

ヤフーファイナンス掲示板を用いたコロナ・ショックにおける 投資家の話題変化と株式指標の関係分析

Analysis of Relationship between Stock Indexes and Changes in Investors' Topics during 2020 Stock Market Crash using Yahoo! Finance Stock Bulletin Board

佐々木 皓大^{1*} 諏訪 博彦^{1,5} 小川 祐樹²
梅原 英一³ 山下 達雄⁴ 坪内 孝太⁴ 安本 慶一¹
Kodai Sasaki,¹ Hirohiko Suwa,¹ Yuki Ogawa,²
Eiichi Umehara,³ Tatsuo Yamashita,⁴ Kota Tsubouchi,⁴ Keiichi Yasumoto¹

¹ 奈良先端科学技術大学院大学

² 立命館大学

³ 東京都市大学

⁴ Yahoo! JAPAN 研究所

⁵ 理化学研究所 革新知能統合研究センター

要旨: 新型コロナウイルス感染症 (COVID-19) の影響により, 全世界の株式市場に多大な影響を及ぼした. これはコロナ・ショック (2020 Stock Market Crash) と呼ばれ, 日本でも 2020 年 2 月中旬から 3 月下旬にかけて, 投資家の心理状態を表す日経平均 VI は上昇し, 日経平均株価が大幅な下落している. 一方ソーシャルメディアには投資家の考えが投稿されており, その話題は日々変わっている. そこで本研究では, ヤフーファイナンス掲示板における話題の変化に注目して分析を行う. コロナ・ショック下におけるファイナンス掲示板での投稿を収集し, 投資家が議論している話題の変化を捉え, 日本市場の株式指標との関係性を明らかにする.

1 はじめに

新型コロナウイルス感染症 (COVID-19) は 2019 年に初めて確認 [1] されてから, 2021 年 1 月現在も世界的に収束の兆しは見えず, 経済的にも大きな影響を及ぼしている. 世界的に流行し始めた 2020 年 2 月下旬から同年 4 月上旬にかけて, 金融業界においても大きな影響を与え, 大規模な株式市場の暴落が発生した. このことはコロナ・ショック (2020 Stock Market Crash) と呼ばれている [2]. アメリカでは, 国内で新型コロナウイルスが初めて確認されたあとの 2020 年 2 月 12 日にダウ工業平均株価が当時の史上最高値を更新し, その後に急落したことから投資家の考え方が大きく変化したことが考えられる. 投資家が市場の先行きに対してどのように想定しているかを表す指数としてボラティリティ・インデックス (VI) があり, ダウ工業平均の VI である VIX は, 2 月の中旬まで 12 から 18 台を推移していたが 2 月 28 日には高値が 49.48 を記録している.

日本でも日経平均株価の VI である日経平均ボラティリティ・インデックス (日経 VI) の推移を見ると, 大幅な上昇を記録しており, アメリカと同様に投資家の考えに変化があったことが推測される.

一方でマイクロブログやソーシャルメディアにおいては特定の話題が大きく取り上げられることが多々あり, この話題の変化や話題に対する投稿における感情を分析することで, 社会の情勢を観測できる [3]. 金融分野においても, ソーシャルメディアに投稿された情報を分析し, 投資行動やトレンドを観測している例が多数存在している [4, 5].

本研究は, コロナ・ショック下におけるファイナンス掲示板での投稿を収集し, そこで議論されている話題の変化と日本市場の株式指標との関係性を明らかにすることを目的とする. ファイナンス掲示板で議論されている話題を抽出するために, トピック分析の一種である LDA [6] を用いて 32 種類の話題を抽出する. この話題の分布が似ている日同士のクラスタリングをするために, 階層的クラスタ分析を用いた. この結果, 5 種類のグループに分類でき, それぞれのグループにお

*連絡先: 奈良先端科学技術大学院大学先端科学技術研究科
〒 630-0192 奈良県生駒市高山町 8916 番地 5
E-mail: sasaki.kodai.sb6@is.naist.jp

ける特徴的な話題と株式指標との関係性を考察する。

2 関連研究

2.1 日経平均株価と日経 VI の関係性

株価の値動きは、効率市場仮説 [7] によりランダムウォークの動きをすることが知られている。これにより株価の値動きを予測することは難しいが、価格変動の変動率 (ボラティリティ) を見ることで、リスク指標に用いることができる。ボラティリティには 2 種類存在しており、過去の価格の変動率をもとにして計算されるヒストリカル・ボラティリティ (HV) と、市場で取引されている実際のオプションの価格 (プレミアム) から逆算して導き出されるインプライド・ボラティリティ (IV) がある。このオプション取引のプレミアムはブラック・ショールズ微分方程式により算出され、この方程式の中には、将来の変動率が含まれている。オプション取引は、将来あらかじめ決められた時点における、派生元の株価を売買する権利のことであり、そこから計算されるボラティリティは、将来時点におけるボラティリティを表すと考えられる [8]。そのためプレミアムから逆算することにより、投資家が将来の派生元の値動きの大きさをどう想定しているかを表す数値を得ることができる。

日本では、日本経済新聞社が日経平均オプションのプレミアムから算出している日経平均ボラティリティ・インデックス (日経 VI) がある。つまり日経 VI は、日経平均株価を対象とした日経平均オプションの価格を用いて算出される、30 日後のボラティリティを表す指数である [8]。加えて、日経平均株価と日経 VI は負の相関をもつことが知られている [9]。日経平均株価と日経 VI の推移を図 1 に示す。図 1 を見ても、日経平均株価の下落時に日経 VI が上昇していることが観測できる。

日経 VI の上昇は、投資家が将来の日経平均株価の先行きに不安があることを示している。株価の値動きは社会のあらゆるイベントによって変化するため、社会イベントに対する投資家の感情や考えなどの変化から日経 VI の変化を捉えられることができれば、リスクマネジメントへの応用可能性が検討できると考える。

2.2 ソーシャルメディアを用いた社会イベントに対する分析

ソーシャルメディアを用いて社会イベントに対する世情の分析が行われている。例えば、大阪都構想に対しての分析 [10] では、LDA を用いて計画への支持・反対を基準にツイートやアカウントを分類し、新聞記事

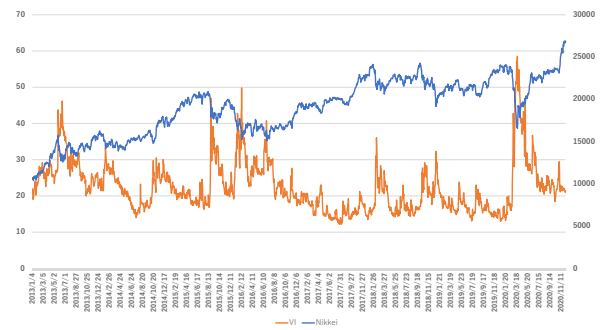


図 1: 日経平均株価と日経 VI の推移

との比較を行っている。またソーシャルネットワーク分析を用いてリツイートネットワークを調査している。この結果、支持と反対が明確に分かれているネットワーク構造が見られ、オピニオンリーダー効果やアナウンス効果の検証が可能であることを示している。Twitter を用いて新型コロナ渦における感情の変化を分析した例 [11] では、新型コロナウイルス関係のツイート群を収集し、ユーザのコミュニティの偏りについて分析した後、ツイートを 10 種類の感情に分類して感情の時間的な推移を分析している。その結果、北海道で緊急事態宣言がなされた 2 月 28 日以降はユーザの偏りはなくなり、一般的な話題になったことを明らかにした。さらにソーシャルメディアから「気の緩み」の存在を推定できる可能性を示唆している。このようにソーシャルメディアの投稿は、社会イベントと関係している。本研究においてもこれらのアプローチを用いて分析を行う。

2.3 本研究の目的

ソーシャルメディアを用いて様々なイベントに対する分析が行われているが、コロナ・ショック下における日本市場に対する分析については、筆者の知る限り存在しない。そこで本研究では、株の取引や社会情勢についての議論が活発に行われているファイナンス掲示板で議論されている話題の変化を分析することで、日経 VI の変動との関係性を捉えることを目指す。

3 データセット

本研究では、コロナ・ショック下における日本市場におけるファイナンス掲示板の話題と日経 VI との関係性を明らかにすることを目的とする。分析のためのデータセットとして、日経平均株価および日経 VI の時系列データは、JPX データクラウド [12] から収集する。対象期間は、2020 年 1 月 6 日から 2020 年 5 月 29 日としている。

分析期間中の日経平均株価の推移を図2に、日経VIの日足四本値の推移を図3に示す。日経平均株価の推移を見ると、1月6日から2月21日にかけての平穏な期間と、2月25日から3月25日までの急落期間、3月26日から5月29日までの回復期間の大きく3つの期間に分けることができる。



図 2: 分析期間中の日経平均株価の推移



図 3: 分析期間中の日経 VI の推移

投稿データは、ヤフーファイナンス掲示板の日経平均株価スレッドを使用する。ヤフーファイナンス掲示板は個別銘柄ごとにスレッドが分かれているが、本研究の分析対象である日経VIは日経平均株価に関連した指数であるため、日経平均株価スレッドを対象とする。分析対象は、2020年1月1日から2020年5月31日までの1,142,408件の投稿文書とする。ヤフーファイナンス掲示板の日経平均株価スレッドの投稿数の推移を図4に示す。なお分析はYahoo! JAPAN 研究所内のサーバにて行った。

4 分析手法

本研究では、次のアプローチを用いる。まずファイナンス掲示板の投稿にどのような話題が含まれているのかを明らかにするため、LDA[6]を用いてトピックを抽出する。次に抽出した話題の分布が時系列でどのように変化しているかを明らかにするために、鳥海ら[11]

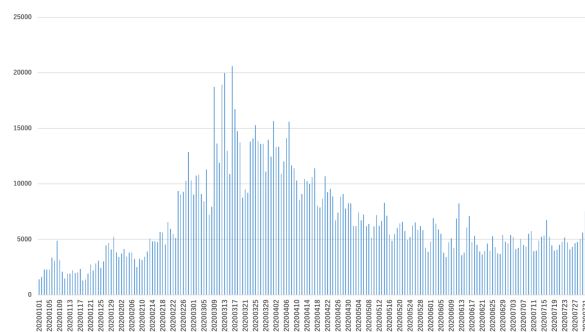


図 4: 投稿数の推移

の手法に習い、日別に集計した話題の時間変化に対する分析を行う。

4.1 トピック分析

最初にファイナンス掲示板の投稿から話題を抽出するために、LDAを用いて「各文書におけるトピックの所属確率」を獲得する。ニュース記事を対象とした研究では、ニュース記事の内容ではなく、見出しを用いることが多い[13, 14, 15]。これはニュースの内容は語彙数が多く、ノイズが発生してしまうことから避けられている。ファイナンス掲示板の投稿においては、内容に荒らしのコメント等がありノイズが多く発生することが予想される。しかしファイナンス掲示板を分析した既存の研究[16]において、荒らしコメントが株価の予測に一定の効果を示すことが報告されているため、本研究では投稿の内容すべてを対象とする。

前処理としてソーシャルメディアの投稿から、URL、HTML、改行コードを除去し、半角ひらがなおよびカタカナを全角に変換する。さらに全角英数字を半角に変換し、形態素解析を行う。形態素解析にはMecab[17]を使用し、辞書にはNeologd[18]を使用する。ソーシャルメディアの分析では、流行りや話題の変化を察知することが重要である。よって毎週辞書を更新し、新規語や固有表現に強いNeologdを使用する。抽出する形態素は、名詞、動詞、形容詞でかつSubtypeが数値、非自立、代名詞、接尾を除く。さらにストップワードの除去を行う。ストップワードのリストは、京都大学が公開しているSlothlibのテキストデータ¹に「ある、する、ちやう、ない、なる、やる」を追加したものを使用する。LDAの実装はgensimを用いて行い、トピック数は32、パラメータはデフォルト値を用いる。

¹<http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

4.2 話題の時間変化

LDA の出力である「各文書におけるトピックの所属確率」を日次にとめることで、日次の話題ベクトルを獲得する。日次のとめる方法は、既存研究 [13, 19] において、日別に各投稿の単純平均を用いているため、本研究においてもこれを適応する。ここで本研究では、株式市場との関係性について分析を行うため、営業日のみの日次話題ベクトルに限定する。話題ごとの変化の度合いを同等に比較するために、鳥海ら [11] と同様に各話題ベクトルの時系列データを平均 0、分散 1 に正規化を行う。

加えて話題の分布の相関を図 5 に示す。図 5 から相関係数の正負が入れ替わっている点があるのがわかる。そこで話題の分布が似通っている日を明らかにするため、階層的クラスター分析を用いて、話題の分布が似通っている日をクラスタリングする。なお、コロナウイルス感染症の集団感染のことを一般的にクラスターと呼ぶことを考慮し、本研究においてクラスター分析を用いて分類されたクラスターのことをこれからグループと呼ぶ。グループ間の距離の定義に ward 法、サンプル間の距離の定義にユークリッド距離を用いて作成したデンドログラムを図 6 に示す。

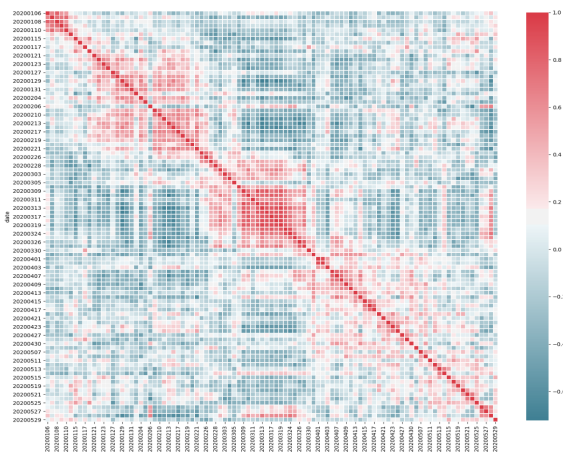


図 5: 日ごとの話題の相関

5 日経 VI と話題の関係

本節では、抽出された話題と日経 VI の関係について分析する。図 6 のデンドログラムを距離 0.06 で分類した結果、5 つのグループに分類された。このときの各グループとその期間を表 1 に示す。また各グループの特徴のある話題の推移を図 7 に示す。

グループ 1 の話題とその上位特徴語の例を表 2 に示す。グループ 1 に所属する日時は、1 月 6 日から 10 日であり、時期的にまだコロナ感染症に関する話題は見

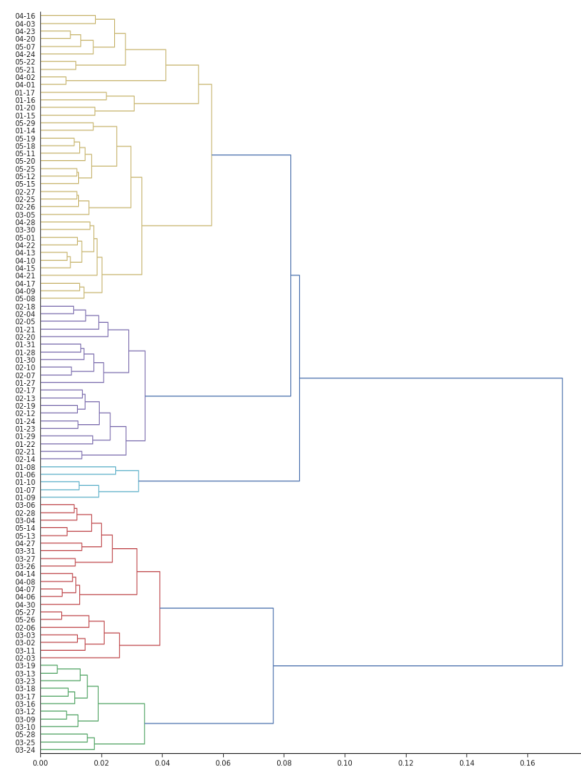


図 6: 日ごとのデンドログラム

表 1: 各グループと所属する日にち

グループ	期間
1	1/6~1/10
2	1/14~1/20, 2/25~2/27, 3/5, 3/30, 4/1~4/3, 4/9~4/13, 4/15~4/24, 4/28, 5/1~5/12, 5/15~5/25, 5/29
3	1/21~1/31, 2/4~2/5, 2/7~2/21
4	2/3, 2/6, 2/28~3/4, 3/6, 3/11, 3/26~3/27, 3/31, 4/6~4/8, 4/14, 4/27, 4/30, 5/13~5/14, 5/26~5/27
5	3/9~3/10, 3/12~3/25, 5/28

当たらない時期である。特徴語からもわかる通り、政治批判等が多く、コロナウイルス感染症に関連した話題は目立っていない。

グループ 2 の話題とその上位特徴語の例を表 3 に示す。特徴語としては、「企業」や「会社」などの個別銘柄に関わることや直近のニュースに対する話題なども多く見受けられた。グループ 2 に所属する日時は、広範囲に分かれており、中でも 4 月 1 日から 5 月 25 日の間に多く存在する。この期間の日経 VI の値としては、20~40 程度が大半を占めており、不安の程度が中程度の時期と一致している。

グループ 3 の話題とその上位特徴語の例を表 4 に示す。グループ 3 に所属する日時は、1 月 20 日から 2 月 21 日（2 月 3 日と 6 日を除く）となっている。このグループでは、コロナ関連の話題も増え、「オンライン」

表 2: グループ 1 での話題と上位特徴語

話題 ID	特徴語
4	批判, 朝日新聞, 森, 政府
16	トランプ, 米国, 強い, 反対
20	売る, 退職金, 発言, 国内

表 3: グループ 2 での話題と上位特徴語

話題 ID	特徴語
15	会社, 世界中, 企業, 失う
22	聞く, ネット, 思う, ニュース
29	投資, 金融, 銀行, 責任

や「自宅」などの生活様式の変化やオリンピックの開催可否の話題なども増えている。一方で政治や国際関係の話題も多く、株取引関連の話題はあまり出てこない。この期間の日経 VI は、これまでの期間よりも上昇しているが、20 台に到達した日は 2 日間ほどであり、市場の不安視は低いと考えられる。

表 4: グループ 3 での話題と上位特徴語

話題 ID	特徴語
1	払う, オンライン, Twitter, 自宅
2	政府, オリンピック, 対策, 辞める
10	国民, 野党, 国会, 立憲

グループ 4 については、全期間にまばらに存在している。話題については、特に多い話題があるわけではなく、平均的な分布となっている。これらのことから、トピックが入れ替わる時期が抽出されていると考える。

グループ 5 の話題と特徴語を表 5 に示す。表 5 の通り、崩壊や暴落など日経平均株価や他国の市場などがあれていることを示す話題が大半を占めている。グループ 5 に所属する日時は、3 月 9 日から 25 日（3 月 11 日を除く）および 5 月 28 日となっている。この期間は、日経平均株価が 1 万円台にまで落ち込み、日経 VI は 40 前半から分析期間中の最高値である 60.86 に達している。

6 おわりに

本研究では、新型コロナウイルス感染症 (COVID-19) の影響で世界の株価が急落したコロナ・ショック下における投資家の考えの変化を明らかにするために、ファイナンス掲示板の話題と日経 VI を分析した。話題の日次の変化を調べるために、ヤフーファイナンス掲示板の投稿を LDA を用いて 32 トピックに分類し、日次にまとめることで、日別の話題ベクトルを獲得した。次に投稿されている話題の分布が似ている日を調査するために、階層的クラスタ分析を用いて 5 種類のグループに分割した。この 5 種類のグループの中で、特徴的

表 5: グループ 5 での話題と上位特徴語

話題 ID	特徴語
5	崩壊, 終わる, 危機, マジで
12	上がる, 下がる, ダウ, 日経平均
32	買う, 株, 暴落, 相場

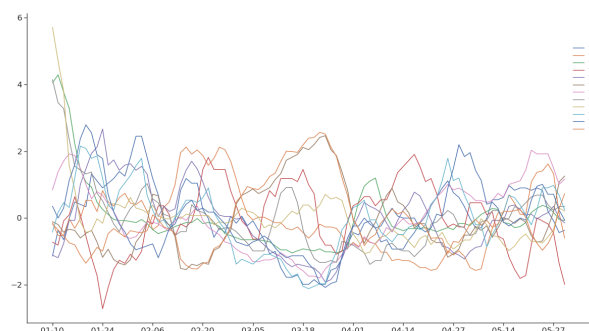


図 7: 各グループの特徴のある話題の推移

な話題と日経 VI の変化を調査した結果、日経 VI が低く、平穏な日には政治的な話題が多いことが明らかとなった。さらにまた、休日における話題の調査や話題の切り替わるタイミングに存在するグループの調査は、今後の課題とする。研究の発展として、話題の分布が切り替わるタイミングを活用し、今後の日経 VI の動向が予測することができれば、投資リスクに対する意思決定支援システムの構築できる可能性が考えられる。

謝辞

本研究の一部は科研費基盤 C(20K01863) による助成を受けて行われた。

参考文献

- [1] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature*, Vol. 579, No. 7798, pp. 270–273, 2020.
- [2] Scott R Baker, Nicholas Bloom, Steven J Davis, Kyle Kost, Marco Sammon, and Tasaneeya Viratyosin. The unprecedented stock market reaction to covid-19. *The Review of Asset Pricing Studies*, Vol. 10, No. 4, pp. 742–758, 2020.
- [3] Peter Sheridan Dodds, Kameron Decker Harris, Isabel M Kloumann, Catherine A Bliss, and

- Christopher M Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one*, Vol. 6, No. 12, 2011.
- [4] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, Vol. 2, pp. 1–8, 2011.
- [5] Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, Vol. 59, No. 3, pp. 1259–1294, 2004.
- [6] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, Vol. 3, No. Jan, pp. 993–1022, 2003.
- [7] Eugene Fama. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, Vol. 25, No. 2, pp. 383–417, 1970.
- [8] 柴田舞. 日経平均ボラティリティ・インデックスの現物と先物の関係—期待仮説による実証分析. 先物・オプションレポート 2019 年 2 月号, 2019.
- [9] 小林寛司, 宮崎浩一. 資産評価モデルにおける短期・長期ボラティリティの影響. 情報処理学会論文誌数理モデル化と応用 (TOM), Vol. 5, No. 1, 2012.
- [10] Kouki Hayashi, Eiichi Umehara, and Yuuki Ogawa. Analysis of twitter messages about the osaka metropolis plan in japan. In *2017 IEEE International Conference on Big Data (Big Data)*, pp. 3064–3070. IEEE, 2017.
- [11] 鳥海不二夫, 榑剛史, 吉田光男. ソーシャルメディアを用いた新型コロナ禍における感情変化の分析. 人工知能学会論文誌, Vol. 35, No. 4, 2020.
- [12] Jpx data cloud. available from <http://db-ec.jpx.co.jp/item/C430509.html>. Tokyo Stock Exchange, Inc.
- [13] Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. Deep learning for event-driven stock prediction. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.
- [14] Xin Du and Kumiko Tanaka-Ishii. Stock embeddings acquired from news articles and price history, and an application to portfolio optimization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3353–3363. Association for Computational Linguistics, July 2020.
- [15] 五島圭一, 高橋大志, 寺野隆雄. ニュースのテキスト情報から株価を予測する. 人工知能学会全国大会論文集 第 29 回全国大会 (2015). 一般社団法人人工知能学会, 2015.
- [16] 山下達雄, 坪内孝太. 株価掲示板情報における煽り情報の検出. 人工知能学会全国大会論文集 第 29 回全国大会. 一般社団法人人工知能学会, 2015.
- [17] Taku Kudo. Mecab: Yet another part-of-speech and morphological analyzer. <http://mecab.sourceforge.jp>, 2006.
- [18] 佐藤敏紀, 橋本泰一, 奥村学. 単語分かち書き辞書 mecab-ipadic-neologd の実装と情報検索における効果的な使用方法の検討. 言語処理学会第 23 回年次大会 (NLP2017). 言語処理学会, 2017.
- [19] Kodai Sasaki, Hirohiko Suwa, Yuki Ogawa, Eiichi Umehara, Tatsuo Yamashita, and Kota Tsubouchi. Evaluation of vi index forecasting model by machine learning for yahoo! stock bbs using volatility trading simulation. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020.