

Em Qual Portfólio Investir? Análise e Seleção a partir de dados do StockTwits

Ozório J. S. Camargos¹, Adriano C. M. Pereira¹, Michele A. Brandão²

¹Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Belo Horizonte, MG - Brasil.

²Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG)
Ribeirão das Neves, MG - Brasil

ozorio.camargos@gmail.com, adrianoc@dcc.ufmg.br,
michele.brandao@ifmg.edu.br

Abstract. *Information present in online social networks can be used to solve various problems. This is no different for the financial market. This paper proposes three strategies for selecting an investment portfolio based on the use of StockTwits data. The results show promising performance for our strategies and indicate that investors may have higher returns than market indexes that were used as benchmarks and baselines.*

Resumo. *Informações presentes em redes sociais online podem ser utilizadas para solucionar diversos problemas. Isso não é diferente para o mercado financeiro. De fato, este trabalho propõe três estratégias para seleção de Portfólio de investimentos a partir do uso de dados do StockTwits. Os resultados mostram desempenho promissor para nossas estratégias e indicam que investidores podem ter retornos superiores aos índices de mercado e baselines.*

1. Introdução

Investir não é tarefa fácil, principalmente, por diversos fatos poderem provocar oscilações no mercado financeiro. Para auxiliar em tal decisão, diferentes estudos consideram dados de cotações históricas dos ativos negociados em bolsas de valores para analisar variações de preço e volume ao longo do tempo. Isso possibilita identificar padrões de comportamento e estratégias de negociação [Aldridge 2013, Jabbur and Pereira 2018].

Outro tipo de estudo é a análise de dados de volumes que são coletados em redes sociais e na *Web*, e se referem aos comentários dos ativos negociados em bolsas de valores. Esses dados podem representar pensamentos e ideias que abrangem aspectos fundamentalistas de notícias de várias partes do mundo. Por ser uma base de dados relevante, plataformas de integração em tempo real estão surgindo, por exemplo, o *StockTwits*¹.

Nesse contexto, este trabalho tem como principal objetivo gerar *insights* e proporcionar melhor entendimento sobre o comportamento de usuários e características dos ativos para auxiliar investidores na tomada de decisão sobre a seleção de portfólios² do mercado financeiro. Assim, este estudo busca responder às seguintes questões: (i) é possível

¹StockTwits: <https://stocktwits.com/>, uma plataforma de comunicações para a comunidade financeira e de investimento. Em tal rede são feitas postagens que incluem ideias, links, gráficos e outros dados financeiros importantes, resumidos no limite de 140 caracteres (*tweets* ou *twits*).

²Portfólio de investimentos é o conjunto de aplicações do investidor, seja pessoa física ou jurídica, ele reúne todos os ativos financeiros que escolhidos pelo cliente, tanto em renda fixa quanto variável.

a seleção de portfólios de investimento a partir da caracterização de dados de uma rede social?; e (ii) qual é o desempenho do portfólio criado frente a outro índice de mercado?

2. Trabalhos Relacionados

Diferentes estudos realizam análise de sentimentos, que inclui análise de opiniões, sentimentos e emoções expressas em textos. Existem muitas tarefas relacionadas a essa área, como a extração de elementos de texto de opinião, à classificação quanto ao seu caráter (positivo, negativo e neutro), à comparação de sentenças quanto a suas opiniões. Nesse contexto, [Agrawal et al. 2019] demonstram que a atividade de mídia social pode afetar significativamente a liquidez no dia a dia, e [Audrino et al. 2020] descrevem o impacto das variáveis de sentimento e medidas de atenção na volatilidade do mercado de ações.

Diferentemente deste trabalho, há pesquisas que focam em entender o comportamento dos ativos. Por exemplo, [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 dos ativos que são citados juntos em uma postagem. Em seguida, é feita uma análise das principais características e pares isolados na rede. Os resultados mostram que ativos citados juntos estão correlacionados com variações no mercado financeiro, o que pode servir de suporte à decisão para investimento.

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*. Os resultados mostram que os eventos e notícias sobre o mercado de ações são capazes de gerar picos de publicação pelos usuários. A avaliação desta pesquisa utilizou treze meses de mensagens registradas no *Twitter*, fornecendo uma base de padrões comportamentais para a elaboração de indicadores sociais de apoio à decisão.

Diante disso, este trabalho investiga a utilização dos dados da rede social para melhoria da estratégia de análise do comportamento de ativos. Também objetiva auxiliar na tomada de decisão ao operar no mercado financeiro.

3. Metodologia

A metodologia utilizada possui quatro etapas: coleta de dados, limpeza de dados, caracterização de dados e seleção dos portfólios. Ademais, essa metodologia é uma expansão da publicada em [Ferreira et al. 2019], principalmente, a última etapa.

Os dados utilizados são postagens da rede social *StockTwits*, dados disponibilizados devido a parceria com essa empresa, e os dados de preços e volume dos ativos via API, *Application Programming Interface*, da plataforma do *Yahoo Finance*³ nos anos de 2017 e 2018. Para a limpeza e preparação dos dados, foram desenvolvidos *scripts*, que extraem das bases de dados algumas características como: *hashtags* de ativos, *replies*, comentários, sentimentos, *likes*, preço ajustado e volume negociado, além de retirar alguns registros de mensagens postadas com atributos com ausências de valores.

Na etapa de caracterização e análise dos dados são utilizados três conceitos de avaliação: análise estática, temporal e de correlação. Na análise estática, são consolidadas

³Yahoo Finance: <https://finance.yahoo.com/>

as seguintes informações: estatística básica dos dados coletados de mensagens, mensagens que mencionam ativos, sentimento marcado pelo usuário na mensagem como *Bullish* (Positivo) e *Bearish* (Negativo) e a distribuição de mensagens por ativos. Por outro lado, na análise temporal, é estudada a evolução dos atributos dos *tweets* ao longo do tempo, para os anos de 2017 e 2018. Já a análise de correlação investiga se existe correlação entre o volume de mensagens postados na rede social e a oscilação das variáveis relacionadas aos preços dos ativos no mercado financeiro. O coeficiente utilizado foi o de *Spearman*, pois avalia quão bem uma função monotônica arbitrária pode descrever a relação entre duas variáveis. E para cálculo considerou-se o número de mensagens publicadas nos dias de operações no mercado financeiro e o volume negociado diariamente.

Após a realização da caracterização dos dados da *StockTwits* e normalização dos dados advindos do *Yahoo Finance* foram desenvolvidas as seguintes estratégias:

Estratégia 1 - (E1). Selecionar os cinco ativos mais citados na rede nos últimos seis meses e com maiores Rendimentos nos últimos cinco anos de observação. Logo após, realizar testes e avaliação de desempenho nos próximos três meses. Para identificar os ativos mais citados na rede social no período avaliado foi utilizada a técnica de Mineração de Itens Frequentes, introduzida por [Agrawal et al. 1993]. A partir deste conjunto, foram selecionados, com base na frequência absoluta, os cinco ativos mais citados para compor o portfólio.

Estratégia 2 - (E2). Selecionar os cinco ativos mais citados na rede positivamente nos últimos seis meses e com maiores Rendimentos nos últimos cinco anos de observação; e realizar testes e avaliação de desempenho nos próximos três meses. Para identificar os ativos mais citados positivamente utilizamos um indicador social que é composto pela quantidade de mensagens positivas e negativas postadas durante o dia na rede social. Esse indicador é chamado de ($S - Score$) conforme [Ferreira et al. 2019].

Estratégia 3 - (E3). Selecionar os cinco ativos mais citados de um determinado setor mais correlacionado na rede nos últimos seis meses e com maiores Rendimentos nos últimos cinco anos de observação. Em seguida, realizar testes e avaliação de desempenho nos próximos três meses. 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*.

Após a seleção dos ativos, é realizada uma otimização dos pesos de cada ativo na composição do Portfólio. Para isso, é utilizado o índice de Sharpe, indicador que mede o retorno excedente de uma aplicação financeira em relação a uma aplicação livre de risco [Neto 2016], com o auxílio da Simulação de Monte Carlo, uma metodologia estatística que se baseia em uma grande quantidade de amostragens aleatórias para se chegar em resultados próximos de reais [Ribeiro and Ferreira 2005].

4. Resultados e Discussões

Esta seção apresenta os resultados dos experimentos realizados para a seleção de portfólios. Entretanto, para entender tal seleção são apresentados os resultados da caracterização dos dados considerados do *StockTwits*.

De fato, a Tabela 1 apresenta os resultados da análise estática. Tal tabela mostra

Tabela 1. Informações dos dados coletados das mensagens publicadas

Descrição	Quantidade	Percentual
Mensagens publicadas entre 2017 e 2018	81.848.148	100%
Mensagens que mencionam ativos	51.605.258	63,05%
Mensagens que mencionam sentimentos	24.407.118	29,82%



Figura 1. Retorno Acumulado e Volatilidade - E1 - Portfólio e SP500 - 2018.

que 63,05% das mensagens fazem menção direta a um determinado código de ativo e, do total das mensagens postadas, 29,82% são marcadas com um sentimento do usuário em relação à postagem em questão. Em seguida, observa-se uma média de 3.369.154 mensagens por mês no período de 2017 e 2018, o que mostra o quanto a *StockTwits* é frequentemente utilizado e pode ser fonte de informações para investidores.

Conforme visto em [Ferreira et al. 2019] e seguindo a escala de [Cohen 1988], a maioria dos ativos observados tem uma correlação alta (maior que 0,50) ou uma correlação muito alta (maior que 0,70) entre o número de mensagens postadas e o volume negociado. Assim, é possível observar uma correlação significativa do número de mensagens com o volume das negociações dos ativos. O mesmo observa-se no conjunto de dados utilizado neste trabalho.

Após a realização da caracterização dos dados, será realizado a Seleção de Portfólios utilizando as três estratégias. Para cada uma delas, são apresentados o gráfico de retorno acumulado entre Portfólio e o Índice de renda variável (SP500)⁴ e a avaliação de desempenho frente às métricas utilizadas.

Análise Estratégia 1. Verifica-se na Figura 1 o retorno acumulado e a volatilidade do Portfólio indicado frente ao SP500 para o ano de 2018. Observa-se que o segundo trimestre apresentou o melhor resultado em 2018, com um retorno acumulado de 23% frente a 5% do índice SP500. Nos demais trimestres, o rendimento do Portfólio foi inferior ao índice da bolsa norte-americana. 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.

Análise Estratégia 2. A Figura 2 mostra que o trimestre 2018/3T apresentou o melhor resultado em 2018, com um retorno acumulado de 24% frente a 5% do índice SP500. No primeiro e quarto trimestre o rendimento do Portfólio foi inferior ao índice da bolsa norte-americana, porém, verificamos também que o índice da bolsa (SP500) sofreu

⁴SP500 é a abreviação de Standard & Poor's 500, trata-se de um índice composto por quinhentos ativos cotados nas bolsas de NYSE ou NASDAQ dos EUA.

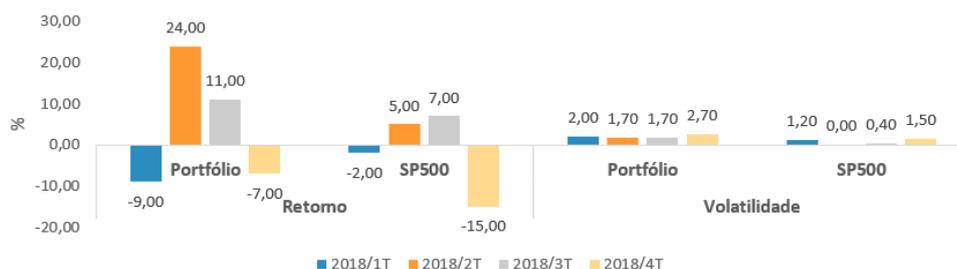


Figura 2. Retorno Acumulado e Volatilidade - E2 - Portfólio e SP500 - 2018

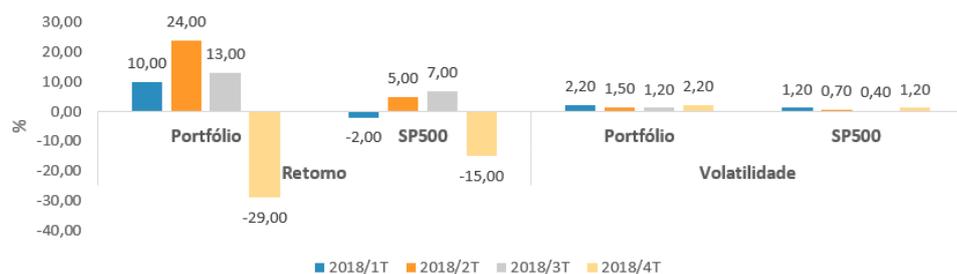


Figura 3. Retorno Acumulado e Volatilidade - E3 - Portfólio e SP500 - 2018.

desvalorização nestes trimestres. Conclui-se que a indicação do Portfólio com base nos ativos mais citados positivamente rendeu ao final do ano 19% ao investidor frente a -5% do SP500. Percebe-se então uma evolução na seleção de portfólio utilizando esta segunda estratégia. 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.

Análise Estratégia 3. Observa-se, na Figura 3 que o melhor resultado em 2018 foi o segundo trimestre de 2018 com um retorno acumulado de 24% frente a 5% do índice SP500. No quarto trimestre de 2018, o rendimento do Portfólio foi bem inferior ao índice da bolsa norte-americana, que também sofreu desvalorização neste trimestre. Assim, conclui-se que a indicação do Portfólio com base nos ativos mais citados e correlacionado da rede de um determinado setor da economia rendeu ao final do ano 18% ao investidor frente a -5% do SP500. Percebemos então uma evolução na seleção de Portfólio utilizando esta terceira estratégia, apesar de no quarto trimestre o Portfólio apresentar um rendimento bem inferior ao índice. Isso se deve à desvalorização dos ativos deste setor no último trimestre de 2018. A *Estratégia 3 – E3* indica que ativos que estejam em alta podem levar investidores a retornos superiores aos índices de mercado. Verifica-se então que o Portfólio conseguiu retornar um rendimento considerável ao investidor e com menores perdas ao longo do ano. Dentre os resultados apresentados, observa-se que os ativos mais citados e com maior correlação obtiveram, de forma satisfatória durante o ano de 2018 melhores resultados de taxa de retorno em relação aos investimentos de renda variável.

Limitações. Uma das limitações deste trabalho é realizar os experimentos com apenas um conjunto de dados. Assim, planeja-se considerar, futuramente, diferentes fontes de dados, como a *Leadr*⁵. Também espera-se utilizar uma estratégia para identificar postagens feitas por bots, apesar do *StockTwits* possuir um público específico e não ser muito

⁵Lead: <https://www.leadr.com.br/>

suscetível a isso. Outra limitação é a utilização de apenas um índice de renda variável para a comparação e validação dos resultados. Portanto, o desafio é combinar várias técnicas e fontes de dados para a melhoria contínua ao suporte de tomada de decisão do investidor.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou três estratégias para realizar seleção de Portfólios de investimento. Os resultados obtidos usando dados reais mostram que a rede social de finanças tem uma correlação significativa com o mercado financeiro. 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 para serem investidos no mercado financeiro. Como trabalhos futuros, pretendemos realizar o rebalanceamento dos portfólios a cada mês e explorar a composição dos portfólios com diferentes cardinalidades.

Agradecimentos. Agradecemos o apoio do CEFET-MG, IFMG, UFMG e *StockTwits*.

Referências

- Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Procs. of the ACM SIGMOD International Conference on Management of Data, Washington DC (USA)*, pages 207–216.
- Agrawal, S., Azar, P. D., Lo, A. W., and Singh, T. (2019). Practical applications of momentum, mean-reversion, and social media: Evidence from stocktwits and twitter. *Practical Applications*, 6(3):1–4.
- Aldridge, I. (2013). *High-frequency trading: a practical guide to algorithmic strategies and trading systems*. John Wiley & Sons.
- Alves, G., Bastos, J. P., Brandão, M., and Pereira, A. C. M. (2018). Uma análise do mercado de ações baseada na correlação entre ativos no stocktwits. In *Procs. of the 7st Brazilian Work. on Social Network Analysis and Mining – BraSNAM*, Natal, Brazil.
- Audrino, F., Sigrist, F., and Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *Int. Journal of Forecasting*, 36(2):334–357.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, second edition.
- Ferreira, R. S., Pereira, A. C. M., Camargos, O. J. S., and Brandão, M. A. (2019). Data science in financial markets: Characterization and analysis of stocktwits. In *Pros. of the 25th Brazillian Symposium on Multimedia and the Web – WebMedia*, pages 393–400, Rio de Janeiro, Brazil.
- Jabbur, E. and Pereira, A. C. M. (2018). Proposal and implementation of machine learning models for stock markets using web data. In *Procs. of the 24st Brazilian Symposium on Multimedia and the Web – Webmedia*, pages 61–64, Salvador, Brasil.
- Neto, A. A. (2016). *Mercado Financeiro*. Atlas.
- Ribeiro, C. and Ferreira, L. (2005). Uma contribuição ao problema de composição de carteiras de mínimo valor em risco. *Gestão & Produção*, 12.
- Santos, H. S. (2016). Um estudo sobre o mercado brasileiro de ações a partir de dados do twitter. Master's thesis, Departamento de Ciências da Computação - UFMG.