

## **Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações**

### **Sentiment analysis and technical indicators: a correlation analysis of asset prices and polarity of stock market news**

DOI: 10.34140/bjbv3n1-029

Recebimento dos originais: 20/11/2020

Aceitação para publicação: 20/12/2020

#### **Wagner Igarashi**

Doutor em Engenharia e Gestão do Conhecimento pela Universidade Federal de Santa Catarina - UFSC

Instituição: Universidade Estadual de Maringá - UEM

Endereço: Av Colombo, n. 5790, bloco C56, sala 107 – Jd. Universitário, Maringá – Paraná, Brasil

E-mail: wigarashi@uem.br

#### **Guilherme Soares Valdevieso**

Graduando em Bacharel em Informática pela Universidade Estadual de Maringá - UEM

Instituição: Universidade Estadual de Maringá - UEM

Endereço: Av Colombo, n. 5790, bloco C56, sala 107 – Jd. Universitário, Maringá – Paraná, Brasil

E-mail: ra92859@uem.br

#### **Deisy Cristina Corrêa Igarashi**

Doutora em Engenharia de Produção pela Universidade Federal de Santa Catarina - UFSC

Instituição: Universidade Estadual de Maringá - UEM

Endereço: Av Colombo, n. 5790, bloco C56, sala 107 – Jd. Universitário, Maringá – Paraná, Brasil

E-mail: dccigarashi@uem.br

### **RESUMO**

Atualmente o movimento do mercado financeiro é um dos principais objetos de estudo para a predição de preços e análise de tendência. Concomitantemente tem-se um evidente crescimento do campo da computação voltado a reprodução da inteligência humana através de técnicas de inteligência artificial. O presente estudo baseou-se na possibilidade de aliar técnicas de análise juntamente com algoritmos computacionais de inteligência artificial voltados ao processamento de linguagem natural (PLN). Com isso, buscou-se construir conhecimento acerca da influência de notícias e opiniões na constante movimentação do mercado, confrontando a polaridade destes sentimentos com a tendência sugerida pela aplicação de técnicas de análise como as médias móveis. Optou-se por utilizar neste estudo, fontes de dados de origem confiável visando a diminuição de ruídos ou das ambiguidades presentes na linguagem em certas situações de informalidade. Para tanto, utilizou-se de portais especializados em publicações de notícias com foco no mercado financeiro. Os resultados obtidos pelo estudo demonstram que, a partir do recorte analisado, a correlação existente entre os sentimentos e a movimentação do preço da ação é, predominantemente, moderada.

**Palavras-chave:** análise de sentimentos; análise técnica; processamento de linguagem natural, mercado de ações.

## ABSTRACT

Currently, the movement of the financial market is one of the main objects of study for price prediction and trend analysis. Concomitantly, there is an evident growth in the field of computing aimed at the reproduction of human intelligence through artificial intelligence techniques. The present study was based on the possibility of combining analysis techniques together with computational artificial intelligence algorithms focused on natural language processing (PLN). With this, we sought to build knowledge about the influence of news and opinions in the constant movement of the market, confronting the polarity of these feelings with the trend suggested by the application of analysis techniques such as moving averages. We chose to use in this study, data sources of reliable origin aiming to reduce noise or ambiguities present in language in certain informal situations. To this end, we used specialized portals in news publications with a focus on the financial market. The results obtained by the study demonstrate that, based on the analyzed profile, the correlation between feelings and the movement of the share price is predominantly moderate.

**Keywords:** sentiment analysis; technical analysis; natural language processing, stock market.

## 1 INTRODUÇÃO

Com a evolução das redes de computadores e o aumento considerável de informações disponíveis, a Internet se tornou fonte de dados para pessoas consultarem e compreenderem opiniões alheias (PANG; LEE, 2008). Junto a isso, o avanço da inteligência artificial na área de processamento de linguagem natural (PLN) possibilitou a automatização deste comportamento utilizando técnicas como a de Análise de Sentimentos em manifestações de linguagem verbal.

Uma das aplicações desta técnica é a de rastreamento do sentimento público em relação ao comportamento do preço de determinado ativo na bolsa de valores. Estudos neste campo, habitualmente buscam coletar opiniões textuais das mais diversas fontes e por fim correlacioná-las com outras métricas, visando inferir um possível movimento no preço de ativos.

Por outro lado, a capacidade de prever o comportamento variante do preço de ações impacta diretamente no sucesso de uma estratégia de investimento sendo que, do ponto de vista econômico, tal estratégia normalmente pode estar baseada em duas escolas: a análise técnica e a análise fundamentalista. Enquanto a análise técnica propõe uma abordagem estatística com base em gráficos e cálculos matemáticos, a análise fundamentalista se pauta em um estudo aprofundado dos fundamentos do negócio, tais como demonstrações financeiras, posicionamento estratégico ou cenário macroeconômico do ativo em questão. Ambas as escolas possuem qualidades e deficiências devendo ser utilizadas de acordo com a estratégia do investidor, como por exemplo, *day trade* ou *buy and hold*.

Neste contexto, o estudo tem como objetivo a análise da correlação do comportamento variante do preço de ativos da bolsa de valores, por meio de análise técnica conjuntamente com o indicador de polaridade das notícias publicadas por portais com foco no mercado financeiro.

Esta pesquisa motiva-se pela possibilidade de evidenciar que os aspectos subjetivos interferem em ambas as escolas de análise de mercado, coletando os dados através de opiniões, notícias publicadas em portais especializados, considerando inclusive aspectos ideológicos como a movimentação de investidores que são tomados como inspiração para outros acionistas.

Nota-se também que a consideração apenas de dados históricos para realizar a predição do preço de ações associados a indicadores técnicos nem sempre se mostra eficiente, assim como descrito por Galdão e Famá (1998), sobre a hipótese de mercado eficiente em sua forma fraca, onde o investidor é incapaz de obter ganhos em excesso desenvolvendo regras de negociação baseadas em históricos de preços ou em informações sobre retornos.

Ressalta-se o fato de que optar pelo uso de notícias publicadas por portais especializados contribui para a diminuição do ruído e da ambiguidade presente em mídias sociais (ALVES, 2015), o que também se mostra como um diferencial a outros trabalhos realizados nesta área, como o de Mittal e Goel (2012). Notadamente, a quantidade de trabalhos voltados ao estudo do mercado de ações mostra-se satisfatória quando se trata de áreas como economia, administração ou contabilidade, porém, o mesmo não ocorre quando entramos no âmbito computacional ou estudos relacionados a tecnologia, fazendo com que isso se torne um dos pilares para a construção deste estudo.

Analisar a influência das opiniões e recomendações coletadas por meio da internet com alguns indicadores técnicos, podem servir de insumo para que futuras ferramentas possam ser desenvolvidas, com o intuito de analisar a movimentação do mercado a partir de fatos e opiniões que sejam coletados via internet.

## **2 FUNDAMENTAÇÃO TEÓRICA**

Este estudo fundamenta-se em conceitos existentes na área de inteligência artificial e conseqüentemente em teorias e definições presentes no mercado financeiro. Apesar de ambas as áreas serem de igual importância para o sucesso do estudo, ressalta-se que o mercado financeiro caracteriza-se com foco do problema e as técnicas de inteligência artificial, mais especificamente a análise de sentimento, como ferramental para a análise do problema.

## 2.1 MERCADO FINANCEIRO

Podemos entender o mercado financeiro como um conjunto de instituições e instrumentos financeiros com foco na possibilidade da transferência de recursos ofertados para tomadores, viabilizando condições de liquidez no mercado, (ANDREZO A. F.; LIMA, 2000). Neste estudo, enfocaremos o mercado de ações e a predição do movimento de preços.

As ações são a menor parcela de capital de uma empresa. Elas são títulos que não garantem retorno dos recursos investidos pois sua remuneração é determinada pela capacidade da empresa em gerar lucros (BERTOLO, 2012). Ao adquirir ações o investidor torna-se sócio da empresa, podendo negociar posteriormente os títulos comprados por intermédio da bolsa de valores cujo funcionamento é regido pela comissão de valores mobiliários.

O preço dos ativos está relacionado diretamente com a oferta e a demanda, ou seja, quanto mais visada for uma ação, seu preço tende a aumentar e quanto menor o interesse do mercado sobre ela, os preços tendem a cair. Esta variação de preço pode estar relacionada a vários aspectos como política, estratégias empresarias, inovações e aumento de competitividade, (BERTOLO, 2012).

Assim como descrito por Gomes (1997), a especulação na bolsa de valores se caracteriza quando grandes investidores se aproveitam de altas e baixas nas cotações de títulos a fim de obter lucros extraordinários ou conseguir controle acionário de determinada empresa, isso resulta em boatos relativos à situação político-econômica do país, por exemplo, o que faz com que as bolsas oscilem, provocando queda ou alta nas cotações.

Tais oscilações provocam oportunidades de investimento, mas também, possíveis riscos de perda de investimento. Como exposto por Knight (1967), o risco é verificado quando o futuro não é inteiramente conhecido, mas a distribuição de probabilidade de possibilidades futuras é conhecida. A incerteza, no entanto, ocorre quando até mesmo a distribuição de probabilidade é desconhecida.

Neste cenário, a análise técnica define um conjunto de técnicas utilizadas para a negociação e previsão de valores mobiliários. Estas técnicas visam prever o comportamento do mercado baseando-se em estudos de séries históricas de preços. Um investidor que utiliza apenas análise técnica, acredita que é possível detectar possíveis padrões em preços de ativos por meio da análise de gráficos que expõem a movimentação de preços dos ativos do mercado (BROCK; LAKONISHOK; LEBARON, 1992).

Segundo Belmont (2010), a análise técnica surgiu por meio das teorias propostas por Charles Dow, no início do século XX. Dow sugeriu alguns princípios em sua teoria que até hoje são válidas, segundo ele: os preços refletem tudo, ou seja, todas as opiniões sobre um determinado ativo, bem como todas as notícias publicadas são levadas em consideração e estão expressas nos preços desse ativo; o mercado segue tendências (primária, secundária e terciária), e dentro dessas três tendências,

podem assumir outras três: alta, baixa ou neutra; o volume de negociações acompanha a tendência, ou seja, em um mercado de alta, maior força compradora, o volume de negociações aumenta quando o preço do ativo aumenta e o volume diminui quando o preço do ativo diminui. Em um mercado onde a força vendedora é superior à força compradora, ou seja, um mercado de baixa o volume tem uma relação inversa com o preço, isto é, o volume de negociações aumenta quando o preço do ativo diminui e o volume diminui quando o preço do ativo aumenta; a tendência um dia se reverterá e que, para que tenhamos uma confirmação dessa reversão de tendência, é necessário que esse movimento seja evidenciado em gráficos e índices de preços simultaneamente, considerado assim como o princípio da confirmação.

O conceito de tendência é absolutamente essencial para a abordagem técnica de análise de mercado. Em grande maioria, as ferramentas utilizadas por analistas técnicos possuem como proposta o auxílio na medição da tendência de mercado com o intuito de participar desta tendência. Genericamente, a tendência é simplesmente a direção do mercado, a forma como ele se move. Estes movimentos raramente são lineares visto que se caracterizam por frequentes oscilações. A tendência é um espelho destas oscilações que constituem a movimentação do mercado (MURPHY, 1999).

Em relação a tendências, as médias móveis são indicadores de preço que expressam o valor médio de uma amostra de determinado dado. Esta técnica é seguramente o tipo de indicador mais popular e um dos mais antigos da análise técnica. Sua principal utilidade é indicar através da inclinação da média de preços a direção da força do mercado. Este tipo de indicador também é classificado como *trend following*, ou seja, um tipo de rastreador de tendência para o mercado. Existem diversos métodos para o cálculo de médias móveis, dentre eles, os mais populares são o cálculo aritmético e o cálculo exponencial (BISI, 2009).

O cálculo de média móvel aritmética (MMA) é dado como uma das médias mais simples de serem calculadas, podendo ser expressa pela seguinte fórmula:

$$\text{MMA} = \frac{P_1 + P_2 + P_3 + \dots + P_n}{n} \quad (2.1)$$

Em que:

- MMA: valor da média móvel aritmética calculada
- Pn: preço relacionado a cotação n do ativo
- n: número de cotações

Como pode ser observado (2.1),  $n$  é a quantidade de valores a serem considerados no cálculo da MMA. Em um exemplo hipotético, poderia ser calculado o MMA(15) que representa a média móvel aritmética de  $n = 15$  dias do preço de um determinado ativo.

Sob outro ponto de vista, a análise fundamentalista estuda os fatores que afetam as situações de oferta e demanda de um mercado. Esta análise está geralmente atrelada a grandes movimentos no longo prazo e não prediz o tempo certo para entrar ou sair do mercado. Este tipo de técnica é composta por uma série de análises como: a análise estratégica do negócio, análise financeira, análise prospectiva, análise estratégica, análise contábil, entre outros (LIMA et al., 2009). Cabe ressaltar que os conceitos de análise fundamentalista não foram foco deste estudo.

## 2.2 ANÁLISE DE SENTIMENTOS

A análise de sentimentos combina processamento de linguagem natural com mineração de texto e utiliza técnicas de aprendizado de máquina para classificar textos como positivos ou negativos (SINGH V.; DUBEY, 2014). O processamento de linguagem natural tem como objetivo adquirir conhecimento ainda que parcial sobre a linguagem utilizada por um humano, porém, a própria linguagem possui características ambíguas que mudam constantemente (RUSSELL, 2013).

Destaca-se também que a técnica de análise de sentimentos tem o objetivo de detectar e extrair informações subjetivas a partir de documentos de texto sendo que executar esta tarefa utilizando dados não estruturados pode desencadear uma série de desafios (BHATIA; BHATIA, 2015).

A análise de sentimentos possui como principal tarefa a classificação de sentimentos (GUELLIL; BOUKHALFA, 2015), a qual define diferentes níveis, tais como: nível de documento: que tem como objetivo classificar como positivo, negativo ou neutro um documento completo; nível de sentença: cujo objetivo é classificar o sentimento de cada sentença de um determinado documento; nível de aspecto: que identifica textos e sentenças que contenham determinado aspecto que deseja-se buscar e após isto o classifica como positivo, negativo ou neutro. O presente estudo terá como foco a classificação dos dados em nível de aspecto, a partir da filtragem dos dados para a posterior classificação.

O processo global de análise de sentimentos divide-se em quatro passos: extração dos dados, pré-processamento, classificação de sentimento e identificação do conhecimento útil. A primeira etapa é a extração dos dados obtidos das fontes de interesse, podendo ser obtida das seguintes maneiras (GUELLIL; BOU-KHALFA, 2015): 1. a partir de um corpus já existente (*existing corpus*) que utiliza de textos já existentes anteriormente; 2. a partir de um corpus manual extraído

diretamente da fonte de dados; 3. a partir de um corpus automático (*automatic corpus*) que é extraído automaticamente da fonte utilizando-se as APIs (*Application Programming Interfaces*) disponíveis.

Na etapa de pré-processamento são realizados tratamentos nos dados para que se torne possível a classificação. Esta etapa pode ser executada utilizando-se algumas das técnicas citadas em Gupellil e Boukhalifa (2015), como: quebra das sentenças em palavras, tradução, *lemmatization* ou *stemming* que têm como objetivo reduzir as palavras para uma forma mais primitiva, *stop word removal* cujo objetivo é tirar palavras comuns a língua (e.g. 'e', 'ou', 'para', 'de'). O *stemming* é o processo de reduzir a inflexão de palavras em suas formas raiz, como mapear um grupo de palavras para o mesmo radical, mesmo que o radical em si não seja uma palavra válida no idioma. *Lemmatization*, diferentemente do *Stemming*, reduz as palavras flexionadas adequadamente, garantindo que a palavra raiz pertença ao idioma.

Na penúltima etapa a classificação do sentimento é efetivamente realizada, podendo utilizar técnicas como aprendizagem de máquina. Por fim, na última etapa o conhecimento útil é obtido, gerado a partir de todos os processos anteriores, possibilitando assim o alcance de ganhos com a informação.

### 2.2.3 Fatos e opiniões

As informações disponíveis na Web podem ser classificadas em dois tipos principais: fatos e opiniões (LIU, 2007). Apesar de opiniões não necessariamente representarem um acontecimento ou uma verdade absoluta (em contraponto a um fato), iremos considerá-la como um dado absolutamente relevante, por conta da característica especulativa presente nas oscilações diárias durante as negociações na bolsa. A coleta de opiniões fará com que esta análise trabalhe um aspecto subjetivo, gerando um valor classificável que posteriormente poderá ser utilizado para o cruzamento com outras informações geradas pela aplicação das mais diversas técnicas. Entretanto, não serão considerados dados provindos de mídias sociais por conta da incerteza acerca do conhecimento dos indivíduos em relação ao mercado acionário assim como pelo alto grau de ambiguidade presente em textos publicados nas redes sociais, o que busca se mitigar coletando as informações de fontes confiáveis como sites e portais de notícias especializados do mercado financeiro.

## 2.3 ESTUDOS CORRELATOS

Existem alguns trabalhos que visam explorar a mesma área de estudo desta pesquisa.

Em um deles, Shibukawa (2017) expõe os desafios e as conclusões alcançadas com o estudo onde nota-se que a falta de diversificação da fonte de dados colaborou para a baixa correlação entre a polaridade dos sentimentos e a real variação no preço das ações. A dificuldade de exploração dos



dados disponíveis na internet também é um ponto evidenciado pelo autor que sugere algumas ferramentas como o ScappingHub para auxiliar na extração dos dados. Por fim, o autor sugere que a escolha da fonte é fundamental para a análise durante o processo de classificação.

No trabalho de Lopes et al. (2008), conclui-se que a padronização do formato de divulgação de notícias, como o RSS, auxiliou na padronização do formato de leitura dos dados que serviram como insumo para o estudo, o que serve como inspiração para esta pesquisa que busca utilizar o mesmo padrão. Os autores também citam a dificuldade em encontrar notícias com real vínculo às empresas que habitualmente são citadas como o código do ativo na bolsa de valores. Por fim, os autores também destacam a existência de um certo grau de correlação entre o cálculo de orientação realizado e a variação do valor da ação.

Em, Lima (2016), buscou-se desenvolver e apresentar um modelo para predição na bolsa de valores, utilizando-se da técnica de análise de sentimentos. A pesquisa apresenta como limitações a utilização de apenas um ativo para o estudo, bem como alguns problemas ao se utilizar a rede social Twitter como fonte de dados. O autor também ressalta a grande dificuldade em se processar a linguagem natural, o que serve como ponto de atenção para esta pesquisa. O autor recomenda a criação de um mecanismo que impeça a interferência de postagens mal-intencionadas, provindas de robôs ou de forma manual, que visem burlar o teor natural das postagens. Com isto, pode-se vislumbrar um ganho ao utilizar-se portais especializados para a coleta de dados.

### 3 MATERIAIS E MÉTODOS

#### 3.1 COTAÇÕES HISTÓRICAS BOVESPA

A série histórica de cotações traz todo o histórico de preços dos títulos negociados na Bolsa desde 1986. As cotações são fornecidas na moeda e forma de cotação da época, sem nenhum ajuste para a inflação ou proventos (dividendos, bonificações, direitos de subscrição, etc.) distribuídos pelas empresas emissoras. Estes dados contemplam as principais informações dos ativos, como: nome e código da empresa, código da ação, código ISIN, tipo de mercado (a vista, termo, opções), especificação (ON/PN), preços (anterior, abertura, mínimo, médio, máximo, fechamento), quantidade de negócios e volume negociado com o papel, dentre outros dados disponíveis.

#### 3.2 PORTAIS ESPECIALIZADOS EM MERCADO FINANCEIRO

O presente estudo irá embasar-se nos dados publicados por mídias especializadas ou com grande foco em mercado financeiro. Dentre os portais utilizados pode-se citar: [www.istoedinheiro.com.br](http://www.istoedinheiro.com.br), [www.valor.globo.com](http://www.valor.globo.com), [www.moneytimes.com.br](http://www.moneytimes.com.br), [www.infomoney.com.br](http://www.infomoney.com.br), [www.exame.abril.com.br](http://www.exame.abril.com.br), [www1.folha.uol.com.br/mercado/](http://www1.folha.uol.com.br/mercado/),



www.economia.uol.com.br, www.br.advfn.com, www.br.investing.com, entre outros portais classificados como relevantes pelo próprio Google.

### 3.3 ARTEFATOS COMPUTACIONAIS

O presente estudo visa coletar dados relacionados estritamente ao mercado de ações, os quais devem possuir opiniões ou fatos referentes a movimentação do preço de ativos, estratégias executadas pelo mercado, recomendações ou críticas negativas referentes a empresas, entre outros conteúdos que possam respaldar um indicador relevante para a pesquisa. Para tal, foram utilizados os seguintes artefatos computacionais: Python como linguagem de programação de alto nível e open source, utilizada por muitas empresas para projetos reais, que estão em produção; FEED RSS, um formato de distribuição de informações em tempo real pela internet, no qual um conjunto de dados no formato XML agrega o conteúdo do documento, podendo ser acessado mediante programas ou sites agregadores, e usado principalmente em sites de notícias; GitHub como plataforma de hospedagem de código-fonte com controle de versão usando o Git; Amazon Comprehend para a aplicação da técnica da análise de sentimentos nas notícias coletadas, ressalta-se que utiliza um modelo genérico de análise de sentimentos; MongoDB como software de banco de dados orientado a documentos.

## 4 RESULTADOS E DISCUSSÃO

Para coletar os dados utilizados na análise foi necessário manter em execução o software durante o período de negociação na bolsa de valores. Como explicado no capítulo de desenvolvimento, a aplicação se mantinha em execução durante o pregão da bolsa e realizava a coleta dos dados, análise dos documentos, classificação dos sentimentos e posteriormente o cálculo da correlação das polaridades retornadas com o movimento do preço da ação.

Para uma maior assertividade na busca pelos documentos a serem analisados e a eliminação da possibilidade de ausência de notícias publicadas, optou-se pela utilização de uma ação que compusesse o ranking de blue chips brasileiras, no caso deste estudo, optou-se pela ação da Petrobras (PETR4).

Durante o experimento, optou-se por utilizar um período de análise de uma hora, ou seja, quando houvesse a publicação de uma notícia, a média móvel era calculada a 30 minutos antes e 30 minutos após a publicação da notícia, garantindo assim que houvesse margem para que se pudesse analisar o movimento pré publicação e pós publicação da notícia. Este intervalo foi definido buscando-se considerar os preços anteriores a publicação da notícia e também a movimentação posterior a sua publicação, considerando que a movimentação do preço do ativo pode refletir o

impacto da notícia neste período de tempo. O período de execução dos experimentos compreendeu o intervalo de 18/11/2019 a 02/12/2019

## 5 ANÁLISE GRÁFICA

Partindo dos dados coletados e dos experimentos realizados, pode-se comparar as polaridades retornadas a partir da análise de sentimentos das notícias com a variação do preço do ativo, que é representado a partir do cálculo da média móvel no intervalo de uma hora a partir da publicação da notícia.

Para comparar-se a variação das polaridades com a média móvel aritmética foi necessário a aplicação da normalização linear na variável, fazendo com que os valores se adequassem ao intervalo 0 e 1.

$$f(x) = \frac{x - \text{Min}}{\text{Max} - \text{Min}} \quad (5.1)$$

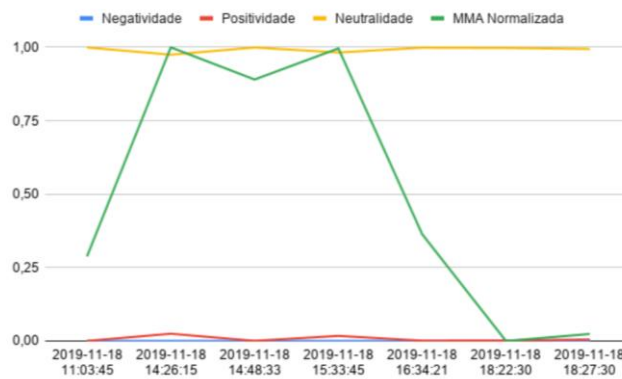
Em que:

X: valor a ser normalizado; Max: valor máximo dentro do intervalo; Min: valor mínimo dentro do intervalo

Durante a execução do experimento no dia 18/11/2019 a média móvel aritmética correspondente ao período foi uma mínima de R\$ 27,68 e uma máxima de R\$ 27,96, o que é equivalente a uma variação média de R\$ 0,28 no preço do ativo. Segundo os dados coletados através do software, o sentimento de neutralidade foi o mais significativo dentre os demais. Nota-se a polaridade neutra sempre próxima do valor 1,00 (Figura 1). O sentimento de positividade demonstra pequenos picos durante a execução, porém, essas oscilações se mostram praticamente irrelevantes em relação ao sentimento de neutralidade.

Nota-se também que o sentimento de negatividade se mantém estável próximo de 0,00. Em contrapartida, observa-se que no início do experimento a média móvel possui tendência de alta até 14:26, quando apresenta uma pequena queda e cresce novamente até 15:33. No entanto, a partir de seu último pico de crescimento, a média móvel cai para seu valor mínimo até o final do experimento.

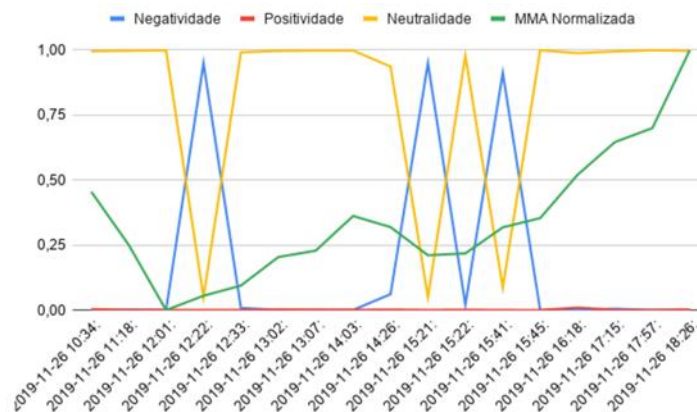
Figura 1 – Gráfico de polaridade x média móvel aritmética normalizada do preço do ativo no dia 18 de novembro de 2019



A partir dos dados coletados, pode-se verificar uma relativa correlação entre o sentimento de positividade e a média móvel, e correlação inversa entre a neutralidade e a média móvel. Nota-se que apesar do sentimento ser em grande maioria de polaridade neutra, que a ação apresenta uma variação considerável entre 11:03 até 14:26 e entre 15:33 até 18:22.

No dia 26 há uma oscilação maior dos sentimentos coletados através da API disponibilizada pela ferramenta. A variação da média móvel se deu entre R\$ 27,32 (valor mínimo) e R\$ 28,00 (valor máximo) (Figura 2).

Figura 2 – Gráfico de polaridade x média móvel aritmética normalizada do preço do ativo no dia 26 de novembro de 2019



O sentimento de neutralidade se apresenta extremamente variável no decorrer do tempo, se apresentando com seu valor máximo em alguns períodos como 12:33 até 14:03 e quedas bruscas como em 12:22, 15:21 e 15:41. O sentimento de negatividade também apresenta alguns picos chegando próximo de seu valor máximo nos horários 12:22, 15:21 e 15:41.

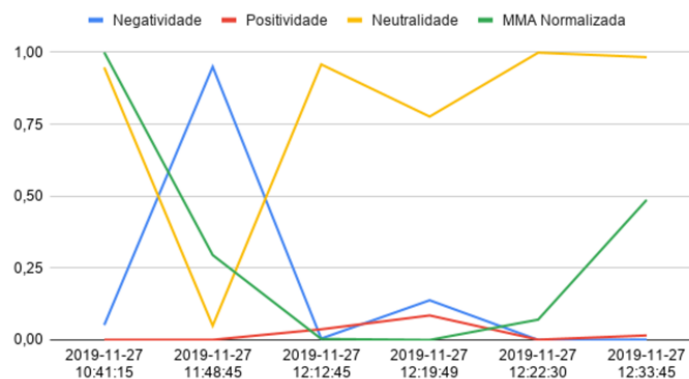
No decorrer do gráfico, apesar dos picos, o sentimento de negatividade se mantém na maior parte do tempo próximo do valor 0,00. O sentimento de positividade praticamente não apresenta

oscilações durante o experimento, mantendo-se estável e próximo do valor 0,00. A média móvel apresenta uma tendência de queda no início do experimento, voltando para um movimento de alta em 12:21 até 14:03 onde volta a cair e cresce até 17:57.

A partir destes valores pode-se notar que apesar da polaridade negativa apresentar um pico em 12:22 a média móvel aritmética continua crescendo até 14:03, não apresentando correlação entre o sentimento e o movimento do ativo. Porém, nota-se que a partir das 14:26 as altas em relação ao sentimento negativo acompanham a tendência de baixa da média móvel até 15:22. Por fim, o sentimento de neutralidade se perdura entre as 15:45 até o fim do experimento onde no mesmo período a média móvel se apresenta em alta.

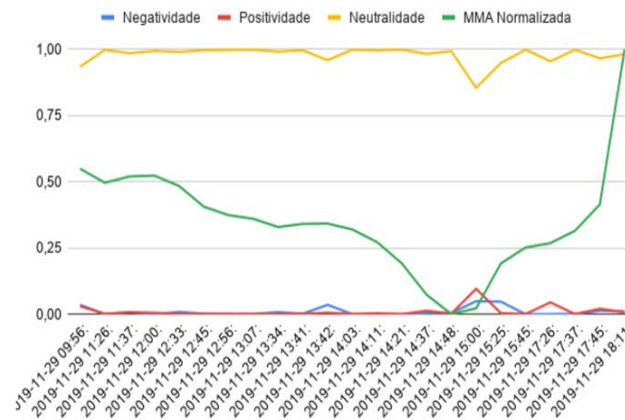
No dia 27/11/2019 (Figura 3) é verificado que a neutralidade e a negatividade tem correlação negativa em relação a média móvel no primeiro intervalo de tempo. No segundo ao quarto período há variações de todos os sentimentos, e quando estes cessam, a média móvel volta a subir.

Figura 3 – Gráfico de polaridade x média móvel aritmética normalizada do preço do ativo no dia 27 de novembro de 2019



O dia 29/11/2019 apresentou uma variação da média móvel entre R\$ 27,92 e R\$ 28,44 sendo esta diferença de R\$ 0,52 centavos (Figura 4). O início do experimento apresenta o sentimento de neutralidade estável, possuindo pequenas variações de queda em 13:41 e posteriormente as 15:00.

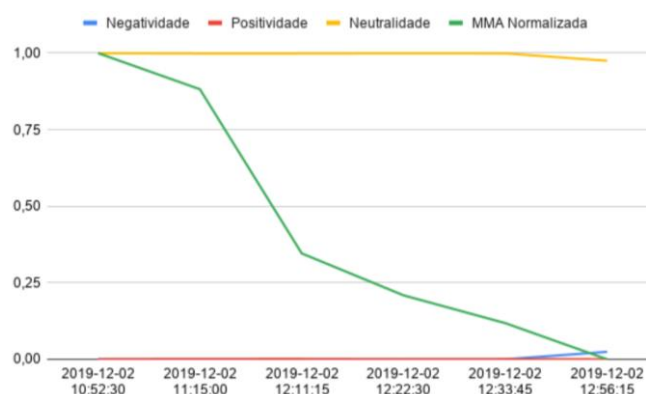
Figura 4 – Gráfico de polaridade x média móvel aritmética normalizada do preço do ativo no dia 29 de novembro de 2019



De modo geral pode-se perceber que há uma tendência de correlação entre a neutralidade e a média móvel, e a partir do decréscimo do pico da negatividade percebe-se a reversão da média móvel. Logo após, percebe-se outro pico de negatividade e uma suavização da média móvel. Por fim, quando a negatividade desce a média móvel ganha força e cresce rapidamente.

No dia 02/12/2019 nota-se uma variação da média móvel aritmética entre R\$ 28,43 e R\$ 28,65 o que equivale a uma variação de R\$ 0,22 no período (Figura 5). Durante o experimento o sentimento de neutralidade se manteve estável e muito próximo do valor 1,00. Os demais sentimentos também se mostraram invariantes com um pequeno pico de negatividade ao final do experimento.

Figura 5 – Gráfico de polaridade x média móvel aritmética normalizada do preço do ativo no dia 02 de dezembro de 2019



Na Figura 5 é possível afirmar que não houve relação entre a variação da MMA com os sentimentos de negatividade ou positividade, pois ambos se mantiveram praticamente estáticos durante o período.

Com base nos dados levantados foi possível realizar o cálculo de correlação entre a MMA e o indicador de sentimento. Os valores de correlação podem ser observados no Quadro 1.

Quadro 1. Nível de correlação

	MMA 18/11/2019	MMA 18/11/2019	MMA18/11/2019	MMA 18/11/2019
Negatividade	-0,81	-0,30	-0,03	-0,13
Positividade	0,66	0,23	-0,53	-0,26
Neutralidade	-0,65	0,30	0,08	0,23

Com base na Quadro 1, pode-se perceber que em relação à negatividade foi encontrada uma correlação negativa em relação à MMA. No entanto, em relação à positividade, o mesmo não ocorreu sempre, pois apresentou correlação positiva nos dois primeiros dias e negativa nos dois últimos dias. Pode-se destacar que a neutralidade apresentou predominantemente correlação positiva, e que pode-se inferir que sempre que há variações de neutralidade haverá mudanças no movimento da MMA.

## 6 CONCLUSÕES

O presente estudo teve como objetivo analisar a correlação da variação do preço de um ativo utilizando a técnica de cálculo da média móvel aritmética em conjunto com uma ferramenta de análise de sentimentos disponível no mercado, sendo que o conteúdo analisado foi coletado de portais com grande foco no mercado financeiro.

A utilização do padrão feed para coletar e centralizar os dados provindos da internet mostrou-se efetiva pois foi possível aplicá-lo em um buscador (Google) que automaticamente realizou a classificação de relevância das notícias publicadas pelos portais, eliminando o desafio de que alguns portais não dispunham deste padrão para a entrega do conteúdo.

A ferramenta Amazon Comprehend se mostrou interessante no começo da pesquisa por conta de fornecer uma interface amigável para o consumo e também resultados promissores durante alguns testes prévios. Porém, no decorrer do estudo notou-se que a ferramenta não obteve resultados satisfatórios pois grande parte de seus retornos demonstravam o sentimento de neutralidade com o maior valor.

No que tange aos custos, pela opção de utilizar uma ferramenta paga abriu-se mão da gratuidade de alguns serviços já disponíveis na internet, na esperança de que a utilização de uma ferramenta provinda de uma empresa privada seria mais efetiva que ferramentas mantidas pela própria comunidade sem fins lucrativos. Infelizmente, o custo para a utilização da API se mostrou proporcional ao seu uso mas relativamente elevado para os resultados que forneceu ao estudo.

Analisando manualmente o corpus de algumas notícias extraídas, notou-se também algumas dificuldades durante o estudo. Dentre estas podemos citar a existência de trechos relacionados a propagandas ou direitos autorais das publicações, trechos da linguagem de marcação HTML, em suma, conteúdos que não foram totalmente removidos durante a limpeza dos textos extraídos e que não agregavam valor para a análise do sentimento.

O fato da ferramenta implementada no estudo consumir e gerar informações em um período de tempo mais granular também se mostrou uma contribuição importante do estudo pois foi possível realizar análises em termos de horas e não mensais ou diários como visto em muitos estudos.

Esta contribuição só foi possível visto que havia em mãos uma API que disponibilizava as cotações com um curto período de atraso. Esta decisão de projeto contribuiu para um *feedback* mais rápido dos experimentos mas trouxe consigo a necessidade de uma execução contínua durante os dias para que os dados fossem coletados e posteriormente analisados.

Vale ressaltar que a opção por analisar apenas uma ação (ação da Petrobras) durante o estudo teve efeitos positivos sobre a pesquisa, pois foi possível especializar as buscas salvas na ferramenta de geração de feeds tornando os resultados mais assertivos e direcionados no momento da extração.

Em alguns momentos, nota-se uma correlação entre os sentimentos analisados e a variação do preço do ativo, porém, predominantemente, com base no recorte analisado, a correlação observada é moderada. Desta forma, entende-se que o estudo alcançou seu objetivo, analisando sentimentos de notícias e correlacionando-os com a MMA da análise técnica.

O estudo também possibilita espaço para a geração de novos trabalhos visando seu refinamento e melhoria na acurácia dos dados coletados, na extração do conteúdo e da análise de sentimentos. O fato de utilizar-se uma ferramenta de mercado para a análise de sentimentos também contribui tanto com o meio empresarial quanto com a academia para possíveis decisões de projeto em empresas ou pesquisas futuras. Por fim, ressalta-se a possibilidade de construção de um modelo de análise de sentimentos mais direcionado para a área financeira com a utilização da ferramenta Amazon Comprehend ou até mesmo a migração para um algoritmo próprio de classificação de sentimentos.



## REFERÊNCIAS

- Alves, D. S. (2015). *Uso de técnicas de Computação Social para tomada de decisão de compra e venda de ações no mercado brasileiro de bolsa de valores*. Tese de Doutorado, PGEA, Universidade de Brasília, Brasília, DF, Brasil.
- Andrezo, A. F., & Lima, I. S. (2000). *Mercado financeiro: aspectos históricos e conceituais*. São Paulo, Brasil: Pioneira.
- Belmont, D. F. D. S. (2010). *Teoria das ondas de Elliott: uma aplicação ao mercado de ações da bm&fbovespa*. Dissertação de Mestrado, Programa de Pós-Graduação em Economia, Universidade Federal da Paraíba, João Pessoa, PB, Brasil.
- Bertolo, L. A. (2012). *Matemática Financeira*. In: Curso de matemática financeira. Disponível em: <<http://lbertolo.tripod.com/MATEMATICAFINANCEIRA.pdf>>. Acesso em: 28/07/2019.
- Bhatia, S., Sharma, M., & Bhatia, K. K. (2015, March). Strategies for mining opinions: A survey. *International Conference on Computing for Sustainable Global Development, INDIA*, 2, 262-266.
- Bisi, T. C. (2009). *Estratégia operacional para o mercado de ações: média móvel de 09 períodos*. Trabalho de conclusão de graduação, Administração, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brasil.
- Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of finance*, 47(5), 1731-1764.
- Galdão, A., & Famá, R. (1998). Avaliação da eficiência na precificação de ações negociadas no Brasil, por teste de volatilidade. *Revista de Administração da Universidade de São Paulo*, 33(2).
- Gomes, F. R. (1997). A Bolsa de Valores brasileira como fonte de informações financeiras. *Perspectivas em ciência da informação*, 2(2).
- Guellil, I., & Boukhalifa, K. (2015, April). Social big data mining: A survey focused on opinion mining and sentiments analysis. *International symposium on programming and systems (ISPS)*, 12.
- Knight, F. (1967) *Risk, Uncertainty and Profit*. Boston, EUA: Houghton Mifflin Company.
- Lima, M. L. (2016). *Um modelo para predição de bolsa de valores baseado em mineração de opinião*. Dissertação de Mestrado, Programa De Pós-Graduação Em Engenharia de Eletricidade, Universidade Federal do Maranhão, São Luis, MA, Brasil.
- Lima, V. S., Ilha, H. F., Scalzer, R. S., & Galdi, F. C. (2009). *Análise fundamentalista sob a perspectiva do analista de mercado: um estudo de caso na AES TIETÉ comparando os modelos de fluxo de caixa descontado e AEG Ohlson (1995)*. Congresso USP de Iniciação Científica, 6, USP, São Paulo, Brasil.
- Liu, B. (2007). *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.

Lopes, T. J. P., Hiratani, G. K. L., Barth, F. J., Rodrigues Jr, O., & Pinto, J. M. (2008, October). Mineração de opiniões aplicada à análise de investimentos. *Companion Proceedings of the XIV Brazilian Symposium on Multimedia and the Web*, 14, 117-120.

Mittal, A., & Goel, A. (2012). *Stock prediction using twitter sentiment analysis*. Stanford University, CS229. Disponível em: <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>. Acesso em: 2019.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis, *Foundation and Trends in Information Retrieval*, 2, 1-2.

Russell, S. J., & Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno*. Editora Elsevier.  
Shibukawa, E. H. K. (2017). Um estudo da correlação de notícias e os valores das ações de empresas. Trabalho de conclusão de curso, Departamento de Informática, Universidade Estadual de Maringá, Maringá, PR, Brasil.

Singh, V., & Dubey, S. K. (2014, September). Opinion mining and analysis: A literature review. *International Conference-Confluence The Next Generation Information Technology Summit*, 5, 232-239.