



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

Seleção de Portfólios de Ações a partir da Caracterização e Análise de Dados de Redes Sociais

Ozório Júnio Soares Camargos

Orientador: Adriano César Machado Pereira
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Michele Amaral Brandão
Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

BELO HORIZONTE
AGOSTO DE 2021

Ozório Júnio Soares Camargos

Seleção de Portfólios de Ações a partir da Caracterização e Análise de Dados de Redes Sociais

Dissertação apresentado ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Métodos Matemáticos Aplicados

Orientador: Adriano César Machado Pereira
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientadora: Michele Amaral Brandão
Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL
BELO HORIZONTE
AGOSTO DE 2021

C172s Camargos, Ozório Júnio Soares
Seleção de portfólios de ações a partir da caracterização e análise de dados de redes sociais / Ozório Júnio Soares Camargos. – 2021.
59 f.

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.

Orientador: Adriano César Machado Pereira.

Coorientadora: Michele Amaral Brandão.

Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Redes sociais – Processamento de dados – Teses. 2. Mercado financeiro – Processamento de dados – Teses. 3. Ciência da computação – Matemática – Teses. 4. Investimentos – Teses. 5. Comércio eletrônico – Teses. 6. Stocktwits (Rede social) – Teses. I. Pereira, Adriano César Machado. II. Brandão, Michele Amaral. III. Centro Federal de Educação Tecnológica de Minas Gerais. IV. Título.

CDD 519.6



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO CURSO DE MESTRADO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

**“SELEÇÃO DE PORTFÓLIOS DE AÇÕES A PARTIR DA
CARACTERIZAÇÃO E ANÁLISE DE DADOS DE REDES SOCIAIS”**

Dissertação de Mestrado apresentada por **Ozório Júnio Soares Camargos**, em 27 de agosto de 2021, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Adriano César Machado Pereira (Orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof^ª. Dr^ª. Michele Amaral Brandão (Coorientadora)
Instituto Federal de Minas Gerais

Prof. Dr. Daniel Hasan Dalip
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Arthur Rodrigo Bosco de Magalhães
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Prof^ª. Dr^ª. Elizabeth Fialho Wanner
Presidenta do Colegiado do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

À Deus, a minha esposa Larissa, aos meus pais Ozório (*in memoriam*) e Joana, aos meus irmãos Nilton e Rosana, aos meus sobrinhos, ao meu Orientador Adriano César, a minha Coorientadora Michele Brandão, aos professores, aos amigos do CEFET-MG e do Grupo de Estudos em Finanças Computacionais (FICO-DCC-UFMG) e ao Instituto Federal de Minas Gerais (IFMG) dedico este trabalho.

Agradecimentos

Agradeço a Deus pelas oportunidades que me foram dadas na vida, por ter vivido momentos alegres, tristes, calmos e difíceis, momentos estes importantes para o meu aprendizado.

Agradeço aos meus pais Ozório M. Camargos (*in memoriam*) e Joana S. Camargos, que sempre me apoiaram em meus estudos e que me ajudaram a tornar o profissional e a pessoa que sou.

Aos meus irmãos, que desde pequeno me ensinaram diversas coisas.

Aos meus amigos e familiares, que de uma forma ou de outra colaboraram com a minha caminhada.

Ao meu Orientador Adriano César Machado Pereira e minha Coorientadora Michele Amaral Brandão, pelas boas reuniões que tivemos e por suas disponibilidade e dedicação.

Aos professores deste excelente programa de mestrado, que apoiaram e contribuíram para que este trabalho fosse possível.

Aos professores Cristina Murta e Edson Marchetti pelo incentivo e apoio.

A Universidade Federal de Minas Gerais - UFMG.

Ao Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - IFMG.

Ao Centro Federal de Educação Tecnológica - CEFET-MG.

E um agradecimento especial a minha esposa e grande colaboradora, Larissa Sousa, pelo apoio, carinho, amor e compreensão durante todo este tempo em que estamos juntos.

No mais, quero dizer: Obrigado a todos!

“No meio da dificuldade encontra-se a oportunidade.” (Albert Einstein)

Resumo

As redes sociais online fornecem uma variedade de informações úteis que podem ajudar a resolver diversos problemas. Diferentes fatos que ocorrem no mundo podem provocar oscilações no mercado financeiro, tais como: variação do dólar e indicadores, taxas de juros, crise sanitária, além de cenários políticos econômicos. Nesse contexto, esta dissertação apresenta estratégias para seleção de portfólios de ações a partir da caracterização e análise de dados de redes sociais. Especificamente, os objetivos deste trabalho são: realizar a construção de carteiras de ações, também chamadas de portfólios, através do uso de dados de redes sociais, analisar o desempenho da seleção com métricas de avaliação e comparar o desempenho dos portfólios criados com os índices de mercado das bolsas de valores norte-americanas. O trabalho foi desenvolvido em cinco etapas: coleta de dados, limpeza e preparação de dados, caracterização e análise de dados, seleção de portfólios de ações e, por fim, a análise de resultados. Os portfólios foram criados a partir dos dados coletados no período de 2015 a 2018. Os Resultados experimentais apresentaram evidências que nossa proposta de composição de portfólios utilizando dados de rede social mostrou-se útil como estratégia de investimento em ações. Os rendimentos dos portfólios indicados através da análise de dados da rede social foram satisfatórios em vários períodos analisados.

Palavras-chave: Ciência de Dados, Finanças Computacionais, Análise e Caracterização de Dados, Portfólios, Mercado Financeiro, Comércio eletrônico, Redes Sociais, Stocktwits.

Abstract

Online social networks provide a wealth of useful information that can help you solve a variety of problems. Different facts that occur in the world can cause fluctuations in the financial market, such as: dollar variation and indicators, interest rates, health crisis, in addition to economic political scenarios. In this context, this dissertation presents strategies for selecting stock portfolios from the characterization and analysis of data from social networks. Specifically, the objectives of this work are: to carry out the construction of stock portfolios, through the use of data of social networks, analyze the selection performance with evaluation metrics and compare the performance of portfolios created with the market indexes of the North American stock exchanges. The work was carried out in five stages: data collection, data cleaning and preparation, data characterization and analysis, selection of stock portfolios and, finally, the analysis of results. The portfolios were created from data collected from 2015 to 2018. The experimental results presented evidence that our proposal for portfolio composition using social network data proved to be useful as a stock investment strategy. The yields of the portfolios indicated through the analysis of data from the social network were satisfactory in several periods analyzed.

Keywords: Data Science, Financial Computing, Data Characterization and Analysis, Portfolios, Financial Markets, e-Commerce, Social Networks, Stocktwits.

Lista de Figuras

Figura 1 – Painel gráfico do Índice S&P500	9
Figura 2 – Processo Operacional de Mineração de Dados em Redes Sociais	12
Figura 3 – Representação da Fronteira Eficiente	15
Figura 4 – Etapas a serem seguidas na metodologia do trabalho.	24
Figura 5 – Valores de preço de fechamento em (US\$) dos ativos AMD, DCTH, DRYS, JNUG, SPY entre 2017 e 2018.	34
Figura 6 – Valores de log-retorno em percentual (%) dos ativos AMD, DCTH, DRYS, JNUG, SPY entre 2017 e 2018.	35
Figura 7 – Distribuição de ativos com sentimentos positivos e negativos	36
Figura 8 – Distribuição de ativos por percentual de tweets - 2015 a 2016	37
Figura 9 – Distribuição de ativos por percentual de tweets - 2017 a 2018	38
Figura 10 – Distribuição do índice S-Score dentre os ativos mais citados	39
Figura 11 – Distribuição de tweets publicados	39
Figura 12 – Tweets sumarizados por hora do dia	40
Figura 13 – Análise da correlação do coeficiente de Spearman.	41
Figura 14 – Análise da correlação com o coeficiente de Spearman.	42
Figura 15 – Correlação entre os ativos mais citados e positivos da rede social.	43
Figura 16 – Retorno Acumulado - E1 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2018/1S	44
Figura 17 – <i>Max. Daily Drawdown</i> – E1 – Portfólio, SP500, Dow Jones, Nasdaq e CD – 2018/1S	45
Figura 18 – Retorno Acumulado - E1 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2015/2S	46
Figura 19 – <i>Max. Daily Drawdown</i> – E1 – Portfólio, SP500, Dow Jones, Nasdaq e CDD – 2015/2S	46
Figura 20 – Retorno Acumulado - E2 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2016/2S	47
Figura 21 – <i>Daily Drawdown</i> – E2 – Portfólio, SP500 e CD – 2016/2S	48
Figura 22 – Retorno Acumulado - E2 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2016/1S	48
Figura 23 – <i>Máx. Daily Drawdown</i> – E2 – Portfólio, SP500 e CD – 2016/1S	49
Figura 24 – Retorno Acumulado - E3 - Portfólio, SP500 e CD - 2017/1S	49
Figura 25 – <i>Daily Drawdown</i> – E3 – Portfólio, SP500 e CD – 2017/1S	51
Figura 26 – Retorno Acumulado - E3 - Portfólio, SP500 e CD - 2018/2S	51
Figura 27 – <i>Daily Drawdown</i> – E3 – Portfólio, SP500 e CD – 2018/2S	52
Figura 28 – Retorno percentual entre 2015 e 2018	52

Lista de Tabelas

Tabela 1 – Exemplos de Indicadores Financeiros	10
Tabela 2 – Escala de Cohen.	27
Tabela 3 – Dados coletados da Stocktwits Inc.	32
Tabela 4 – Dados coletados da Stocktwits Inc.	33
Tabela 5 – Dados coletados das Bolsas de Valores Americanas.	34
Tabela 6 – Dados preparados - Stocktwits Inc.	34
Tabela 7 – Informações dos dados coletados das mensagens publicadas na <i>Stocktwits</i> .	35
Tabela 8 – Composição dos Portfólios - E1	43
Tabela 9 – Avaliação de Desempenho – E1 – Portfólio x SP 500 x Dow Jones x Nasdaq	44
Tabela 10 – Avaliação de Desempenho – E1 – Portfólio x SP 500 x Dow Jones x Nasdaq	45
Tabela 11 – Composição dos Portfólios - E2	46
Tabela 12 – Avaliação de Desempenho – E2 – Portfólio x SP 500 x Dow Jones x Nasdaq	47
Tabela 13 – Avaliação de Desempenho – E2 – Portfólio x SP 500 x Dow Jones x Nasdaq	47
Tabela 14 – Composição dos Portfólios - E3	49
Tabela 15 – Avaliação de Desempenho – E3 – Portfólio x SP 500 x Dow Jones x Nasdaq	50
Tabela 16 – Avaliação de Desempenho – E3 – Portfólio x SP 500 x Dow Jones x Nasdaq	50

Lista de Abreviaturas e Siglas

AAOI	<i>Applied Optoelectronics Inc.</i>
AAPL	<i>Apple Inc.</i>
AKER	<i>Valeant Pharmaceuticals Intl</i>
AMD	<i>Advanced Micro Devices Inc.</i>
AMZN	<i>Amazon.com Inc.</i>
BABA	<i>Alibaba Group</i>
BBRY	<i>BlackBerry Limited</i>
B3	Brasil Bolsa Balcão
CD	<i>Certificate Of Deposit</i>
CDB	Certificado de Depósito Bancário
CDI	Certificado de Depósito Interbancário
CNPC	Conselho Nacional de Previdência Complementar
CNSP	Conselho Nacional de Seguros Privados
CMN	Conselho Monetário Nacional
CYTR	<i>CytRx Corporation</i>
DCTH	<i>Delcath Systems Inc.</i>
DJIA	<i>Dow Jones Index</i>
DPW	<i>DPW Holdings Inc.</i>
DRYS	<i>DryShips Inc.</i>
EUA	Estados Unidos da América
FB	<i>Facebook Inc.</i>
FICO	Grupo de Estudos em Finanças Computacionais - DCC/UFMG
GEVO	<i>Gevo, Inc.</i>
GLBR	<i>Global Brokerage, Inc.</i>

GPRO	<i>GoPro, Inc.</i>
ÊSPC	<i>S&P500 Index</i>
HME	Hipótese do Mercado Eficiente
HMNY	<i>Helios and Matheson Analytics</i>
IDXG	<i>Interpace Biosciences Inc.</i>
IPCA	Índice de Preços ao Consumidor Amplo
ÊXIC	<i>Nasdaq Index</i>
JNUG	<i>Direxion SHS ET/Daily JR Gold Miner</i>
KNDI	<i>Kandi Technologies Group, Inc.</i>
LC	Letra de Câmbio
LCA	Letra de Crédito do Agronegócio
LCI	Letra de Crédito Imobiliário
MNKD	<i>MannKind Corporation</i>
MPT	<i>Modern Portfolio Theory</i>
MU	<i>Micron Technology</i>
NAK	<i>Northern Dynasty Minerals Ltd.</i>
NASDAQ	<i>National Association of Securities Dealers Automated Quotations</i>
NETE	<i>Net Element Inc.</i>
NFLX	<i>Netflix Inc.</i>
NVCN	<i>Neovasc Inc.</i>
NVDA	<i>Nvidia Corporation</i>
NYSE	<i>New York Stock Exchange</i>
OBV	<i>On Balance Volume</i>
PLUG	<i>Plug Power Inc</i>
ROI	<i>Return on investment</i>
ROR	<i>Rate of return</i>

RSI	<i>Relative Strength Index</i>
SNAP	<i>Snap Inc.</i>
SMA	<i>Simple Moving Average</i>
SMAV	<i>Simple Moving Average of Volume</i>
SMC	<i>Simulação de Monte Carlo</i>
SPY	<i>SPDR S&P 500 ETF Trust</i>
SPX	<i>S&P 500 Index</i>
TOPS	<i>TOP Ships</i>
TRXC	<i>Transenterix Inc.</i>
TSLA	<i>Tesla Inc.</i>
TVIX	<i>CBOE Volatility Index</i>
TWTR	<i>Twitter, Inc.</i>
UGAZ	<i>VelocityShares 3x Long Natural Gas ETN Linked to the S&P GSCI Natural Gas Index ER</i>
VRX	<i>Valeant Pharmaceuticals Intl</i>
XXII	<i>22nd Century Group Inc.</i>

Sumário

1 – Introdução	1
1.1 Motivação	3
1.2 Objetivos	3
1.3 Contribuições	4
1.4 Organização do trabalho	4
2 – Fundamentação Teórica	6
2.1 Introdução ao Mercado Financeiro	6
2.2 Mineração de dados em redes sociais	11
2.2.1 Redes Sociais	11
2.2.2 Análise de Redes Sociais	12
2.3 Seleção de Portfólios	13
2.3.1 Gestão de Portfólios	14
2.3.2 Modelos de Seleção de Portfólios	15
2.3.2.1 Modelo de Markowitz	15
2.3.2.2 Modelo de Sharpe	17
2.3.3 Simulação de Monte Carlo	18
3 – Trabalhos Relacionados	20
4 – Metodologia	24
4.1 Coleta de Dados	24
4.2 Limpeza e Preparação dos Dados	25
4.3 Caracterização e Análise de Dados	25
4.3.1 Análise Estática	26
4.3.2 Análise Temporal	26
4.3.3 Análise de Correlação	26
4.4 Seleção de Portfólios	28
4.5 Análise de Resultados	29
4.6 Considerações Importantes	31
5 – Resultados	33
5.1 Coleta de Dados	33
5.2 Limpeza e Preparação de Dados	34
5.3 Caracterização e Análise de Dados	35
5.3.1 Análise Estática	35

5.3.2	Análise Temporal	38
5.3.3	Análise de Correlação	39
5.4	Análise da Seleção de Portfólios	42
5.4.1	Análise da Estratégia 1	42
5.4.2	Análise da Estratégia 2	45
5.4.3	Análise da Estratégia 3	48
5.5	Considerações Importantes	51
6	– Conclusão	54
6.1	Conclusões	54
6.2	Trabalhos futuros	55
	Referências	56

1 Introdução

As redes sociais online se transformaram em ambientes virtuais de disseminação de novas informações em larga escala, através da popularização da Internet e de dispositivos móveis, como *smartphones* e *tablets*, gerou-se uma adesão e aumento na utilização de novas tecnologias computacionais para comunicação e disseminação da informação.

Diversos fatos que ocorrem no mundo podem provocar oscilações no mercado financeiro como: variação do dólar, indicadores, taxas de juros, crise sanitária além de cenários políticos e econômicos de países e grupos econômicos.

Vários autores com o objetivo de criar subsídios para auxiliar na tomada de decisão ao montar uma carteira de ações, realizaram estudos de correlações de indicadores em séries temporais de preço e volume, através de dados de cotações históricas dos ativos negociados na bolsa de valores, a fim de identificar possíveis padrões de comportamento e estratégias de negociação conforme citado por (ALDRIDGE, 2013) e (FERREIRA, 2010).

Outro tipo de estudo de séries temporais é a análise de dados de volumes que são coletados nas redes sociais e na Web e se referem aos comentários relacionados dos ativos negociados na bolsa de valores. Estes dados podem representar pensamentos e ideias que abrangem aspectos fundamentalistas de notícias de várias partes do mundo.

Por se tratar de uma base de dados relevante, plataformas de previsão em tempo real estão surgindo, como por exemplo, o *Stocktwits*. A rede social *Stocktwits*¹, uma plataforma de comunicações financeiras para a comunidade financeira e de investimento, consiste em postagens que incluem ideias, links, gráficos e outros dados financeiros importantes, resumidos dentro de 140 caracteres.

Estas postagens referentes à rede social *Stocktwits* podem ser chamadas por *tweets* ou *twits*. Dentre os usuários podemos citar analistas de mercado financeiro, profissionais da mídia e investidores de todos os tipos, bem como empresas públicas. Investidores e outros interessados em ações e mercados, podem seguir facilmente ações individuais, contribuidores específicos, além de visualizar a postagem do ativo em vários sites financeiros que integram as postagens, incluindo *Yahoo! Finanças*, *CNNMoney*, *Reuters*, *TheStreet.com*, *Bing.com* e *The Globe and Mail*.

Esta plataforma também é integrada com as principais redes sociais, incluindo *Twitter*, *Facebook* e *LinkedIn*, que contém mensagens de conversas, coletadas em tempo real, sobre as condições do mercado e valores de ativos específicos.

Ao contrário de outras redes sociais, o *Stocktwits* é focado somente no público que

¹ Stocktwits, Inc.: <https://stocktwits.com/>

deseja discutir sobre mercado financeiro, dando possibilidade aos usuários amadores interagirem livremente com profissionais. Além disso, o mercado de ações ainda é pouco explorado pela área de computação, tendo muita pesquisa ainda a ser feita. Atualmente, é possível encontrar pesquisas que utilizam análise de sentimento para tentar prever movimentações no mercado de ações (OH; SHENG, 2011). Alguns pesquisadores também se envolvem em esforços para encontrar os usuários mais experientes e que podem oferecer *tweets* mais precisos e úteis para serem utilizados em modelagem de aprendizado de máquina para prever o mercado de ações (BAR-HAIM et al., 2011a).

Segundo (LOPES; BARTH; MARACCINI, 2008), os agentes investidores, antes de tomarem uma decisão para a negociação de compra e venda de ações, utilizaram recursos de diversos meios de comunicação e informação, com destaque para a Web, que nos tempos atuais tornou-se um importante meio de produção de dados e divulgação de informações para o mercado financeiro.

A teoria moderna do portfólio, ou simplesmente teoria do portfólio, explica como investidores racionais usarão o princípio da diversificação para otimizar as suas carteiras de investimentos, e como um ativo arriscado deve ser precificado. O desenvolvimento de modelos de otimização de portfólio tem origem na área econômico-financeira. O trabalho pioneiro na área de otimização de portfólio foi o modelo média-variância por (MARKOWITZ, 1952).

Esta teoria estabelece que decisões relacionadas à seleção de investimentos devam ser tomadas com base na relação risco-retorno. Para auxiliar neste processo, modelos de otimização de portfólio têm sido desenvolvidos. De modo a serem efetivos, tais modelos devem ser capazes de quantificar os níveis de risco e retorno dos investimentos.

A teoria do portfólio considera a rentabilidade do ativo como uma variável aleatória, e uma carteira como uma combinação ponderada de ativos, de modo que o retorno de uma carteira é a combinação ponderada dos retornos dos ativos. Além disso, o retorno da carteira é uma variável aleatória e, conseqüentemente, tem um valor esperado e uma variância. O risco, neste modelo, é o desvio-padrão do retorno (MARKOWITZ, 1952).

Este trabalho propõe realizar a caracterização e análise dos dados de séries temporais de publicações da rede social de finanças *Stocktwits*. Para isso, são utilizados recursos de análise estática, análise temporal, análise de correlação, mineração de dados e avaliação de desempenho do portfólio criado frente a outros índices de mercado. A finalidade é gerar *insights* e proporcionar melhor entendimento sobre o comportamento de seus usuários e características dos ativos, para auxiliar investidores na tomada de decisão sobre a seleção de portfólios do mercado financeiro.

1.1 Motivação

As grandes organizações, principalmente nos países desenvolvidos, apresentam dentre sua carteira de investimentos ativos de renda variável. O estudo desta área pode trazer resultados que gerem novas informações acerca do comportamento do mercado financeiro e, conseqüentemente, potencializar eventuais ganhos. Estes ganhos geram novos empregos e expansão do setor privado, o que aumenta os recursos arrecadados pelo setor público. Para este trabalho abordaremos o problema de seleção de carteiras de investimento, que consiste na seleção de um conjunto de ativos, portfólio, que dê ao investidor um retorno esperado, minimizando o risco.

Segundo (MARKOWITZ, 1952), o processo de seleção de uma carteira de ações pode ser dividido em dois estágios. O primeiro começa com observação e experiência e termina com opiniões sobre a performance futura dos negócios avaliados. O segundo estágio começa com as opiniões relevantes sobre o futuro e termina com a escolha de uma carteira de ações.

O investidor que esteja no mercado de ações, opções ou obrigações deve, ou pelo menos deveria, buscar o máximo retorno possível de seu investimento, dentro de níveis de risco aceitáveis. O risco e retorno, em geral, possuem uma correlação positiva entre si, quando um aumenta, o outro aumenta junto, quanto maior for o retorno, maior será o risco e vice-versa. Porém essa relação entre risco e retorno pode ser trabalhada de forma que se consiga, através da diversificação, reduzir o risco de uma carteira a níveis algumas vezes menores que o risco do investimento mais seguro que participa da carteira, dependendo da correlação entre os componentes da carteira (BODIE; MERTON, 1999).

O risco que uma pessoa está disposta a assumir pode variar muito dependendo de fatores como idade, nível social, estado civil, número de filhos, perspectivas futuras e vários outros fatores. Desse modo, nunca existirá uma carteira de ações que agrade a todos, já que quanto maior o retorno desejado maior será o risco a se assumir.

A composição de uma carteira pode variar muito devido ao fato de existirem diferentes tipos de investidores que desejam assumir diferentes riscos e retornos. Harry Markowitz (1952) e William Forsyth Sharpe (1964) contribuíram imensamente com o processo de seleção de carteiras de investimentos. Ambos desenvolveram metodologias de avaliação e compensação do risco através da diversificação de investimentos. As teorias desenvolvidas por eles são amplamente utilizadas nos dias de hoje, ambos demonstraram matematicamente que não se deve carregar todos os ovos na mesma cesta (JUNIOR; PAMPLONA; MONTECECHI, 2002).

1.2 Objetivos

A busca por estabelecer a relação ideal entre risco e retorno é constante entre os investidores de ações. No entanto, não existe uma resposta simples para esse dilema. Esse

fato se dá, principalmente, porque diminuir o risco e aumentar o retorno de um investimento implica em a capacidade de prever eventos futuros que acarretarão em resultados desejados, o que é uma tarefa bastante trabalhosa.

Nesse contexto, esta dissertação tem como objetivo geral: (i) realizar a construção de carteiras de investimento por meio do uso de dados de redes sociais, a partir da caracterização e análise destes dados. Alguns dos objetivos específicos e principais diferenciais deste trabalho em relação aos estudados são: (i) caracterizar os dados da *Stocktwits* utilizando da análise de dados de características disponibilizadas nos dados da rede social; (ii) analisar o desempenho da seleção de portfólio com as métricas obtidas da caracterização dos dados; e (iii) comparar o desempenho do portfólio criado com base nos dados sociais com outros índices de mercado, de renda variável e de renda fixa.

1.3 Contribuições

Espera-se com os resultados desta dissertação entender melhor o uso dos dados da rede social para melhoria de estratégias de investimento no mercado financeiro de ações. Além disso, os objetivos propostos tornam possível estruturar um paralelo entre as informações relevantes da *Stocktwits* e o comportamento dos ativos da Bolsa Americana de Valores de forma a compreender como esses dois universos se relacionam.

Assim, as principais contribuições desta dissertação são:

- uma análise do comportamento do usuário em relação às postagens de mensagens com sentimento no *Stocktwits*;
- uma análise da existência de correlação entre postagens realizadas e movimentações no mercado financeiro; e
- uma análise da caracterização, compreensão e seleção de portfólios com o auxílio de informações extraídas de mensagens postadas por investidores utilizando dados disponíveis em uma rede social de finanças.

1.4 Organização do trabalho

Este projeto de pesquisa é organizado em 6 capítulos. O atual trata desta introdução, no qual são expostos o tema, a relevância, uma contextualização histórica e os objetivos deste projeto. A fundamentação teórica é exposta no Capítulo 2, com subseções específicas para discorrer sobre introdução ao Mercado Financeiro, Mineração de dados em redes sociais e Seleção de Portfólios. No Capítulo 3 são apresentados os trabalhos relacionados, com um estudo referente ao tema de redes sociais e carteiras de investimento ou portfólios. No Capítulo 4 é apresentada a metodologia proposta (materiais e métodos), evidenciando principalmente o sequenciamento das atividades e as escolhas da pesquisa. Os Resultados obtidos pelos

experimentos desenvolvidos são apresentados Capítulo 5. Por fim, o Capítulo 6 apresenta a conclusão do trabalho e descreve tópicos que podem ser explorados em trabalhos futuros.

2 Fundamentação Teórica

Este capítulo está organizado em três seções principais, que apresentam uma fundamentação teórica de trabalhos sobre o mercado financeiro (Seção 2.1), a mineração de dados de redes sociais (Seção 2.2) e os conceitos de seleção e otimização de portfólios (Seção 2.3).

2.1 Introdução ao Mercado Financeiro

O mercado financeiro é o local onde podem ser negociados, comprados ou vendidos, bens como valores mobiliários, mercadorias e câmbio. A função do mercado é reunir diversos vendedores em um único lugar, tornando-os acessíveis aos compradores interessados. Mercados são considerados como parte vital de qualquer economia, pois, quanto mais ativos, mais se traduzem em oportunidades para que pessoas e empresas apliquem seus recursos e contribuam para o aquecimento da economia (NETO, 2009).

O mercado financeiro funciona de forma a aproximar agentes, como um investidor e um tomador de recursos. Eles não precisam conversar ou estabelecer contato entre si, pois essa ponte é feita através de aplicações da própria instituição financeira.

Dessa forma, você pode investir seu dinheiro aplicando em um Certificado de Depósito Bancário (CDB), que são investimentos de Renda Fixa de baixo risco emitidos pelos bancos e instituições financeiras, e acabar fornecendo o capital necessário para que a instituição financeira ofereça um empréstimo a um empresário que precisa de capital de giro para o seu negócio. Assim, o resumo é que quem possui recursos em excesso empresta para quem sofre com sua falta e demonstra capacidade de pagar.

Para normatizar o mercado, existem diversos órgãos importantes, entre eles: Conselho Monetário Nacional (CMN), o Conselho Nacional de Seguros Privados (CNSP) e o Conselho Nacional de Previdência Complementar (CNPc). Eles dão as diretrizes operacionais a partir das quais as instituições financeiras se baseiam.

Veja como o mercado financeiro é dividido:

- **Mercado de crédito:** trata dos empréstimos bancários. É o mercado que você acessa ao solicitar um financiamento ou usar o cheque especial.
- **Mercado aberto:** cuida das empresas com capital aberto, ou seja, que negocia suas ações através da bolsa de valores, que regula a oferta e a demanda pelos papéis das companhias.
- **Mercado de câmbio:** é a plataforma de negociação de moedas estrangeiras da relação justa entre as moedas dos países.

O mercado financeiro oferece basicamente dois tipos de investimento, a renda fixa e a renda variável. Ambas são bastante interessantes e não devem ser descartadas pelo investidor (NETO, 2009).

- **Renda fixa:** é o tipo de investimento que oferece uma base de projeção ou o cálculo do retorno exato antes da aplicação. Títulos assim podem ter rendimento prefixado, com um juro anual definido, pós-fixado, atrelado a um indicador como o CDI (Certificado de Depósito Interbancário, referência de rentabilidade), ou híbrido, com um juro fixo mais a variação do IPCA (Índice de Preços ao Consumidor Amplo, considerado a inflação oficial do país). São exemplos de renda fixa boa parte das aplicações que você conhece, como a poupança, o Certificado de Depósito Bancário (CDB), a Letra de Crédito Imobiliário e Letra de Crédito do Agronegócio (LCI/LCA), Tesouro Direto, Debêntures, Letra de câmbio (LC), entre outros.
- **Renda variável:** ainda é pouco explorada pelo investidor pessoa física no Brasil. Em mercados mais desenvolvidos, como os Estados Unidos, a ela representa fatia bem mais ampla dos investimentos. Exemplos de renda variável são ações, opções e derivativos na bolsa de valores, fundos de investimento de ações e multimercados, entre outros. Na comparação com a renda fixa, a variável acarreta maior volatilidade e maior risco de prejuízo, embora ofereça potencial de retornos mais elevados.

O princípio do mercado financeiro é simples, tentar comprar quando acredita-se que o ativo vai subir de preço no futuro e vender quando acredita-se que o preço irá diminuir, de forma a obter lucro. No entanto, realizar essas previsões é um desafio gigantesco, pois existem inúmeros fatores que podem influenciar o valor de um ativo. O preço de uma ação em bolsa é fruto das condições de mercado (oferta e demanda) que refletem as condições estruturais e comportamentais da economia de um país, específicas da empresa e de seu setor econômico (FORTUNA, 2007).

Ações são títulos, conhecidos também como papéis, que representam frações de capital de uma determinada empresa. Uma ação é uma fração da empresa. Quando adquirimos ações de uma empresa, nos tornamos sócios desta, participando de seus resultados. Esses títulos são emitidos quando uma empresa toma a decisão de abrir seu capital para o público, e ser negociada na bolsa de valores. Existem dois tipos de ações (CAVALCANTE; MISUMI, 2005):

- **Ordinárias:** estas ações, além de proporcionarem a participação nos resultados da empresa, conferem o direito de voto nas assembleias da empresa.
- **Preferenciais:** adquirindo este tipo de ações, o investidor terá preferência no recebimento de dividendos, porém, não terá o direito a voto nas assembleias.

Existem 60 grandes bolsas de valores em todo o mundo e 16 delas têm capitalização de mercado de mais de US\$ 1 trilhão. Tanto os grandes *players*, como bancos ou fundos de investimentos e os pequenos investidores participam desse mercado. O mercado de ações opera em um sistema de oferta e demanda e tem oportunidades para investidores de longo prazo e

para *day traders*, investidores que tem por objetivo a obtenção de lucro com a oscilação de preço, ao longo do dia, de ativos financeiros. O mercado de ações dos EUA tem bolsas bem conhecidas como o mercado de ações NASDAQ e a bolsa de valores de Nova York (NYSE). Elas têm a reputação de negociar títulos sólidos e de alta qualidade e apresentam algumas das empresas mais negociadas no mundo, como *Apple*, *Amazon*, *Bank of America*, *General Electric*, *ExxonMobil* e *Johnson&Johnson*¹.

A Bolsa de Nova York (NYSE)² apresenta uma capitalização em torno de US\$ 23 trilhões e 2.400 empresas listadas. Seu principal índice é o *Dow Jones* (.DJI) que serve como referência para os mercados de capitais e para todas as bolsas do mundo. A Bolsa de valores NASDAQ³ é lembrada sobretudo por ser a bolsa das empresas de tecnologia e pela bolha dos pontos com dos anos 2000 (embora tenha sido fundada em 1971). Apresenta uma capitalização por volta de US\$ 10 trilhões e 3.058 empresas listadas. O principal índice é o *Nasdaq Composite* (.IXIC).

Outro índice importante a destacar é o SP500, índice *Standard and Poor's 500 - S&P500* (.GSPC)⁴ - é um índice de capitalização ponderada de 500 ações mais representativas e negociadas na NYSE e na NASDAQ.

Na Figura 1 podemos verificar um gráfico com a evolução do índice ao longo dos últimos 5 anos. Esta visão gráfica é fornecida por uma importante plataforma de análises do mercado de ações global chamada *Trading View*⁵. Este índice foi criado em 1957. Desde então, ele é considerado como o principal indicador do mercado acionário norte-americano. Atualmente, o S&P 500 possui patrimônio em torno de US\$ 20 trilhões e a sua pontuação está na casa dos 2.700 pontos.

No Brasil a Bolsa de Valores, Mercadorias e Futuros de São Paulo (B3) é responsável por administrar todas as negociações de títulos, valores mobiliários e contratos derivativos. A mesma realiza serviços de registro, compensação, liquidação e atua como contraparte garantidora da liquidação financeira das operações realizadas pelos investidores (B3. . . , 2020). Atualmente a B3 possui cerca de 500 empresas cadastradas e em sua página Web são disponibilizados arquivos de séries temporais de cotações históricas desde a década de 80 até os dias atuais. Estes arquivos são públicos e possuem as seguintes informações: nome da empresa, código, tipo de mercado (a vista, termo, opções), preços (abertura, fechamento, máximo, mínimo, médio e anterior), quantidade de negócios, volume negociado, dentre outras informações.

(FAMA, 1970) propõe a hipótese do mercado eficiente (HME) afirmando que os preços das ações refletem instantaneamente toda a informação disponível, não sendo possível um investidor obter vantagens sobre o mercado. Pesquisas envolvendo esta hipótese demonstram que

¹ Mercado de Ações dos EUA - <https://br.tradingview.com/markets/stocks-usa/>

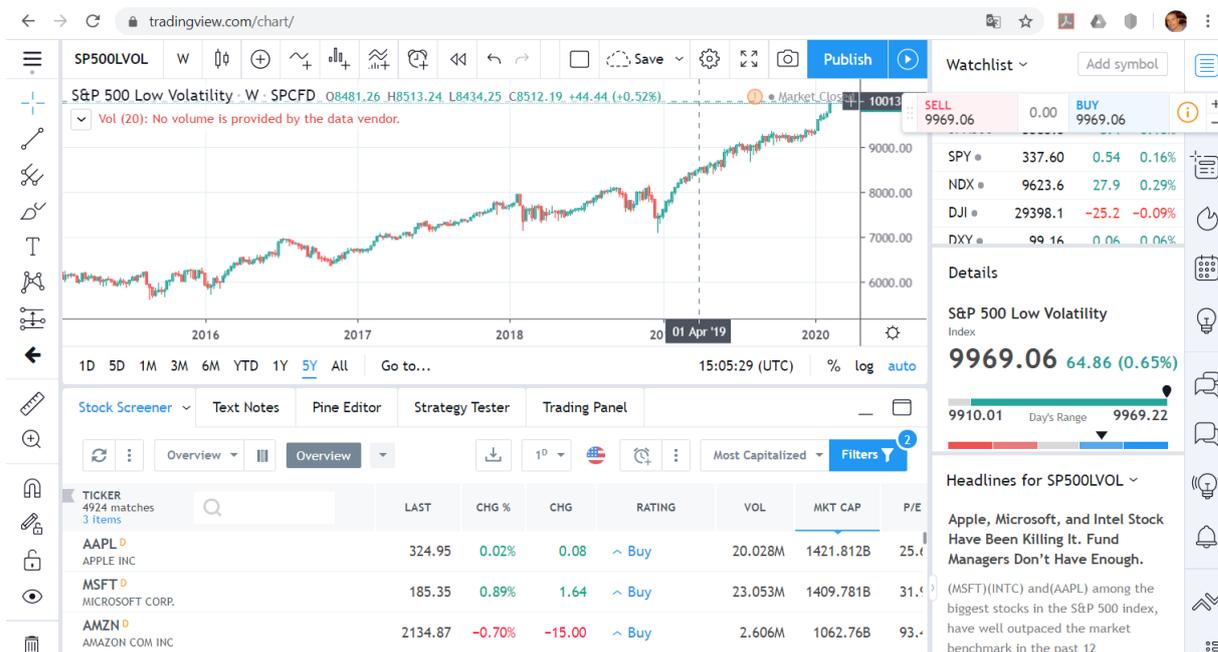
² NYSE - <https://www.nyse.com/index>

³ NASDAQ - <https://www.nasdaq.com/>

⁴ Índice S&P 500 - <https://br.tradingview.com/symbols/SPX/>

⁵ Trading View - <https://www.tradingview.com/>

Figura 1 – Painel gráfico do Índice S&P500



Fonte: *Tradingview.com*

raramente o mercado cria oportunidades possibilitando valores altos de lucro. Para corroborar a hipótese, já foram realizados experimentos demonstrando que uma estratégia de investimento utilizando um agente aleatório consegue produzir resultados melhores do que indicadores técnicos tradicionais (BIONDO et al., 2013). Entretanto, existem anomalias que não podem ser explicadas pela HME, e há uma outra corrente de pesquisadores que acredita ser possível prever variações no comportamento do preço das ações até certo nível. (LO, 2004) propõe a hipótese do mercado adaptativo, em que o mercado na maior parte do tempo é eficiente, mas que há períodos de tempo em que nem todas as informações são refletidas no preço das ações de forma instantânea. Alguns trabalhos empíricos defendem essa segunda corrente, como por exemplo (LO; MACKINLAY, 2011).

Os métodos de análise do mercado financeiro podem se dividir em dois tipos (ANGHEL, 2013):

- **Análise fundamentalista:** o investidor possui como referência parâmetros que definem a situação financeira da empresa como o lucro líquido, grau de endividamento e distribuição de dividendos, dentre outros. Em síntese, a análise fundamentalista parte do princípio de que as ações têm um valor intrínseco, que corresponderia ao seu preço justo. Esse preço, por sua vez, seria determinado pelo fluxo de rendimentos auferido pela ação e efetivamente distribuído em um dado intervalo de tempo, descontando o valor presente.
- **Análise técnica:** o foco principal deste tipo de análise é exclusivamente o valor da ação (preço) e a movimentação ou quantidade de compra e venda em um determinado período (volume). Esse fato permite projetar uma trajetória ou movimentos oscilatórios futuros

Preço	Volume
Média Móvel Simples	Índice de Fluxo de Dinheiro
Média Móvel Exponencial	<i>On Balance Volume</i>
Índice de Força Relativa	Média Móvel Simples de Volume

Tabela 1 – Exemplos de Indicadores Financeiros

prováveis dos preços das ações.

Os Indicadores financeiros representam linhas de oscilações ou tendências, construídas através de fórmulas matemáticas ou estatísticas, cujos objetivos são antecipar, confirmar ou descrever tendências. Os índices de volatilidade representam importantes instrumentos de avaliação dos movimentos de preços em relação à tendência do preço no mercado financeiro. Devido à existência de uma grande quantidade de indicadores, este trabalho ficou limitado ao entendimento e utilização de alguns indicadores de preço e de volume nas simulações, conforme Tabela 1.

Para exemplificar, serão detalhados a seguir os seguintes indicadores financeiros de análise técnica de preço e volume.

- **Média Móvel Simples:** A Média Móvel Simples, *Simple Moving Average*, (SMA) representa a média dos preços de fechamento nos períodos de curto, médio e longo prazo. A SMA suaviza os movimentos voláteis de uma série de preços, identificando melhor a tendência do preço para enxergar oportunidades de compra e venda nas mudanças de padrões de tendência. Matematicamente, a média móvel simples nada mais é que uma média de certo período dos dados passados em série temporal, por exemplo, a média dos preços de fechamento de uma ação dos últimos 20 dias.
- **Índice de Força Relativa:** O Índice de Força Relativa, *Relative Strength Index*, (RSI) verifica o enfraquecimento de uma tendência e a sinalização de um rompimento de suporte ou resistência antes que esses movimentos ocorram. O RSI retorna uma escala de variação fixa, que varia de 0 a 100 indicando regiões de sobrecompra (ação sobrevalorizada) quando o retorno é > 70 ou sobrevenda (ação subvalorizada) quando o retorno é < 30 .
- **On Balance Volume:** O Volume de Balança, *On Balance Volume*, (OBV) é um indicador que relaciona mudanças no preço com as variações de quantidade de ações negociadas, que identifica a direção do fluxo do volume, ou seja, se está entrando ou saindo volume do ativo. Quando o fechamento do dia for superior ao do dia anterior todo o volume é considerado volume de alta. Caso contrário, todo o volume é considerado volume de baixa.
- **Média Móvel Simples de Volume:** A Média Móvel Simples de Volume, *Simple Moving Average of Volume* (SMAV) utiliza o volume transacionado (compras e vendas executadas) de certo ativo financeiro em determinado ponto do tempo juntamente com a média móvel desse mesmo volume ao longo do tempo. A SMAV identifica rompimentos de suportes e

resistências, quebras de tendências nos preços e a tendência do volume transacionado. O volume é computado pelo número de transações para qualquer ativo financeiro. Pode ser exibido como o número de ações transacionadas, onde o mais comum é, o número de negócios feitos ou volume financeiro movimentado.

2.2 Mineração de dados em redes sociais

O objetivo desta seção é apresentar uma visão geral de redes sociais, a análise de dados a partir de rede sociais, bem como algumas aplicações que podem ser utilizadas nesta área.

2.2.1 Redes Sociais

As redes sociais estão inseridas no cotidiano das pessoas, alcançando um grande número de possíveis clientes. Dados recentes indicam que o *Twitter* obteve um rápido e intenso crescimento, passando de 1,2 milhão de usuários em 2008 para 200 milhões de contas em janeiro de 2011 (OLIVEIRA; CAMPOS; SMARZARO, 2012). Sendo assim, tal alternativa mostra-se uma forma interessante de coleta de dados, pois possibilita a extração de opiniões de forma automatizada do grande volume de dados.

Redes sociais tornaram-se especialmente relevantes devido à grande variedade de sites que utilizam o conceito, como *Twitter*⁶, *Facebook*⁷, *Instagram*⁸ e *Stocktwits*⁹. Seus usuários formam bases de dados que provêm um importante meio de compartilhar, organizar e encontrar conteúdo e contatos. Devido ao uso intenso, estes sites reúnem material suficiente para subsidiar estudos de características de redes sociais em larga escala.

Para (EMIRBAYER; GOODWIN, 1994) uma rede social é um grafo, orientado ou não, que mapeia uma realidade ou um mundo restrito, no qual os nodos representam as entidades (indivíduos ou classes de indivíduos – também chamados atores) e as arestas representam os relacionamentos entre essas entidades. Os relacionamentos podem ser o compartilhamento de um ou mais atributos. A realidade representada pelas redes sociais são fontes de dados heterogêneos e multi-relacionais, cujos relacionamentos podem ser unidirecionais e não necessariamente precisam ser binários.

(HAN; KAMBER; PEI, 2011) definem que, do ponto de vista de implementação, uma rede social é um conjunto de dados heterogêneos e multi-relacionais representados por um grafo. O grafo é tipicamente muito grande, com nós correspondendo aos objetos e às arestas correspondendo as ligações que representam relacionamentos ou interações entre os objetos.

⁶ Twitter: <http://www.twitter.com/>

⁷ Facebook: <http://www.facebook.com/>

⁸ Instagram: <http://www.instagram.com/>

⁹ Stocktwits: <http://www.stocktwits.com/>

Os objetos e as ligações possuem atributos, sendo que os objetos podem ter rótulos de classe e as ligações podem ser unidirecionais.

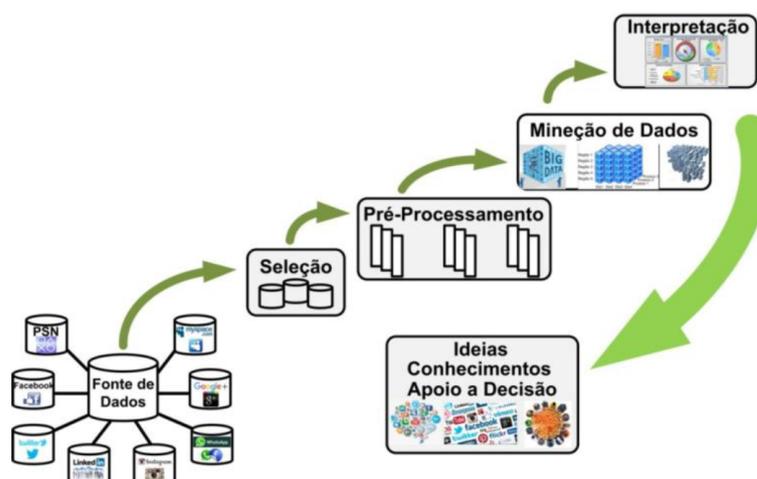
Portanto, considerando-se o grande alcance das redes sociais e as informações que podem ser obtidas delas, propõe-se com este trabalho a coleta de dados de uma rede social em especial e o uso destes para obtenção de informações, através da utilização de técnicas de mineração de dados (OLIVEIRA; CAMPOS; SMARZARO, 2012).

2.2.2 Análise de Redes Sociais

Redes sociais podem ser aplicadas no estudo de diferentes situações e questões, nas diferentes áreas do conhecimento. A análise de redes sociais é uma abordagem oriunda da Sociologia, da Psicologia Social e da Antropologia (FREEMAN, 1994). Tal abordagem estuda as ligações relacionais entre atores sociais. Os atores, na análise de redes sociais, cujas ligações são analisadas, podem ser tanto pessoas e empresas, analisadas como unidades individuais, quanto unidades sociais coletivas.

As aplicações desenvolvidas para a extração de informações são classificadas como sistemas de descoberta de conhecimento, também chamado de KDD (*knowledge Discovery in Databases*), e poderão atender a gestores, políticos, diretores e empresários. A análise sistemática a ser estudada tem como premissa a capacidade forte de integração em redes sociais (ARAUJO; CRUZ, 2017). Um esboço de cenário operacional de mineração de dados em redes sociais é ilustrado na Figura 2 .

Figura 2 – Processo Operacional de Mineração de Dados em Redes Sociais



Fonte:(SILVA; CRUZ, 2016)

Na figura 2 é mostrada a aquisição de dados de possíveis diversas fontes geradoras de dados em redes sociais. Os dados extraídos das fontes são selecionados de acordo com nível de interesse em análise. Após a etapa operacional de seleção, os dados são transformados e adaptados para formatos computacionais tratáveis o que permite a compatibilidade com os

algoritmos e técnicas de mineração de dados a serem aplicadas sobre os dados em redes sociais. Feito esta etapa, pode-se contribuir para gerar interpretações personalizadas e dar apoio à geração de conhecimento e tomadas de decisão (SILVA; CRUZ, 2016).

Com a popularização de redes sociais, a participação de usuários se comunicando, navegando ou mesmo acessando conteúdos e atividades publicados, tem possibilitado a geração de um histórico destas interações (ARAUJO; CRUZ, 2017).

A mineração de dados ou *data mining* surgiu da junção de três áreas científicas que se relacionam: Banco de dados, Estatística e Inteligência Artificial (LIN; CHUNG, 2015). Ela oferece estratégias para análises de Big data, encontrando padrões e correlações entre os dados, que ajudam na tomada de decisões e desenvolver estratégias.

Entender a natureza distribuída e conexa dos dados nestes ambientes, é algo fundamental para que se possa atingir metas de negócio, e garantir uma maior abrangência de personificação e inovação de conhecimento sobre hábitos e padrões de interação humanas em redes sociais. A mineração de dados aplicada nas redes sociais busca obter resultados de comportamentos ou tendência social, usando como base de pesquisa as informações de cultura, costumes, localização (ARAUJO; CRUZ, 2017).

Atualmente, com a base informacional pública produzida em ambientes interacionais online, como as mídias sociais, temos a possibilidade de resgatar os rastros deixados pelos usuários, compreender um pouco mais alguns comportamentos, cenários e hábitos existentes para, com base nisso, realizar tomadas de decisão. É nesse contexto que a Mineração de Dados dialoga fortemente com as mídias sociais (YAKUSHEV; MITYAGIN, 2014), permitindo que organizações possam obter insumos para a inteligência competitiva.

2.3 Seleção de Portfólios

O processo de seleção de carteiras de investimento em ativos de risco permanece como um dos problemas centrais em economia financeira, tanto do ponto de vista acadêmico, como também para os praticantes de mercado.

A teoria base da seleção de portfólios teve início com (MARKOWITZ, 1952), em que a seleção de portfólio é baseada sob o paradigma da análise média-variância introduzida por Harry Markowitz em 1952, a qual transformou o processo de alocação de ativos em um problema de otimização com base no *trade-off* fundamental entre retorno esperado e risco.

Dessa forma, o investidor deve escolher a carteira com a menor variância entre um infinito número de carteiras que proporcionassem um determinado retorno ou, de forma equivalente, para um determinado nível de aversão ao risco, escolher a carteira que maximize o retorno esperado.

2.3.1 Gestão de Portfólios

Gestão de portfólio consiste na aplicação de recursos limitados, nos investimentos que melhor atendem às necessidades dos investidores, sob o ponto de vista de seu anseio por maior retorno em comparação com sua aversão ao risco.

Os gestores de portfólio podem optar por dois tipos básicos de investimento: renda fixa e renda variável. Os investimentos em renda fixa podem ser divididos em pré e pós-fixados, sendo que os pré-fixados apresentam remuneração determinada no momento da aplicação, enquanto que os pós-fixados somente no momento final da aplicação, pois seu rendimento se encontra atrelado a algum índice.

Existem diversos investimentos de renda fixa, como por exemplo, caderneta de poupança, Certificado de Depósito Bancário (CDB), dentre vários outros títulos de dívida pública ou privada, porém, a discussão acerca dos mesmos não faz parte do escopo deste trabalho. Já os investimentos de renda variável possuem retorno totalmente incerto, como por exemplo, o investimento em ações. Existem inúmeras possibilidades de investimento, cada qual apresentando certa expectativa de retorno para um determinado nível de risco. Porém, somente após os estudos de (MARKOWITZ, 1952) a seleção de portfólio passou a ser tratada de maneira científica, no que diz respeito à combinação dos investimentos com melhor relação risco-retorno. Antes do desenvolvimento da teoria de portfólio, os investidores analisavam simplesmente a relação risco-retorno de ativos individuais na composição de suas carteiras de investimento.

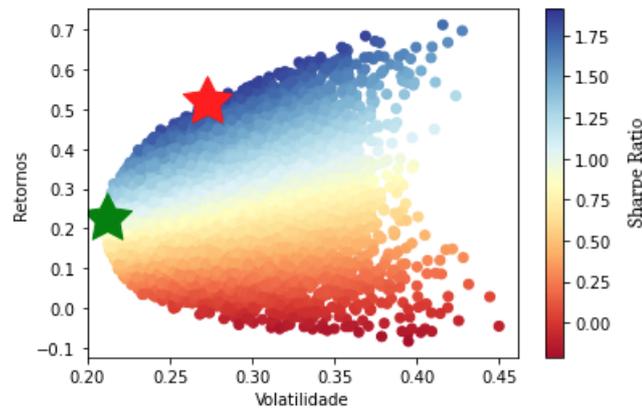
Na Gestão de portfólio são utilizados os conceitos de diversificação e de criação de uma fronteira eficiente que auxiliam os investidores no momento de escolher combinações de ativos com base nas metas de investimento; isto é, a troca entre o nível desejado de retorno do portfólio e o nível desejado de risco do portfólio.

A Teoria Moderna do Portfólio, *Modern Portfolio Theory*, (MPT) uma hipótese apresentada por Harry Markowitz (MARKOWITZ, 1952) em seu artigo *Portfolio Selection*, publicado em 1952 pelo *Journal of Finance*, é uma teoria de investimento baseada na ideia de que investidores avessos ao risco podem construir portfólios para otimizar ou maximizar o retorno esperado com base em um determinado nível de risco de mercado, enfatizando que o risco é uma parte inerente da maior recompensa. É uma das teorias econômicas mais importantes e influentes que lidam com finanças e investimentos.

Também chamada de teoria do portfólio ou teoria do gerenciamento de portfólio, o MPT sugere que é possível construir uma fronteira eficiente de portfólios ótimos, oferecendo o máximo retorno possível para um determinado nível de risco. Isso sugere que não é suficiente olhar para o risco esperado e o retorno de uma ação específica. Ao investir em mais de uma ação, um investidor pode colher os benefícios da diversificação, particularmente uma redução no risco da carteira.

Na Figura 3 podemos verificar que a estrela vermelha representa o portfólio com retornos

Figura 3 – Representação da Fronteira Eficiente



Fonte:Elaborada pelo autor

mais altos ajustados ao menor risco. Já a estrela verde representa o portfólio de variação mínima, que é o portfólio com a menor volatilidade e menor taxa de retorno.

2.3.2 Modelos de Seleção de Portfólios

Nesta sub-seção será apresentados os dois modelos utilizados para selecionar e otimizar a formação dos portfólios construídos a partir da análise de dados da *Stocktwits*. Serão apresentados os modelos de (MARKOWITZ, 1952) e o de (SHARPE, 1964).

2.3.2.1 Modelo de Markowitz

O modelo de (MARKOWITZ, 1952) representa matematicamente o problema da composição de uma carteira de investimentos que possibilite a minimização do risco para um determinado nível de retorno, sendo o risco representado pela variância. No intuito de se encontrar o portfólio com variância mínima, são definidos valores esperados para o retorno dos ativos, os quais devem resultar em uma carteira com retorno maior ou igual ao valor mínimo definido pelo tomador de decisão. Além disso, a variância de cada ativo, bem como a correlação entre os mesmos são definidos com base em dados históricos. Dessa forma, Markowitz definiu o retorno esperado e o risco do portfólio.

Pela Equação 1, o Retorno Esperado:

$$R = \sum_{i=1}^N w_i r_i \quad (1)$$

Pela Equação 2, a Variância:

$$\sigma^2(A) = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} w_i w_j \quad (2)$$

Onde:

w_i : composição de cada ativo no portfólio;

r_i : retorno esperado para o ativo i ;

σ_{ij} : covariância entre os ativos i e j .

A função da Variância pode ser representada vetorialmente pela Equação 3 :

$$\sigma^2 = w' \Sigma w \quad (3)$$

Sendo que:

w : vetor composição dos ativos no portfólio;

Σ : matriz de covariância.

A matriz de covariância consiste em uma matriz simétrica, na qual cada valor a_{ij} representa a covariância do ativo i com o ativo j . Desta forma, o problema de determinação da composição ótima de uma carteira de investimentos de n ativos $W = (w_1, w_2, \dots, w_N)$, com retorno esperado $R = (r_1, r_2, \dots, r_N)$ pode ser formulado da seguinte maneira através do modelo de Média-Variância de Markowitz:

$$\min \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} w_i w_j \quad (4)$$

$$s.t. \sum_{i=1}^N w_i \bar{r}_i = \mu \quad (5)$$

$$\sum_{i=1}^N w_i = 1 \quad (6)$$

$$x_i \geq 0 \quad (7)$$

$i = 1, 2, 3, \dots, N$

Onde:

N = número de ativos candidatos a compor o portfólio;

w_i = fração do capital a ser investido no ativo i ;

σ_{ij} = covariância entre os retornos dos ativos i e j ;

\bar{r}_i = valor esperado dos retornos do ativo i ;

μ = valor esperado do retorno do portfólio dado pelo investidor.

Neste modelo, a intenção é construir um portfólio com risco mínimo, dado o retorno esperado (μ). Dessa forma, a função objetivo (Equação 4) do modelo tem como finalidade

minimizar o risco (variância) do portfólio. A restrição (Equação 5) determina que o valor do retorno esperado do portfólio é igual a μ , meta do retorno esperado dada pelo investidor. A restrição (Equação 6) descreve que todo o capital disponível seja investido em 100%. Por fim, a restrição (Equação 7) garante que não haja um percentual de investimento negativo em nenhum dos ativos.

Um ponto importante a ser ressaltado é a dificuldade existente para a aplicação da teoria proposta por (MARKOWITZ, 1952), uma vez que sua formulação apresenta um elevado número de operações e certa complexidade dos cálculos necessários ao seu desenvolvimento. Em decorrência disso, William Sharpe, desenvolveu um modelo simplificado, chamado Modelo de Índice Único, (MIU) (SHARPE, 1963).

2.3.2.2 Modelo de Sharpe

Em seu modelo, (SHARPE, 1963) parte de um índice único chamado de índice beta (β), que compara a volatilidade do ativo para com o mercado, ao invés da utilização de covariância.

O índice de Sharpe, determinado pela Equação 8 abaixo:

$$IS = \frac{R_i - R_f}{\sigma_i} \quad (8)$$

Sendo que:

IS = Índice de Sharpe;

R_i = Retorno do Ativo;

R_f = Retorno Livre de Risco (*Risk-Free*);

σ_i = Risco do Ativo (a letra grega sigma representa volatilidade).

Conforme (ROMERO, 2019) o índice de Sharpe é um indicador que mede o retorno excedente de uma aplicação financeira em relação a uma aplicação livre de risco, isso implica que quanto mais alto for o índice de Sharpe, melhor foi a performance desse investimento em relação ao risco que ele oferece.

Destaca-se o fato de que o Índice deve ser normalizado, adotando-se aqui o período de um ano. Para isso, divide-se a rentabilidade esperada e a rentabilidade livre de risco pela quantidade de anos de análise do portfólio.

Destaca-se também o fato de que os portfólios são compostos por diferentes períodos acoplados, logo, faz-se necessário o cálculo da volatilidade. Para isso, é necessário o cálculo da variação média (Equação 9), a partir da qual obtemos a variância amostral (Equação 10) e então o desvio padrão (Equação 11), que corresponde a volatilidade. Conforme (BARBETTA REIS; BORNIA, 2004), a variação média é dada pela média da variação nos períodos, a variância amostral é a soma do quadrado da diferença entre o valor de cada período e a variação média,

dividido pelo número de amostras menos um e o desvio padrão é a raiz da variância amostral. Logo, temos:

$$\bar{R} = \frac{\sum_{t=1}^N R_t}{n} \quad (9)$$

$$Var = \frac{\sum_{i=1}^N (R_i - \bar{R})^2}{n - 1} \quad (10)$$

$$\sigma = \sqrt{Var} \quad (11)$$

Sendo que:

\bar{R} = Retorno médio amostral;

R_t = Retorno obtido no período t ;

n = Número de períodos no intervalo de estudo;

Var = Variância amostral;

σ = Desvio padrão amostral.

É importante aqui também definir o método de Monte Carlo, uma vez que esse é utilizado para a plotagem do gráfico da fronteira eficiente. Segundo (RIBEIRO; FERREIRA, 2005), o método baseia-se na geração de amostragens aleatórias em larga escala para obtenção de resultados numéricos, ou seja, sucessivas simulações em um elevado número de vezes, para calcular probabilidades heurísticamente. Para a geração do gráfico da fronteira eficiente, varia-se os pesos dos ativos aleatoriamente, em larga escala, e calcula-se o retorno e o risco de cada um dos portfólios para plotagem.

2.3.3 Simulação de Monte Carlo

A Simulação de Monte Carlo (SMC) envolve o uso de números aleatórios e probabilidades para analisar e resolver problemas. Esse método surgiu no Projeto Manhattan, nos Estados Unidos, durante a Segunda Guerra Mundial, sendo desenvolvido pelos cientistas John Von Neumann e Stanislaw Ulam. A denominação “Monte Carlo” foi cunhada em referência aos jogos de azar que fazem uso constante de sorteios e de dados, uma atração popular na cidade de Monte Carlo, Mônaco (METROPOLIS; ULAM, 1949).

A simulação de Monte Carlo consiste em um método que utiliza a geração de números aleatórios para atribuir valores às variáveis do sistema que se deseja investigar. Os números são obtidos de artifícios aleatórios (por exemplo: tabelas, roletas, sorteios) ou diretamente de softwares, através de funções específicas. A cada iteração, o resultado é armazenado e, ao final de todas as repetições, a sequência de resultados gerados é transformada em uma distribuição de frequência que possibilita calcular estatísticas descritivas, como média (valor esperado),

valor mínimo, valor máximo e desvio-padrão, cabendo ainda ao executor das simulações a prerrogativa de projetar cenários futuros de operação do sistema em análise. O método de simulação de Monte Carlo pode ser aplicado em problemas de tomada de decisão a qual envolva risco e incerteza, ou seja, situações nas quais o comportamento das variáveis envolvidas com o problema não é de natureza determinística (CORRAR; THEOPHILO, 2004).

Para uma correta operacionalização da SMC, a simulação deve ser replicada mais de cem vezes para que se obtenha uma amostra representativa. No entanto, não há recomendação quanto ao número máximo de simulações a serem realizadas. Como instrução básica, deve-se aplicar o maior número de simulações possíveis levando em consideração o poder de processamento do equipamento a ser utilizado, pois o equilíbrio entre precisão e tempo de computação é uma característica importante das simulações da SMC (ESCUDEIRO, 1973).

3 Trabalhos Relacionados

Este capítulo apresenta as principais pesquisas relacionadas a este trabalho. É então apresentada uma revisão da literatura sobre a caracterização e a análise de dados da *Web* e de redes sociais, em especial, os dados da rede social *StockTwits*.

Grande parte dos trabalhos relacionados realizam análise de sentimentos em sua metodologia, que é o estudo de opiniões, sentimentos e emoções expressas em textos. Existem muitas tarefas relacionadas a essa área, como a extração de elementos do texto referente à opinião, à classificação quanto ao seu caráter (positivo, negativo e neutro), à comparação de sentenças quanto a suas opiniões, sendo possível categorizar informações de texto como fatos ou opiniões (LIU, 2010). Oliveira et al. (OLIVEIRA; CORTEZ; AREAL, 2013a) utilizam indicadores de sentimento associados a um modelo de regressão para prever três variáveis do mercado: retorno, volatilidade e volume negociado. Nesse trabalho, eles chegam à conclusão de que prever o andamento do mercado é uma tarefa muito complexa e que são necessários modelos muito bem embasados.

Estudos mais recentes, como o realizado por Li et al. (LI; SHAH, 2017), consideram métodos mais específicos de análise de sentimentos, através de indicadores léxicos associados a redes neurais, obtendo resultados melhores na predição de sentimento das postagens (*tweets*). Nesse contexto, (AGRAWAL et al., 2019) demonstram que a atividade de mídia social pode afetar significativamente a liquidez no dia a dia, e (AUDRINO; SIGRIST; BALLINARI, 2020) descrevem o impacto das variáveis de sentimento e medidas de atenção na volatilidade do mercado de ações.

Similarmente a este trabalho por utilizarem dados do *Stocktwits* e usar análise de sentimentos para predição no mercado financeiro, Oh and Sheng (OH; SHENG, 2011) estudam o poder preditivo de tal *microblog* por meio da análise de sentimentos de dados financeiros sobre os movimentos direcionais dos preços das ações futuras. Foi construído um conjunto robusto de modelos baseados em algoritmos de análise de sentimento e mineração de dados, utilizando 72.221 micro publicações de *blog* para referências de ações de 3.874 autores distintos. Além disso, foi verificado que a análise de sentimentos de dados do *microblog* possui uma precisão preditiva e consistente com a hipótese de sub-reação observada nas finanças comportamentais, fornecendo suporte para o modelo de sentimento do investidor irracional.

Nessa mesma linha, Oliveira et al. (OLIVEIRA; CORTEZ; AREAL, 2013b) utilizam o volume de postagens e informações de sentimento para desenvolver indicadores para prever o retorno, a volatilidade e o volume de negociações. Foram coletados da plataforma *Stocktwits* postagens de 1 junho de 2010 até 31 de outubro de 2012, totalizando 605 dias de postagens. Nesse trabalho, eles chegam à conclusão de que prever o andamento do mercado é uma tarefa

muito complexa e que são necessários modelos muito bem embasados. Considerando ainda a volatilidade do mercado de ações, Atkins et al. (ATKINS; NIRANJAN; GERDING, 2018) estudaram como as notícias impactam nesse aspecto. Em tal pesquisa, utilizaram métodos de aprendizado de máquina aplicado a notícias de jornais e concluíram que a predição do mercado de ações apresenta maior acurácia quando notícias midiáticas são consideradas nos modelos.

Ademais, diferentemente deste trabalho, existem linhas de pesquisa que investigam a influência dos usuários em relação ao mercado financeiro. Por exemplo, Wang et al. (WANG et al., 2015) utilizam dados do Stocktwits e do SeekingAlpha¹ para determinar quem são os usuários especialistas e se suas postagens têm utilidade para investidores individuais.

Similarmente, Tu et al. (TU et al., 2016) investigam os especialistas no Stocktwits por meio de técnicas de aprendizado de máquina e utilizam as postagens de usuários *experts* para gerar recomendações de portfólio. Similarmente, Bar-Haim et al. (BAR-HAIM et al., 2011b) criaram um modelo para a identificação de investidores *experts*, através de dados da *Stocktwits* coletados durante 2 períodos (25 de abril até 01 de Novembro de 2010 e 14 de dezembro até 03 de Fevereiro de 2011), aproximadamente 1 ano de dados de postagens, fornecendo uma análise aprofundada do conteúdo e de potencialidade de utilização das postagens desta rede social de investimentos.

Em Ferreira (FERREIRA, 2016), avaliou-se a influência de usuários do *Stocktwits* dedicados aos mercados financeiros via uma abordagem de mineração de dados. Utilizou-se um conjunto de dados contendo cerca de 340.000 mensagens relativas a 10.000 usuários em um período de 2 anos e 9 meses, entre junho de 2010 e março de 2013. Sobre esse conjunto de dados foram testadas 8 métricas para medir a influência de utilizadores dentro de uma rede social, designada de *Parent*, em que se contabiliza o número de mensagens (*posts*) originais de um dado utilizador, dentro do conjunto de conversações associado ao período em análise.

Diferentemente, Cortez et al. (CORTEZ; OLIVEIRA; FERREIRA, 2016) modelam a interação entre usuários em grafo, considerando os *retweets*, compartilhamentos e *replies*. Assim, o objetivo é medir a influência dos usuários no Stocktwits, partindo do pressuposto que usuários mais influentes são mais úteis para criar modelos de previsão no mercado financeiro.

Também há pesquisas que focam em entender o comportamento dos ativos e, por isso, se assemelham a este trabalho. Por exemplo, Alves et al. (ALVES et al., 2018) investigam o porquê de empresas distintas serem citadas juntas no Stocktwits e avaliam como se pode utilizar essa informação para auxiliar na tomada de decisão no mercado de ações. Especificamente, é proposta uma modelagem em grafo na qual os ativos são os nós e as arestas se formam quando os ativos são citados juntos em uma postagem. Em seguida, é feita uma análise das principais características e pares isolados da rede. Os resultados mostram que ativos citados juntos estão correlacionados com as variações no mercado financeiro. Tais resultados podem

¹ Seeking Alpha: <https://seekingalpha.com/>

servir de suporte à decisão para investimento.

Fundamentado na Teoria das Ondas de *Elliott* e associando às estratégias de compra e venda de ações, utilizando análise técnica, correlacionou-se a dois movimentos de mercado (Rompimento e Correção). Jabbur et al. (JABBUR et al., 2015) implementaram 32 estratégias de negociação em um simulador realístico do mercado financeiro de alta frequência, para tentar antecipar certas fases e ciclos da Bolsa de Valores. Foram avaliados 5 códigos de ativos (BBDC4, BBAS3, USIM5, VALE5 e um índice), no período de setembro a outubro de 2013, com dados de granularidade de 15 minutos, com o objetivo de diminuir a quantidade de gatilhos executados com a eliminação de possíveis ruídos. Assim, associou-se indicadores financeiros de volume (i.e., MFI, OBV e SMA) a dados coletados de redes sociais e da *Web*, para verificar se o que é comentado nestes meios de comunicação, em termos de volume de dados, pode ou não influenciar no preço de um determinado código de ativo.

Mesmo utilizando uma mídia social diferente, (SANTOS, 2016) também estuda o mercado brasileiro de ações, baseando-se em uma ampla caracterização e análise de dados do *Twitter*, mostrando que os eventos e notícias sobre o mercado de ações são capazes de gerar picos de publicação pelos usuários do *Twitter* e que a frequência de postagem segue o início do pregão de negociação e se mantém por aproximadamente três horas após o encerramento do mercado de ações. A avaliação desta pesquisa utilizou 13 meses de postagens no *Twitter*, fornecendo uma base de padrões comportamentais, para a elaboração de indicadores sociais de apoio à decisão.

Um método supervisionado, para classificação de polaridade baseado em múltiplos modelos com o intuito de classificar documentos financeiros com múltiplas entidades, foi proposto em Silva et al. (SILVA et al., 2014). Foi estudado estratégias de segmentação em texto que usam heurísticas de casamento de *string* e resolução de anáfora, além de proporem um método de classificação hierárquica baseada em detecção de subjetividade para o mercado financeiro.

Assim, o desafio de caracterizar, compreender e prever o comportamento de investidores e, conseqüentemente, a direção do fluxo das ações em bolsa de valores através de dados disponíveis na *Web*, é tema de muitos estudos citados anteriormente. Portanto, percebe-se que a utilização dos dados coletados da plataforma *Stocktwits* (BAR-HAIM et al., 2011b; ALVES et al., 2018; OH; SHENG, 2011; FERREIRA, 2016; OLIVEIRA; CORTEZ; AREAL, 2013b; WANG et al., 2015; TU et al., 2016), ainda com muito espaço para ser explorada, apresenta um grande potencial para análise de padrões e comportamentos relacionados ao mercado financeiro.

Esse estudo diferencia-se de (ATKINS; NIRANJAN; GERDING, 2018; OH; SHENG, 2011; ALVES et al., 2018; JABBUR et al., 2015; LI; SHAH, 2017; OLIVEIRA; CORTEZ; AREAL, 2013b; WANG et al., 2015) por realizar uma caracterização e análise estática, temporal e de correlação de mensagens postadas por usuários do *Stocktwits* que permitem identificar os

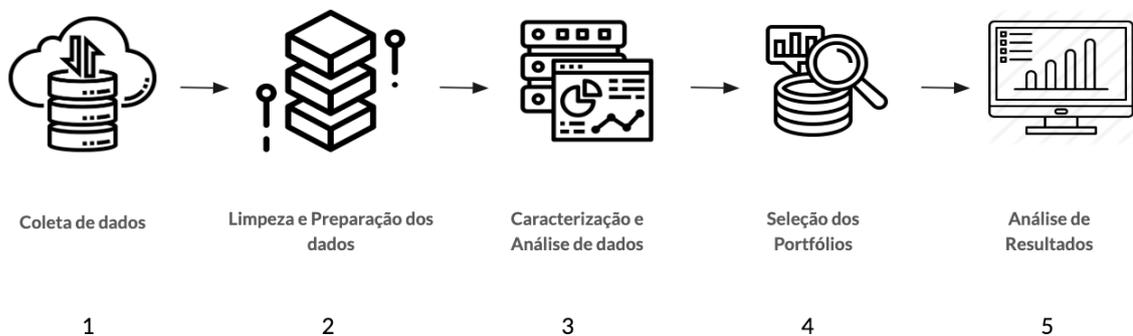
ativos mais citados, o volume de mensagens por mês e a correlação das mensagens com preço e volume dos ativos.

Dessa forma, esta dissertação propõe uma metodologia que aborda uma caracterização dos dados da rede social em diferentes níveis: desde o levantamento de dados estáticos e de correlação dos dados das postagens. Foram acrescentados também dados de séries temporais de cotações históricas com o objetivo de identificar padrões para a seleção de carteiras de ações com o intuito de auxiliar investidores na tomada de decisões.

4 Metodologia

A metodologia utilizada neste trabalho está dividida em cinco etapas: Coleta de Dados, Limpeza e Preparação dos Dados, Caracterização e Análise de Dados (Estática, Temporal e de Correlação), Seleção dos Portfólios e definição dos pesos dos ativos e Análise de Resultados, conforme Figura 4. As etapas de coleta, limpeza e preparação dos dados são discutidas na mesma seção por serem relacionadas.

Figura 4 – Etapas a serem seguidas na metodologia do trabalho.



Fonte: Elaborada pelo autor.

4.1 Coleta de Dados

Coleta de dados para pesquisa é o processo de recolhimento de dados para uso secundário por meio de técnicas específicas. A classificação dos dados é dividida em três tipos: contínua, periódica e ocasional.

- **Coleta de dados contínua:** os eventos são registrados à medida que acontecem, durante um determinado período de tempo.
- **Coleta de dados periódica:** são eventos de ciclo em ciclo, por exemplo, as pesquisas do censo realizadas no Brasil.
- **Coleta de dados ocasional:** são as coletas realizadas sem determinação de continuidade ou período.

Os dados utilizados neste estudo são obtidos a partir de postagens da rede social Stocktwits. Esses dados foram disponibilizados pela instituição *Stocktwits* após contato prévio e autorização. Após a disponibilização, realizamos a coleta dos dados mês a mês do período de 2015 a 2018.

Além dos dados da rede social *Stocktwits*, também foram coletados dados dos ativos das bolsas norte-americanas via plataforma do *Yahoo Finance*. Os dados coletados são arquivos

de cotações históricas com dados agrupados por intervalo de tempo diário, ou seja, coleta de dados contínuos. Os atributos desses dados são código do ativo, preços (abertura, máximo, mínimo, fechamento anterior, fechamento atual, fechamento ajustado), quantidade de negócios e volume financeiro das ações negociadas nas bolsas norte-americanas.

4.2 Limpeza e Preparação dos Dados

Frequentemente, os dados são encontrados com diversas inconsistências: registros incompletos, valores errados e dados inconsistentes. A etapa de limpeza e preparação dos dados visa eliminar estes problemas de modo que eles não influam no resultado dos algoritmos usados. As técnicas usadas nesta etapa vão desde a remoção do registro com problemas ou atribuição de valores padrões. Foram realizados a exclusão ou substituição de valores duplicados para evitar redundâncias na base de dados que poderiam gerar algum conflito na análise e a identificação, exclusão e tratamento de valores nulos ou inconsistentes.

Após a etapa de Coleta de dados da rede social, dentre os quarenta atributos existentes na base de dados, extraiu-se os atributos mais relevantes para a pesquisa que estavam valorados, pois vários atributos não continham valores que poderiam nos auxiliar em nossas análises. As características das postagens são: *username*, *name*, *tweets*, *date*, *stock*, *sentiment*, *URLs*, *asset hashtags*, *replies*, *reshares*, *comments*, *post feeling* e *likes*.

Após a Coleta de dados dos ativos da bolsa de valores, foram extraídos apenas os atributos relevantes para este trabalho, são eles: *asset code*, *adjusted closure*, *amount of business* e *financial volume*.

Ademais, para a limpeza e preparação dos dados, foram desenvolvidos *scripts* na linguagem de programação *Python*, que extrai das mensagens na base de dados as características listadas, além de retirar alguns registros de mensagens postadas contendo atributos com valores ausentes.

4.3 Caracterização e Análise de Dados

Na etapa de caracterização e análise de dados são utilizados três conceitos de avaliação: análise estática, temporal e de correlação. Essa etapa consiste em realizar um levantamento dos atributos *URLs*, *asset hashtags*, *replies*, *comments*, *sentiment* e *likes* com o propósito de entender se o que é comentado no Stocktwits com foco em investimentos pode auxiliar na implementação de uma seleção de portfólios para o mercado financeiro.

4.3.1 Análise Estática

Nesta primeira etapa são construídos algoritmos para extrair atributos estáticos dos *tweets* em nossa base de dados.

Para explicar a ocorrência de alguns eventos, realizou-se uma análise estática de mensagens disponibilizadas pela *Stocktwits*, contabilizando e consolidando as seguintes informações: estatística básica dos dados coletados de mensagens, mensagens que mencionam ativos, sentimento identificado pela marcação realizada pelo usuário na mensagem como *Bullish* (Positivo) e *Bearish* (Negativo), e distribuição de mensagens por ativos.

4.3.2 Análise Temporal

A segunda etapa da construção da nossa visão é a análise temporal. Conforme mencionado anteriormente, nessa etapa estudamos a evolução dos atributos dos *tweets* ao longo do tempo.

Para identificar padrões de comportamento dos usuários, realizou-se uma análise temporal dos dados de mensagens disponibilizados pela *Stocktwits* com as informações sobre a quantidade de mensagens postadas por mês e a quantidade de mensagens postadas por sentimento, para os anos de 2015, 2016, 2017 e 2018.

4.3.3 Análise de Correlação

A análise de correlação investiga se existe alguma relação entre o volume de mensagens postados na rede social e a oscilação das variáveis relacionadas aos preços dos ativos no mercado financeiro, como: *Open*, preço de abertura; *High*, preço máximo; *Low*, preço mínimo; *Close*, preço de fechamento e *Adj. Close*, preço de fechamento ajustado¹. Esse conhecimento é de grande valia tanto a existência de uma relação linear entre as variáveis quanto para o grau de intensidade desta.

Para representar a correlação entre as variáveis x e y para cálculo em uma amostra de $r(x,y)$ utiliza-se a Equação 12:

$$r(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y} \quad (12)$$

Como uma medida da força da associação entre o número de publicações diárias de alguns ativos (x_i) e o volume diário de negociação destes (y_i) em um conjunto de dados de

¹ O preço de fechamento ajustado corrige o preço de fechamento de uma ação para refletir o valor dessa ação após a contabilização de quaisquer ações corporativas. Geralmente é usado ao examinar retornos históricos ou fazer uma análise detalhada do desempenho anterior.

preços no período (N), foi utilizado o coeficiente de *Spearman*, conforme a Equação 13:

$$\rho = 1 - \frac{6 \sum (x_i - y_i)^2}{N^3 - N} \quad (13)$$

O cálculo do coeficiente de correlação de *Spearman* é não paramétrico. O coeficiente avalia quão bem uma função monotônica arbitrária pode descrever a relação entre duas variáveis, sem fazer quaisquer outras suposições sobre a natureza particular da relação entre as duas variáveis, conforme (DIAS; JUNIOR; REIS, 2017), por isso, tal coeficiente foi escolhido. Além disso, em trabalhos futuros planejamos considerar os coeficientes de correlação de *Pearson* (BENESTY et al., 2009) e (ABDI, 2007).

Para calcular o coeficiente de *Spearman*, é considerado o número de mensagens ocorridas em todos os dias úteis de movimentação no mercado financeiro e o volume negociado diariamente. Como exemplo serão apresentadas as análises de correlação dos seguintes ativos: CYTR (CytRx Corporation) e TSLA (Tesla Motors).

O coeficiente de correlação de Spearman entre duas variáveis X e Y considera uma escala contínua de -1 a 1, em que -1 e 1 são valores de limite em que as duas variáveis podem ser interpretadas como perfeitamente correlacionadas. Assim, se X ou Y assumem -1 elas são negativamente correlacionadas, ou seja, caso o valor de X cresça, há uma tendência de que o valor de Y diminua e vice-versa. Por outro lado, caso X e Y assumam 1 como valor para o coeficiente de correlação, diz-se que elas são positivamente correlacionadas e, com isso, caso o valor de X cresça, há uma tendência de que o valor de Y também cresça.

A escala de Cohen, por sua vez, é utilizada para avaliar o coeficiente de correlação de forma a determinar a força do relacionamento, ou o tamanho do efeito, onde o valor do coeficiente estando entre 0,10 e 0,29 representa uma pequena associação, entre 0,30 e 0,49 representa uma associação média e acima de 0,50 representa uma associação maior ou mesmo um relacionamento. A Tabela 2.3 descreve em detalhe a escala de Cohen.

Com base neste conceito, definimos uma escala de correlação ilustrada na Tabela 2 com o objetivo de identificar a correlação como positiva, negativa ou inexistente, sem, necessariamente, considerar maiores variações na magnitude da correlação.

Tabela 2 – Escala de Cohen.

Coefficiente de Correlação	Descrição
0.0-0.1	trivial, muito baixo, insignificante
0.1-0.3	pequeno, baixo, menor
0.3-0.5	moderado, médio
0.5-0.7	elevado, alto, maior
0.7-0.9	muito elevado, muito alto
0.9-1.0	perfeito, distinto, infinito

4.4 Seleção de Portfólios

Ao selecionar o conjunto de ativos que farão parte dos Portfólios, foram implementadas três estratégias para verificar se a seleção de Portfólios utilizando dados sociais retorna ao investidor um percentual de lucro na operação financeira frente a um determinado risco.

Após a realização dos estudos da caracterização dos dados da *Stocktwits* foram desenvolvidas as seguintes estratégias:

- Estratégia 1 - (E1): Selecionar os k ativos mais citados na rede nos últimos 6 meses de observação e realizar testes e avaliação de desempenho nos próximos 6 meses.
 - Questão de pesquisa a validar: se os 5 ativos mais citados apresentam um retorno financeiro satisfatório.
 - Para identificar os ativos mais citados na rede social no período avaliado foi utilizado a técnica de Mineração de Itens Frequentes (*Frequent Itemset Mining*), introduzido por (AGRAWAL; IMIELINSKI; SWAMI, 1993). Foi utilizada a primeira etapa dessa técnica que consiste em criar um conjunto de conjunto de itens frequentes (*Frequent Itemset*). A partir deste conjunto, foram selecionados, com base na frequência absoluta, os 5 ativos mais citados para compor o portfólio.

- Estratégia 2 - (E2): Selecionar os k ativos mais citados na rede positivamente nos últimos 6 meses de observação e realizar testes e avaliação de desempenho nos próximos 6 meses.
 - Questão de pesquisa a validar: se os 5 ativos mais citados positivamente apresentam um retorno financeiro satisfatório.
 - Para identificar os ativos mais citados positivamente utilizamos um indicador social, o mesmo é composto pela quantidade de mensagens positivas e negativas postadas durante o dia na rede social, ou seja, captura uma visão diferente dos demais indicadores para análise do mercado de ações. Esse indicador chamamos de $(S - Score)^2$, que é definido pela Equação 14:

$$S - score = \frac{Bullish - Bearish}{Bullish + Bearish} \quad (14)$$

Os valores de $(Bullish)$ e $(Bearish)$ significam o número de mensagens positivas e negativas, respectivamente, que foram mencionadas em algum ativo em um determinado dia. O indicador apresenta um intervalo de variação definido entre -1 e +1. Um valor maior que 0 significa que contém mais mensagens positivas no dia, e um valor menor que 0 indica que contém mais mensagens negativas.

² Quantitative Trading - Quantitative Investment and Trading Ideas, Research and Analysis: http://epchan.blogspot.com/2017/09/Stocktwits-sentiment-analysis_7.html

- Estratégia 3 - (E3): Selecionar os k ativos mais citados de um determinado setor mais correlacionado na rede nos últimos 6 meses de observação e realizar testes e avaliação de desempenho nos próximos 6 meses.
 - Questão de pesquisa a validar: se os 5 ativos mais correlacionados de um setor mais citado apresentam um retorno financeiro satisfatório.
 - Nesta estratégia realizou-se a identificação dos ativos mais correlacionados de um setor mais citado no período analisado. Para a análise de correlação foi utilizado o coeficiente de *Spearman* já apresentado na 4.3.3.

Após seleção dos ativos será executada uma otimização dos pesos de cada ativo na composição do Portfólio utilizando a técnica de Markowitz e a técnica de Sharpe com o auxílio da Simulação de Monte Carlo. A composição de ativos escolhidos para a carteira será o que apresentar o maior valor de retorno acumulado no período analisado. Estas técnicas foram apresentadas nas Subseções 2.3.2.1, 2.3.2.2 e 2.3.3.

4.5 Análise de Resultados

A análise dos resultados é feita em quatro etapas distintas, são elas: estática, temporal, de correlação e da seleção de portfólios. Em tais análises, destacam-se as métricas utilizadas para avaliar as estratégias de seleção de portfólios propostas, são elas: retorno financeiro acumulado, volatilidade, *skewness*, *kurtosis*, *Drawdown* e o *Maximum Drawdown*.

Em geral, o retorno sobre o investimento (ROI) (KEEN; DIGRIUS, 2003) (em inglês, *return on investment ou ROI*), também chamado taxa de retorno (em inglês, *rate of return ou ROR*), taxa de lucro ou simplesmente retorno, é a relação entre a quantidade de dinheiro ganho (ou perdido) como resultado de um investimento e a quantidade de dinheiro investido.

A fórmula do ROI é apresentada na Equação 15:

$$ROI = \frac{(R - C)}{C} \cdot 100 \quad (15)$$

Onde:

ROI = taxa de retorno;

R = receita;

C = custo.

A Volatilidade, na área financeira, é uma medida de dispersão dos retornos de um título ou índice de mercado ³. Quanto mais o preço de uma ação varia em um período curto de

³ Investopedia: <http://www.investopedia.com/terms/v/volatility.asp>

tempo, maior o risco de se ganhar ou perder dinheiro negociando esta ação, e, por isso, a volatilidade é uma medida de risco.

A fórmula da Volatilidade é apresentada pela Equação 16:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (r_i - \bar{r})^2}{N - 1}} \quad (16)$$

Sendo que:

σ = Volatilidade;

N = número de observações;

\bar{r} = retorno médio;

r_i = retorno em período i .

A *skewness* ou assimetria é uma medida de simetria, ou mais precisamente, a falta de simetria. Uma distribuição, ou conjunto de dados, é simétrica se parecer igual à esquerda e à direita do ponto central.

A seguir a fórmula da *Skewness*, Equação 17:

$$Skew = \frac{1}{N} \sum_{i=1}^N \left[\frac{r_i - \bar{r}}{\sigma} \right]^3 \quad (17)$$

Onde:

$Skew$ = *Skewness*;

N = número de observações;

\bar{r} = retorno médio;

r_i = retorno em período i .

Já a métrica *kurtosis* ou curtose indica se os dados são de cauda pesada ou de cauda clara em relação a uma distribuição normal. Ou seja, conjuntos de dados com alta curtose tendem a ter caudas pesadas ou *outliers*. Conjuntos de dados com baixa curtose tendem a ter caudas leves ou falta de valores extremos. Uma distribuição uniforme seria o caso extremo ⁴.

A fórmula da *Kurtosis* é apresentada pela Equação 18:

$$Kurt = \frac{1}{N} \sum_{i=1}^N \left[\frac{r_i - \bar{r}}{\sigma} \right]^4 \quad (18)$$

Sendo que:

$Kurt$ = *Kurtosis*;

N = número de observações;

\bar{r} = retorno médio;

r_i = retorno em período i .

⁴ Engineering Statistics Handbook: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>

O indicador *Drawdown* (Equação 19)⁵ expressa quanto foi a queda do valor de um ativo em relação a sua cotação máxima. Dessa forma, o indicador pode ser utilizado para representar quanto dinheiro o investidor perdeu, em termos percentuais, em determinado ativo durante um período. Ou seja, é o percentual de queda a partir do último ponto máximo de retorno financeiro cumulativo.

$$D(T) = \frac{(V_{Max.})}{(V_{Min.})} - 1 \quad (19)$$

Sendo que:

$D(T)$ = *Drawdown Time*;

$V_{Max.}$ = valor máximo do Ativo;

$V_{Min.}$ = valor mínimo do Ativo.

Já o *Drawdown* Máximo (MDD) é a maior perda cumulativa a partir de um pico de capital alcançado até um vale posterior no tempo ao referido pico, ou seja, é o percentual máximo de queda em uma determinada série de dados.

O Máximo *Drawdown* é calculado conforme abaixo (Equação 20)⁶:

$$MDD(T) = \frac{(P - L)}{(P)} \quad (20)$$

Onde:

$MDD(T)$ = *Maximum Drawdown Time*;

P = valor de pico antes da maior queda;

L = menor valor antes do novo máximo estabelecido.

4.6 Considerações Importantes

Neste capítulo, foram descritas as etapas a serem realizadas neste trabalho de pesquisa, são elas: a coleta dos conjuntos de dados, a etapa de preparação de dados, a caracterização e análise de dados, a seleção dos portfólios e os pesos de ativos na composição do portfólio e, por último, a descrição da análise de resultados.

Finalmente, a Tabela 3 sumariza as estratégias utilizadas na etapa de seleção dos portfólios, uma das principais contribuições desta dissertação.

No próximo capítulo, a metodologia será instanciada utilizando os dados reais da rede social *Stocktwits* em conjunto com os dados de ativos da Bolsa Americana de Valores. Serão

⁵ Suno - Drawdown: entenda o que esse parâmetro de volatilidade representa: <https://www.suno.com.br/artigos/drawdown/>

⁶ Robeco - The Investment Engineers - The Formula: Maximum drawdown: <https://www.robeco.com/en/insights/2018/04/the-formula-maximum-drawdown.html>

Estratégia	Descrição
Estratégia 1	selecionar os ativos mais citados
Estratégia 2	selecionar os ativos mais citados positivamente
Estratégia 3	selecionar os ativos mais citados do setor mais correlacionado

Tabela 3 – Dados coletados da Stocktwits Inc.

executados experimentos correspondentes a cada etapa e os resultados encontrados serão analisados por meio da comparação com *baselines* do mercado financeiro de ações.

5 Resultados

Este capítulo apresenta os resultados da caracterização e análise dos dados através das análises estática, temporal e de correlação das mensagens postadas pelos usuários na rede social de finanças *Stocktwits*. Também são apresentados os resultados obtidos para as experimentações realizadas com a seleção de portfólios utilizando os dados da rede social.

5.1 Coleta de Dados

Os dados utilizados neste trabalho consistem nas séries históricas de postagens da rede social *Stocktwits* e de preços e volumes de ações negociadas nas Bolsas de Valores norte-americanas no período de 1 de janeiro de 2016 até 31 de dezembro de 2019, com periodicidade diária de coleta.

A Tabela 4 ilustra o conjunto de dados obtido por meio do processo de coleta disponibilizado pela empresa *Stocktwits Inc.* Por exemplo, o usuário com *username* *nvoky* publicou um *tweet* na data de 01/01/2017 utilizando uma tag para o ativo *HRTX* e obteve nove respostas e seis compartilhamentos.

<i>username</i>	<i>tweet</i>	<i>date</i>	<i>stock</i>	<i>replies</i>	<i>reshares</i>
<i>nvoky</i>	@gotgains @michaelmcl Not guaranteed ...	2017-01-01T00:00:00Z	<i>HRTX</i>	9.0	6.0
<i>Candigit</i>	@J457 shale supply is overrated. Also ...	2017-01-01T00:00:15Z	NaN	1.0	2.0
...
...
<i>Fixafoot9</i>	@rodneystanger Neovasc had a tough ...	2018-12-31T23:59:18Z	<i>AAPL</i>	3.0	0.0
<i>HeatSeeker82</i>	@Dig4Gold Bitcoin ... beast #Beastmode	2018-12-31T23:59:32Z	<i>FB</i>	12.0	8.0

Tabela 4 – Dados coletados da *Stocktwits Inc.*

Já a Tabela 5 apresenta e ilustra o conjunto de dados referentes aos ativos de Bolsas de Valores Norte-Americanas que foram coletados para subsidiar as análises frente aos índices de renda variável e de renda fixa. Cada entrada está no seguinte formato:

- *Ticker*: código do ativo.
- *Date*: data de negociação.
- *High*: preço máximo negociado.
- *Low*: preço mínimo negociado.
- *Open*: preço de abertura .
- *Close*: preço de fechamento.
- *Volume*: volume financeiro negociado.
- *Adj. Close*: preço de fechamento ajustado.

<i>ticker</i>	<i>date</i>	<i>high</i>	<i>low</i>	<i>open</i>	<i>close</i>	<i>volume</i>	<i>adj. close</i>
AMD	2017-01-03	11.6500000	11.020000	11.420000	11.430000	55182000.0	11.4300000
...
JNUG	2018-11-15	34.849998	33.299999	33.500000	34.599998	1853100.0	34.359299
...
SPY	2018-12-31	250.190002	247.470001	249.559998	249.919998	144299400.0	246.481415

Tabela 5 – Dados coletados das Bolsas de Valores Americanas.

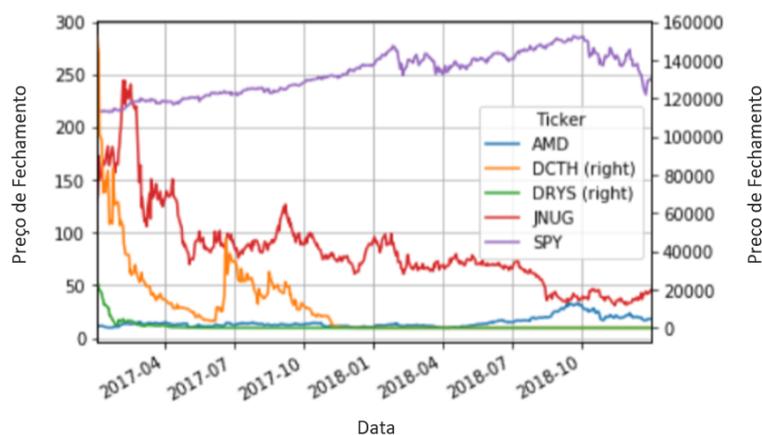
5.2 Limpeza e Preparação de Dados

Após a coleta de dados, foi realizado um processo de limpeza das bases de dados. A Tabela 6 ilustra a base de dados resultante com informações da rede social e as Figuras 5 e 6 mostram o comportamento do preço de fechamento (*Close*) e do log-retorno das ações AMD, DCTH, DRYS, JNUG, SPY, de diferentes setores do mercado, entre 2017 e 2018, como exemplo.

<i>date</i>	<i>tweet</i>	<i>stock</i>	<i>replies</i>	<i>comment</i>	<i>sentiment</i>	<i>likes</i>	<i>urls</i>
2017-01-01	@gotgains Not guara...	HRTH	9.0	@gotgains what calls...	Bullish	5.0	sec.gov/Archives/edgar/data...
...
...
2018-12-31	\$FB sell shares now before...	FB	8.0	@MLR pump...	Bearish	6.0	chartmill.com/stock/quote/F...

Tabela 6 – Dados preparados - Stocktwits Inc.

Figura 5 – Valores de preço de fechamento em (US\$) dos ativos AMD, DCTH, DRYS, JNUG, SPY entre 2017 e 2018.

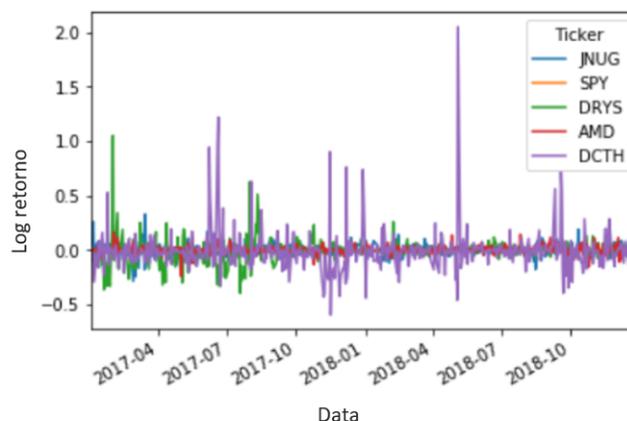


Fonte: Elaborado pelo autor.

Os picos representam a valorização do preço de fechamento do ativo e os vales representam uma desvalorização. Portanto podemos observar períodos de tendência de alta e de baixa. E períodos de tendência lateral como é o caso dos ativos SPY e AMD, como podemos observar na Figura 5.

Todas as séries de preços, nas Figuras 5 e 6, correspondentes aos das Bolsas Norte-Americanas foram tratadas e normalizadas, obtendo-se ao final as séries consistentes relativas

Figura 6 – Valores de log-retorno em percentual (%) dos ativos AMD, DCTH, DRYS, JNUG, SPY entre 2017 e 2018.



Fonte: Elaborado pelo autor.

aos log-retornos financeiros a partir do preço de fechamento ajustado *Adj. Close*. Pode-se notar a diferença na escala dos gráficos para os valores de preço de fechamento (Figura 5) e do log-retorno financeiro (Figura 6).

5.3 Caracterização e Análise de Dados

Esta seção apresenta os resultados para da análise estática (Seção 5.3.1), análise temporal (Seção 5.3.2) e análise de correlação (Seção 5.3.3).

5.3.1 Análise Estática

A plataforma *Stocktwits* permite que os usuários marquem a mensagem como *Bullish* (Positivo) e *Bearish* (Negativo). De fato, a Tabela 7 mostra que 64,08% das mensagens fazem menção direta a um determinado código de ativo e, do total das mensagens postadas, 27,17% são marcadas com um sentimento do usuário em relação à postagem em questão.

Tabela 7 – Informações dos dados coletados das mensagens publicadas na *Stocktwits*

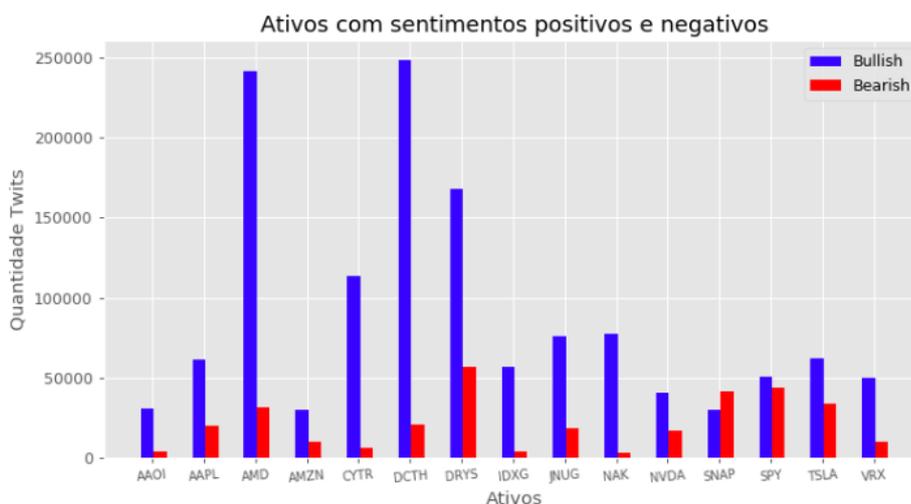
Descrição	Quantidade	Percentual
Mensagens publicadas entre 2015 e 2018	124.664.797	100%
Mensagens que mencionam ativos	79.885.202	64,08%
Mensagens que mencionam sentimentos	33.865.603	27,17%

Considerando tais mensagens, em 2017, por exemplo, os seguintes códigos de ativos são apresentados na Figura 7: AAOI, AAPL, AMD, AMZN, CYTR, DCTH, DRYS, IDXG, JNUG, NAK, NVDA, SNAP, SPY, TSLA, VRX.

É possível notar que a maioria os ativos receberam mais mensagens positivas (*Bullish*) do que mensagens negativas (*Bearish*). Isso significa que naquele momento o sentimento de

positividade em relação a esses ativos estava em alta. O destaque é para o ativo SNAP que obteve mais mensagens negativas do que positivas, que é justificado pelo fato do prejuízo líquido no ano de 2017 que incluiu 2,6 bilhões de dólares de despesas com remuneração baseada em ações¹.

Figura 7 – Distribuição de ativos com sentimentos positivos e negativos



Fonte: Elaborada pelo autor.

Ademais, conforme Figura 7, há ativos com picos de mensagens com sentimento positivo, é o caso das empresas AMD, CYTR, DCTH e DRYS. Em relação a AMD, o pico de mensagens positivas acompanhou os resultados positivos da companhia no quarto trimestre. Os resultados vieram mais fortes do que o esperado e isso incluiu um enorme salto de 60% em receita em seus negócios de computação e negócios gráficos, devido aos novos chips Ryzen e Radeon em desktops, laptops e Macs².

Além disso, o otimismo dos usuários na rede foi devido ao interesse em novos chips e placas gráficas por pessoas que queriam minerar criptomoedas³. Já para empresa CytRx (CYTR), segundo o CEO (*Chief Executive Officer*), 2017 foi crucial para CytRx com a concessão estratégica de medicamentos para desenvolvimento e comercialização⁴.

Por outro lado, o otimismo na rede social para a empresa Delcath (DCTH) se dá pelas mudanças e resultados positivos na empresa durante 2017. A DCTH aumentou as vendas em 35% se comparado com o ano anterior, devido a mudanças nas diretrizes internas da empresa

¹ Snap Investor: <https://investor.snap.com/news-releases/2018/02-06-2018-211639653>

² Marketwatch: <https://www.marketwatch.com/story/amd-earnings-cap-a-stunning-2017-but-there-is-still-plenty-to-do-2018-01-30>

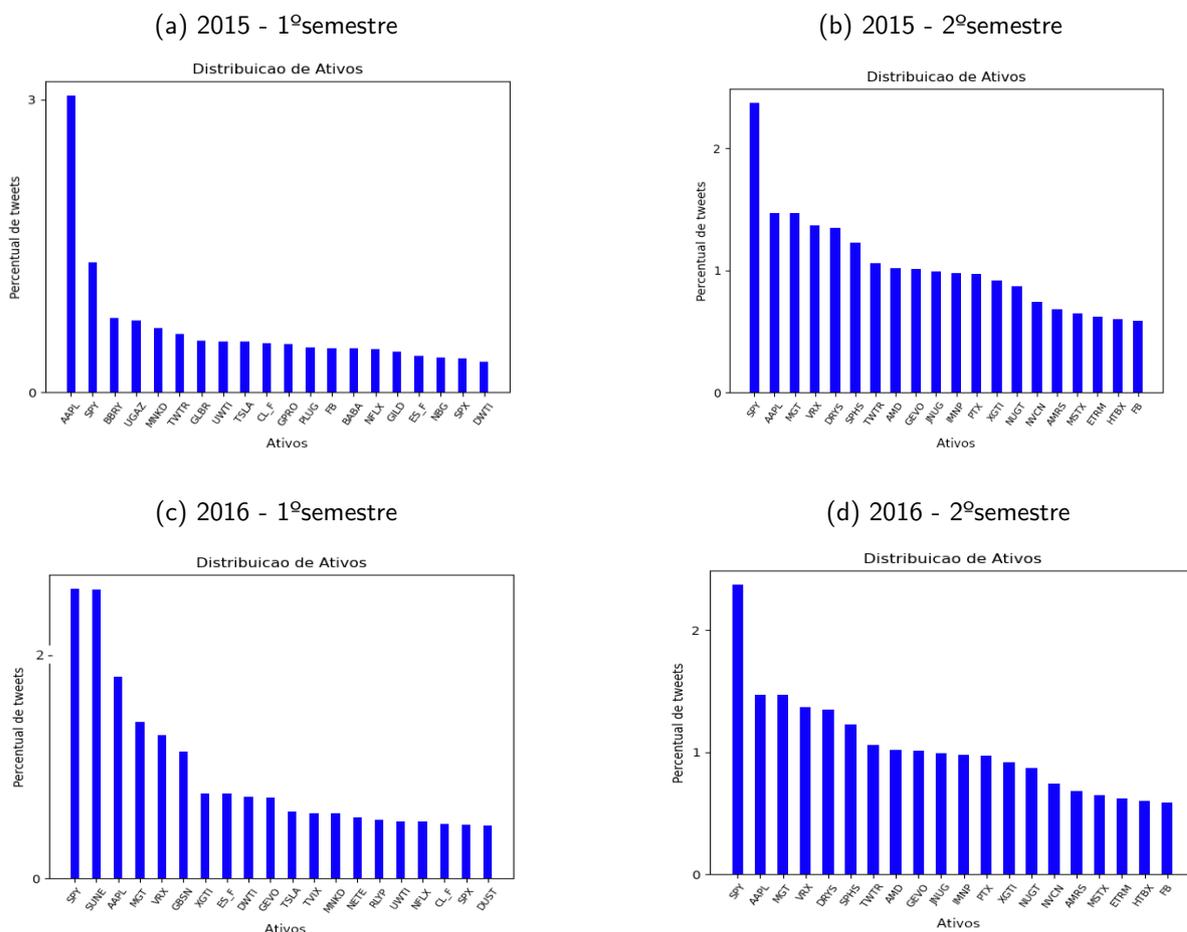
³ Fool: <https://www.fool.com/investing/2018/07/09/why-advanced-micro-devices-stock-gained-59-so-far.aspx>

⁴ Marketchameleon: <https://marketchameleon.com/PressReleases/i/594238/CYTR/cytrx-reports-2017-financial-results>

⁵. Por fim, para a empresa DryShips (Drys), os resultados da companhia mostram aumento no lucro líquido e linhas de crédito ⁶.

Com essa análise, observa-se que existem empresas com investidores muito mais otimistas que outras. Isso pode ter alguma relação com notícias que são tráfegadas na rede.

Figura 8 – Distribuição de ativos por percentual de tweets - 2015 a 2016



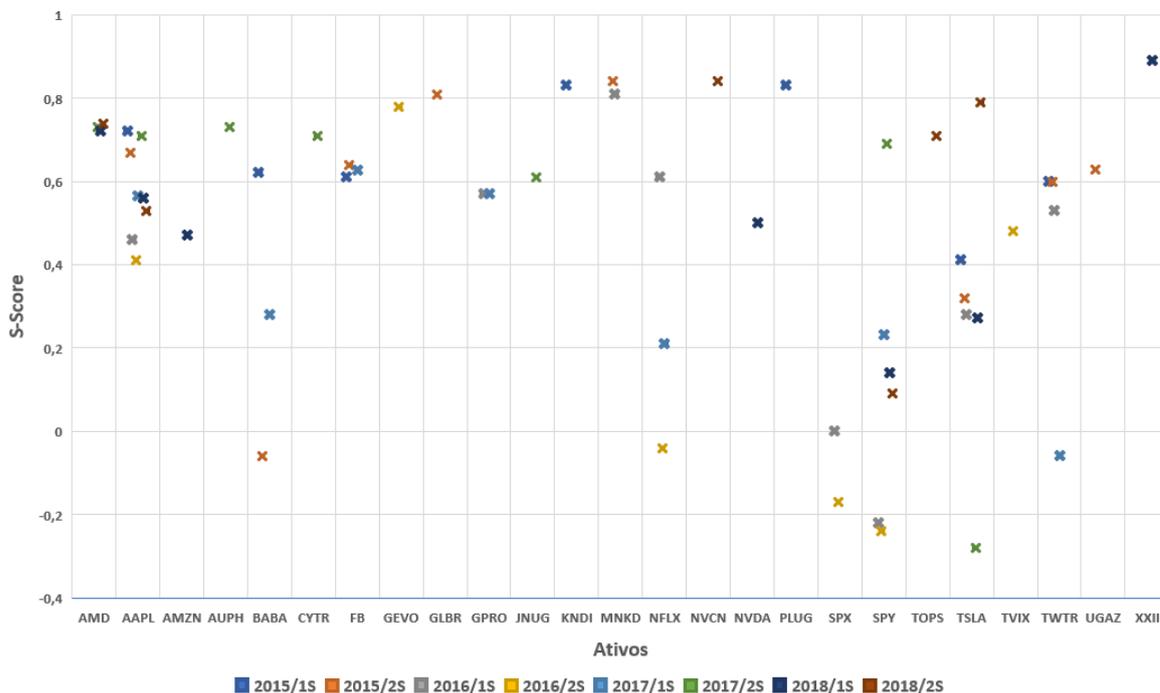
Em complemento a esta análise, as Figuras 8 e 9, segmentadas por semestres, demonstra os ativos mais citados no período entre o ano de 2015 e o ano de 2018 em uma escala percentual aos *tweets* publicados na rede social e que serão utilizados como base nos estudos da Seção 5.4.1.

A Figura 10 representa os ativos mais citados e positivos durante os anos de 2015 e 2018 e que foram base de estudos para a execução da Estratégia 2 - E2, na Seção 5.4.2. Verifica-se que ativos mais positivos da rede neste período apresentaram, $S - Score > 0,80$ em algum semestre. Estes ativos foram: GLBR *Global Brokerage Inc*, KNDI *Kandi Technologies Group Inc*, MNKD *MannKind Corporation*, NVCN *Neovasc Inc*, PLUG *Plug Power Inc*, XXI

⁵ Globenewswire: <https://www.globenewswire.com/news-release/2018/03/16/1441408/0/en/Delcath-Announces-2017-Financial-Results.html>

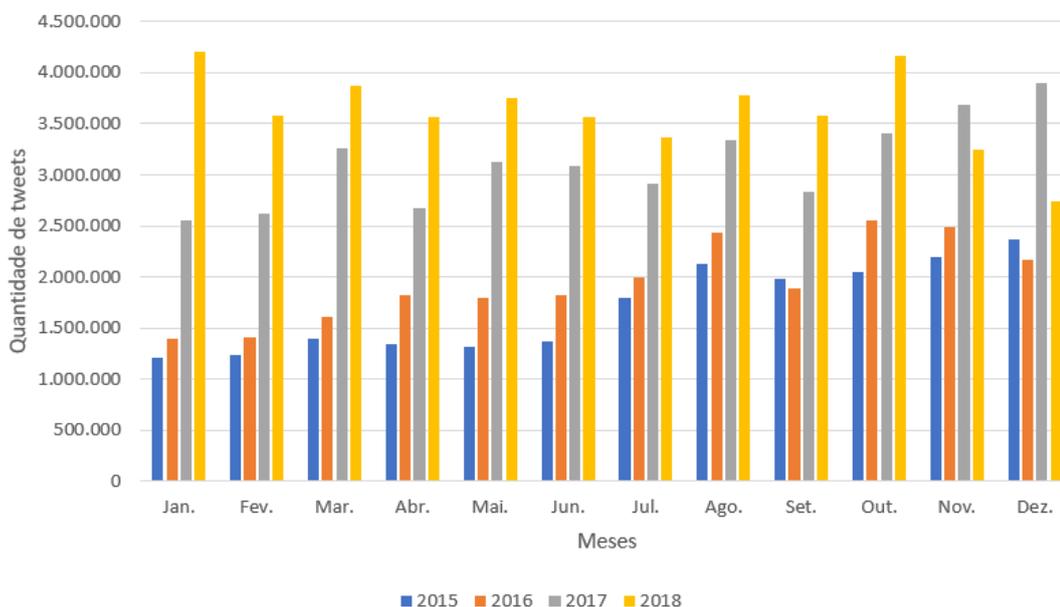
⁶ CNBC: <https://www.cnbc.com/2018/02/27/globe-newswire-dryships-inc-reports-financial-and-operating-results-for-the-fourth-quarter-of-2017.htm>

Figura 10 – Distribuição do índice S-Score dentre os ativos mais citados



Fonte: Elaborada pelo autor.

Figura 11 – Distribuição de tweets publicados

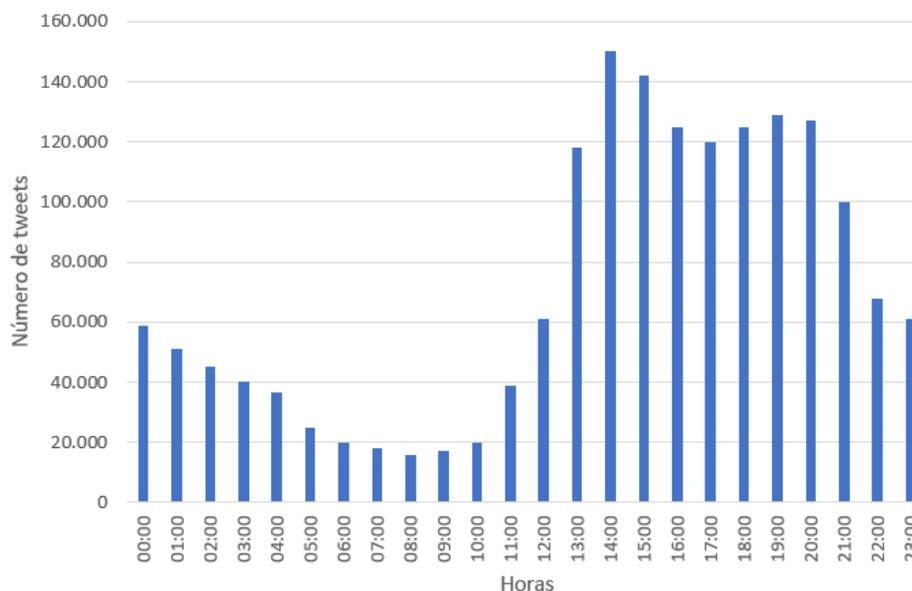


Fonte: Elaborada pelo autor.

5.3.3 Análise de Correlação

A fim de validar se a rede social tem alguma correlação com o mercado, foram coletados dados de preços dos 20 ativos com maior presença na rede. Por exemplo, dentre os ativos citados, foram escolhidos ativos de diferentes setores como: tecnologia, automobilístico, biofarmacêutico e índice de bolsa de valores.

Figura 12 – Tweets sumarizados por hora do dia



Fonte: Elaborada pelo autor.

São calculados o número de mensagens ocorridas em todos os dias do ano para os ativos com mais mensagens publicadas. Por exemplo: AAPL (*Apple*), TSLA (*Tesla*), SPY (*SPDR SP 500 ETF Trust*), CYTR (*CytRx Corporation*), TOPS (*TOP Ships*), XXII (*22nd Century Group*), AMD (*Advanced Micro Devices*) e FB (*Facebook*).

Em seguida, o coeficiente de Spearman (4.3.3) já apresentado, é calculado com o objetivo de verificar se há alguma correlação positiva ou negativa com os dados de preços, volume negociado e volume *tweets* dos ativos, com destaque para os quatro ativos já mencionados.

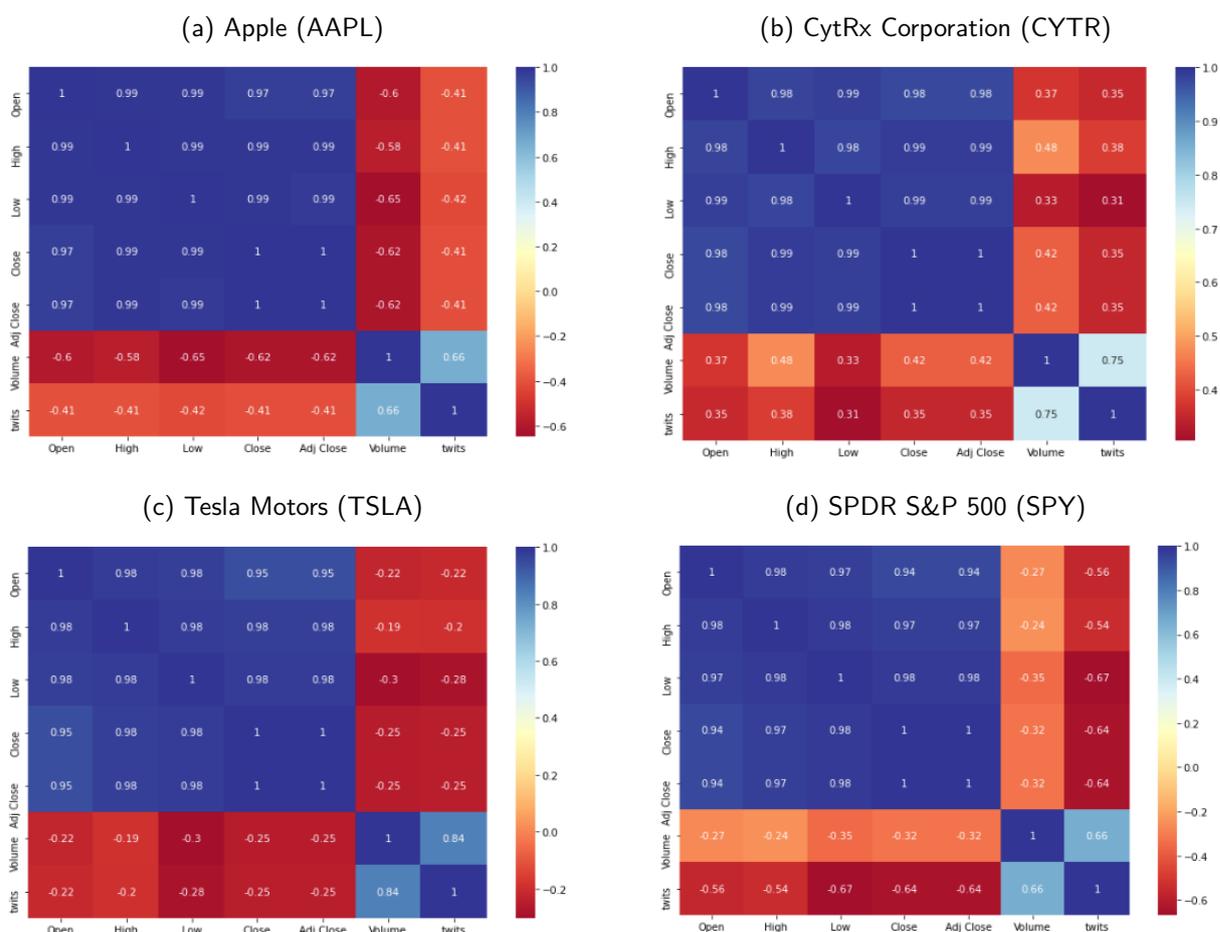
Conforme mostra a Figura 13, existe uma alta correlação positiva entre o número de mensagens diárias e o volume de negócios diários das empresas. Isso mostra que quando há uma alta volatilidade nos preços dos ativos, o número de mensagens trafegadas na rede tende a aumentar também.

Neste estudo, foram calculadas a correlação das variáveis *Open*, *High*, *Low* e *Close* e *Adj Close* com a quantidade de mensagens postadas durante o dia sobre cada ativo. A maioria dos ativos observados tem uma correlação alta (maior que 0,50), como exemplo temos as Figuras 13a, 13d e 14a ou uma correlação muito alta (maior que 0,70), que podemos verificar nas Figuras 13b, 13c, 14b, 14c e 14d, entre o número de mensagens postadas e o volume negociado, o que pode nos dizer que o número de postagens acompanha o volume de ações negociadas, visto que aplicamos estas etapas em outros códigos de ativos mais citados da rede.

Observa-se, no caso da empresa CYTR Figura 13b, que há uma correlação muito alta (0,84) do número de mensagens postadas com o volume de negociações diários do ativo após o anúncio de novas estratégias de comercialização de medicamentos.

Os ativos AAPL, TSLA e SPY nas Figuras 13a, 13c, 13d também mostram um

Figura 13 – Análise da correlação do coeficiente de Spearman.



Fonte: Elaborado pelo autor.

comportamento semelhante com uma alta correlação entre o número de mensagens postadas e o volume de negociações.

É possível observar ainda que a maioria dos ativos têm uma correlação positiva entre os preços e o número de mensagens. Isso também mostra que a rede social pode influenciar os preços dos ativos ou que alterações nos preços causam uma movimentação na rede social aumentando o volume de postagens.

Assim, é possível observar uma correlação significativa do número de mensagens com o volume das negociações dos ativos. Percebemos então que o volume de mensagens postadas acompanha o volume de negociação dos ativos. Portanto, o uso de informações extraídas dessas mensagens irá nos auxiliar na construção e seleção dos portfólios.

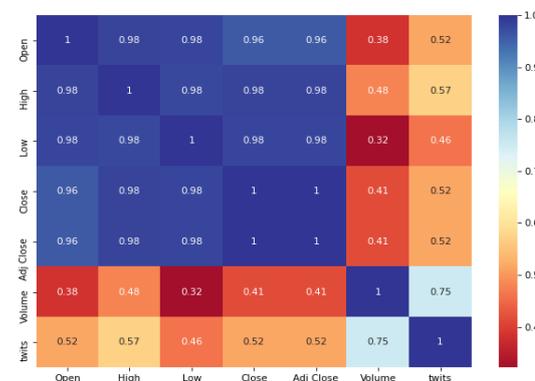
A Figura 15 apresenta os ativos mais citados e positivos da rede social no período analisado. Conforme já explicado, pela Correlação de Spearman, quanto mais próximo de 1, mais correlacionados positivamente são os ativos. Portanto, os ativos utilizados na seleção de Portfólios da 5.4.3 foram os ativos mais correlacionados quanto ao seu preço de fechamento

Figura 14 – Análise da correlação com o coeficiente de Spearman.

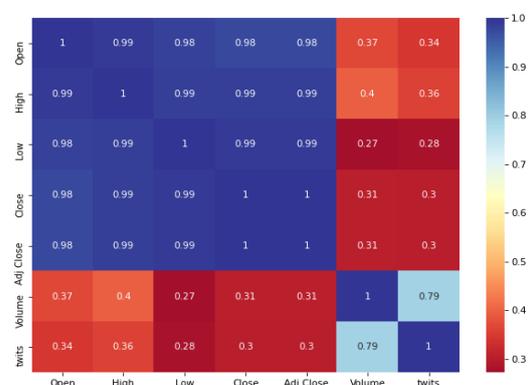
(a) TOP Ships (TOPS)



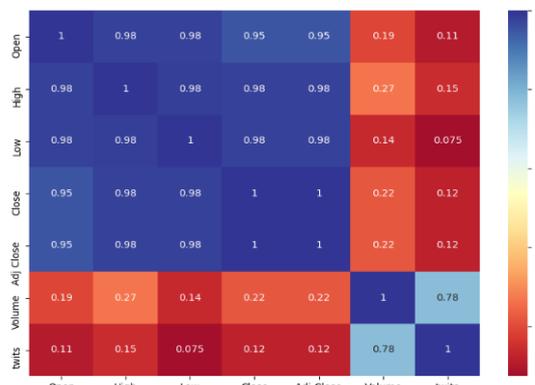
(b) 22nd Century Group (XXII)



(c) Advanced Micro Devices (AMD)



(d) Facebook (FB)



Fonte: Elaborado pelo autor.

ajustado.

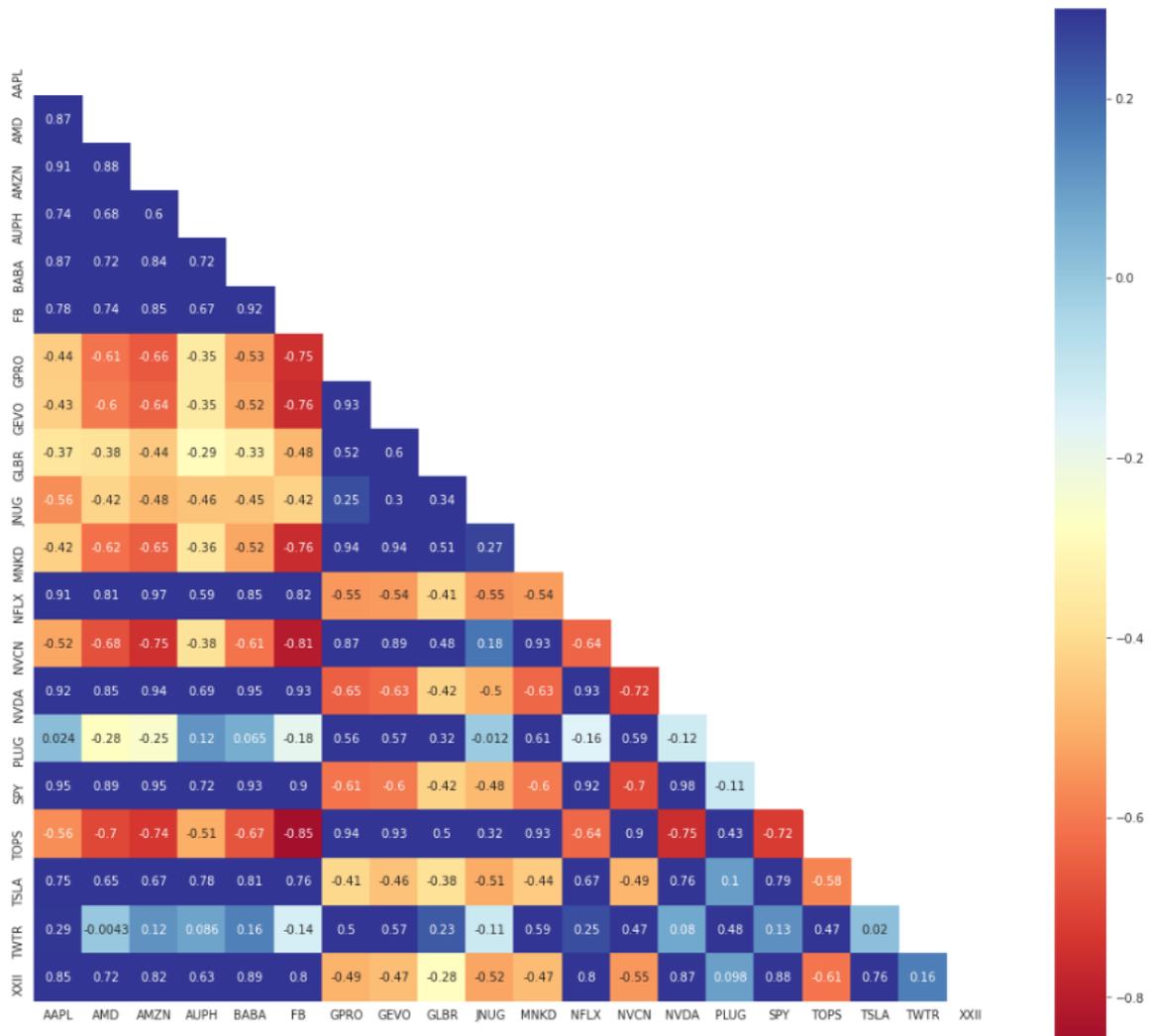
5.4 Análise da Seleção de Portfólios

Esta seção analisa os resultados dos testes realizados com a indicação dos portfólios através dos dados da *Stocktwits*. Assim, são apresentados: a composição de cada portfólio de cada estratégia para cada semestre de 2015 a 2018, o gráfico de retorno acumulado entre portfólio, índices de renda variável (S&P500), (Dow Jones) e (Nasdaq), além do índice de renda fixa, Certificado de Depósito, (CD), a avaliação de desempenho frente às métricas utilizadas e o gráfico de *Drawdown* de alguns períodos apresentados nas Tabelas a seguir.

5.4.1 Análise da Estratégia 1

A Estratégia 1 – E1 consiste em um portfólio com os 5 ativos mais citados da rede nos últimos seis meses de observação e aplicados em teste nos próximos seis meses. A Tabela

Figura 15 – Correlação entre os ativos mais citados e positivos da rede social.



Fonte: Elaborada pelo autor.

8 apresenta a composição dos Portfólios da Estratégia 1 – E1. Sendo S_n o ativo e W_{S_n} o respectivo peso do ativo na composição do portfólio.

	S_1	W_{S_1}	S_2	W_{S_2}	S_3	W_{S_3}	S_4	W_{S_4}	S_5	W_{S_5}
2015/1S	AAPL	0,287	FB	0,146	GPRO	0,309	KNDI	0,062	TSLA	0,196
2015/2S	AAPL	0,411	GLBR	0,001	SPY	0,186	TSLA	0,400	TWTR	0,002
2016/1S	AAPL	0,198	GPRO	0,006	MNKD	0,068	NFLX	0,033	SPY	0,695
2016/2S	AAPL	0,245	GEVO	0,004	SPY	0,397	TSLA	0,353	TVIX	0,001
2017/1S	AAPL	0,113	BABA	0,001	FB	0,295	NFLX	0,290	SPY	0,301
2017/2S	AAPL	0,245	GEVO	0,004	SPY	0,397	TSLA	0,353	TVIX	0,001
2018/1S	AMD	0,019	AMZN	0,268	NVDA	0,509	TSLA	0,130	XXII	0,074
2018/2S	AMD	0,101	AAPL	0,242	HMNY	0,018	TSLA	0,290	SPY	0,349

Tabela 8 – Composição dos Portfólios - E1

Verifica-se na Figura 16 o gráfico referente ao retorno acumulado diário do Portfólio indicado frente aos índices S&P500, Dow Jones, Nasdaq e ao CD para o primeiro semestre de

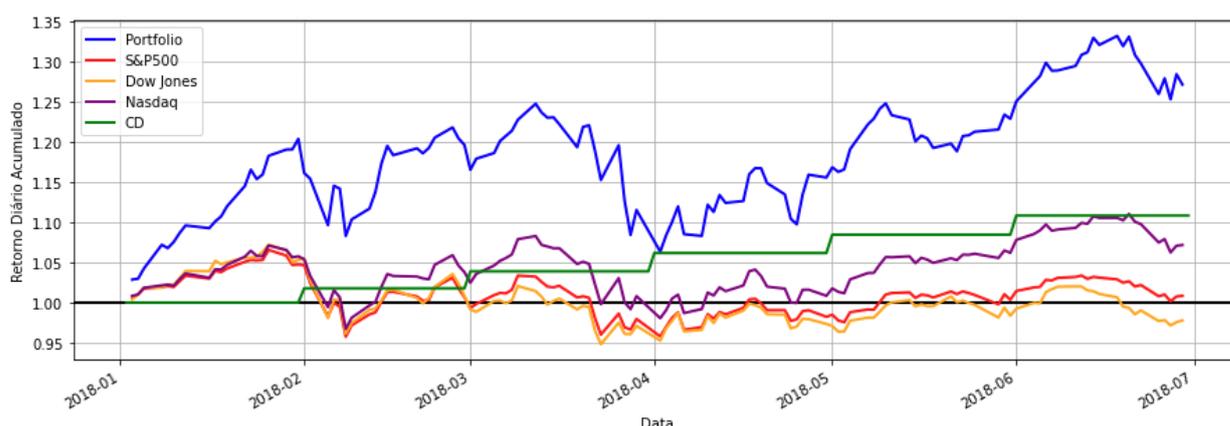
2018. Saliento que estas Figuras de retorno acumulado temos uma escala em decimal indicando em quantas vezes o retorno acumulado é maior que o investimento inicial, ou seja, para o valor 1,15, temos uma valorização de 15% e para um valor de 0,95, temos uma desvalorização de 5%.

Observa-se juntamente com a Tabela 9 e Tabela 10 que este semestre apresentou o melhor resultado, com um retorno acumulado de 22% frente a 1% do índice SP500, 3% do índice Dow Jones e 8% do índice Nasdaq . Em seis semestres do período analisado o rendimento do Portfólio foi inferior aos índices da bolsa norte-americana.

Sendo assim, conclui-se que a indicação do portfólio com base somente nos ativos mais citados pela rede social não garante um retorno consistente para o investidor.

A Estratégia 1 – E1 pode indicar ativos que estejam em baixa e que podem levar o investidor a retornos inferiores aos índices de mercado.

Figura 16 – Retorno Acumulado - E1 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2018/1S



Fonte: Elaborada pelo autor.

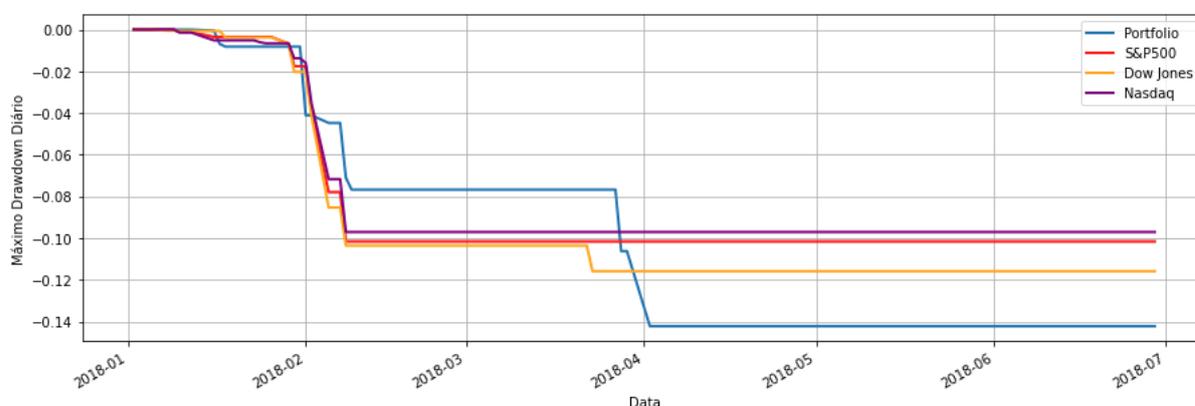
	Retorno				Volatilidade			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	0,04	0,01	-0,01	0,06	0,016	-0,975	0,008	0,008
2015/2S	-0,08	0,03	0,02	0,08	0,017	0,011	0,011	0,526
2016/1S	0,04	0,05	0,05	0,01	0,013	0,010	0,010	0,012
2016/2S	0,08	0,06	0,10	0,10	0,009	0,006	0,005	0,007
2017/1S	0,18	0,07	0,07	0,13	0,008	0,004	0,004	0,006
2017/2S	0,08	0,06	0,10	0,06	0,009	0,006	0,006	0,006
2018/1S	0,22	0,01	0,03	0,08	0,020	0,010	0,011	0,012
2018/2S	-0,09	-0,08	-0,04	-0,13	0,020	0,011	0,011	0,015

Tabela 9 – Avaliação de Desempenho – E1 – Portfólio x SP 500 x Dow Jones x Nasdaq

Na Figura 17, verifica-se através do gráfico de *Drawdown* que as perdas ao logo do primeiro semestre foi bem significativa em relação aos índices de renda variável. Saliento que as próximas Figuras de *Daily Drawdown* e *Max. Daily Drawdown* temos uma escala em decimal, ou seja, para o valor -0,10, temos uma desvalorização de 10%.

	<i>Skewness</i>				<i>Kurtosis</i>			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	-0,197	-0,071	-0,037	-0,435	0,749	-0,114	-0,156	0,283
2015/2S	-0,009	-0,187	-0,096	0,000	0,446	1,650	1,260	1,413
2016/1S	-0,130	-0,382	-0,425	-0,468	0,171	1,085	0,919	1,173
2016/2S	-0,091	-0,101	0,115	-0,146	-0,213	3,668	2,995	1,443
2017/1S	-0,996	-0,121	0,030	-1,170	3,240	2,824	2,689	3,759
2017/2S	-0,091	-0,101	0,115	-0,146	-0,213	3,668	2,995	1,443
2018/1S	-0,616	-0,990	-0,981	-0,762	1,168	2,675	2,770	1,588
2018/2S	0,483	0,066	0,105	0,004	1,151	3,677	3,251	2,501

Tabela 10 – Avaliação de Desempenho – E1 – Portfólio x SP 500 x Dow Jones x Nasdaq

Figura 17 – *Max. Daily Drawdown* – E1 – Portfólio, SP500, Dow Jones, Nasdaq e CD – 2018/1S

Fonte: Elaborada pelo autor.

A Figura 18 pode-se verificar que foi o pior desempenho do período analisado para a Estratégia 1 - E1. Observa-se que as perdas foram maiores que os índices de renda fixa, que também não tiveram uma bom desempenho ao longo do segundo semestre de 2015.

Pode-se confirmar este mau desempenho com a Figura 19 que nos revela um máximo de perda (*Drawdown*) por volta de 17.5%.

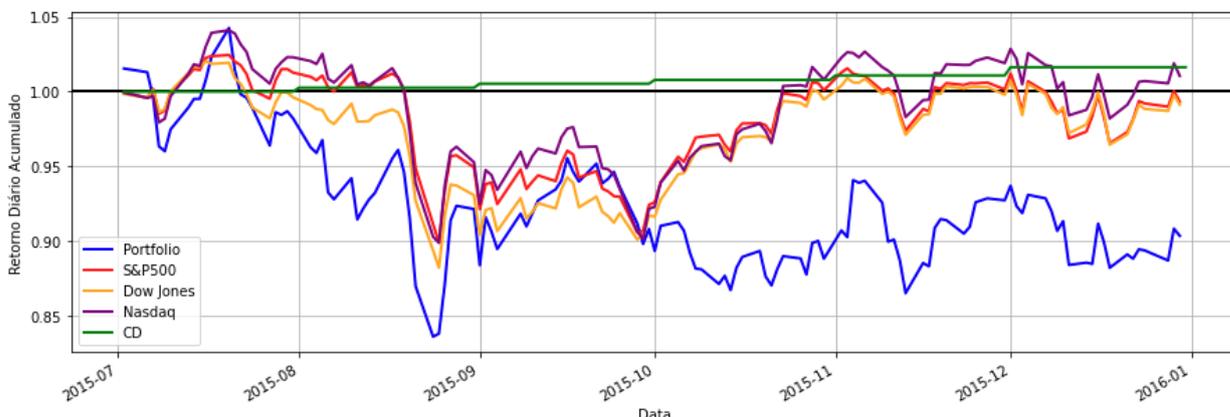
5.4.2 Análise da Estratégia 2

A Estratégia 2 – E2 consiste em um portfólio com os 5 ativos mais citados da rede positivamente nos últimos seis meses de observação nos últimos cinco anos e aplicados em teste nos próximos seis meses. A Tabela 11 apresenta a composição dos Portfólios da Estratégia 2 – E2.

Na Figura 20 é demonstrado o gráfico referente ao retorno acumulado do Portfólio indicado frente aos índices de renda variável e ao índice de renda fixa CD para o segundo semestre de 2016.

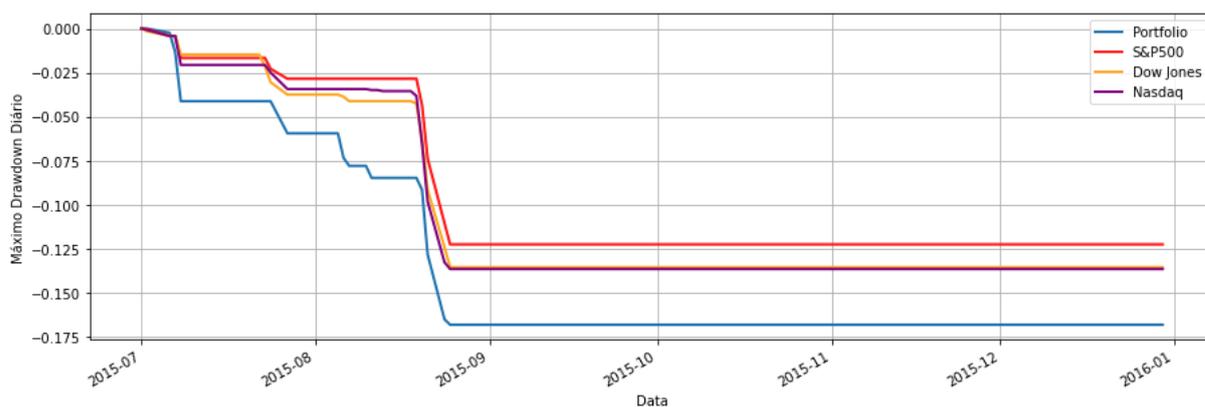
Observa-se juntamente com a Tabela 12 e com a Tabela 13 que este semestre apresentou

Figura 18 – Retorno Acumulado - E1 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2015/2S



Fonte: Elaborada pelo autor.

Figura 19 – Max. Daily Drawdown – E1 – Portfólio, SP500, Dow Jones, Nasdaq e CDD – 2015/2S



Fonte: Elaborada pelo autor.

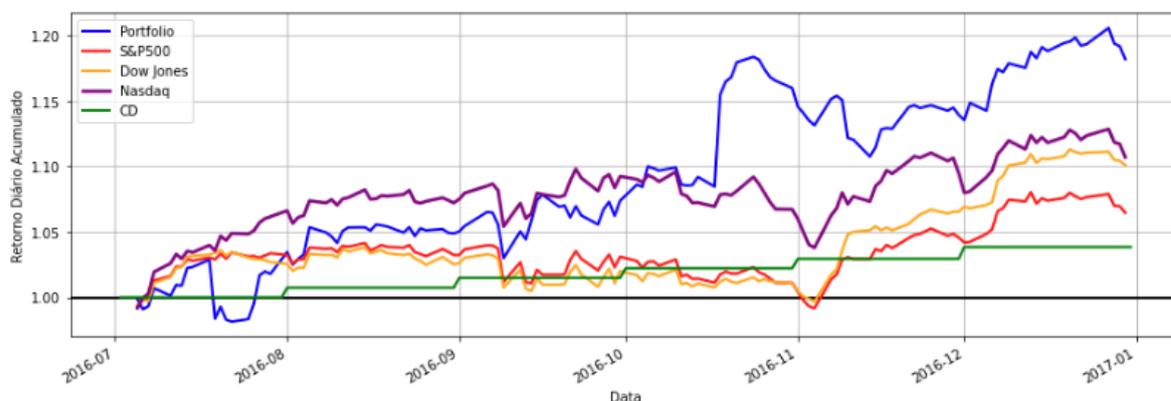
	S1	W _{S1}	S2	W _{S2}	S3	W _{S3}	S4	W _{S4}	S5	W _{S5}
2015/1S	AAPL	0,570	BABA	0,250	FB	0,146	KNDI	0,030	PLUG	0,005
2015/2S	AAPL	0,512	FB	0,433	GLBR	0,001	MNKD	0,047	UGAZ	0,007
2016/1S	AAPL	0,333	GPRO	0,001	MNKD	0,081	NFLX	0,571	TWTR	0,015
2016/2S	AAPL	0,309	GEVO	0,005	NFLX	0,331	SPY	0,349	TVIX	0,006
2017/1S	AAPL	0,104	BABA	0,012	FB	0,315	GPRO	0,004	SPY	0,565
2017/2S	AMD	0,245	AAPL	0,004	AUPH	0,397	JNUG	0,353	SPY	0,001
2018/1S	AMD	0,027	AAPL	0,110	AMZN	0,262	NVDA	0,508	XXII	0,093
2018/2S	AMD	0,002	AAPL	0,354	NVCN	0,001	TOPS	0,002	TSLA	0,642

Tabela 11 – Composição dos Portfólios - E2

o segundo melhor resultado no período analisado, com um retorno acumulado de 17% frente a 6% do índice SP500 e 10% dos índices Dow Jones e Nasdaq.

No segundo semestre de 2015 e no primeiro semestre de 2016, o rendimento do Portfólio foi inferior aos índices das bolsas norte-americana. Conclui-se que a indicação do portfólio com base nos ativos mais citados pela rede social positivamente rendeu ao final do ano 65% ao

Figura 20 – Retorno Acumulado - E2 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2016/2S



Fonte: Elaborada pelo autor.

investidor frente aos índices: 22% do SP500, 30% do Dow Jones e 38% do Nasdaq no período de quatro anos analisado .

Percebe-se então uma evolução na seleção de portfólio utilizando esta segunda estratégia, pois verificamos que houve retornos positivos em 6 semestres e apenas 2 semestres com retornos percentuais negativos.

	Retorno				Volatilidade			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	0,04	0,01	-0,01	0,06	0,011	-0,975	0,008	0,008
2015/2S	-0,03	0,03	0,02	0,02	0,016	0,011	0,011	0,012
2016/1S	-0,07	0,05	0,05	0,01	0,022	0,010	0,010	0,123
2016/2S	0,17	0,06	0,10	0,10	0,011	0,006	0,006	0,007
2017/1S	0,16	0,07	0,07	0,13	0,006	0,004	0,004	0,006
2017/2S	0,15	0,10	0,14	0,12	0,010	0,004	0,004	0,006
2018/1S	0,21	0,01	-0,01	0,08	0,019	0,010	0,011	0,018
2018/2S	0,02	-0,08	-0,04	-0,13	0,031	0,011	0,011	0,015

Tabela 12 – Avaliação de Desempenho – E2 – Portfólio x SP 500 x Dow Jones x Nasdaq

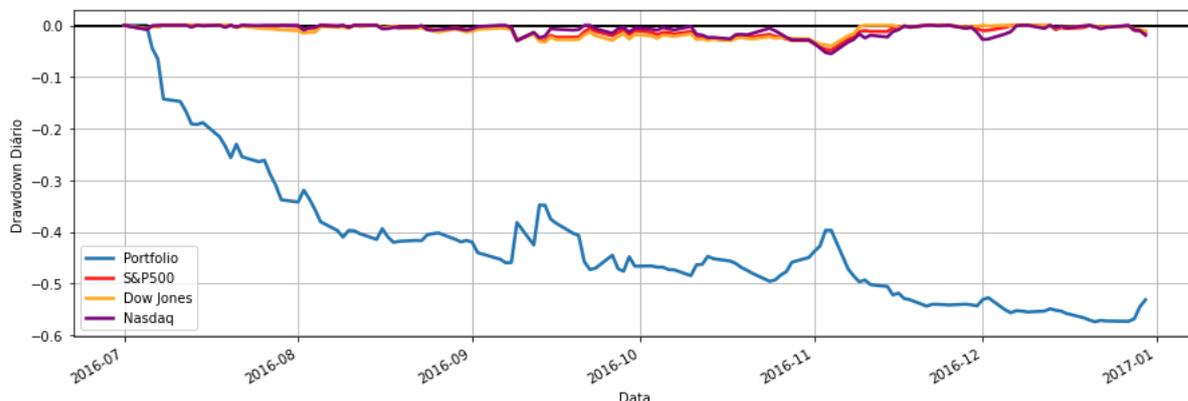
	Skewness				Kurtosis			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	-0,109	-0,071	-0,037	-0,435	0,749	-0,114	-0,156	0,283
2015/2S	-0,070	-0,187	-0,096	-0,120	0,997	1,649	1,526	1,413
2016/1S	-0,659	-0,382	-0,425	-0,468	1,175	1,081	0,919	1,173
2016/2S	1,053	-0,101	0,115	-0,146	1,082	3,667	2,995	1,443
2017/1S	-0,740	-0,361	-0,122	-1,170	3,292	2,824	2,689	3,759
2017/2S	-0,091	-0,553	-0,981	-0,221	1,411	3,352	2,004	2,741
2018/1S	-0,396	-0,990	-0,981	-0,762	1,144	2,675	2,770	1,588
2018/2S	1,020	0,066	0,105	0,0438	3,129	3,677	3,251	2,501

Tabela 13 – Avaliação de Desempenho – E2 – Portfólio x SP 500 x Dow Jones x Nasdaq

A Estratégia 2 – E2 pode nos indicar ativos que estejam em alta, ou seja, positivos, e que podem levar o investidor a retornos superiores aos índices de mercado.

Na Figura 21, verificamos através do gráfico de *Drawdown* que as perdas máximas calculadas na variação dos preços dos ativos que compõe o Portfólio foram superiores às perdas dos índices ao longo do segundo semestre de 2016, porém mesmo diante de destas perdas superiores, o portfólio conseguiu retornar um rendimento considerável ao investidor.

Figura 21 – *Daily Drawdown* – E2 – Portfólio, SP500 e CD – 2016/2S



Fonte: Elaborada pelo autor.

Já na Estratégia 2 - E2 temos como o pior desempenho o primeiro semestre de 2016, com o menor rendimento do Portfólio e com perdas máximas por volta de 27%, conforme veremos nas Figuras 22 e 23.

Figura 22 – Retorno Acumulado - E2 - Portfólio, SP500, Dow Jones, Nasdaq e CD - 2016/1S



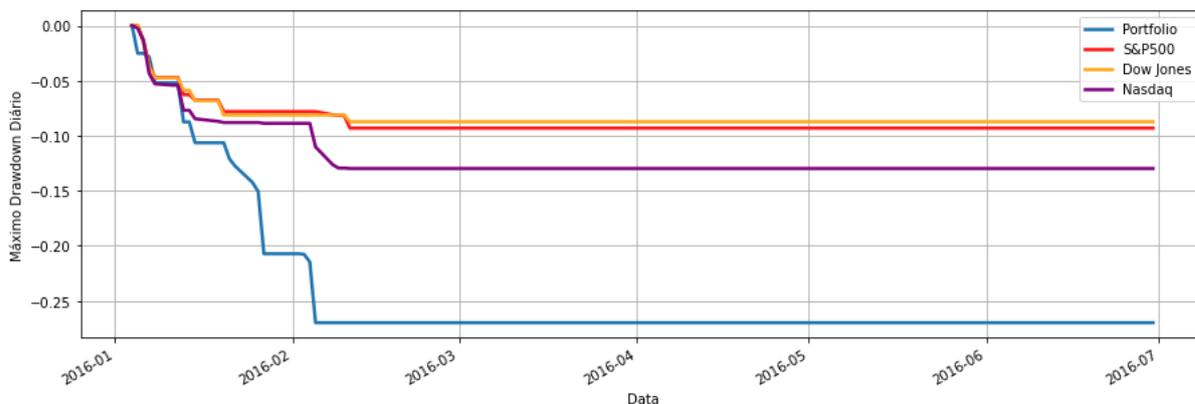
Fonte: Elaborada pelo autor.

5.4.3 Análise da Estratégia 3

A Estratégia 3 – E3 consiste em um portfólio com os 5 ativos mais citados e correlacionados da rede de um determinado setor da economia nos últimos seis meses de observação e aplicados em teste nos próximos seis meses. A Tabela 14 apresenta a composição dos Portfólios da Estratégia 3 – E3 no período de 2015 a 2018.

O setor mais citado e correlacionado nos períodos de observação foi o da tecnologia, conforme demonstrado na Figura 15.

Figura 23 – Máx. Daily Drawdown – E2 – Portfólio, SP500 e CD – 2016/1S



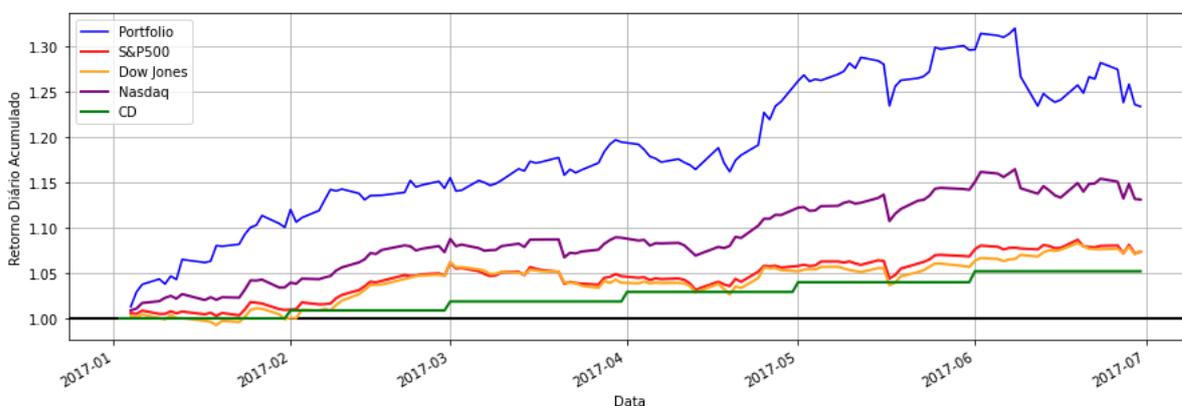
Fonte: Elaborada pelo autor.

	S1	W _{S1}	S2	W _{S2}	S3	W _{S3}	S4	W _{S4}	S5	W _{S5}
2015/1S	AAPL	0,600	BABA	0,108	FB	0,114	TSLA	0,043	TWTR	0,134
2015/2S	AAPL	0,634	FB	0,136	NFLX	0,134	TSLA	0,051	TWTR	0,044
2016/1S	AAPL	0,528	BABA	0,180	FB	0,191	NFLX	0,081	TWTR	0,020
2016/2S	BABA	0,006	FB	0,593	NFLX	0,380	TSLA	0,015	TWTR	0,005
2017/1S	AAPL	0,145	BABA	0,001	FB	0,396	NFLX	0,456	TWTR	0,003
2017/2S	AMD	0,029	AAPL	0,497	AMZN	0,302	NVDA	0,112	TSLA	0,060
2018/1S	AMD	0,016	AAPL	0,119	AMZN	0,259	NVDA	0,447	TSLA	0,160
2018/2S	AMD	0,213	AAPL	0,147	AMZN	0,039	FB	0,037	TSLA	0,565

Tabela 14 – Composição dos Portfólios - E3

A Figura 24 apresenta o gráfico referente ao retorno acumulado do Portfólio indicado frente aos índices de renda variável e de renda fixa para o primeiro semestre de 2017.

Figura 24 – Retorno Acumulado - E3 - Portfólio, SP500 e CD - 2017/1S



Fonte: Elaborada pelo autor.

Juntamente com a Tabela 15 e com a Tabela 16 observa-se que este trimestre apresentou um retorno acumulado de 22% frente a 7% e 13% dos índices de renda variável.

No quarto segundo semestre de 2015, o rendimento do Portfólio foi inferior ao índice da bolsa norte-americana, que também sofreu desvalorização neste semestre. Sendo assim

conclui-se que a indicação do portfólio com base nos ativos mais citados e correlacionados da rede de um determinado setor da economia rendeu ao final do ano 72% ao investidor, frente a 38% do índice Nasdaq, que apresentou o desempenho entre os índices.

Percebemos então uma evolução na seleção de portfólio utilizando esta terceira estratégia, apesar de no segundo semestre de 2018 o portfólio apresentar um rendimento bem inferior. Isso se deve à desvalorização dos ativos deste setor em 2018, após uma alta entre o segundo semestre de 2016 e o primeiro semestre de 2018.

	Retorno				Volatilidade			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	0,10	0,01	-0,01	0,06	0,012	0,008	0,008	0,008
2015/2S	-0,04	0,03	0,02	0,02	0,017	0,011	0,011	0,012
2016/1S	-0,01	0,05	0,05	0,01	0,015	0,010	0,010	0,012
2016/2S	0,12	0,06	0,10	0,10	0,014	0,006	0,006	0,007
2017/1S	0,22	0,07	0,07	0,13	0,010	0,004	0,004	0,006
2017/2S	0,19	0,10	0,14	0,12	0,011	0,004	0,004	0,006
2018/1S	0,23	0,01	-0,01	0,08	0,019	0,010	0,011	0,012
2018/2S	-0,09	-0,08	-0,04	-0,13	0,030	0,011	0,011	0,014

Tabela 15 – Avaliação de Desempenho – E3 – Portfólio x SP 500 x Dow Jones x Nasdaq

	Skewness				Kurtosis			
	Portfólio	S&P500	Dow Jones	Nasdaq	Portfólio	S&P500	Dow Jones	Nasdaq
2015/1S	-0,523	-0,071	-0,037	-0,435	0,927	-0,114	-0,156	-0,156
2015/2S	-0,125	-0,187	-0,096	-0,120	1,549	1,649	1,526	1,413
2016/1S	-0,553	-0,382	-0,425	-0,468	1,395	1,086	0,919	1,173
2016/2S	0,989	-0,101	0,115	-0,146	8,017	3,668	2,995	1,443
2017/1S	0,934	-0,361	-0,122	-1,170	3,395	2,824	2,689	3,759
2017/2S	0,710	-0,553	-0,990	-0,221	5,423	3,352	2,004	2,741
2018/1S	-0,693	-0,990	-0,981	-0,762	1,046	2,675	2,770	1,588
2018/2S	0,683	0,105	0,105	0,0438	3,252	3,677	3,251	2,501

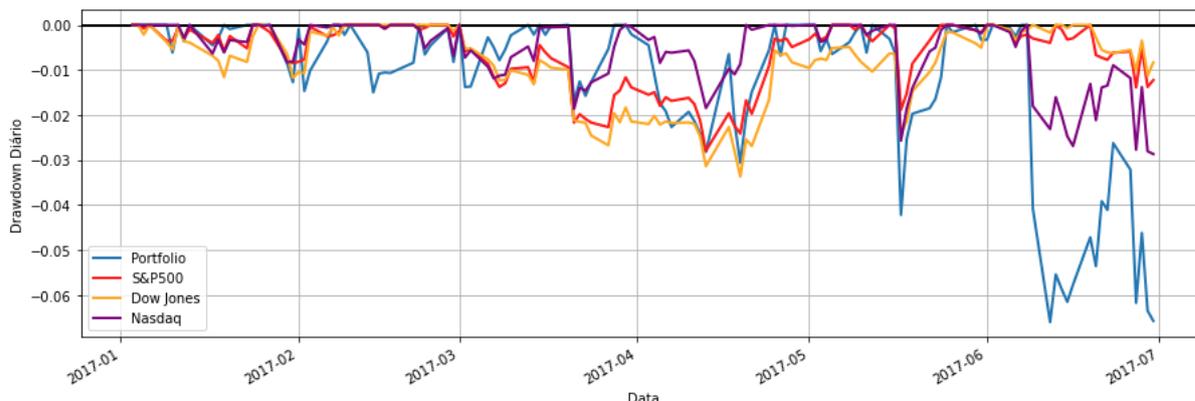
Tabela 16 – Avaliação de Desempenho – E3 – Portfólio x SP 500 x Dow Jones x Nasdaq

A Estratégia 3 – E3 indica que ativos que estejam em alta, ou seja, com alta correlação e valorização, podem levar o investidor a retornos superiores aos índices de mercado, conforme observado no período estudado.

Na Figura 25, verifica-se através do gráfico de *Drawdown* que as perdas ao logo do primeiro semestre de 2017 foram inferiores em relação ao SP500. Porém mesmo diante de algumas perdas inferiores frente ao SP500, apenas ao final do primeiro semestre de 2017, as perdas do portfólio foram maiores do que as perdas de todos os índices.

Ao analisar o pior semestre de desempenho da Estratégia 3 - E3 verificamos uma desvalorização do Portfólio quanto dos índices ao final do segundo semestre de 2018, que coincide com a desvalorização dos ativos do setor de tecnologia da informação.

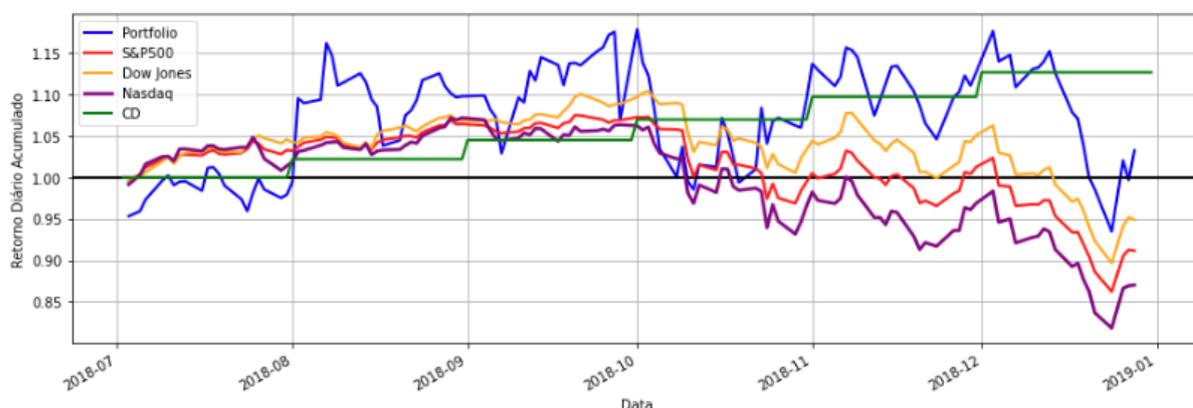
Figura 25 – Daily Drawdown – E3 – Portfólio, SP500 e CD – 2017/1S



Fonte: Elaborada pelo autor.

Estes ativos base para a formação do Portfólio e também são parte constituinte da formação dos índices das Bolsas de Valores norte-americanas. Vejamos as Figuras 26 e 27 que demonstram uma desvalorização frente ao índice de renda fixa (CD) daquela época.

Figura 26 – Retorno Acumulado - E3 - Portfólio, SP500 e CD - 2018/2S



Fonte: Elaborada pelo autor.

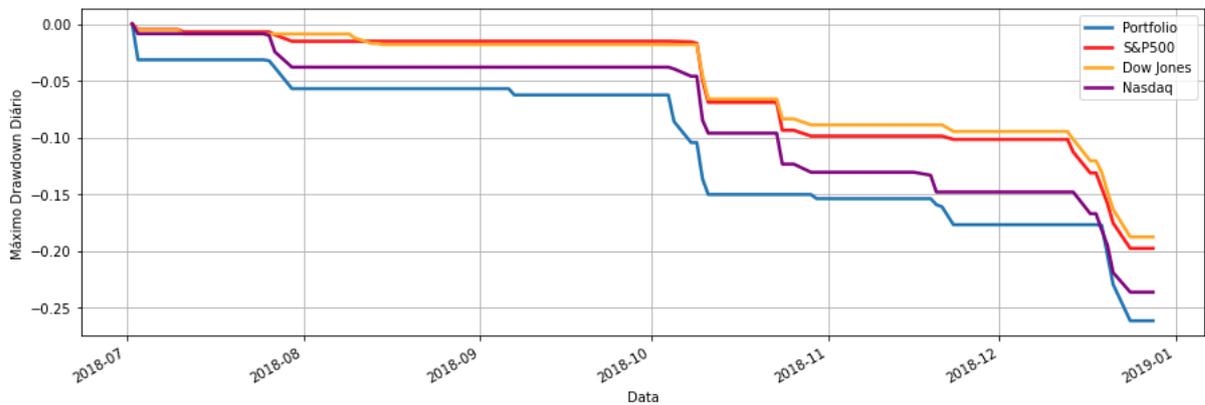
Verifica-se então que o portfólio conseguiu retornar um rendimento considerável ao investidor e com menores perdas ao longo desse semestre.

5.5 Considerações Importantes

Neste Capítulo foram detalhadas todas as etapas dos resultados experimentais, ou seja, como realizou-se as seleções dos portfólios de cada estratégia desenvolvida. Foram analisadas três estratégias com o auxílio do Índice de Sharpe, Modelo de Markowitz e da Simulação de Monte Carlo para definição dos pesos de cada ativo no portfólio.

Dentre os resultados experimentais apresentados pode-se verificar que os ativos mais citados e com maior correlação obtiveram, de forma satisfatória, durante os meses finais de

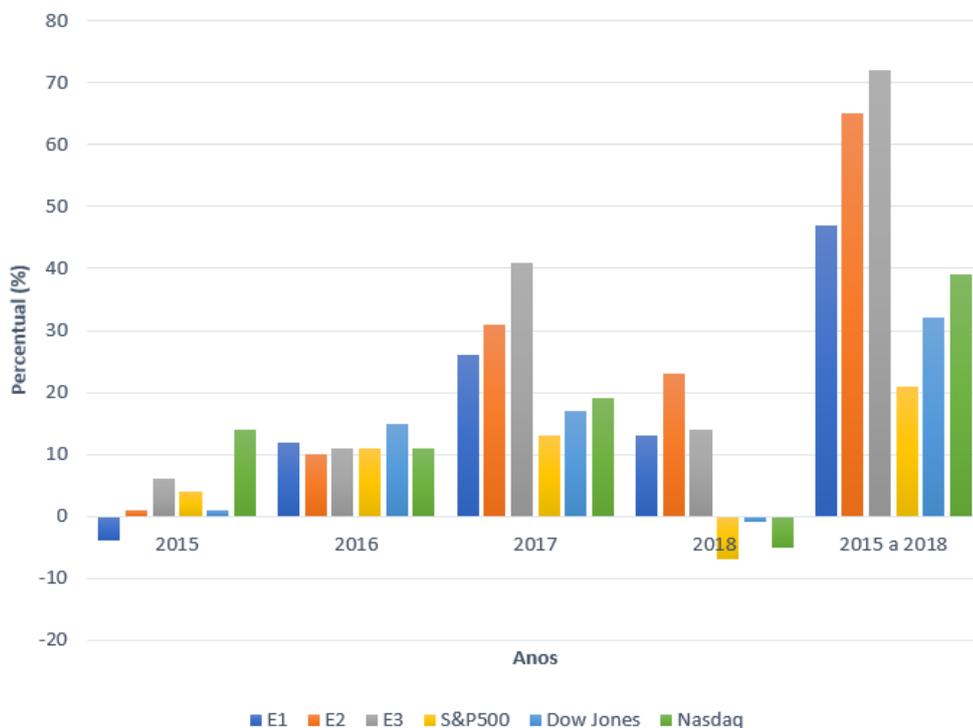
Figura 27 – Daily Drawdown – E3 – Portfólio, SP500 e CD – 2018/2S



Fonte: Elaborada pelo autor.

2016 até os meses iniciais de 2018, melhores resultados de taxa de retorno ao investidor em relação aos investimentos de renda fixa e de renda variável conforme podemos verificar na Figura 28 que resume o percentual de retorno para cada Estratégia e para cada Índice no decorrer dos anos de 2015, 2016, 2017 e 2018.

Figura 28 – Retorno percentual entre 2015 e 2018



Fonte: Elaborada pelo autor.

Em relação a Volatilidade, a *Skewness* e a *Kurtosis* mencionadas nas Tabelas 9, 10, 12, 13, 15, 16, que retratam sobre a Avaliação de Desempenho dos Portfólios, não se verificou nenhuma grande distorção entre o Portfólio e os índices.

Percebe-se que a Volatilidade do Portfólio apresentou uma pequena variação em relação a Volatilidade dos índices, ou seja, os Portfólios indicados neste estudo apresentam um maior risco e uma maior variação de retornos conforme vimos ao longo desta Seção.

6 Conclusão

Neste capítulo descrevemos a conclusão desta dissertação de mestrado, apresentando os principais resultados encontrados (Seção 6.1) e, também, os trabalhos futuros que representam os direcionamentos para novas pesquisas (Seção 6.2).

6.1 Conclusões

Neste trabalho foram apresentadas três estratégias de seleção de portfólios de ações a partir da caracterização e análise dos dados da rede social de finanças *Stocktwits*, são elas: *Estratégia 1* - selecionar os ativos mais citados; *Estratégia 2* - selecionar os ativos mais citados positivamente; e *Estratégia 3* - selecionar os ativos mais citados do setor mais correlacionado. Também foram considerados dados de ativos da bolsa de valores norte-americana para complementar as análises da seleção de portfólios. Com as análises realizadas foi possível perceber o comportamento dos usuários, bem como avaliar o padrão de conteúdo que é veiculado na rede social e quais ativos são mais postados na rede social. As estratégias de seleção de portfólios de ações também foram comparadas com outros índices de mercado: renda variável e renda fixa.

Dessa forma, além da contextualização e motivação deste trabalho apresentadas no Capítulo 1, foi descrita uma fundamentação teórica e um levantamento de trabalhos relacionados, apresentados no Capítulo 2 e Capítulo 3, respectivamente. Nesses dois capítulos, foram abordados os temas: introdução ao mercado financeiro; análise de sentimentos, que é o estudo de opiniões, sentimentos e emoções expressas em textos; predição de preços das ações por meio do poder preditivo da análise de sentimentos; análise da influência dos usuários em relação ao mercado financeiro; e modelagem da interação de usuários ou ativos por meio de grafos. Vários destes trabalhos serviram como fundamentação para a pesquisa, incluindo técnicas, metodologia e experimentos descritos pelos autores.

A metodologia do trabalho, descrita no Capítulo 4, descreve todas as etapas necessárias para a elaboração das caracterizações e estratégias de seleção de portfólios, bem como sua avaliação. As etapas da metodologia foram: coleta de dados, limpeza e preparação dos dados, caracterização e análise de dados, seleção de portfólios e análise de resultados.

Por meio dos experimentos e análise dos resultados, apresentados no Capítulo 5, foram alcançados os objetivos propostos neste trabalho. Foi possível concluir que a rede social de finanças tem uma correlação significativa com o mercado financeiro. Especificamente, a análise estática e a análise temporal mostram que a *Stocktwits* é frequentemente utilizado e as interações na rede aumentam a cada ano, o que proporciona uma grande fonte de informações

para investidores. Já a análise de correlação dos ativos mostrou: um valor alto de correlação positiva entre o número de mensagens diárias e o volume dos ativos. O indicador social foi útil na Estratégia 2 (E2) que tinha como base o levantamento dos ativos mais comentados positivamente na rede.

Os resultados para as três estratégias de seleção de portfólios de investimento em ações foram promissores, pois evidenciou que os dados da rede social podem ser utilizados como suporte à tomada de decisão no mercado financeiro. Também observou-se que os portfólios indicados pela análise de dados apresentam retornos financeiros satisfatórios, quando comparado aos índices *S&P500*, *Dow Jones* e *Nasdaq*.

Em geral, os experimentos realizados mostraram que os dados da rede social podem ser utilizados como apoio e suporte para a seleção de um conjunto de ativos a serem investidos no mercado de ações.

Finalmente, algumas limitações do trabalho foram: a utilização de apenas um conjunto de dados. Assim, planeja-se considerar, futuramente, diferentes fontes de dados, como por exemplo a *Leadr*¹. Também espera-se utilizar uma estratégia para identificar postagens feitas por bots, apesar do *StockTwits* possuir um público específico e não ser muito suscetível a isso. Portanto, o desafio é combinar várias técnicas e fontes de dados para a melhoria contínua ao suporte de tomada de decisão do investidor.

6.2 Trabalhos futuros

Assim, foi apresentado um estudo completo sobre a seleção de Portfólios a partir de dados da rede social *Stocktwits* composta por análises estáticas, temporais, correlacionais e de seleção de ativos. Apesar da abrangência de nossas análises, existem muitos outros aspectos promissores a serem investigados com outras fontes de dados de outras redes sociais e o Mercado Financeiro de Ações.

Como trabalhos futuros, alguns aspectos podem ser considerados de forma a melhorar os resultados apresentados nesse trabalho:

- **Periodicidade:** Ampliar o período de dados analisado. Explorar dados de períodos maiores, com uma base de dados de 10 anos para validar períodos de comportamentos diferentes da economia e dos ativos;
- **Fontes de dados:** Utilizar outras fontes de dados advindas de outras redes sociais de finanças;
- **Desempenho:** a fim de melhorar o desempenho das estratégia de seleção de ativos utilizando outros algoritmos de otimização ou de inteligência artificial.

¹ Leadr: <<https://www.leadr.com.br/>>

Referências

- ABDI, H. The kendall rank correlation coefficient. **Encyclopedia of Measurement and Statistics**. Sage, Thousand Oaks, CA, Citeseer, p. 508–510, 2007. Citado na página 27.
- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Mining association rules between sets of items in large databases. In: **IN: PROCEEDINGS OF THE 1993 ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, WASHINGTON DC (USA)**. [S.l.: s.n.], 1993. p. 207–216. Citado na página 28.
- AGRAWAL, S. et al. Practical applications of momentum, mean-reversion, and social media: Evidence from stocktwits and twitter. **Practical Applications**, Institutional Investor Journals Umbrella, v. 6, n. 3, p. 1–4, 2019. Citado na página 20.
- ALDRIDGE, I. **High-frequency trading: a practical guide to algorithmic strategies and trading systems**. [S.l.]: John Wiley & Sons, 2013. Citado na página 1.
- ALVES, G. et al. Uma análise do mercado de ações baseada na correlação entre ativos no stocktwits. In: **BraSNAM**. Natal, Brazil: [s.n.], 2018. Citado 2 vezes nas páginas 21 e 22.
- ANGHEL, M. G. Technical analysis versus fundamental analysis of securities. **Romanian Statistical Review Supplement**, v. 61, n. 2, p. 257–262, 2013. Disponível em: <<https://EconPapers.repec.org/RePEc:rsr:supplm:v:61:y:2013:i:2:p:257-262>>. Citado na página 9.
- ARAUJO, M. M.; CRUZ, E. A. Análise e mineração de dados sociais com a ferramenta weka. **FAMETRO 2017**, 2017. ISSN 2357-8645. Citado 2 vezes nas páginas 12 e 13.
- ATKINS, A.; NIRANJAN, M.; GERDING, E. Financial news predicts stock market volatility better than close price. **The Journal of Finance and Data Science**, 2018. Citado 2 vezes nas páginas 21 e 22.
- AUDRINO, F.; SIGRIST, F.; BALLINARI, D. The impact of sentiment and attention measures on stock market volatility. **Int. Journal of Forecasting**, Science Direct, v. 36, n. 2, p. 334–357, 2020. Citado na página 20.
- B3 - Brasil, Bolsa e Balcão. 2020. Disponível em: <http://www.b3.com.br/pt_br/>. Citado na página 8.
- BAR-HAIM, R. et al. Identifying and following expert investors in stock microblogs. In: **Procs. of EMNLP**. Edinburgh, United Kingdom: [s.n.], 2011. p. 1310–1319. Citado na página 2.
- BAR-HAIM, R. et al. Identifying and following expert investors in stock microblogs. In: **Procs. of the Conference on Empirical Methods in Natural Language Processing**. Scotland – UK: [s.n.], 2011. p. 1310–1319. Citado 2 vezes nas páginas 21 e 22.
- BARBETTA REIS, M. M.; BORNIA, A. C. **Estatística: para cursos de engenharia e informática**. [S.l.]: Atlas, 2004. ISBN 9788522459940. Citado na página 17.
- BENESTY, J. et al. Pearson correlation coefficient. In: **Noise reduction in speech processing**. [S.l.]: Springer, 2009. p. 1–4. Citado na página 27.

BIONDO, A. E. et al. Are random trading strategies more successful than technical ones? **PloS ONE**, v. 8, n. 7, p. e68344, 2013. Citado na página 9.

BODIE, Z.; MERTON, R. C. **Finanças**. Bookman, 1999. ISBN 85-7307-535-X. Disponível em: <<https://www.estantevirtual.com.br/livros/zvi-bodie-robert-c-merton/financas/3932857026>>. Citado na página 3.

CAVALCANTE, F.; MISUMI, J. Y. **Mercado de Capitais: o que é, como funciona**. [S.l.]: Campus, 2005. ISBN 9788535216738. Citado na página 7.

CORRAR, L.; THEOPHILO, C. R. **Pesquisa operacional para decisão em contabilidade e administração: contabilometria**. [S.l.]: Atlas, 2004. Citado na página 19.

CORTEZ, P.; OLIVEIRA, N.; FERREIRA, J. a. P. Measuring user influence in financial microblogs: Experiments using stocktwits data. In: **Procs. of WIMS**. Nîmes, France: [s.n.], 2016. p. 23:1–23:10. Citado na página 21.

DIAS, W. L.; JUNIOR, E. R.; REIS, A. L. N. Analysis of the federal district's parliament in the information age. **Revista Brasileira de Ciência Política**, 2017. Citado na página 27.

EMIRBAYER, M.; GOODWIN, J. Network analysis, culture, and the problem of agency. **American Journal of Sociology**, University of Chicago Press, v. 99, n. 6, p. 1411–1454, 1994. ISSN 00029602, 15375390. Disponível em: <<http://www.jstor.org/stable/2782580>>. Citado na página 11.

ESCUADERO, L. F. **La simulación en la empresa**. [S.l.]: Ediciones Deusto, 1973. ISBN 9788423402403. Citado na página 19.

FAMA, E. Efficient capital markets: A review of theory and empirical work. **Journal of Finance**, v. 25, n. 2, p. 383–417, 1970. Disponível em: <<https://EconPapers.repec.org/RePEc:bla:jfinan:v:25:y:1970:i:2:p:383-417>>. Citado na página 8.

FERREIRA, E. B. A. **Análise de sentimento em redes sociais utilizando influência das palavras**. Dissertação (Mestrado) — Universidade Federal de Pernambuco, www.cin.ufpe.br/tg/2010-2/ebaf.pdf, 2010. Citado na página 1.

FERREIRA, J. C. P. **Análise de Influência de Utilizadores e Redes Sociais em Microblogs sobre Mercados Financeiros**. Dissertação (Mestrado) — Engenharia e Gestão de Sistemas de Informação - Escola de Engenharia - Universidade do Minho, <http://repositorium.sdum.uminho.pt/handle/1822/42114>, 2016. Citado 2 vezes nas páginas 21 e 22.

FORTUNA, E. **Mercado Financeiro: produtos e serviços**. [S.l.]: Qualitymark, 2007. ISBN 9788573035391. Citado na página 7.

FREEMAN, L. C. Some antecedents of social network analysis. **CiteSeerx**, Connections, v. 19, n. 1, p. 39–42, 1994. Citado na página 12.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: Concepts and Techniques**. [S.l.]: Morgan Kaufmann, 2011. ISBN 978-0123814791. Citado na página 11.

JABBUR, E. et al. Proposal and implementation of new trading strategies for stock markets using web data. In: **Procs. of the 21st Brazilian Symposium on Multimedia and the Web**. Manaus, Brasil: [s.n.], 2015. p. 113–120. Citado na página 22.

JUNIOR, C. G.; PAMPLONA, E. O.; MONTECECHI, J. A. B. Seleção de carteiras através do modelo de markowitz para pequenos investidores. In: **SIMPEP**. Bauru, Brazil: [s.n.], 2002. Citado na página 3.

KEEN, J. M.; DIGRIUS, B. **Making technology investments profitable : ROI road map to better business cases**. [S.l.]: John Wiley & Sons, 2003. ISBN 0471227331 9780471227335. Citado na página 29.

LI, Q.; SHAH, S. Learning stock market sentiment lexicon and sentiment-oriented word vector from stocktwits. In: **CoNLL**. Vancouver, Canada: [s.n.], 2017. p. 301–310. Citado 2 vezes nas páginas 20 e 22.

LIN, K. W.; CHUNG, S.-H. A fast and resource efficient mining algorithm for discovering frequent patterns in distributed computing environments. **Future Gener. Comput. Syst.**, Elsevier Science Publishers B. V., NLD, v. 52, n. C, p. 49–58, nov. 2015. ISSN 0167-739X. Disponível em: <<https://doi.org/10.1016/j.future.2015.05.009>>. Citado na página 13.

LIU, B. **Sentiment Analysis and Subjectivity**. [S.l.]: second edition, 2010. 627–666 p. Citado na página 20.

LO, A.; MACKINLAY, A. A non-random walk down wall street. **A Non-Random Walk Down Wall Street**, p. 1–424, 11 2011. Citado na página 9.

LO, A. W. The adaptive markets hypothesis. **The Journal of Portfolio Management**, Institutional Investor Journals Umbrella, v. 30, n. 5, p. 15–29, 2004. ISSN 0095-4918. Disponível em: <<https://jpm.pm-research.com/content/30/5/15>>. Citado na página 9.

LOPES, T.; BARTH, F.; MARACCINI, J. Mineração de opiniões aplicada a análise de investimentos. **Sociedade Brasileira de Computação**, 2008. Citado na página 2.

MARKOWITZ, H. Portfolio selection. **Journal of Finance**, p. 77–91, 1952. Citado 6 vezes nas páginas 2, 3, 13, 14, 15 e 17.

METROPOLIS, N.; ULAM, S. The monte carlo method. **Journal of the American Statistical Association**, [American Statistical Association, Taylor & Francis, Ltd.], v. 44, n. 247, p. 335–341, 1949. ISSN 01621459. Disponível em: <<http://www.jstor.org/stable/2280232>>. Citado na página 18.

NETO, A. A. **Mercado Financeiro**. [S.l.]: Atlas, 2009. ISBN 9788522454365. Citado 2 vezes nas páginas 6 e 7.

OH, C.; SHENG, O. Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement. In: **Procs. of Thirty Second International Conference on Information Systems, Shanghai**. Shanghai – China: [s.n.], 2011. p. 1–19. Citado 3 vezes nas páginas 2, 20 e 22.

OLIVEIRA, L.; CAMPOS, G.; SMARZARO, R. Mineração de dados em redes sociais. In: . [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 11 e 12.

OLIVEIRA, N.; CORTEZ, P.; AREAL, N. On the predictability of stock market behavior using stocktwits sentiment and posting volume. In: . Berlin, Germany: [s.n.], 2013. v. 854, p. 355–365. Citado na página 20.

- OLIVEIRA, N.; CORTEZ, P.; AREAL, N. On the predictability of stock market behavior using stocktwits sentiment and posting volume. In: **Procs. of Portuguese Conference on Artificial Intelligence**. Açores, Portugal: [s.n.], 2013. p. 355–365. Citado 2 vezes nas páginas 20 e 22.
- RIBEIRO, C.; FERREIRA, L. Uma contribuição ao problema de composição de carteiras de mínimo valor em risco. **Gestão & Produção**, v. 12, 08 2005. Citado na página 18.
- ROMERO, M. **Índice de Sharpe: comparando as Carteiras Magnetis aos fundos multimercado**. 2019. Disponível em: <<https://blog.magnetis.com.br/indice-de-sharpe/>>. Acesso em: 10 jan. 2020. Citado na página 17.
- SANTOS, H. S. **Um Estudo sobre o Mercado Brasileiro de Ações a partir de dados do Twitter**. Dissertação (Mestrado) — Mestrado em Ciências da Computação da Universidade Federal de Minas Gerais, 2016. Citado na página 22.
- SHARPE, W. F. A simplified model for portfolio analysis. **Management Science**, INFORMS, v. 9, n. 2, p. 277–293, 1963. ISSN 00251909, 15265501. Disponível em: <<http://www.jstor.org/stable/2627407>>. Citado na página 17.
- SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. **The Journal of Finance**, [American Finance Association, Wiley], v. 19, n. 3, p. 425–442, 1964. ISSN 00221082, 15406261. Disponível em: <<http://www.jstor.org/stable/2977928>>. Citado na página 15.
- SILVA, A. C.; CRUZ, E. A. Análise e mineração de dados sociais com a ferramenta tableau e pentaho. **FAMETRO 2016**, 2016. ISSN 2357-8645. Citado 2 vezes nas páginas 12 e 13.
- SILVA, J. R. da et al. Multi-entity polarity analysis in financial documents. In: **Procs. of the 20st Brazilian Symposium on Multimedia and the Web**. João Pessoa, Brasil: [s.n.], 2014. p. 115–122. Citado na página 22.
- TU, W. et al. Investment recommendation using investor opinions in social media. In: **SIGIR**. Pisa, Italy: [s.n.], 2016. p. 881–884. Citado 2 vezes nas páginas 21 e 22.
- WANG, G. et al. Crowds on wall street: Extracting value from collaborative investing platforms. In: **CSCW**. Vancouver, Canada: [s.n.], 2015. p. 17–30. Citado 2 vezes nas páginas 21 e 22.
- YAKUSHEV, A.; MITYAGIN, S. Social networks mining for analysis and modeling drugs usage. **Procedia Computer Science**, v. 29, p. 2462–2471, 12 2014. Citado na página 13.