**INSPER**

**GRADUAÇÃO EM CIÊNCIAS ECONÔMICAS**

**ALLAN DE LIMA BIRMAN**

# REDES SOCIAIS, EXPECTATIVAS DE MERCADO E PREÇOS DE ATIVOS FINANCEIROS: UM ESTUDO DE EVENTOS

**SÃO PAULO**

**2020**

Allan de Lima Birman

**Redes sociais, expectativas de mercado e preços de ativos financeiros: um estudo de eventos**

Monografia apresentada ao curso de Ciências Econômicas, como requisito parcial para obtenção do Grau de Bacharel do Insper Instituto de Ensino e Pesquisa.

Orientador: Prof. Sérgio Martins – Insper

**São Paulo**

**2014**

Birman, Allan de Lima
Redes sociais, expectativas de mercado e preços de ativos financeiros:
um estudo de eventos / Allan de Lima Birman – São Paulo, 2020
f.
Monografia: Faculdade de Economia e Administração.
Insper Instituto de Ensino e Pesquisa.

Orientador: Prof. Sérgio Martins

1.Redes Sociais 2.Mercado Financeiro 3.Finanças Comportamentais

Allan de Lima Birman

**Redes sociais, expectativas de mercado e preços de ativos financeiros: um estudo de eventos**

Monografia apresentada ao curso de Ciências Econômicas, como requisito parcial para obtenção do Grau de Bacharel do Insper Instituto de Ensino e Pesquisa.

**EXAMINADORES**

Prof. Sérgio Martins

Prof. Dra. Andrea Maria Accioly Fonseca Minardi

## RESUMO

O objetivo desse trabalho é explorar se o sentimento presente em redes sociais é capaz de indicar movimentos subsequentes nos preços de ações e índices. Especificamente, é investigado se o conteúdo publicado no twitter pode fornecer informações sobre movimentações futuras no preço de uma amostra de 12 ações americanas. É considerado um período de 75 dias. Os tweets publicados nesse período que fazem referência ao grupo amostral são coletados e classificados como positivos, negativos ou neutros. Essa classificação é feita usando um algoritmo de *machine learning* conhecido como Algoritmo de classificação por Regressão Logística Multinomial com Gradiente Descendente Estocástico. Se conduz então um Estudo intradiário de Eventos para capturar a complexa relação entre as redes sociais e os preços de ativos financeiros. Os dados são estudados em intervalos de 5 minutos entre cada observação . Espera-se comprovar a hipótese de que o Twitter é um bom instrumento para identificar mudanças nas expectativas do mercado antes que essas sejam completamente incorporadas nos preços dos ativos financeiros.

Palavras-chave: Redes sociais. Mercado Financeiro. Twitter. Ações. Expectativas. Finanças comportamentais. Sentimento. Humor. Machine Learning. Regressão Logarítmica. Gradiente Descendente Estocástico. Estudo de eventos.

## ABSTRACT

This research paper explores wether the sentiment of social media is capable of indicating subsequent movements in prices of stocks and indices . In particular, it focuses on the effects of Twitter postings on the future movements in stock prices of a sample of 12 U.S. companies. A time window of 75 days is considered. The tweets published within that time period and which reference any of the sampled companies are colected and classified as either positive, negative or neutral. This is accomplished using a machine learning algorithm known as Stochastic Gradient Descent Logistic Regression Classifier. Then, an intraday Event Study is conducted in order to capture the unorthodox relationship between social media and asset prices. A high frequency 5-minute-interval time series is considered. It is expected to confirm the hypothesis that Twitter is a good proxy for identifying market expectations before they get completely incorporated in the stock prices.

Key words: Social Media. Stock Market. Twitter. Stocks. Expectations. Behavioral Finance. Sentiment. Humor. Machine Learning. Logistic Regression. Stochastic Gradient Descent. Event Study.

# SUMÁRIO

**1. Introdução ..... 10**

**2. Revisão da literatura ..... 14**

**3. Bases de Dados .....19**

**3.1 Ativos financeiros ..... 19**

**3.2 Twitter ..... 21**

**4. Metodologia ..... 24**

**4.1 Análise de sentimento dos *tweets* ..... 24**

**4.2 Estudo de eventos ..... 33**

**5. Resultados ..... 39**

**5.1 Análise de sentimento dos *tweets* ..... 39**

**5 Resultados ..... 44**

**6. Conclusão..... 50**

**Referências ..... 51**

**Anexos ..... 54**

## Lista de Tabelas

**Tabela 1 –** Empresas componentes do índice Dow Jones(2019) ..... **20**

**Tabela 2 –** Lista de ações consideradas no estudo ..... **22**

**Tabela 3 –** Excerto da Base de Dados Final .....**26**

**Tabela 4 –** Métricas de Performance do Algoritmo Classificador..... **42**

**Tabela 5 –** Valores de CAR e $\hat{\theta}$ referentes a menções ao Ticker das empresas. ..... **46**

**Tabela 6 –** Valores de CAR e $\hat{\theta}$ referentes a menções textuais às empresas. ..... **47**

## Lista de Figuras

**Figura 1 –** Classificação dos tweets, por amostra, entre negativo, positivo e neutro ..... **26**

**Figura 2 –** Função Sigmoide ..... **30**

**Figura 3 –** Frequência Absoluta de termos no vetor vocabulário, por categoria ..... **39**

**Figura 4 –** Termos com maior valor $tfidf$ no vetor vocabulário, por categoria ..... **40**

**Figura 5 –** Parameter Hyper Tunning ..... **41**

**Figura 6 –** Matriz de Confusão para tweets da amostra de teste classificados..... **43**

**Figura 7 –** Volume de Tweets para menções textuais ao Walmart, por 5 minutos ..... **44**

**Figura 8 –** Distribuição de polaridade de sentimento para os picos detectados..... **45**

**Figura 9 –** Retorno Anormal Acumulado para os eventos detectados referentes a menções ao Ticker das empresas..... **46**

**Figura 10 –** Retorno Anormal Acumulado para os eventos detectados referentes a menções ao textuais às empresas..... **47**

# 1.INTRODUÇÃO

Nos últimos anos, a convergência das taxas de juros das principais economias mundiais para valores próximos de zero tornou, cada vez mais, o mercado de renda variável uma opção atrativa para o investidor. Em 2017, as transações desse mercado movimentaram mais de 75 trilhões de dólares pelo mundo. Comparativamente, em 1985 essas mesmas transações somavam pouco menos de 2 trilhões de dólares, de acordo com dados do Banco Mundial. Nos Estados Unidos, por exemplo, a porcentagem de famílias detentoras de ações foi de 40% em 1989 para quase 60% em 2016 (St. Louis Fed, 2018). Com o aumento da relevância da renda variável no cenário de investimentos global, é natural que tenha se intensificado o debate sobre estratégias de investimento eficientes, levando em conta retorno, risco e liquidez à que se sujeitam qualquer indivíduo.

Nesse cenário, as estratégias de investimento costumam se enquadrar, de forma geral, em dois grandes grupos: (i) A análise fundamentalista projeta o valor das empresas com base no entendimento de fluxos financeiros e dinâmicas operacionais da companhia, bem como da estratégia de crescimento futura da empresa. Enquanto (ii) a análise técnica se baseia no estudo de padrões gráficos e dados históricos de mercado para inferir o valor justo dos ativos.

No entanto, Burton Malkiel, autor de *A Random Walk Down Wall Street,* defende a ideia de que são ineficientes quaisquer estratégias que objetivem precificar ações com base em sua performance passada, dado a natureza aleatória que se verifica nessas séries temporais. Para Malkiel, 1973, “um macaco vendado lançando dardos na página financeira produzirá uma carteira da qual se pode esperar performance igual a de qualquer outra carteira gerenciada por analistas profissionais de títulos.”

E, de certa forma, as evidencias empíricas apoiam a ideia defendida por Malkiel em seu livro. Nos Estados Unidos, por exemplo, 91,6% dos fundos de investimento tiveram performance inferior ao índice de mercado (S&P 500) no retorno acumulado dos últimos 15 anos (CNBC, 2019).

Nesse contexto, surgem algumas teorias que buscam explicar o movimento dos preços de ativos financeiros a partir de outros fatores. Estudiosos do ramo de

finanças comportamentais parecem concordar que o preço de ativos financeiros é, em grande parte, guiado por fenômenos psicológicos (SHILLER, 2000). Para eles, o humor e o otimismo dos investidores afeta o preço das ações. Sendo assim, as expectativas do mercado parecem ter um papel importantíssimo na formação dos preços.

Se esse for o caso, o monitoramento das expectativas de mercado e do humor dos investidores poderia conter informações a respeito de preços de ativos financeiros mais valiosas que as estratégias tradicionais mais comumente utilizadas. Estudos mostram que, hoje em dia, é possível fazer esse tipo de monitoramento através das redes sociais (CHOI; VARIAN, 2012).

Nos últimos anos houve um crescimento expressivo no uso de redes sociais. Nelas, a sociedade encontrou uma forma de se expressar e se conectar com o mundo. Esses desdobramentos implicam que, pela primeira vez na história, é possível monitorar emoções, pensamentos e opiniões de uma grande parcela da população em tempo real - e com custos de aquisição de dados praticamente inexistentes (ZHELUDEV et al.,2014). Dessa maneira, o objetivo deste trabalho será investigar se as opiniões e o humor dos usuários das redes sociais, referidos aqui como o “sentimento" das redes sociais e quantificado através de métodos estatísticos de classificação semântica, podem conter informações relevantes, ex-ante, sobre o desempenho futuro de ações e índices. Aqui, não se objetivará a criação de uma estratégia de compra e venda nem de um algoritmo de previsão. O objetivo aqui é, simplesmente, investigar este tema relativamente novo e pouco estudado mas que parece ser tão promissor na constituição da precificação de ativos financeiros.

A rede social escolhida nesse estudo é o Twitter. Nele, milhares de indivíduos e organizações do mundo inteiro publicam pequenas mensagens de até 140 caracteres. Essas mensagens não possuem um recipiente específico. Elas são disponibilizadas em tempo real para todos os usuários. O crescimento da plataforma desde sua criação em 2006 foi expressivo. Em 2019, o Twitter reportou 326 milhões de usuários ativos por mês e 500 milhões de tweets feitos por dia. Ao longo dos anos, o Twitter se tornou uma plataforma poderosa de compartilhamento de notícias e opiniões variadas.

Não por acaso, empresas começaram a fazer uso da plataforma para comunicarem-se com clientes e investidores. Vários anúncios passaram a ser feitos pelo Twitter. Um exemplo proeminente disso foi quando o CEO da Netflix, Redd Hastings, anunciou em sua conta pessoal que sua empresa havia ultrapassado a marca de 1 bilhão de horas assistidas por mês. A notícia se espalhou primeiro nas redes sociais e demorou para ser reportada pela mídia tradicional. Isso causou um aumento de 6% no preço da ação no período imediatamente posterior ao anúncio. Hastings foi acusado de transmitir informação privilegiada, visto que as redes sociais não eram consideradas meios de divulgação oficiais pela SEC, a comissão de valores mobiliários americana. Iniciou-se então uma investigação por parte da SEC, com o objetivo de avaliar se as redes sociais eram um canal passível de divulgação de conteúdo sensível sobre as empresas. Por fim, no dia 2 de abril de 2013, a SEC legislou à favor da inclusão das redes sociais na lista de canais oficiais de divulgação de resultados. Isso só aumentou a relevância do conteúdo publicado nesse tipo de plataforma.

Desde então, muitos outros tweets polêmicos já foram capazes de provocar grandes movimentações nos preços de algumas ações. O presidente dos Estados Unidos, Donald Trump, é famoso por suas declarações polêmicas no twitter. Em Janeiro de 2017, após twittar sobre a empresa Toyota, afirmando que aumentaria os impostos pagos pela montadora, o valor da empresa caiu em mais de 1.2 bilhões de dólares. O bilionário e investidor Carl Icahn, após fazer uso da sua conta para revelar sua posição otimista nas ações da Apple, causou um aumento de mais de 5% no valor da empresa minutos depois. Em 2020, o empresário Elon Musk revelou pelo Twitter que acreditava que as ações da Tesla, companhia que é CEO, estavam supervalorizadas. Nos minutos seguintes o valor de mercado da companhia caiu 10%.

É bastante comum também que o mercado financeiro responda a mudanças mais generalizadas no sentimento das redes sociais. Em 2018, a blogueira e *influencer* digital, Kylie Jenner publicou no Twitter que tinha abandonado a rede social SnapChat. Milhares de pessoas responderam à pergunta de Jenner, publicando *tweets* afirmando que também haviam feito o mesmo. Os usuários, de modo geral, também se revelaram pessimistas sobre o assunto. Esse pessimismo foi refletido no

preço das ações da companhia controladora, Snap Co., que foi caindo ao longo do dia. No fechamento das negociações, o valor de mercado da companhia havia caído em quase 7%.

Essas evidências motivam a realização de um estudo mais cuidadoso. A análise conduzida aqui terá um foco em dados intradiários. Espera-se comprovar a hipótese de que o twitter é um bom instrumento para identificar mudanças nas expectativas do mercado antes que essas sejam completamente incorporadas nos preços dos ativos financeiros. Nesse contexto, será realizado um estudo de eventos com o objetivo de entender se o sentimento expressado em tweets é um fator relevante para explicar retornos anormais em perídos subsequentes.

## 2.REVISÃO DA LITERATURA

Existem diversos modelos teóricos para descrever a forma com que o mercado precifica ativos financeiros. Dentre elas, uma das teorias mais disseminadas e amplamente aceitas pela comunidade acadêmica é a Hipótese do Mercado Eficiente (HME), desenvolvida por Eugene Fama, 1970. Segundo Fama, assumindo que todos os investidores sejam racionais, o preço das ações reflete inteiramente as informações disponíveis, sendo assim, o mercado sempre negocia as ações pelo seu valor justo. A suposição de eficiência dos mercados postula que informações novas são incorporadas nos preços imediatamente, impedindo, portanto, que o investidor tenha tempo de negociar com base nessas informações, algo que resultaria em arbitragem (FAMA, 1970). Sob esse conjunto de premissas, os coeficientes de correlação serial devem ser próximos a zero e o preço de ativos financeiros é regido por um passeio aleatório (ROSS et al., 2012). Isso implica que os preços futuros desses ativos não são determinado por preços passados, mas sim por informações novas (QIAN; RASHEED, 2007), como notícias e outros desdobramentos. No entanto, mesmo que notícias não possam ser previstas antes que os eventos noticiados aconteçam, pesquisas recentes indicam que sinais prematuros podem ser extraídos das redes sociais (ASUR et al., 2010; CHOI et al., 2012). De fato, é cada vez mais comum que veículos de comunicação tradicionais reportem acontecimentos que foram originalmente publicados em redes sociais.

E se as redes sociais são capazes de influenciar as notícias, certamente os sentimentos e opiniões dos usuários também o são. Pesquisas recentes em psicologia indicam que emoções e humor afetam o processo de tomada de decisão (DAMASIO, 1994). Essa conclusão também foi verificada por estudos de finanças comportamentais (NOFSINGER, 2005). A partir disso, podemos inferir que o “sentimento” dos indivíduos nas redes sociais pode influenciar preços de ações de maneira significativa.

Além disso, estudos mais recentes trazem evidências de que os mercados não são totalmente eficientes tampouco os investidores são racionais (SHILLER, 2000). Pesquisadores de psicologia cognitiva verificaram que indivíduos desviam significativamente da racionalidade quando defronte a decisões que contemplam incertezas(MALKIEL, 1973). Kahneman e Tversky, em seu artigo “Prospect Theory:

An Analysis of Decision Under Risk", apontam que, frente a situações que envolvem risco, indivíduos apresentam padrões de comportamento inconsistentes com preceitos clássicos de maximização de utilidade esperada. Os autores propõem que perdas causam um impacto emocional maior do que ganhos, fazendo com que indivíduos tomem decisões que desafiem a racionalidade. Notadamente, indivíduos são avessos à risco quando confrontados com a possibilidade de ganhos, preferindo opções menos arriscadas mesmo que com utilidade esperada menor. Já quando confrontados com escolhas que envolvem perdas, indivíduos são amantes de risco, preferindo opções que podem levar a uma perda de utilidade maior desde que com potencial de eliminar qualquer perda (KAHNEMAN, TVERSKY, 1979).

Os *players* do mercado financeiro possuem, também, viés no seu julgamento. Orgulho, por exemplo, influencia a tomada de decisão dos investidores, que podem optar por manter em sua carteira uma posição com desempenho ruim, na esperança de que eventualmente a ação vai lhe render retornos positivos, postergando um inevitável arrependimento (MALKIEL, 1973; STAW, 1976). O conservadorismo também possui um importante papel nesse sentido. As pessoas são lentas para ajustar suas crenças à nova informação. Estudos relatam, por exemplo, que preços parecem ajustar-se lentamente após a divulgação de resultado de uma empresa (ROSS et al., 2012). De modo geral, as evidências apontam que o mercado não incorpora todas as informações relevantes imediatamente, como é defendido pela hipótese da eficiência. É possível, portanto, que em dado momento no tempo, um conjunto de informação e opiniões da população, cuja proxy pode ser obtida pelo conjunto de informação presente em uma rede social (RANCO et al., 2015), ainda não tenha sido incorporado completamente na série de preços dos ativos. É concebível, assim, que redes sociais ajudem a identificar mudanças nas expectativas do mercado antes que essas sejam completamente incorporadas nos preços de ativos financeiros.

O exposto acima, por si só, já é evidência suficiente para estudar mais a fundo a relação das redes sociais e do desempenho de ativos financeiros. Entretanto, existem outras grandes correntes de pensamento que podem embasar essa mesma premissa. Uma delas, e talvez a mais importante dentro do escopo proposto aqui, é conhecida como *Castle-in-the-air Theory*, proposta pelo economista John Maynard

Keynes (KEYNES, 1936). “To *Build a Castle-in-the-air*” é uma expressão inglesa que denota algum desejo, visão ou plano improvável de se materializar.

De acordo com Keynes, lucros devem ser obtidos quando prevemos mudanças um pouco antes do mercado como um todo. Para ele, princípios fundamentalistas não são tão importantes, ao menos no curto prazo, para determinar o valor intrínseco dos ativos. Pelo contrário, ele defende que o maior determinante do preço dos ativos é a percepção e a expectativa do mercado como um todo. É importante, portanto, que o investidor seja capaz de analisar como os outros investidores devem se comportar no futuro e como, durante períodos de otimismo, estes investidores tendem a construir suas expectativas fundadas em premissas irreais (daí a analogia *Castles in the air*) (MALKIEL, 1973). O investidor de sucesso, então, tenta estimar quais investimentos são mais propensos ao fenômeno de “*Castle-building*” para, assim, comprar antes da maior parte dos investidores. No curto prazo, portanto, ser capaz de prever as expectativas do mercado em relação a um ativo é mais valioso que ser capaz de avaliar o valor justo do mesmo. Keynes escreveu: “Não é razoável pagar $25 por um investimento cujo valor justo, você acredita, é de $30, se você também acredita que o mercado vai precifica-lo em $20 daqui 3 meses”

A teoria *Castle-in-the-sky* enfatiza, portanto, os aspectos psicológicos do comportamento do mercado. Um investimento tem determinado valor para um comprador apenas porque ele acredita que em algum momento ele irá vender este mesmo investimento para alguém por um preço mais alto. O novo comprador, por sua vez, antecipa que compradores futuros pagarão ainda mais caro. Sendo assim, tudo que o investidor precisa ser capaz de fazer é se antecipar em relação ao mercado: Entrar na onda de otimismo o mais cedo possível e sair antes que este otimismo se torne realismo, destruindo o castelo. De acordo com Burton Malkiel, em seu livro *A Random Walk Down Wall Street*, “É perfeitamente aceitável pagar 3 vezes o valor justo de alguma coisa, desde que você consiga achar alguém inocente o bastante para pagar 5 vezes esse valor.”

O Nobel em economia Robert Shiller, no seu livro Irrational Exuberance, argumenta que vários eventos do mercado financeiro só podem ser explicados em

termos das expectativas, opiniões e otimismo do mercado. Para Schiller, esses fenômenos são psicológicos, e contrários as premissas tradicionais de racionalidade.

Acontece que nem sempre foi possível medir as expectativas, opiniões e o otimismo de forma quantitativa. Por isso, estratégias de precificação de ativos financeiros que se baseavam nos princípios de *Castle-in-the-sky* poderiam se provar pouco eficientes (MALKIEL, 1973). No entanto, o crescimento expressivo das redes sociais nos últimos anos implica que pela primeira vez na história, é possível monitorar emoções, pensamentos e opiniões de uma grande parcela da população em tempo real - e com custos de aquisição de dados praticamente inexistentes (ZHELUDEV; SMITH; ASTE, 2014).

Com base nesta suspeita de que as opiniões e o humor dos usuários das redes sociais são uma boa proxy para estimar as expectativas do mercado, alguns estudos já foram realizados. Na verdade, as expectativas refletidas nas redes sociais se mostram correlacionadas com vários indicadores econômicos. Um desses estudos, por exemplo, mostra que a atividade de bate papos online pode ajudar a prever a venda de livros (GRUHL et al., 2005). Outro, usa o conteúdo de blogs e outras redes sociais para prever o faturamento de filmes nas bilheterias (ASUR et al., 2010; GILAD et al., 2010). Um terceiro estudo verifica que as buscas mais frequentes do Google contém indicadores preliminares de surtos de doenças e da propensão a consumir dos consumidores (CHOI; VARIAN, 2012). Além disso, um estudo realizado durante as eleições britânicas mostrou que o sentimento dos usuários no Twitter estava positivamente correlacionado com a performance das ações que compõem o índice FTSE 100 (NISAR; YEUNG, 2018).

Finalmente, um estudo realizado em 2015 por Ranco et al. com os 30 ativos que compõem o índice Dow Jones, ao longo de 15 meses, revelou que o sentimento presente em tweets indicava movimentos subsequentes nos preços dos ativos. O artigo faz uma contribuição relevante ao reconhecer que a relação entre tweets e ações não se faz presente a todo tempo, e portanto sua dinâmica não pode ser verificada através de medidas convencionais de correlação. Por esse motivo, os autores optam por realizar um estudo de eventos, analisando apenas retornos anormais observados em momentos de alto volume de menções a uma determinada ação no twitter - chamados picos. O estudo concluiu que ao restringir a análise a

períodos de tempo mais curtos – janelas de evento – existe uma relação estatisticamente significante entre sentimento do Twitter e os retornos anormais observados nas ações investigadas, em um período de até 8 dias após a ocorrência do evento. (RANCO et al., 2015).

O presente trabalho busca adicionar ao que foi demonstrado por Ranco et al. Assim como no artigo base, será feito um estudo de eventos com as ações que compõem o índice Dow Jones. No entanto, objetivando entender como o sentimento do Twitter se relaciona com os movimentos da bolsa de forma mais imediata, analisaremos os dados intra-day, com intervalos de 5 minutos entre cada observação. Além disso, enquanto Ranco et al. apenas considera tweets com menções a símbolos de ticker ($cashtags) da companhia, esse estudo analisará também o efeito de menções ao perfil oficial da empresa (@perfil) no Twitter e citações ao nome da empresa de forma geral. O período analisado é entre 04 de fevereiro de 2019 e 20 de abril de 2019. Assim como no artigo base, a hipótese nula $H_0$ é de que eventos externos não possuem impacto futuro no comportamento dos retornos observados, ou seja, os mercados são de fato eficientes e os preços reagem instantaneamente, não sendo possível observar retornos anormais. A hipótese alternativa $H_A$ é de que eventos externos, aqui representados pela proxy de sentimento de mercado extraída do Twitter, possuem impacto no comportamento futuro desses retornos.

# 3. BASES DE DADOS

A viabilização deste trabalho depende de um conjunto de bases de dados, muitas delas não disponíveis publicamente. Por se tratar de um estudo que requer dados intradiários, existem inúmeras restrições quanto aos dados disponíveis, que serão discutidas mais à frente.

O estudo dos preços intradiários, por sua vez, é apropriado para investigar a capacidade das redes sociais de capturarem uma mudança nas expectativas do mercado antes que essa mudança seja completamente incorporada no preço dos ativos. A análise diária provavelmente não seria suficiente para este tipo de estudo.

A seguir, são descritas as duas principais bases de dados utilizadas neste trabalho, bem como suas limitações.

## 3.1 Ativos financeiros

Alguns fatores foram determinantes na escolha das empresas participantes deste estudo. Primeiramente, só é possível realizar a análise intradiária para ativos com alto volume de negociação. Além disso, a relevância de uma empresa no cenário cotidiano influencia a quantidade de informação disponível no Twitter sobre ela. Em um estudo de 2013 realizado por Forbergskog et al., argumenta-se que a falta de informações sobre *small caps* no Twitter torna análises de sentimento praticamente impossíveis. Por estas razões, o presente estudo, tal como outros neste tema[1], opta por analisar as empresas componentes do Índice Industrial Dow Jones (DJIA). Essas empresas não só respeitam as restrições acima como também formam um portfolio de mercado bastante diversificado. O DJIA é composto por 30 das maiores e mais influentes empresas americanas. Ao todo, as companhias integrantes do índice representam em torno de 25% do valor total do mercado de ações americano. A Tabela 1 lista as empresas componentes do índice, bem como alguns indicadores relevantes[2]:

---

[1] Dentre os estudos similares que também usam o DJIA, estão: BOLLEN; MAO; ZENG, 2011; FORBERGSKOG; BLOM, 2013; MITTAL 2011; RANCO et al., 2015

[2] Acesso feito em 30/05/2019. Note que a empresa Dow Inc, resultante da fusão da Dow Chemical com a DowDuPont, foi listada na bolsa de valores em abril de 2019. Por esse motivo, a Dow Inc. não apresenta dados históricos suficientes e não foi considerada neste estudo.

**Tabela 1 –** Empresas componentes do índice Dow Jones (2019)

| Ticker | Empresa | Setor de Atuação | Valor de mercado | Volume Médio | Índice P/L |
|---|---|---|---|---|---|
| **JNJ** | Johnson & Johnson | Farmacêutico | 351B | 6.6M | 24.48 |
| **DIS** | The Walt Disney Company | Entretenimento | 238B | 14.2M | 14.79 |
| **MSFT** | Microsoft | Tecnologia | 963B | 24.7M | 27.93 |
| **PG** | The Procter & Gamble Company | Bens de consumo | 264B | 7.4M | 25.05 |
| **PFE** | Pfizer Inc. | Farmacêutico | 233B | 24M | 21.52 |
| **IBM** | International Business Machines | Tecnologia | 115B | 2.6M | 13.63 |
| **INTC** | Intel Corporation | Tecnologia | 200B | 24.1M | 10.10 |
| **MMM** | 3M Company | Químico | 93B | 3.1M | 16.97 |
| **MRK** | Merck & Co | Farmacêutico | 205B | 10.5M | 25.16 |
| **WBA** | Walgreens Boots Alliance | Varejo | 46B | 7.3M | 9.49 |
| **CSCO** | Cisco System | Tecnologia | 229B | 20.5M | 18.59 |
| **AAPL** | Apple | Tecnologia | 820B | 29.8M | 15.00 |
| **XOM** | Exxon Mobil Corporation | Energia | 305B | 11.6M | 16.58 |
| **UNH** | UnitedHealth Group Incorporated | Saúde | 231B | 6M | 18.91 |
| **KO** | The Cola-Cola Company | Bebidas | 210B | 13.7M | 31.31 |
| **CAT** | Caterpillar | Industrial | 70B | 4.3M | 11.33 |
| **GS** | The Goldman Sachs Group | Financeiro | 69B | 2.8M | 7.80 |
| **NKE** | NIKE | Vestuário | 125B | 6.8M | 30.88 |
| **MCD** | McDonald's Corporation | Restaurantes | 152B | 2.7M | 26.40 |
| **TRV** | The Travelers Companies | Seguros | 38B | 1.3M | 14.81 |
| **V** | Visa | Financeiro | 356B | 7.7M | 33.63 |
| **CVX** | Chevron Corporation | Energia | 220B | 8.2M | 15.95 |
| **UTX** | United Technologies Corporation | Aeroespacial | 111B | 3.2M | 19.95 |
| **VZ** | Verizon Communications | Telecomunicação | 235B | 13.6M | 14.68 |
| **WMT** | Walmart Inc. | Varejo | 293B | 4.7M | 35.76 |
| **HD** | The Home Depot | Varejo | 211B | 4.2M | 19.24 |
| **AXP** | American Express | Financeiro | 97B | 3.3M | 14.87 |
| **JPM** | JPMorgan Chase | Financeiro | 347B | 12.4M | 11.55 |
| **DOW** | Dow Inc. | Químico | 36B | 6.3M | - |
| **BA** | The Boeing Company | Aeroespacial | 197B | 7.9M | 20.03 |

Fonte: Bloomberg

Finalmente, os preços históricos intradiários foram coletados através da plataforma Bloomberg. O período considerado é de 04 de fevereiro outubro de 2019

a 20 de abril de 2019. Os dados serão analisados minuto a minuto, levando em conta apenas o preço de fechamento realizado à cada minuto.

### 3.2 Twitter

Esse trabalho, tal como alguns outros já existentes(ASUR; HUBERMAN, 2010; BOLLEN; MAO; ZENG, 2011; MAKREHCHI; SHAH; LIAO, 2009; RANCO et al., 2015; ZHELUDEV; SMITH; ASTE, 2014), irá se limitar ao Twitter como a única rede social estudada. Tal restrição é motivada por algumas particularidades do Twitter que facilitam o monitoramento de "humor" dos seus usuários. Dentre elas, o enfoque em textos curtos de até 140 caracteres é um grande ponto positivo para coleta dos dados. Além disso, a cultura de micro-blogging presente na plataforma incentiva que usuários opinem objetivamente sobre diversos tópicos.

Um estudo recente realizado por Smith et al (2012) fornece ainda mais evidências à favor do supracitado. Ao comparar o conteúdo disseminado no Twitter com o conteúdo de outras redes sociais, como o Facebook e o Youtube, identificou-se que o os usuários do Twitter são os que mais fazem uso da plataforma para engajar em discussões e disseminar informações. Além disso, Chat et al (2010) verificou que tweets cujo conteúdo está atrelado à notícias são os que produzem mais engajamento dentro da plataforma. Estes estudos reforçam a crença de que, dentre as redes sociais, o Twitter seja o que forneça o semblante mais verossímil do humor do mercado em dado momento.

Por meio do serviço Twitter Historical Powertrack, foram coletados 13.564.221 tweets referentes às ações do índice Dow Jones, publicados entre 04 de fevereiro de 2019 a 20 de abril de 2019. Com o objetivo de capturar o maior número possível de referências às empresas estudadas, recuperou-se todos os tweets que faziam menção aos nomes das empresas, aos tickers de mercado ou aos perfis oficiais das empresas. Variantes comuns também foram consideradas, como por exemplo "Johnson & Johnson" e "Johnson and Johnson". É importante entender que uma parcela dos tweets filtrados dessa maneira não é relevante para este estudo. Existem, por exemplo, menções à palavra "apple" para se referir à fruta e não à empresa. Para evitar capturar esse ruído, não foram considerados, por exemplo, tweets com menções à palavra "apple" que também mencionassem palavras como "fruta",

“banana”, “comida” etc. A quantidade remanescente desses tweets atípicos foi considerada irrelevante, o que torna possível tratá-los como ruído amostral. O Anexo I fornece a lista completa dos filtros utilizados para o levantamento desses tweets. Para o período do estudo, todos os tweets feitos na plataforma que se aplicavam aos filtros especificados foram recuperados[3].

Por fim, foram excluídas empresas cujo volume de menções totais no twitter dentro do período era inferior a 50 mil. Isso é feito para garantir que valores insignificantes nos resultados obtidos nas estimações não sejam exclusivamente decorrentes de pequenos tamanhos amostrais. A Tabela 3 lista as empresas restantes para o estudo, bem como o volume de menções por categoria e o volume total de tweets para cada uma delas:

**Tabela 2 –** Lista de ações consideradas no estudo

| Ticker | Empresa | Tweets $cashtag | Tweets @perfil | Tweets “Texto” | Total |
|---|---|---|---|---|---|
| **DIS** | The Walt Disney Company | 24.420 | 681.075 | 559.548 | 1.265.043 |
| **MSFT** | Microsoft | 26.922 | 190.520 | 160.885 | 378.367 |
| **IBM** | International Business Machines | 7.765 | 97.182 | 56.538 | 161.485 |
| **AAPL** | Apple | 91.221 | 332.238 | 1.231.085 | 1.654.544 |
| **XOM** | Exxon Mobil Corporation | 8.836 | 27.389 | 15.171 | 51.396 |
| **GS** | The Goldman Sachs Group | 18.307 | 23.063 | 18.182 | 59.552 |
| **NKE** | NIKE | 9.234 | 348.024 | 246.917 | 604.175 |
| **MCD** | McDonald's Corporation | 5.846 | 207.773 | 86.655 | 300.274 |
| **VZ** | Verizon Communications | 5.801 | 96.845 | 28.050 | 130.696 |
| **WMT** | Walmart Inc. | 12.423 | 230.086 | 148.878 | 391.387 |
| **JPM** | JPMorgan Chase | 17.816 | 64.807 | 24.268 | 106.891 |
| **BA** | The Boeing Company | 46.611 | 111.944 | 170.000 | 328.555 |

Fonte: Elaboração Própria

O resultado final é uma base de dados com a classificação aproximada do “sentimento” dos tweets na forma de série temporal. São contabilizados o número tweets positivos, negativos ou neutros (além do volume total) para cada minuto *t,*

[3] Exceto menções textuais aos nomes das companhias, que foram limitados a 10% do total de tweets feitos no período, devido ao grande volume existente.

para cada companhia considerada no estudo. A outra base de dados é uma série temporal com os retornos das ações do índice escolhido.

O estudo dos preços *intra-day* é apropriado porque reflete o objetivo do trabalho de demonstrar a capacidade das redes sociais de capturarem uma mudança nas expectativas do mercado antes que essa mudança seja completamente incorporada no preço dos ativos. A análise diária provavelmente não seria suficiente para este tipo de estudo.

# 4. METODOLOGIA

## 4.1 Análise de sentimento dos Tweets

A análise de sentimento é uma técnica amplamente disseminada na literatura. Ela consiste, fundamentalmente, na classificação do conteúdo dos tweets como positivo, negativo ou neutro. O sentimento dos tweets, obtido por meio desse procedimento, pode ser interpretado como uma proxy para o humor geral do mercado, que por sua vez deve afetar o preço dos ativos financeiros.

No entanto, existem algumas complicações inerentes à realização de uma análise de sentimento para esse tipo de estudo. Classificar manualmente os milhões de tweets contidos na base de dados deste trabalho seria inviável. E mesmo se tal procedimento fosse possível, a precisão da classificação não seria absoluta. Um experimento conduzido por Mozetic et al. (2016), no qual múltiplos indivíduos foram encarregados de rotular uma mesma amostra de tweets, mostrou que só havia consenso entre os avaliadores para cerca de 75% dos tweets classificados. Outro estudo, realizado por Nar S. (2012), também verificou que, na média, indivíduos só concordam em cerca de 70% dos textos que classificam.

Opta-se aqui, portanto, em utilizar um método de classificação automático para os tweets. Tal procedimento, baseado em métodos estatísticos de Machine Learning supervisionado, é capaz de classificar os milhões de tweets da amostra em poucas horas. O método de classificação escolhido neste estudo é um Algoritmo de classificação por Regressão Logística Multinomial com Gradiente Descendente Estocástico, tal como descrito no livro “Speech and Language Processing” do professor de Stanford, Daniel Jurafsky. A escolha deste método se justifica por diversos estudos que comprovam a superioridade de Regressões Logísticas para classificação textual, quando comparado à outros métodos de classificação, como Naive Bayes Classifier, K-Nearest-Neighbor e Conditional Random Fields (ASIAEE T. et al., 2012; PRANCKEVIČIUS et al. 2017). Por esse motivo, o Algoritmo de Classificação por Regressão Logística é um dos mais difundidos na literatura para classificação de Tweets.

Contudo, para que a classificação forneça altos graus de precisão, um número suficiente de tweets classificados à mão precisa ser fornecido ao algoritmo. Esse procedimento é conhecido na literatura como processo de treinamento do modelo. De modo geral, quanto maior o número de anotações manuais fornecidas, maior a precisão verificada no modelo (NARR; ULFENHAUS; ALBAYRAK, 2012). Normalmente, 10.000 anotações manuais é tido como o mínimo necessário para produzir resultados relevantes. No entanto, é importante ressaltar que um número excessivo de anotações manuais torna o modelo super especificado. Nesse caso, o algoritmo se torna demasiadamente ajustado aos choques e ao ruído da amostra, prejudicando a precisão de classificação (JAMES et al., 2013).

Para os fins deste estudo, serão combinadas 8 bases de dados distintas, disponíveis publicamente. Essa combinação é importante porque cada base de dados traz características linguísticas diferentes. A mistura de excertos de texto mais formais e econômicos com textos de linguagem mais informal torna o algoritmo de classificação mais generalista, o que é apropriado para se analisar uma rede social tão heterogenia como o Twitter. Em todas essas bases, a classificação dos tweets foi feita manualmente, por múltiplos indivíduos, eliminando parte da inconsistência individual de classificação. A base de dados final é dividida em duas amostras menores, uma de treinamento e outra de teste. A amostra de treinamento, com 75% de todos os tweets, é utilizada para treinar o algoritmo de classificação. A amostra de teste, com os 25% dos tweets restantes, é utilizada para averiguar a performance do algoritmo, depois de treinado. Os tweets da amostra de teste são classificados pelo algoritmo de classificação e as decisões desse algoritmo são comparadas com as classificações manuais originais, estabelecendo-se métricas de acurácia e precisão. Uma breve descrição de cada uma das bases de dados que compõem as amostras de treinamento e teste, bem como o peso relativo de cada uma, podem ser encontrados no anexo II.

A amostra final de treinamento consiste de 24.600 tweets classificados à mão, como positivos, negativos ou neutros, de acordo com o sentimento transmitido pelos usuários. A amostra final de teste consiste de 10.500 tweets. Um pequeno excerto dessa base bem como a distribuição de classes de sentimento em cada uma das amostras é apresentado abaixo:

**Tabela 3 –** Excerto da Base de Dados Final

| *Tweet* | Sentimento |
|---|---|
| *@Apple siri is amazing* | Positivo |
| *Wow the new #google #nexus is just beautiful. Totally gonna boost google's market share in the smart phone market.* | Positivo |
| *So, what is behind that giant silver door at the @apple store? #maclife* | Neutro |
| *Bill gates to testify in novel v. #microsoft antitrust lawsuit http://t.co/h0gscvzr* | Neutro |
| *@Apple - #siri is not working due to "network problem"? Seriously? #iphone4s #fail* | Negativo |
| *#Skype is now part of stupid #microsoft it seems* | Negativo |

Fonte: Elaboração Própria

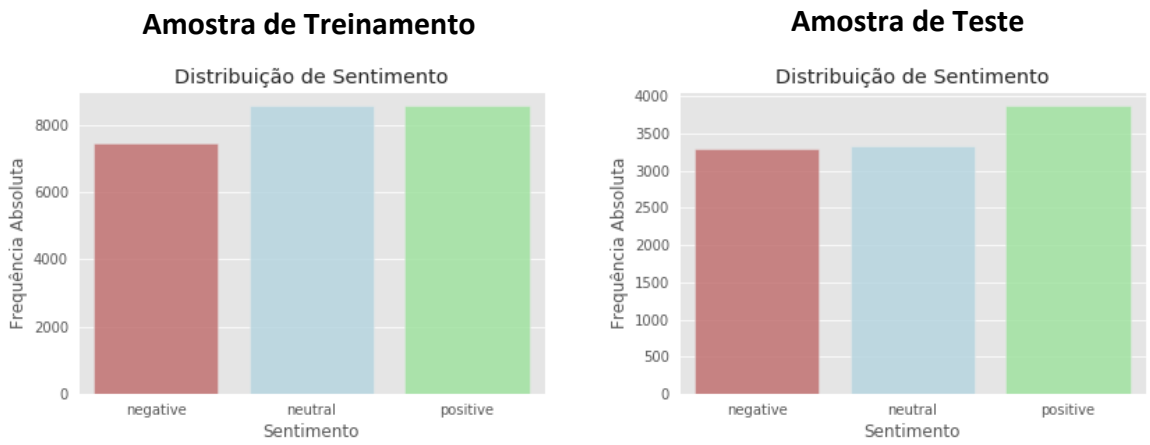



**Figura 1 –** Classificação dos tweets, por amostra, entre negativo, positivo e neutro

Fonte: Elaboração Própria

É importante ressaltar que a quantidade de tweets “negativos”, “positivos” e “neutros” é muito parecida. Esse equilíbrio é desejável e foi determinante ao se compor essas amostras. Se esse não fosse o caso, os estimadores gerados pelo algoritmo seriam viesados. É muito fácil, por exemplo, obter precisão de classificação de 99% quando 99% da sua base pertence a mesma categoria.

Além disso, a fim de facilitar a classificação, foram removidos os componentes textuais que não contribuem para a análise de sentimento. Esse processo foi realizado tanto para a amostra de treinamento como para a amostra de teste. Dentre o que foi suprimido estão: URLs, emojis, nomes de usuário, pontuação, preposições, conectivos e artigos. O texto restante passou por um processo de uniformização.

Todos os caracteres foram padronizados em minúsculo, todas as *hashtags* foram desassociadas do seu símbolo identificador padrão (#) e as palavras escritas com caracteres repetidos foram normalizadas para ortografia usual. As palavras foram também deflexionadas em um processo conhecido como lematização. Essa técnica consiste em ignorar tempos verbais, gênero, sufixos etc. A título de exemplo, o verbo "foram" se torna a palavra "ir". Isso reduz o vocabulário da língua inglesa consideravelmente, facilitando a classificação.

Ademais, vídeos e imagens também foram desconsiderados. Embora esses últimos sejam muito valiosos para análise de sentimento, interpretá-los via inteligência artificial é incrivelmente complexo e vai além do escopo deste estudo.

Por fim, o conteúdo restante de cada tweet é desagregado e as palavras extraídas são armazenadas em vetores denominados Vetores Normalizados. Abaixo, é fornecido um exemplo de um tweet antes e depois do processo de normalização:

| ***Tweet* original** | **Vetor Normalizado** |
|---|---|
| *#Google is the best invention eveeer!! @google #amazing* | {"google", "best", "invention", "ever", "amazing"} |

Em cada Vetor Normalizado, é atribuído um valor numérico, para cada palavra, inversamente proporcional a frequência com que o termo aparece no total. Isso é feito para atribuir mais valor a palavras raras e incomuns e menos valor a palavras triviais da língua inglesa, usadas em quase toda frase. Essa abordagem, conhecida como Term Frequency - Inverse Document Frequency ($tfidf$) é difundida na literatura. Buckley et al. explora esse assunto em seu artigo escrito em 1988. A equação abaixo ilustra o cálculo de $tfidf$ para um termo qualquer $i$:

$$tfidf_i = tf_i \times \left( \ln \left[ \frac{(N+1)}{df_i} \right] + 1 \right) \tag{1}$$

Onde $tf_i$ é a frequência do termo em um tweet, $df_i$ é a frequência absoluta do termo no Vetor Vocabulário e $N$ é o número total de tweets.

Os Vetores Normalizados da amostra de treinamento são então agregados em um só vetor multidimensional, usado como parâmetro no algoritmo durante o processo de classificação. A esse vetor se dá o nome de Vetor Vocabulário, pois ele

agrega todo o vocabulário disponível na amostra de treinamento, junto a respectiva frequência de cada palavra.

Desta maneira, quando o algoritmo classifica um Tweet, são comparados o Vetor Normalizado resultante do Tweet com o Vetor Vocabulário da amostra de treinamento, e identificam-se palavras comuns aos dois. É gerado então um outro vetor, denominado Vetor de Atributos, que contém as palavras (atributos) comuns ao Vetor Normalizado e ao Vetor Vocabulário, além dos respectivos valores $tfidf_i$ de cada termo. Esse procedimento pode ser feito apenas com unigramas (1 termo, ex. "gosto"), apenas com bigramas (combinação de 2 termos, ex. "não gosto"), ou uma combinação de unigramas com bigramas. Embora evidências empíricas (TAN; WANG; LEE, 2002) mostrem que a combinação dos dois gera algoritmos classificadores mais precisos, isso será testado empiricamente aqui também.

O Vetor de Atributos é utilizado como parâmetro pelo algoritmo de classificação, cujo funcionamento é fundamentado em regressões logísticas e otimizado com o método de gradiente descendente estocástico. No entanto, visto que regressões logísticas inerentemente produzem resultados binários, uma adaptação é necessária para criar um algoritmo classificador multiclasse (positivo, negativo e neutro). A estratégia adotada aqui é conhecida como "one vs. all", e também é utilizada por Ranco et al. no artigo que serve de base para esse estudo. O procedimento consiste em criar 3 classificadores binários distintos, como ilustrado abaixo:

$$Classificador\ I: \begin{cases} 1 = Positivo \\ 0 = Neutro\ ou\ Negativo \end{cases}$$

$$Classificador\ II: \begin{cases} 1 = Neutro \\ 0 = Positivo\ ou\ Negativo \end{cases}$$

$$Classificador\ III: \begin{cases} 1 = Negativo \\ 0 = Neutro\ ou\ Positivo \end{cases}$$

Escolhe-se então a classe cujo classificador retornou a maior probabilidade de que o tweet pertença àquela respectiva classe. O procedimento padrão para criação de um classificador binário genérico é descrito abaixo.

Para um tweet qualquer, $x$, os termos são consolidados em um Vetor de Atributos $X = (x_1, x_2, x_3, \ldots, x_n)$, em que $x_i$ corresponde ao valor $tfidf_i$ do termo no tweet. O algoritmo de classificação produz um resultado, $y$, tal que $y = 1$ se o tweet

pertence aquela categoria e $y = 0$ se o tweet pertence a qualquer outra categoria. Queremos, portanto, calcular a probabilidade $P(y = 1|x)$ de um determinado tweet pertencer à determinada categoria.

Para calcular essa probabilidade, partimos da concepção que cada termo, $x_i$, do Vetor de Atributos tem um peso, $w_i$, que representa a importância daquele termo para decisão de classificação. Esse peso pode ser positivo, indicando que o termo está associado aquela classe, ou negativo, sinalizando que o termo não está associado aquela classe. Para o $Classificador\ I$, por exemplo, devemos esperar que palavras como "ótimo" tenham um peso bastante positivo, enquanto palavras como "ruim" tenham um peso bastante negativo. Além disso, consideramos também que existe um termo de viés, refletido no intercepto $b$, que é somado à equação.

Temos então que a soma ponderada das evidências de pertencimento à classe, $z$, é dada por:

$$z = \left(\sum_{i=1}^{n} w_i x_i\right) + b \tag{2}$$

Em que $z$ é expresso pela somatória, nos $n$ termos existentes no Vetor Vocabulário, do valor $tfidf_i$ do termo $x_i$ (>0 se o termo aparece ao menos uma vez no tweet e igual a 0, se o termo não aparece) multiplicado pelo seu peso $w_i$ e adicionado ao intercepto $b$.

Utilizando-se de álgebra linear, podemos reescrever a equação (2) pelo produto escalar dos dois vetores $w$ e $x$, onde $w \cdot x$ é a soma dos produtos dos elementos correspondentes de cada vetor:

$$z = w \cdot x + b \tag{3}$$

No entanto, nada na equação (3) força que $z \in [0,1]$. Na verdade, como os pesos são números reais, $z$ pode assumir qualquer valor entre $-\infty$ e $\infty$, não havendo como garantir que $z$ seja uma probabilidade.

Para criar uma probabilidade, $y$, vamos aplicar uma função sigmoide (logística) em $z$:

$$y = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

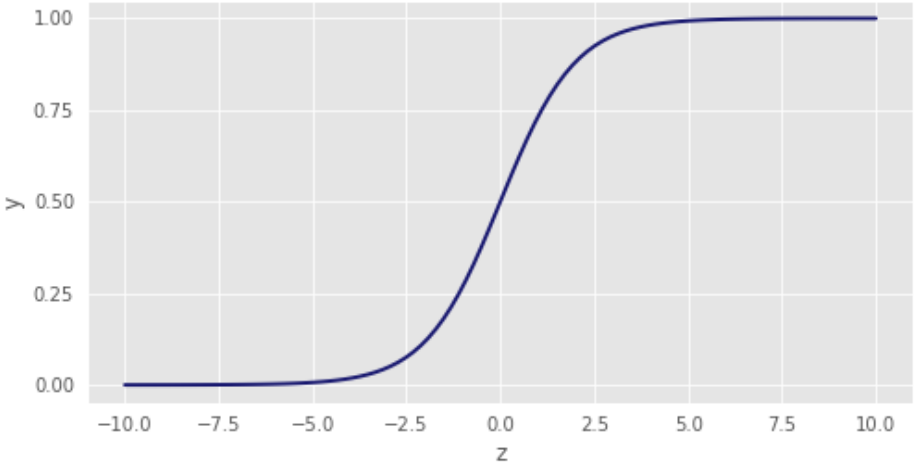



**Figura 2 –** Função Sigmoide

Fonte: Elaboração Própria

A função sigmoide traz uma série de vantagens. Ela transforma um número real em um número dentro do intervalo $[0,1]$, o que é desejável no nosso caso. Além disso, por ser quase linear ao redor do zero, mas ter uma inclinação mais acentuada em direção às extremidades da curva, ela tende a alocar outliers em direção a 0 ou 1.

Finalmente, para tornar $z$ uma probabilidade, precisamos garantir que os dois casos, $P(y = 1)$ e $P(y = 0)$, somem 1. Isso pode ser feito da seguinte forma:

$$P(y = 1) = \sigma(w \cdot x + b) = \frac{1}{1 + e^{-(w \cdot x + b)}}$$
$$P(y = 0) = 1 - \sigma(w \cdot x + b) = 1 - \frac{1}{1 + e^{-(w \cdot x + b)}} = \frac{e^{-(w \cdot x + b)}}{1 + e^{-(w \cdot x + b)}} \quad (5)$$

Temos, então, um algoritmo que, dado um tweet $x$, calcula a probabilidade $P(y = 1|x)$ de $x$ pertencer a determinada categoria. Como já dito anteriormente, o classificador com maior valor $P(y = 1|x)$ determina o sentimento do tweet.

No entanto, a regressão logística nos fornece o valor de $\hat{y}$, que é um estimador do $y$ observável:

$$\hat{y} = \sigma(w \cdot x + b) \quad (6)$$

Ainda precisamos de uma forma de determinar os valores dos parâmetros $w$ e $b$. Para isso, precisamos treinar o modelo para que a diferença entre o $\hat{y}$ estimado e o $y$ observado seja a menor possível para todas as observações. Cria-se, portanto, uma função de perda. Em especial, a função de perda mais utilizada em sistemas de rede neural é a função de entropia cruzada.

A função de entropia cruzada escolhe os parâmetros $w$ e $b$ que maximizem a log probabilidade do $y$ observado na amostra de treinamento dado as observações $x$. Isso é chamado de estimação por máxima verossimilhança condicional. Em outras palavras, gostaríamos de estimar pesos $w$ que maximizem a probabilidade da classe correta $P(y|x)$. Como há apenas duas classes (1 ou 0), temos uma distribuição de Bernoulli, e podemos expressar a probabilidade $P(y|x)$ que o classificador produz para um tweet como:

$$P(y|x) = \hat{y}^{y}(1-\hat{y})^{1-y} \tag{7}$$

Aplicando o log dos dois lados, temos que:

$$log[P(y|x)] = log[\hat{y}^{y}(1-\hat{y})^{1-y}] = ylog(\hat{y}) + (1-y)\log(1-\hat{y}) \tag{8}$$

Acima, se descreve uma log verossimilhança que deve ser maximizada. Para transformar a equação (8) em uma função de perda $L_{CE}$ (algo que precisa ser minimizado), basta mudarmos o sinal:

$$L_{CE}(\hat{y}, y) = -\,log[P(y|x)] = -[ylog(\hat{y}) + (1-y)\log(1-\hat{y})] \tag{9}$$

Agora, podemos substituir $\hat{y}$ pela igualdade exposta na equação (6):

$$L_{CE}(w, b) = -[ylog\big(\sigma(w \cdot x + b)\big) + (1-y)\log\big(1-\sigma(w \cdot x + b)\big)] \tag{10}$$

O log negativo dessa probabilidade é conveniente como métrica de perda porque ele varia de 0 ($-\log(1)$, quando não há perda e tudo é classificado

corretamente) a infinito (($-\log(0)$, perda infinita). Além disso, a função (10) garante simultaneamente que a probabilidade da classificação correta seja maximizada e a probabilidade da classificação errada seja minimizada, pois ambas sempre somam 1 e qualquer aumento na probabilidade de uma necessariamente diminui a probabilidade da outra.

Finalmente, precisamos ajustar conjuntamente o intercepto $b$ e as dezenas de milhares de pesos $w$ de forma a minimizar $L_{CE}(w,b)$. Para isso, se faz necessário entender os componentes direcionais de inclinação máxima para todas as $N$ dimensões de $L_{CE}(w,b)$. O Vetor Gradiente é um vetor que contém a inclinação de $b$ e de todos os pesos $w_i$ em $w$:

$$\nabla_{w,b} = \begin{bmatrix} \frac{\partial L_{CE}(w,b)}{\partial b} \\ \frac{\partial L_{CE}(w,b)}{\partial w_1} \\ \frac{\partial L_{CE}(w,b)}{\partial w_2} \\ \vdots \end{bmatrix} \tag{11}$$

A derivação da função de perda $L_{CE}(w,b)$ vai além do escopo deste trabalho, mas pode ser encontrada em detalhes no livro de Daniel Jurafsky, citado anteriormente. Temos, então, que:

$$\nabla_{w,b} = \begin{bmatrix} \frac{\partial L_{CE}(w,b)}{\partial b} \\ \frac{\partial L_{CE}(w,b)}{\partial w_1} \\ \frac{\partial L_{CE}(w,b)}{\partial w_2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \sigma(w \cdot x + b) - y \\ (\sigma(w \cdot x + b) - y)x_1 \\ (\sigma(w \cdot x + b) - y)x_2 \\ \vdots \end{bmatrix} \tag{12}$$

Sendo assim, para um vetor $\theta_t$ qualquer composto por números reais de $w$ e $b$, o vetor gradiente $\nabla_{w,b}$ nos informa em quais direções devemos mover cada um dos componentes de $\theta_t$ de forma a nos aproximarmos de um mínimo global. O ponto de partida $\theta_0$ é escolhido de forma aleatória e a magnitude da movimentação dos parâmetros de $\theta_t$ para $\theta_{t+1}$ é determinada por uma taxa de aprendizado $\eta$. Quanto

maior o $\eta$, maior o incremento nos parâmetros de $\theta_t$ de uma iteração para a outra. O Gradiente Descendente Estocástico, portanto, é um algoritmo que, através de movimentos iterativos, encontra os parâmetros $w$ e $b$ que resultem no menor valor para função de perda $L_{CE}(w, b)$. A equação que dita os movimentos realizados pelo algoritmo é, portanto:

$$\theta_{t+1} = \theta_t - \eta \nabla_{w,b} \tag{13}$$

O valor de $\eta$ é relativo e varia de caso a caso. Precisaremos, portanto, testar empiricamente vários valores de $\eta$ e encontrar o valor que produz a classificação mais precisa da amostra de teste.

Por fim, encontramos um vetor $\theta_{min}$ que otimiza os valores de $w$ e $b$ de forma a maximizar as chances de que o tweet seja classificado corretamente. Os valores de $\theta_{min}$ são plugados na equação (6), gerando uma probabilidade $P(y = 1|x)$ de $x$ pertencer a determinada categoria.

O resultado final é uma base de dados com a classificação aproximada do sentimento dos tweets na forma de série temporal. Foram contabilizados o número tweets positivos, negativos ou neutros para determinado período de tempo, para cada companhia considerada no estudo. A outra base de dados consiste em uma série temporal com os retornos das ações do índice escolhido.

### 4.2 Estudo de Eventos

Conforme demonstrado por Ranco et al. (2015), a relação entre o “sentimento" do twitter e o preço das ações é muito complexa. Por esse motivo, simples medidas de correlação não são capazes de capturar a dinâmica da interação entre os dois sistemas. Logo, é razoável supor que as duas séries temporais são dependentes apenas em certos momentos do tempo, e não durante todo o período do estudo. De fato, o preço de ativos financeiros depende de muitas outras variáveis, e não seria correto assumir que sempre haverá informações nas redes sociais em quantidade suficiente para que seja feita qualquer inferência.

Por esse motivo, essa relação será investigada a partir de um método econométrico chamado Estudo de Eventos. Esse método é originário da econometria financeira e consiste da análise dos retornos anormais (CAR - Cumulative Abnormal Returns) observados durante eventos específicos. Tal estudo envolve 3 etapas:

**1-Identificação de um conjunto de eventos anormais para cada ação:** picos de atividade no twitter são tratados como eventos. O procedimento para identificação de janelas temporais adotado é uma adaptação intradiária do método introduzido por Lehmann et al., 2012. Em primeiro lugar, os *tweets* sobre cada empresa são agrupados em períodos de 5 minutos, sendo $TW_t$ o volume de publicações dentro desse período. Verifica-se empiricamente que 5 minutos é o intervalo mínimo necessário para que uma quantidade relevante de novos tweets sobre as empresas sejam publicados. Isso evita que o sentimento predominante de $TW_t$ seja viesado em função de um baixo número amostral. Em seguida, para cada intervalo de 5 minutos, $t_0$ da série temporal, considera-se a mediana $\widetilde{TW_{[0,\text{-}12]}}$ dos volumes observados nos 12 intervalos anteriores, ou seja, nos intervalos da última hora. Então se define o coeficiente de discrepância $\phi(t_0)$ como:

$$\phi(t_0) = \frac{[TW_t - \widetilde{TW_j}]}{\max\left(\overline{\mathrm{TW}}, n_{min}\right)} \tag{14}$$

Em que $n_{min} = 1$ é um valor mínimo de atividade usado para padronizar a definição de $\phi(t_0)$ para janelas de baixa atividade e $\overline{\mathrm{TW}}$ de atividade nos 75 dias estudados.

Por fim, pode-se identificar um pico de atividade em $t_0$ quando $\phi(t_0) > \phi_r$, com $\phi_r = 1{,}5$. Onde $\phi_r$ é um valor de referência que controla o número de picos detectados na amostra e a quantidade de sobreposições entre esses picos. O objetivo é maximizar o número de eventos detectados e minimizar o número de sobreposições (MACKINLAY, 1997). Em seguida, removem-se picos que estejam separados por menos de 1 hora, dando preferência sempre ao primeiro pico identificado, em ordem cronológica. Esse procedimento garante a eliminação de qualquer sobreposição entre os eventos.

Além disso, picos que ocorreram simultaneamente a uma divulgação de resultados por parte da empresa são desconsiderados, visto que divulgações de resultados já causam retornos anormais por si só, como observado por Boehmer et al., 1991 e por Campbell et al., 1998.

**2-Classificação da polaridade dos eventos selecionados:** cada evento é classificado como positivo, negativo ou neutro. Essa categorização é derivada da média $\mu_t$ do sentimento dos tweets feitos no intervalo de 5 minutos $t_0$ de um evento. Os intervalos de classificação são descritos a seguir:

- Se $\mu_t \in [-1,\ -0{,}2)$, o evento é classificado como negativo
- Se $\mu_t \in [-0{,}2,\ 0{,}2]$, o evento é classificado como neutro
- Se $\mu_t \in (0{,}2,\ 1]$, o evento é classificado como positivo

Notadamente, os intervalos de classificação são distribuídos de forma simétrica. Isso é feito porque uma amostra significativamente grande de eventos tende a apresentar uma distribuição normal, com a maioria dos eventos tendo sentimento médio $\mu_t$ próximo de zero e poucos eventos cujo sentimento médio $\mu_t$ é próximo de -1 ou 1. A determinação dessas distribuições de polaridade, no entanto, é alvo de debates na literatura. Sprenger et al., 2014, fez uso de intervalos simétricos. Enquanto isso, Ranco et al., 2015, usou intervalos assimétricos, com viés negativo. No entanto, esse último verificou que, sob essa distribuição de polaridade, os picos de tweets neutros não foram considerados significantes, a um nível de confiança de 5%. O autor apontou isso como uma limitação do seu estudo, sugerindo que outros parâmetros de classificação fossem testados.

**3-Estimação dos retornos normais das ações nos períodos analisados:** para compreender o efeito do evento no retorno dos ativos, precisamos medir os retornos anormais. Para realizar esse cálculo, subtrai-se do retorno real para o período do evento um retorno esperado caso o evento não tivesse acontecido. Para o cálculo do retorno real, considera-se $P_{i,t}$ como a última transação ocorrida no minuto $t$ para a ação $i$:

$$R_{i,t} = \frac{P_{i,t} - P_{i,t-5}}{P_{i,t-5}} \tag{15}$$

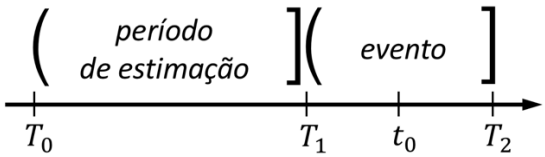
É possível, então, calcular o retorno anormal para o intervalo de 5 minutos $t$ como:

$$AR_{i,t} = R_{i,t} - E[R_{i,t}] \tag{16}$$

Onde $R_{i,t}$ é o retorno efetivo da ação $i$ no minuto $t$ e $E[R_{i,t}]$ é o retorno esperado. Existem várias *proxy's* para o retorno esperado. O retorno apropriado ao mercado/setor da companhia, ou o retorno linear médio da própria companhia são exemplos disso. Neste estudo, será considerado o retorno apropriado para o mercado, que sugere uma relação linear entre o retorno da empresa e o retorno do mercado. Tal metodologia também foi utilizada nos trabalhos de Campell et al., 1998 e Fama, 1970. O índice DJIA é utilizado como proxy para carteira de mercado, tal como em Ranco et al.. Essa escolha evita que o modelo tenha mais variáveis que o necessário. O DJIA é computado a partir da média ponderada dos preços de todas as ações que o compõe. Tem-se, portanto, que para uma ação $i$, no momento $t$, o modelo de índice é dado por:

$$\begin{gathered} R_{i,t} = \alpha_i + \beta_i R_{DJIA,t} + \epsilon_{i,t} \\ E(\epsilon_{i,t}) = 0, \quad var(\epsilon_{i,t}) = \sigma^2_{\epsilon_{i,t}} \\ E[R_{i,t}] = \hat{\alpha}_i + \hat{\beta}_i R_{DJIA,t} \end{gathered} \tag{17}$$

Em que $R_{DJIA,t}$ é o retorno do portfolio de mercado e $\epsilon_{i,t}$ é o erro e tem média zero. A fim de estimar os parâmetros do modelo, $\alpha_i$, $\beta_i$ e $\sigma^2_{\epsilon_{i,t}}$, para cada ação e seus eventos, foi utilizado um período de estimação anterior à ocorrência do evento $L = 180$ minutos, baseado em outros estudos que trabalham com dados intraday(ANDERSEN et al., 2007; CAMPBELL; WÓJTOWICZ et al., 2016). A figura abaixo explicita a linha temporal do estudo de eventos:




A partir dos parâmetros da figura acima, a variância estimada do erro é dada por:

$$\hat{\sigma}^2_{\epsilon_{i,t}} = \frac{1}{L-2} \sum_{t=T_0+1}^{T_1} (R_{i,t} - \hat{\alpha}_i - \hat{\beta}_i R_{DJIA,t})^2 \tag{18}$$

Em que $\hat{\alpha}_i$ e $\hat{\beta}_i$ são parâmetros estimados por MQO a partir da equação:

$$E[R_{i,t}] = \hat{\alpha}_i + \hat{\beta}_i R_{DJIA,t} \tag{19}$$

Por fim, o retorno anormal da ação $i$, no momento $t$ é o resíduo:

$$AR_{i,t} = R_{i,t} - \hat{\alpha}_i - \hat{\beta}_i R_{DJIA,t} \tag{20}$$

Os resultados obtidos pelo procedimento descrito acima precisam ser validados. A hipótese nula $H_0$ é que eventos externos não impactam o comportamento futuro dos retornos (média e variância), ou seja, os mercados são de fato eficientes e os preços reagem instantaneamente, não sendo possível observar retornos anormais. Sob $H_0$, a amostra de retornos anormais para um evento deve seguir uma distribuição normal:

$$AR_{i,\tau} \sim \mathcal{N}(0,\ \sigma^2_{\epsilon_{i,t}}) \tag{21}$$

Para que seja possível inferir alguma conclusão, se faz necessário a agregação dos retornos anormais em duas dimensões: ao longo do tempo e por ações. Agregando na dimensão das ações se obtém:

(22)

$$\overline{AR_\tau} = \frac{1}{N}\sum_{i=1}^{N} AR_{i,t}$$

Enquanto a dimensão temporal fornece o retorno anormal acumulado (CAR) para o período do evento ($T_1$ até $T_2$) pela somatória:

$$CAR(T_1, T_2) = \frac{1}{N}\sum_{t=T_1}^{T_2} \overline{AR_\tau} \tag{23}$$

Cuja variância pode ser calculada, assumindo $\sigma_{AR}^2 = \sigma_{\epsilon_{i,t}}^2$ [4], a partir da fórmula:

$$var(CAR(\tau_1, \tau_2)) = \frac{1}{N^2}\sum_{i=1}^{N} (\tau_2 - \tau_1 + 1)\, \sigma_{\epsilon_{i,t}}^2 \tag{24}$$

Em que $N$ é o número total de eventos, para todas as equações acima. Finalmente, é possível introduzir a estatística de teste $\hat{\theta}$, que possibilita testar se o retorno medido é, de fato, anormal:

$$\frac{CAR(\tau_1, \tau_2)}{\sqrt{var(CAR(\tau_1, \tau_2))}} = \hat{\theta} \sim \mathcal{N}(0, 1) \tag{25}$$

Esse procedimento será realizado para o grupo completo de todas as 12 ações.

[4] Demonstrado por Campbell, 1997 e por Fama 1970

# 5. RESULTADOS

## 5.1 Análise de sentimento dos Tweets

Todos os tweets colhidos sobre as 12 ações, além dos tweets da amostra de treinamento e da amostra de teste, passaram pelo processo de normalização descrito anteriormente. O Vetor Vocabulário foi criado a partir da amostra de treinamento. A figura 3 apresenta as palavras que aparecem com mais frequência no Vetor Vocabulário, por categoria. Como podemos observar, essas palavras não necessariamente são indicativas de qualquer sentimento. Tal conclusão reforça a ideia de que a frequência dos termos no Vetor Vocabulário, por si só, não é uma boa métrica para que se façam quaisquer inferências.

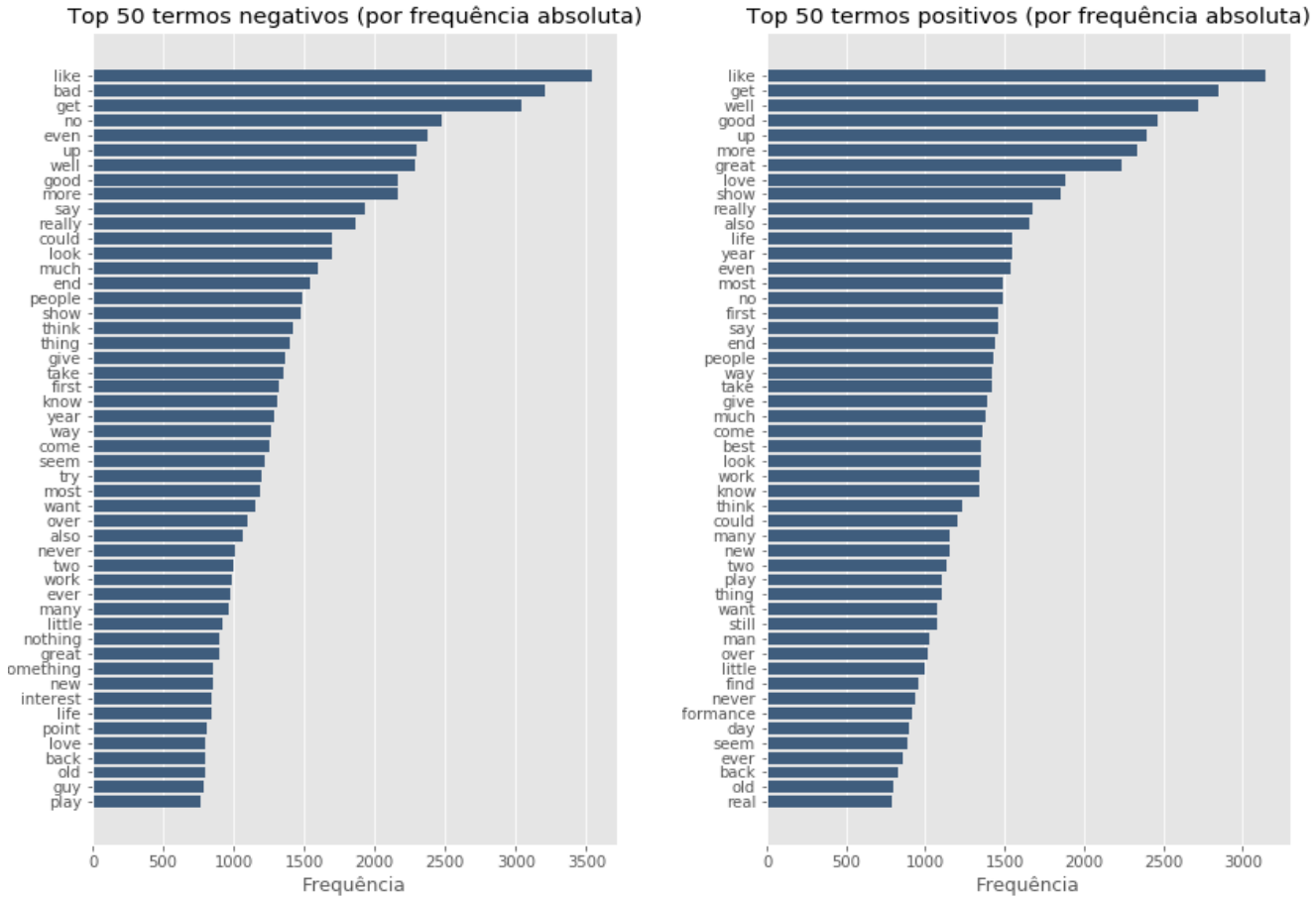



**Figura 3 –** Frequência Absoluta de termos no vetor vocabulário, por categoria
Fonte: Elaboração Própria

Aplicamos, então, a transformação $tfidf$ proposta. A figura 4 ilustra as palavras com maior valor médio de $tfidf$ no Vetor Vocabulário, por categoria. Dessa vez, os termos que aparecem com maior $tfidf$ já são, comparativamente, muito mais indicativos de sentimento do que eram na Figura 3. Fica evidente que, mesmo ainda não tendo determinado os pesos $w$ de cada termo, a transformação por $tfidf$ já atribui, proporcionalmente, mais valor para palavras que possuam relevância na classificação de sentimento. Isso ocorre porque, conforme foi explicado anteriormente, a transformação $tfidf$ atribui mais valor a palavras raras e incomuns e menos valor a palavras triviais da língua inglesa. Isso garante que palavras que sejam muito indicativas de um sentimento específico sejam mais favorecidas.

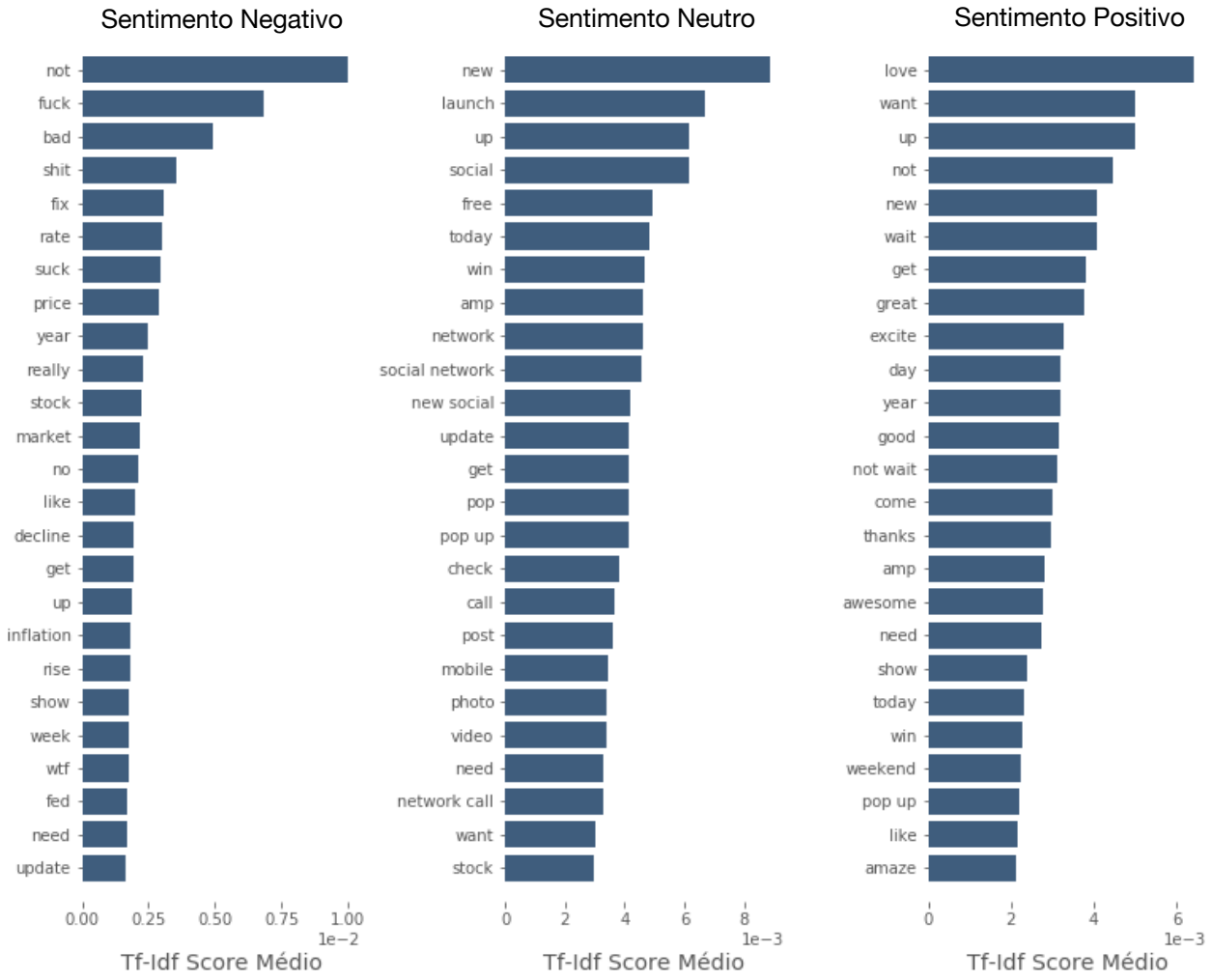



**Figura 4 –** Termos com maior valor $tfidf$ no vetor vocabulário, por categoria
Fonte: Elaboração Própria

O exposto acima já nos traz evidências suficientes de que a transformação $tfidf$ contribui para a criação de um algoritmo de classificação textual mais preciso. No entanto, ainda precisamos comprovar isso empiricamente. Conforme mencionado na seção 4, existem alguns parâmetros do algoritmo que precisam ser testados iterativamente até que se encontre uma combinação ótima que minimize a função de perda. São eles: (i) o uso do $tfidf$ vs. uso de frequência absoluta simples; (ii) o uso de unigramas, bigramas, ou unigramas e bigramas; (iii) o valor da taxa de aprendizado $\eta$.

As possíveis combinações entre esses parâmetros se multiplicam de forma fatorial. A natureza contínua de $\eta$ faz com que um intervalo finito de valores discretos precise ser determinado. Foram testados milhares de algoritmos de regressão logarítmica com gradiente descendente estocástico, cada um com uma combinação diferente desses 3 parâmetros. O módulo Sckit Learn, do Python, possui uma função chamada *GridSearchCV* que auxilia na otimização desses parâmetros. Em machine learning, esse processo é conhecido como *Parameter Hyper Tunning*. A figura 5 ilustra esse processo, bem como a acurácia de previsão obtida, para cada combinação de parâmetros testada.

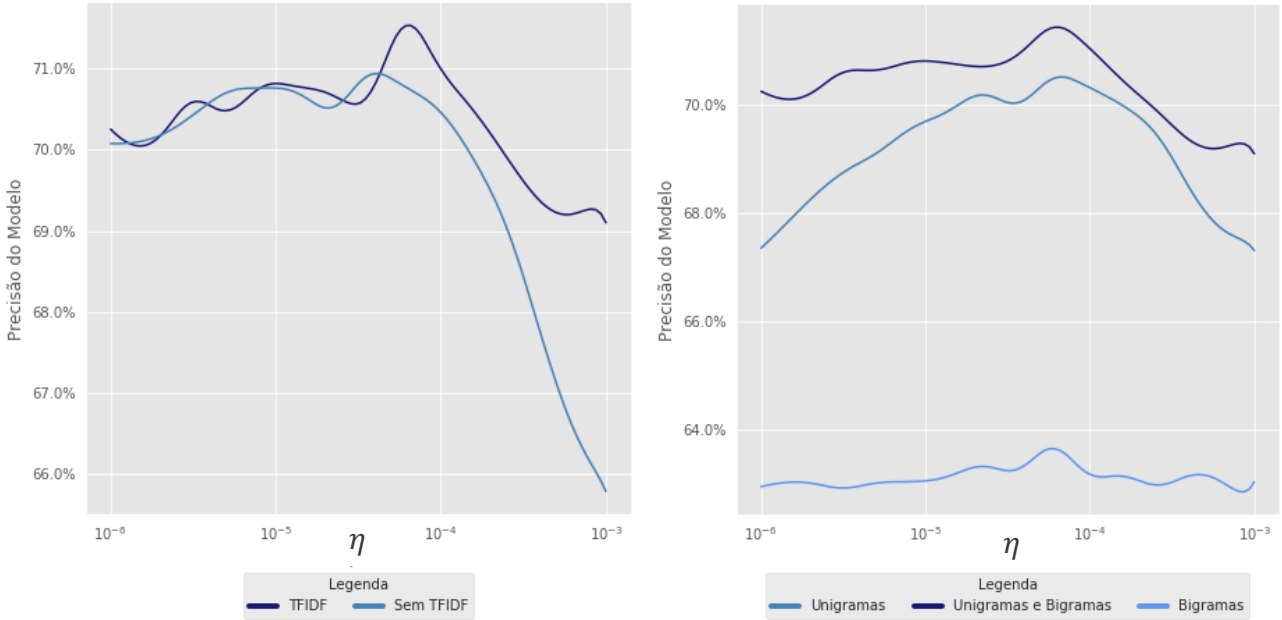



**Figura 5 –** Parameter Hyper Tunning

Fonte: Elaboração Própria

A acurácia de previsão do eixo $y$ é medida através da performance do algoritmo na classificação da amostra de teste, cujos tweets também foram classificados manualmente. Esse procedimento é feito comparando-se a classe prevista pelo algoritmo com a classe verdadeira. Através do exposto na figura 5, podemos concluir que, no ponto de máxima acurácia: (i) o uso do $tfidf$ produz classificadores mais precisos; (ii) a combinação de unigramas com bigramas é preferível ao uso de qualquer um dos dois isoladamente; (iii) o valor ótimo da taxa de aprendizado $\eta$ é de $8 \times 10^{-5}$.

Finalmente, o algoritmo de regressão logarítmica com gradiente descendente estocástico otimizado por *Parameter Hyper Tunning* é o algoritmo que será usado para classificar os tweets referentes às 18 ações estudadas. A performance desse algoritmo, medida através da amostra de teste, pode ser encontrada abaixo:

**Tabela 4 -** Métricas de Performance do Algoritmo Classificador

| | Precisão | Recall | Score F1 |
|---|---|---|---|
| Negativo | 0.73 | 0.76 | 0.74 |
| Neutro | 0.72 | 0.72 | 0.72 |
| Positivo | 0.72 | 0.70 | 0.71 |
| Acurácia | | | 0.72 |
| Média macro | 0.72 | 0.73 | 0.73 |
| Média Ponderada | 0.72 | 0.72 | 0.72 |

Fonte: Elaboração Própria

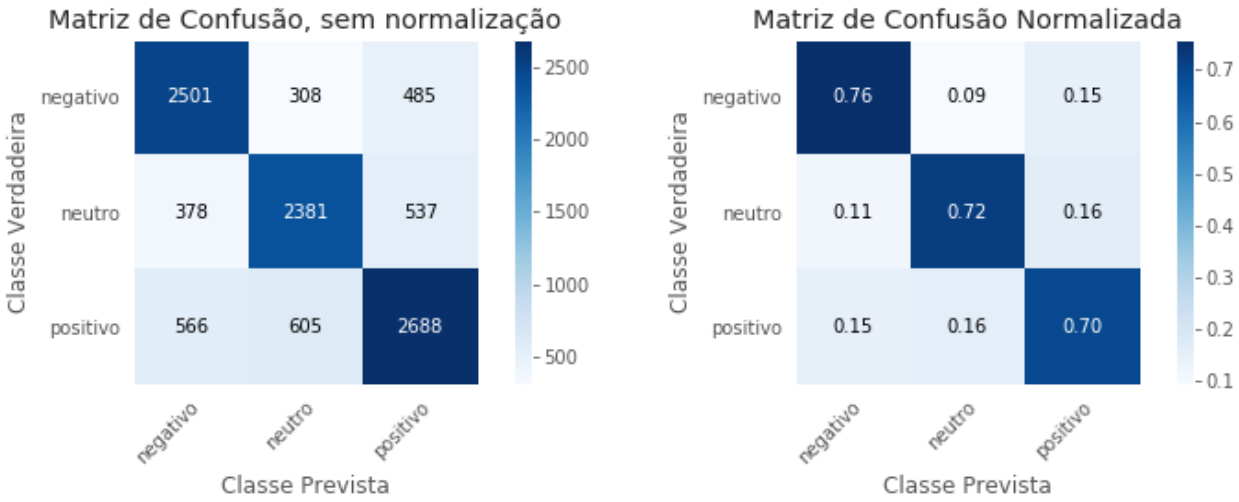


**Figura 6 -** Matriz de Confusão para tweets da amostra de teste classificados
Fonte: Elaboração Própria

No exposto acima, podemos ver que nosso algoritmo classificador foi capaz de gerar uma acurácia de classificação de 72,44%. Esse é um patamar bastante expressivo, principalmente quando se considera que o próprio limite no consenso entre avaliadores humanos é de cerca de 75%. Além disso, o algoritmo classificador se mostra consistente em todas as classes, evitando erros do tipo I e erros do tipo II, conforme exposto na diagonal principal da matriz de confusão, na figura 6. Notadamente, o classificador parece ter uma facilidade maior em classificar tweets negativos (76%), embora seja também consiga acertar corretamente a maioria dos tweets neutros e positivos (72% e 70%, respectivamente).

## 5.2 Estudo de Eventos

Após classificar os tweets de acordo com o sentimento que expressam, são determinados os picos de atividade que serão tratados como eventos. O volume $TW_t$ de tweets em um intervalo de 5 minutos é levado em consideração. Para cada uma das 12 ações, os picos são identificados levando em conta o coeficiente de discrepância $\phi(t_0)$, conforme descrito em (14). A Figura 7 mostra a detecção de picos para menções textuais ao Walmart ("Walmart", "WalmartInc", "Wal-Mart") no período estudado. O mesmo procedimento foi realizado para todas as outras 11 ações, para menções de seus "tickers" na bolsa, menções de seu perfil no Twitter e menções textuais ao nome da companhia, seguindo os filtros do Anexo I. Picos detectados durante o dia de divulgação de resultado da respectiva empresa não foram considerados.

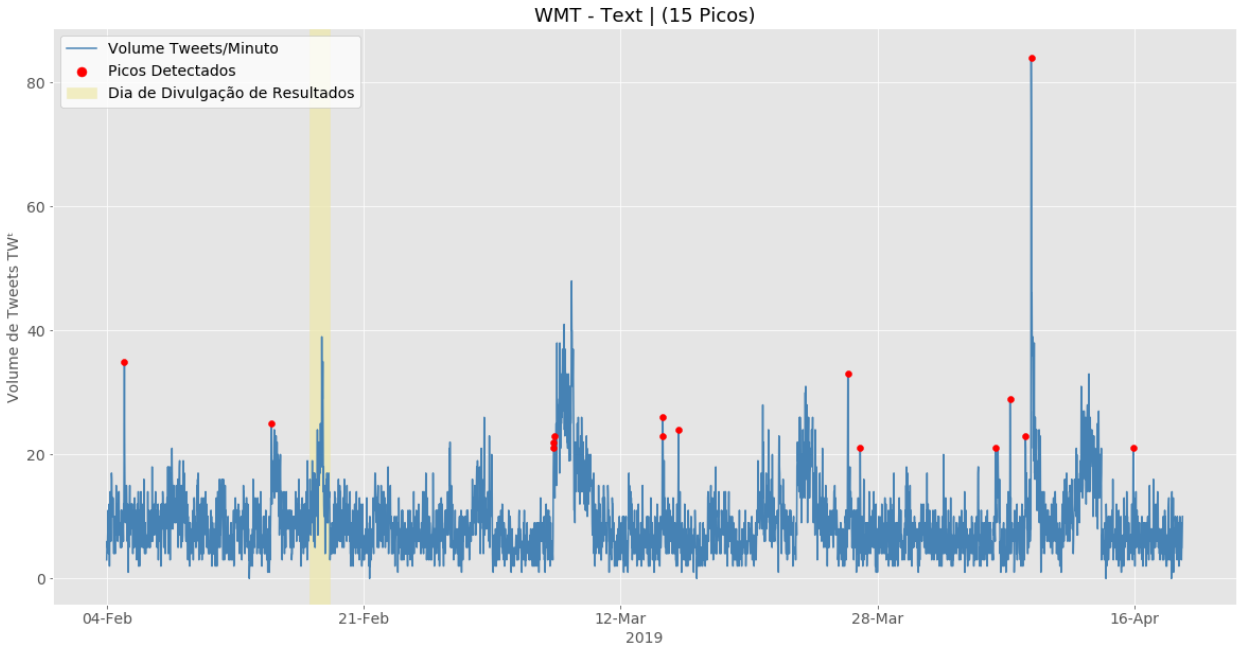



**Figura 7 –** Volume de Tweets para menções textuais ao Walmart, por intervalo de 5 minutos.
Fonte: Elaboração Própria

Os eventos detectados foram então agregados dentro das suas respectivas categorias. O histograma apresentado na Figura 8 corresponde ao total de eventos detectados e suas respectivas polaridades.

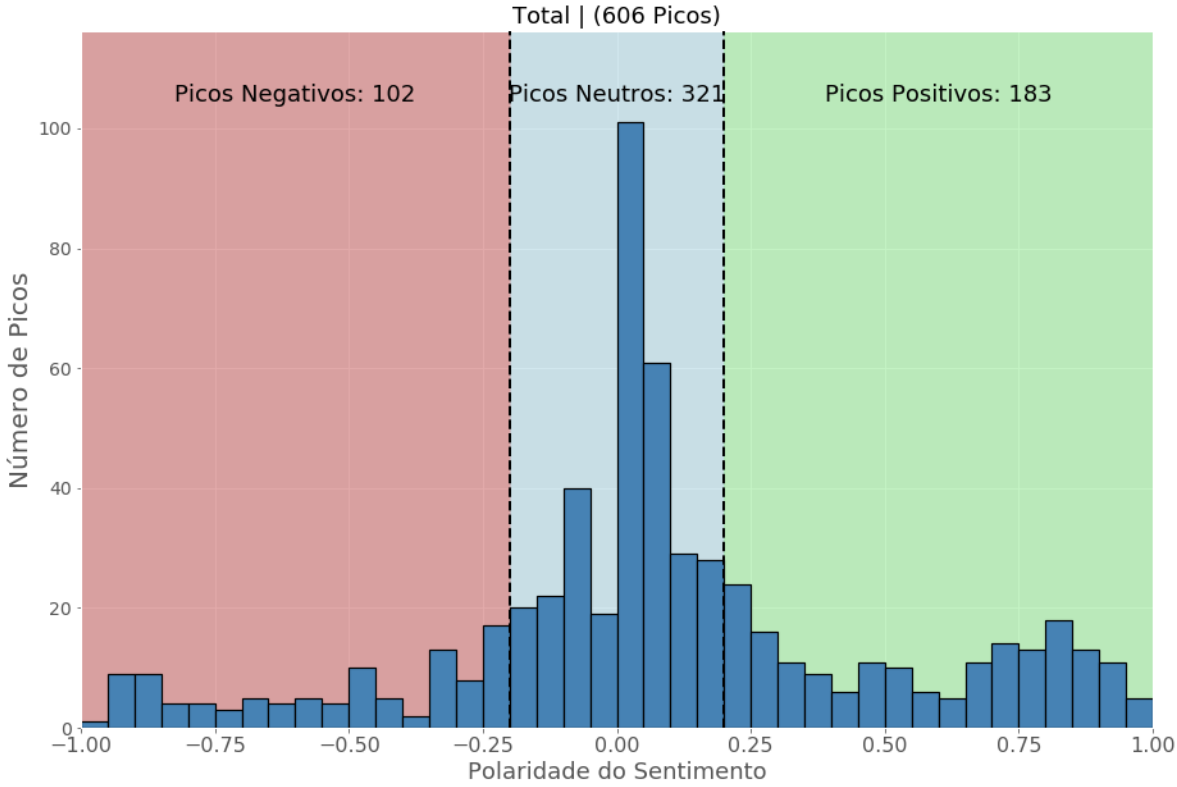



**Figura 8 –** Distribuição de polaridade de sentimento para os 606 picos detectados. As áreas coloridas indicam a categoria de cada pico de acordo com sua polaridade.
Fonte: Elaboração Própria

Por fim, para as empresas consideradas, a Figura 9 mostra o Retorno Anormal Acumulado (CAR) dos eventos detectados apenas para menções aos Tickers das empresas na bolsa (ex. $AAPL). Cada lag corresponde a um intervalo de 5 minutos. É, portanto, analisada uma janela que abrange 1 hora antes e 1 hora depois da ocorrência do evento (25 lags). O CAR médio dos eventos classificados como positivos aumenta em cerca de 1 ponto percentual imediatamente depois do pico de atividade no Twitter. Em contrapartida, diminui em até 2 pontos percentuais após a ocorrência de eventos considerados negativos. Eventos classificados como neutros apresentam CAR levemente positivo, embora significativamente inferior ao CAR de eventos efetivamente positivos. Essas variações são extremamente expressivas quando se considera o intervalo de tempo em questão. Pontos destacados em preto foram considerados significantes a um nível de significância de 1%. Sendo assim, os retornos anormais acumulados são significativos por até 1 hora depois da ocorrência

dos eventos. Os valores do gráfico, bem como as estatísticas de teste $\hat{\theta}$, estão detalhados na Tabela 6.

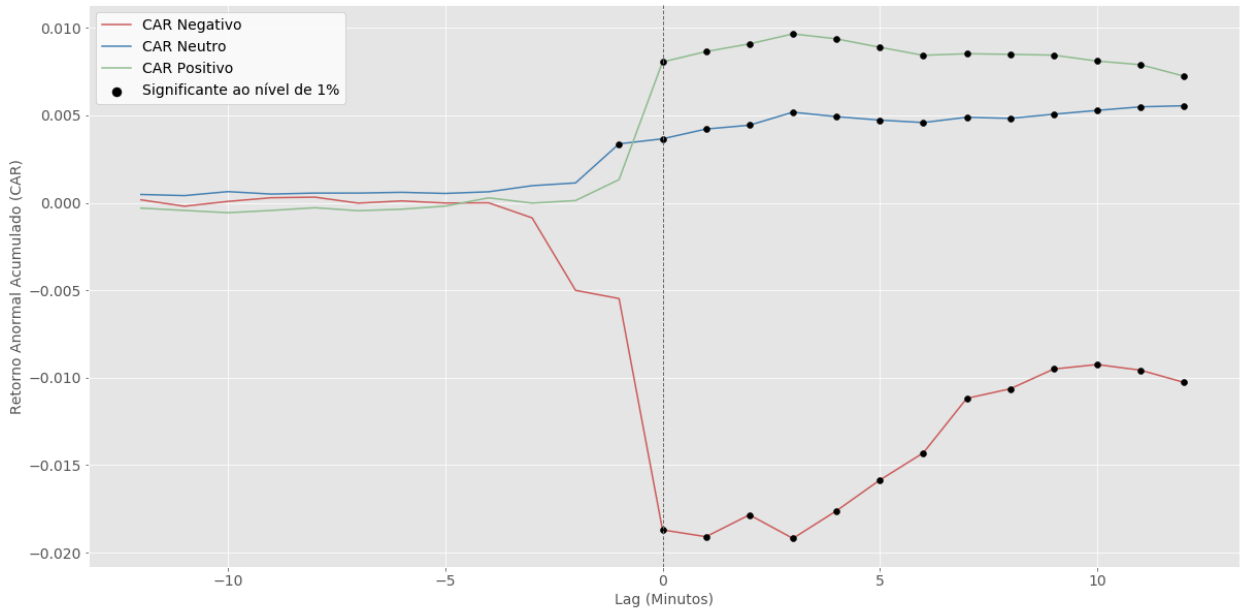



**Figura 9 –** Retorno Anormal Acumulado para os eventos detectados referentes a menções ao Ticker das empresas. Fonte: Elaboração Própria

**Tabela 5 –** Valores de CAR e $\hat{\theta}$ referentes a menções ao Ticker das empresas. CAR significantes à um nível de 1% ($|\widehat{\theta}| > 2{,}58$) são denotados por **. Resultados significantes à um nível de 5% ($|\widehat{\theta}| > 1{,}96$) são denotados por *.

| Lag | CAR | | | Theta | | |
|---|---|---|---|---|---|---|
| | Negativo | Neutro | Positivo | Negativo | Neutro | Positivo |
| -12 | 0.02% | 0.05% | -0.03% | 0.074248 | 0.561521 | -0.246094 |
| -11 | -0.02% | 0.04% | -0.04% | -0.085210 | 0.484885 | -0.353672 |
| -10 | 0.01% | 0.06% | -0.06% | 0.034925 | 0.751854 | -0.459611 |
| -9 | 0.03% | 0.05% | -0.04% | 0.126152 | 0.587598 | -0.354449 |
| -8 | 0.03% | 0.06% | -0.03% | 0.140768 | 0.655011 | -0.227648 |
| -7 | -0.0% | 0.06% | -0.04% | -0.007931 | 0.653988 | -0.365498 |
| -6 | 0.01% | 0.06% | -0.04% | 0.047562 | 0.705609 | -0.297363 |
| -5 | -0.0% | 0.05% | -0.02% | -0.005887 | 0.630340 | -0.153815 |
| -4 | 0.0% | 0.06% | 0.03% | 0.000050 | 0.741248 | 0.231530 |
| -3 | -0.09% | 0.1% | -0.0% | -0.373316 | 1.150797 | -0.013088 |
| -2 | -0.5%* | 0.11% | 0.01% | -2.170099 | 1.336638 | 0.106216 |
| -1 | -0.55%* | 0.34%** | 0.13% | -2.373597 | 3.959923 | 1.080732 |
| 0 | -1.87%** | 0.37%** | 0.81%** | -8.109183 | 4.298621 | 6.549887 |
| 1 | -1.91%** | 0.42%** | 0.86%** | -8.277937 | 4.954061 | 7.031599 |
| 2 | -1.78%** | 0.44%** | 0.91%** | -7.736150 | 5.212066 | 7.389738 |
| 3 | -1.92%** | 0.52%** | 0.97%** | -8.320145 | 6.086833 | 7.853869 |
| 4 | -1.76%** | 0.49%** | 0.94%** | -7.637247 | 5.786353 | 7.623883 |
| 5 | -1.59%** | 0.47%** | 0.89%** | -6.873361 | 5.555465 | 7.232397 |
| 6 | -1.43%** | 0.46%** | 0.84%** | -6.200149 | 5.391209 | 6.858304 |
| 7 | -1.12%** | 0.49%** | 0.85%** | -4.847863 | 5.745499 | 6.932384 |
| 8 | -1.06%** | 0.48%** | 0.85%** | -4.611586 | 5.670215 | 6.903764 |
| 9 | -0.95%** | 0.51%** | 0.84%** | -4.121108 | 5.951097 | 6.865816 |
| 10 | -0.92%** | 0.53%** | 0.81%** | -4.009509 | 6.207304 | 6.590152 |
| 11 | -0.96%** | 0.55%** | 0.79%** | -4.149653 | 6.446677 | 6.421533 |
| 12 | -1.03%** | 0.55%** | 0.72%** | -4.452994 | 6.518041 | 5.888250 |

Resultados similares podem ser observados para menções textuais às empresas (ex. Apple). A Figura 10 mostra o Retorno Anormal Acumulado (CAR) dos eventos detectados nessa categoria. Mais uma vez, o CAR de eventos positivos sobe logo após a ocorrência dos eventos e cai após eventos negativos. Embora a amplitude desses movimentos seja um pouco menor, aproximadamente 0,5 pontos percentuais, esse efeito ainda é bastante relevante levando em consideração o período analisado. Além disso, o efeito também é considerado significante a um nível de 1%, por até 1 hora depois para eventos positivos e 40 minutos para eventos negativos. O presente estudo é o primeiro do tipo a verificar que menções textuais às companhias no Twitter, não necessariamente tão direcionadas quanto menções ao Ticker de negociação, também são capazes de originar eventos com CAR significativo nos períodos subsequentes. A Tabela 7 detalha os valores de CAR e as estatísticas de teste $\hat{\theta}$.

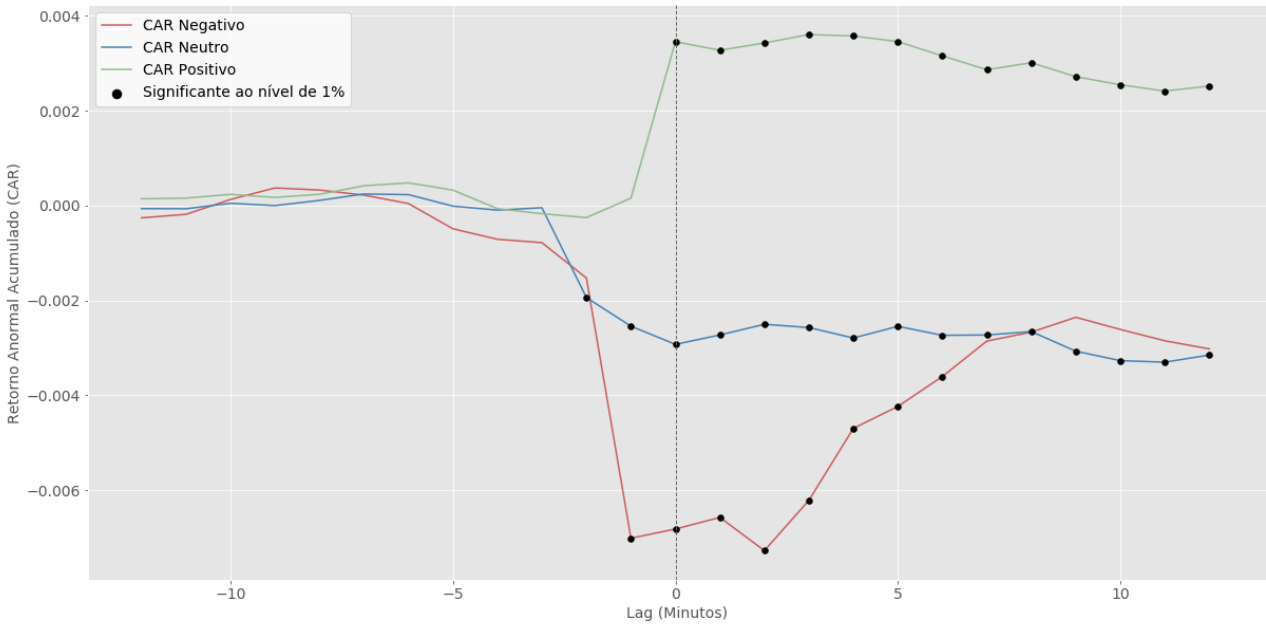



**Figura 10 –** Retorno Anormal Acumulado para os eventos detectados referentes a menções textuais às empresas. Fonte: Elaboração Própria

**Tabela 6 –** Valores de CAR e $\hat{\theta}$ referentes a menções textuais às empresas. CAR significantes à um nível de 1% ($|\widehat{\theta}| > 2{,}58$) são denotados por **. Resultados significantes à um nível de 5% ($|\widehat{\theta}| > 1{,}96$) são denotados por *.

| Lag | CAR | | | Theta | | |
|---|---|---|---|---|---|---|
| | Negativo | Neutro | Positivo | Negativo | Neutro | Positivo |
| -12 | -0.03% | -0.01% | 0.01% | -0.185655 | -0.090549 | 0.218901 |
| -11 | -0.02% | -0.01% | 0.02% | -0.132471 | -0.095151 | 0.242300 |
| -10 | 0.01% | 0.0% | 0.02% | 0.093300 | 0.070104 | 0.363148 |
| -9 | 0.04% | 0.0% | 0.02% | 0.268301 | 0.000779 | 0.264801 |
| -8 | 0.03% | 0.01% | 0.02% | 0.236603 | 0.157349 | 0.365573 |
| -7 | 0.02% | 0.02% | 0.04% | 0.159524 | 0.350123 | 0.640193 |
| -6 | 0.0% | 0.02% | 0.05% | 0.030978 | 0.331124 | 0.732561 |
| -5 | -0.05% | -0.0% | 0.03% | -0.352248 | -0.016113 | 0.499034 |
| -4 | -0.07% | -0.01% | -0.01% | -0.511251 | -0.135934 | -0.098438 |
| -3 | -0.08% | -0.0% | -0.02% | -0.562952 | -0.064393 | -0.261788 |
| -2 | -0.15% | -0.19%** | -0.03% | -1.099435 | -2.769273 | -0.386702 |
| -1 | -0.7%** | -0.25%** | 0.02% | -5.050329 | -3.630435 | 0.240266 |
| 0 | -0.68%** | -0.29%** | 0.35%** | -4.910516 | -4.178763 | 5.283243 |
| 1 | -0.66%** | -0.27%** | 0.33%** | -4.734598 | -3.895103 | 5.007925 |
| 2 | -0.73%** | -0.25%** | 0.34%** | -5.238550 | -3.573903 | 5.243369 |
| 3 | -0.62%** | -0.26%** | 0.36%** | -4.475056 | -3.671970 | 5.513354 |
| 4 | -0.47%** | -0.28%** | 0.36%** | -3.381821 | -3.986903 | 5.468600 |
| 5 | -0.42%** | -0.25%** | 0.35%** | -3.050001 | -3.635274 | 5.287093 |
| 6 | -0.36%** | -0.27%** | 0.32%** | -2.594994 | -3.905729 | 4.828003 |
| 7 | -0.29%* | -0.27%** | 0.29%** | -2.056830 | -3.896278 | 4.381566 |
| 8 | -0.27% | -0.27%** | 0.3%** | -1.921199 | -3.792808 | 4.606348 |
| 9 | -0.24% | -0.31%** | 0.27%** | -1.696175 | -4.384683 | 4.156169 |
| 10 | -0.26% | -0.33%** | 0.25%** | -1.879967 | -4.668345 | 3.896344 |
| 11 | -0.29%* | -0.33%** | 0.24%** | -2.052557 | -4.714579 | 3.692477 |
| 12 | -0.3%* | -0.32%** | 0.25%** | -2.175464 | -4.499588 | 3.854939 |

Fonte: Elaboração Própria

Os resultados expostos acima permitem rejeitar a hipótese nula $H_0$, de que eventos externos não possuem impacto futuro no comportamento dos retornos observados. Os preços não reagem instantaneamente após a ocorrência dos eventos selecionados, sendo possível observar retornos anormais relevantes até uma hora depois dos acontecimentos. Esse *underreaction* consiste em uma forma de ineficiência de mercado. Além disso, as evidências acima corroboram com outras já existentes de que eventos originidados nas redes sociais, especialmente o Twitter, causam impactos significativos no preço de ativos financeiros.

Nos dois casos anteriores, também é interessante notar a assimetria entre reações negativas e positivas: a reação aos sentimentos negativos é muito mais forte do que aos sentimentos positivos. Essa assimetria pode estar relacionada a forma como indivíduos lidam com a tomada de decisões, atribuindo um impacto emocional muito maior à perdas do que ganhos, fazendo com que investidores tomem decisões que desafiem a racionalidade.

Os resultados para menções ao perfil (ex. @apple) das empresas não foram considerados significantes, para nenhum tipo de evento, a um nível de significância de 5%. Como esse é o primeiro estudo do tipo a explorar os efeitos de outros tipos

de tweets além dos que mencionam os Tickers das ações, não existe uma justificativa clara para tal na literatura. É possível, no entanto, especular que esse fato ocorra em função da natureza do uso de menções ao perfil no twitter: a menção ao perfil aparece como prefixo a todo retweet compartilhado originalmente escrito pelas empresas. É possível, por exemplo, que muitos tweets promocionais e de marketing tenham sido capturados, atrapalhando a leitura do sentimento real do mercado.

## 6. CONCLUSÃO

No presente estudo, existem evidências significativas de dependência entre retorno de ativos financeiros e sentimento das redes sociais, em alguns momentos do tempo. Essas conclusões são consistentes com o que já foi observado na literatura. Os resultados obtidos aqui, no entanto, se diferem à medida que são os primeiros a mostrar que essa relação se faz presente também em intervalos de tempo significativamente menores. As evidências colhidas sugerem que os sentimentos do Twitter podem influenciar o mercado nos minutos subsequentes à sua ocorrência. Embora não haja aqui o estabelecimento de um efeito causal propriamente dito, a resposta gradual dos preços após a ocorrência dos eventos abre caminho para fundamentação de uma série de novas estratégias de investimento, baseadas no acompanhamento do sentimento das redes sociais como sinal de compra e venda. A maior ressalva a ser feita, no entanto, é que a dependência evidenciada aqui só se faz presente em momentos específicos, não sendo constante, muito menos regular, ao longo do tempo, tal como já era esperado. O resultado principal obtido, no entanto, é que o sentimento do Twitter durante eventos selecionados implica na direção da movimentação do mercado.

Os resultados obtidos também estão sujeitos a uma série de limitações. Dentre elas, como já dito anteriormente, são poucas as empresas que permitem esse tipo de análise. É necessário que haja uma grande quantidade de tweets referentes as empresas nas redes sociais. Além disso, as ações precisam ser negociadas em alto volume, para que não existam períodos em que os papéis não sejam negociados. Essas limitações tornam incrivelmente difícil a realização deste estudo para empresas brasileiras, por exemplo. Por fim, é inerente ao estudo de dados intra-diários com base nos preços de fechamento minuto à minuto um fenômeno chamado “non-syncronous trading”. Esse fenômeno ocorre porque os intervalos de tempo entre negociações não são iguais ao longo da série temporal, o que pode interferir nos resultados obtidos. Além disso, o preço de fechamento, simplificação crucial para o desenvolvimento do estudo, não necessariamente é representativo do *spread* ocorrido dentro do intervalo de tempo. Estudos futuros podem incorporar modelos estatísticos que levem em conta esses efeitos indesejados.

## REFERÊNCIAS BIBLIOGRÁFICAS

**Active fund managers trail the S&P 500 for the ninth year in a row in triumph for indexing**. Disponível em: <https://www.cnbc.com/2019/03/15/active-fund-managers-trail-the-sp-500-for-the-ninth-year-in-a-row-in-triumph-for-indexing.html>. Acesso em: 14 abr. 2019.

ANDERSEN, T. G. et al. Real-time price discovery in global stock, bond and foreign exchange markets ☆. **Journal of International Economics**, v. 73, p. 251–277, 2007.

ASIAEE T., A. et al. **If you are happy and you know it... tweet.**, *In:* **CIKM '12: Proceedings of the 21st ACM international conference on Information and Knowledge management,** 2012, nº 21, Maui, 2012, p. 1602-1606

ASUR, S.; HUBERMAN, B. A. **Predicting the future with social media**. *In:* **Proceedings: 2010 IEEE/WIC/ACM International Conference on Web Intelligence**, 2010, nº 3, Washington, 2010, p. 492-499

BOEHMER, E.; MASUMECI, J.; POULSEN, A. B. Event-study methodology under conditions of event-induced variance. **Journal of Financial Economics**, v. 30, p. 253-272, 1991.

BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, p. 1-8, 2011.

CAMPBELL, J. Y. et al. THE ECONOMETRICS OF FINANCIAL MARKETS. **Macroeconomic Dynamics**, v.2, p. 559-562, 1998.

CHA, M. et al. Measuring User Influence in Twitter : The Million Follower Fallacy. **Association for the Advancement of Artificial Intelligence**, 2010.

CHOI, H.; VARIAN, H. Predicting the present with Google Trends - Semantic Scholar. **Economic Record**, v. 88, p. 2-9, 2012.

DAMASIO, A. **Descartes' Error: Emotion, Reason, and the Human Brain**. New York: Avon Books, 1994.

DANIEL, J.; MARTIN, J. H. **Speech and Language Processing**. New Jersey: Prentice Hall, 2008.

WÓJTOWICZ, T; GURGUL, H. The impact of US macroeconomic news on the Polish stock market. **Central European Journal of Operations Research**, v. 22, p. 795-817, 2016.

FAMA, E. F. The Behavior of Stock-Market Prices. **The Journal of Business**, v. 38, n. 1, p. 34–105, 1965.

FAMA, E. F. Efficient Capital Markets: A Review of Theory and Empirical Work. **The Journal of Finance**, v. 25, p. 383-417, 1970.

FORBERGSKOG, J.-O.; BLOM, C. R. **Twitter and Stock Returns BI Norwegian Business School**. 2013, Tese (Mestrado em Admnistração), BI Norwegian Business School, Oslo, 2013.

GILAD, M.; NATALIE, G. **Predicting Movie Sales from Blogger Sentiment**. *In:* **AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs**, nº 14, 2006, Palo Alto, 2006, p. 155-158

GRUHL, D. et al. **The predictive power of online chatter**. *In:* **ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**, nº 11, 2005, Chicago, 2005, p. 78-87

**How Has Stock Ownership Trended in the Past Few Decades? | St. Louis Fed**. Disponível em: <https://www.stlouisfed.org/on-the-economy/2018/april/stock-ownership-trended-past-few-decades>. Acesso em: 14 abr. 2019.

JAMES, G. et al. **An Introduction to Statistical Learning**. New York: Springer, 2013.

KAHNEMAN, D.; TVERSKY, A. Prospect Theory: An Analysis of Decision under Risk. **Econometrica,** v. 47, p. 263-291, 1979.

KEYNES, J. M. **The general theory of employment, interest, and money**. United Kingdom: Macmillan Cambridge University Press, 1936.

LEHMANN, J. et al. **Dynamical Classes of Collective Attention in Twitter**. *In:* **WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web,** nº 21**,** 2012, Lyon, 2012, p. 251-270

MACKINLAY, A. C. Event studies in economics and finance. **Journal of economic literature**, v. 35, p. 13–39, 1997.

MAKREHCHI, M.; SHAH, S.; LIAO, W. **Stock Prediction Using Event-Based Sentiment Analysis**. *In:* **Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT),** nº 6, 2013, Atlanta, 2013, p. 337-342

MALKIEL, B. G. **A Random Walk Down Wall Street.** New York: W.W. Norton, 2016.

MOZETIČ, I.; GRČAR, M.; SMAILOVIĆ, J. Multilingual Twitter Sentiment Classification: The Role of Human Annotators. **PLoS ONE**, v. 11, p. 1-26, 2016.

NARR, S.; ULFENHAUS, M.; ALBAYRAK, S. Language-Independent Twitter Sentiment Analysis. *In:* **Knowledge Discovery and Machine Learning (KDML)**, nº 12, 2012, Dortmund, 2012, p. 12–14.

NISAR, T. M.; YEUNG, M. Twitter as a tool for forecasting stock market movements: A short-window event study. **The Journal of Finance and Data Science**, v. 4, p.101-119, 2018.

NOFSINGER, J. R. Social Mood and Financial Economics. **Journal of Behavioral Finance**, v. 6, p. 144-160, 2005.

PRANCKEVIČIUS, T.; MARCINKEVIČIUS, V. Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. **Baltic J. Modern Computing**, v. 5, n. 2, p. 221–232, 2017.

QIAN, B.; RASHEED, K. Stock market prediction with multiple classifiers. **Applied Intelligence**, v. 26, p. 25-33, 2007.

RANCO, G. et al. The effects of twitter sentiment on stock price returns. **PLoS ONE**, v. 10, p. 1-21, 2015.

ROSS, S. A.; WESTERFIELD, R. W.; BRADFORD D. JORDAN. **Fundamentals of Corporate Finance**., New York: McGraw-Hill Companies , 2013

SHILLER, R. J. The Irrational Exuberance. **Philosophy & Public Policy Quarterly**, New Jersey: Princeton University Press, 2005.

SMITH, A. N.; FISCHER, E.; YONGJIAN, C. How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter? **Journal of Interactive Marketing**, v.26, p. 102-113, 2012.

SPRENGER, T. O. et al. News or noise? Using twitter to identify and understand company-specific news flow. **Journal of Business Finance and Accounting**, v. 41, p. 791-830, 2014.

STAW, B. M. Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action. **ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE**. v. 16, p.27-44, 1976.

TAN, C. M.; WANG, Y. F.; LEE, C. DO. The use of bigrams to enhance text categorization. **Information Processing and Management**, v. 38, p. 529–546, jul. 2002.

ZHELUDEV, I.; SMITH, R.; ASTE, T. When can social media lead financial markets? **Scientific Reports**, v. 27, p.208-213, 2014.

# ANEXOS

**Anexo I.** – Conjunto de filtros utilizados para o levantamento de tweets através da ferramenta Twitter Historical Powertrack

| Ticker | Empresa | Contém $cashtag | Contém @perfil | Contém "Texto" | Não Contém "Texto" |
|---|---|---|---|---|---|
| **DIS** | The Walt Disney Company | $DIS | @WaltDisneyCo ou @Disney | Disney ou TheWaltDisneyCompany ou WaltDisneyCompany ou DisneyCompany ou WaltDisney ou WaltDisneyCo | n/a |
| **MSFT** | Microsoft | $MSFT | @Microsoft | Microsoft ou MicrosoftInc ou MicrosoftCorp ou MicrosoftCorporation | n/a |
| **IBM** | International Business Machines Corporation | $IBM | @IBM | IBM ou InternationalBusinessMachines Corporation ou InternationalBusinessMachines ou "International Business Machines Corporation" ou "International Business Machines" | n/a |
| **MMM** | 3M Company | $MMM | @3M | 3M Company | n/a |
| **AAPL** | Apple | $AAPL | @Apple | Apple | fruit ou fruits ou vegetable ou vegetables ou tree ou trees ou banana ou bananas ou orange ou oranges ou grape ou grapes ou strawberry ou strawberries ou watermelon ou watermelons ou lemon ou lemons ou pineapple ou pineapples ou blueberry ou blueberries ou cranberry ou cranberries ou blackberry ou blackberries ou berry ou berries ou cinnamon ou juice ou juices ou tart ou tarts ou cake ou cakes ou pie ou pies ou sauce ou cider ou food ou foods ou supermarket ou supermarkets ou orchard ou orchards ou farm ou farms ou harvest ou harvests ou eat ou eats |
| **XOM** | Exxon Mobil Corporation | $XOM | @ExxonMobil | Exxon ou (Exon Mobil) ou ExxonMobil ou ExxonMobilCorp ou ExxonMobilCorporation ou ExonMobilCorporation ou ExonMobilCorp ou ExonMobil | n/a |
| **GS** | The Goldman Sachs Group | $GS | @GoldmanSachs | (Goldman Sachs) ou GoldmanSachs ou GoldmanSachsGroup | n/a |
| **NKE** | NIKE | $NKE | @Nike | Nike | n/a |

| | | | | | |
|---|---|---|---|---|---|
| **MCD** | McDonald's Corporation | $MCD | @McDonaldsCorp ou @McDonalds | McDonalds ou McDonaldsCorporation ou McDonaldsCorp | n/a |
| **VZ** | Verizon Communications | $VZ | @Verizon | Verizon ou VerizonWireless ou VerizonInc | n/a |
| **WMT** | Walmart Inc. | $WMT | WalmartInc ou @Walmart | Walmart ou WalmartInc ou "Wal-Mart" | n/a |
| **JPM** | JPMorgan Chase | $JPM | @Chase | JPMorgan ou JPMorganChase ou MorganChase ou JPMorganChaseCo ou "JPMorgan Chase" ou "Morgan Chase" | n/a |
| **BA** | The Boeing Company | $BA | @Boeing | Boeing ou TheBoeingCompany ou BoeingCompany ou BoeingAirplanes | n/a |

**Anexo II.** – Descrição das bases de dados utilizadas na formação das amostras de treinamento e teste

| BASE | DESCRIÇÃO | TREINAMENTO | TESTE |
|---|---|---|---|
| **NIEK** | Base com tweets de tópicos variados, linguagem informal | 1800 | 900 |
| **DAI** | Base com tweets de tópicos variados, linguagem informal | 5100 | 1500 |
| **IMDB** | Base com reviews de filmes, linguagem médio-formal | 7000 | 3500 |
| **BRAND** | Tweets sobre percepção de marcas ou produtos | 3400 | 1500 |
| **ECON1** | Manchetes de jornal, formal, informativo | 2000 | 800 |
| **ECON2** | Frases de jornal, formal, informativo | 800 | 500 |
| **COACH** | Tweets sobre o festival coachella, linguagem informal | 2000 | 00 |
| **APPLE** | Tweets sobre a marca Apple, variados | 2500 | 800 |