os sentimentos anotados, também fruto deste trabalho.

O trabalho de anotação foi um trabalho cuidadoso realizado por três pessoas e cada texto teve pelo menos duas anotações de pessoas distintas. O cuidado na anotação se evidencia pelas altas métricas de concordância, com porcentagem de concordância 90,4, e *Krippendorff's alpha* de 0,88.

Os modelos treinados conseguiram superar as métricas dos principais modelos no estado da arte, com destaque para as métricas do modelo de análise de sentimento, com acurácia e F1-Score 0,76 e 0,73, respectivamente.

Além do treinamento dos modelos, foram apresentadas aplicações relevantes para o mercado financeiro, como a análise de sentimento de notícias e índices de sentimento, incluindo uma análise qualitativa do índice de sentimento com fatos relevantes da economia brasileira. Por fim, foi introduzido um fator de investimento denominado *apostando contra o sentimento*, que consiste em investir em ações que apresentam alta correlação negativa com o índice de sentimento. Essa estratégia apresentou um retorno ajustado ao risco satisfatório e ainda pode ser utilizada para a explicabilidade do desempenho do índice Ibovespa.

Para trabalhos futuros, é possível aprimorar o modelo de análise de sentimentos, utilizando uma base maior e mais específica de textos financeiros para o treinamento do modelo de linguagem. Além disso, é possível aprimorar o modelo de análise de sentimentos, aumentando a quantidade de textos rotulados, e mantendo altas métricas de concordância. Em relação às aplicações, é possível aprimorar a forma de cálculo do índice de sentimento e ainda aplicar a metodologia para setores específicos da bolsa de valores.

E por fim, a análise dos fatos relevantes da economia pode ser expandida para avaliar a relação do índice de sentimentos com dados macroeconômicos, como: inflação, PIB e a taxa de desemprego.

Acknowledgments

The authors would like to thank the Brazilian National Council for Scientific and Technological Development (CNPq Grant N. 310085/2020-9) for the support.

Referências

- Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
- Ardia, D., Chopard, B., and Boudt, K. (2015). Using twitter to model the eur/usd exchange rate. *Economics Letters*, 132:23–26.
- Artstein, R. and Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4):555–596.
- Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market.
- Chen, S. F., Beeferman, D., and Rosenfeld, R. (1998). Evaluation metrics for language models.
- de Souza, V. A., de Souza, F., and Meinerz, G. V. (2021). Análise de sentimento em tempo real de notícias do mercado de ações / real-time sentiment analysis of stock market news. *Brazilian Journal of Development*, 7(1):11084–11091.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Hiew, J. Z. G., Huang, X., Mou, H., Li, D., Wu, Q., and Xu, Y. (2019). Bert-based financial sentiment index and lstm-based stock return predictability. *arXiv preprint arXiv:1906.09024*.
- Januário, B. A., Carosia, A. E. d. O., Silva, A. E. A. d., and Coelho, G. P. (2022). Sentiment analysis applied to news from the brazilian stock market. *IEEE Latin America Transactions*, 20(3):512–518.
- Junjie, Z. and Mengoni, P. (2020). Spot gold price prediction using financial news sentiment analysis. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 758–763.
- Kordonis, J., Symeonidis, S., and Arampatzis, A. (2016). Stock price forecasting via sentiment analysis on twitter. In *Proceedings of the 20th Pan-Hellenic Conference on Informatics, PCI '16*, New York, NY, USA. Association for Computing Machinery.
- Kraaijeveld, O. and De Smedt, J. (2020). The predictive power of public twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, 65:101188.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.

- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, 30(5):15–29.
- Man, X., Luo, T., and Lin, J. (2019). Financial sentiment analysis(fsa): A survey. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pages 617–622.
- Manning, C. D. and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press.
- Medeiros, M. and Borges, V. (2019). Tweet sentiment analysis regarding the brazilian stock market. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*, pages 71–82, Porto Alegre, RS, Brasil. SBC.
- Otabek, S. and Choi, J. (2022). Twitter attribute classification with q-learning on bitcoin price prediction. *IEEE Access*, 10:96136–96148.
- Pagolu, V. S., Reddy, K. N., Panda, G., and Majhi, B. (2016). Sentiment analysis of twitter data for predicting stock market movements. In *2016 Int. Conf. on Signal Processing, Communication, Power and Embedded System*, pages 1345–1350.
- Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics, ACL '04*, pages 271–278.
- Pereira, J. G. (2019). Análise de sentimentos da população brasileira em relação a eleição presidencial de 2018 através da rede social twitter.
- Silva, M. C. A. (2018). Percepções sobre corrupção durante as eleições presidenciais no brasil em 2018: uma análise baseada no twitter.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). Bertimbau: pretrained bert models for brazilian portuguese. In *Brazilian Conference on Intelligent Systems*, pages 403–417. Springer.
- Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to fine-tune bert for text classification? In *China national conference on Chinese computational linguistics*, pages 194–206. Springer.
- Tan, K. L., Lee, C. P., and Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7).
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- Xavier, F., OLENSKI, J. R. W., ACOSTA, A. L., SALLUM, M. A. M., and SARAIVA, A. M. (2020). Análise de redes sociais como estratégia de apoio à vigilância em saúde durante a covid-19. *Estudos Avançados*, 34(99):261–282.