

ソーシャルメディアの感情分析を活用した

株価予測モデルの精度向上

慶應義塾大学 経済学部 長倉大輔研究会

芳賀 貴生 ・ 佐々木 求道 ・ 呉 勝洋 ・ 葛 シュガン

要約

この研究は、ソーシャルメディア、特に Reddit の「wallstreetbets」コミュニティにおける感情分析を用いたダウ平均株価の予測精度の比較から、感情データの株価予測における有用性を検証することを目的としたものである。データ収集には、2022 年から 2024 年までの Reddit のコメントを使用し、感情分析ツールである VADER をベースとした、独自の辞書を用いた感情分析を行った。分析を基に算出した感情スコア、及びダウ平均株価を特徴量とする LSTM を用いた予測モデルの予測精度が、株価データのみを使用したモデルと比較して向上するかを検証した。結果として、感情スコアを用いたモデルは予測精度が高く、特に離散的な特徴を持つ感情スコアを使用したモデルの予測精度が高くなることを示した。

キーワード：機械学習, 株価予測, LSTM, 感情分析

目次

- 1 はじめに
- 2 先行研究
 - 2.1 先行研究の紹介, 及び課題の提示
 - 2.2 今回の研究目的
- 3 分析
 - 3.1 使用したデータ
 - 3.1.1 Reddit のデータ収集
 - 3.1.2 株価のデータ収集
 - 3.2 データ処理
 - 3.2.1 センチメント分析
 - 3.2.2 日次データ化
 - 3.2.3 ダウ平均株価への反映
 - 3.3 分析手法
 - 3.3.1 モデルの構築
 - 3.3.2 モデルの予測精度の検定
- 4 結果
 - 4.1 それぞれのモデルの予測結果及び予測精度
 - 4.2 モデルの予測精度の検定結果
- 5 考察
 - 5.1 モデル 2 について
 - 5.2 モデル 3 について
 - 5.3 モデル 4 について
 - 5.4 モデル 5 について
 - 5.5 モデル 6 について
 - 5.6 モデル 7 について
 - 5.7 モデル 8 について
- 6 終わりに
 - 6.1 今回の研究から得られる示唆
 - 6.2 今後の課題
 - 6.3 謝辞
- 7 参考文献
- 8 付録

1 はじめに

金融市場とソーシャルメディア上の言論との間の関連性は、ソーシャルメディアの黎明期である 2010 年代前半から現代までの間に度々注目を浴びてきた。いくつかの論文では、ソーシャルメディア上の言論が、金融市場に対する社会の動向や世論を反映するセンサーとして機能する可能性を指摘しており、Bollen(2010)は、ソーシャルメディアプラットフォームである Twitter 上のテキストデータを用いた感情分析を用いることで、ソーシャルメディアにおける人々の感情を抽出し、それらが金融市場における「ムード」と関連していることを明らかにしたことに加え、感情データを用いたダウ平均株価の予測モデルを構築することで、株価予測における感情データの有用性を示した。また、近年では機械学習モデルの発達に伴い、ニューラルネットワークなどのアルゴリズムを用いた金融市場の分析の試みが増加しているほか、ニュース記事や感情データ等のオルタナティブデータを用いた株価予測の研究も注目を集めており、感情分析を用いた金融市場の予測に関する研究は、今後ますます活発化していくことが予想される。株価予測に感情データが役立つかどうかは未だ議論が続いているが、Wang(2018)や Adekoya (2021)等に代表される、近年の多くの研究において概ね肯定的な結果が示されていることは確かである。実際、2021 年初頭に米国の GameStop 社の株価が急激に変動した、いわゆる「GameStop 騒動」においては、零細な個人投資家らがソーシャルメディア上で連携し、協調して株式の買い注文を出したことで、株価が前年来安値の約 190 倍にまで暴騰するという事態が発生した。これは、ソーシャルメディアの動向が単に市場の動向を反映するだけでなく、それらの動向が市場の動向そのものに影響を及ぼす可能性があることを示唆している。しかしながら、ソーシャルメディアの感情分析を用いた金融市場の分析を行った過去の研究の多くは、感情分析を行うデータの範囲を数週間、あるいは数か月などの比較的短期間に限定していることに加え、それらが研究の対象とした個別のソーシャルメディアの特徴によっても結果に差異が見られるほか、金融市場に対する人々の感情を分析する上で、それらの感情をある一つの尺度による連続的なデータに落とし込むのか、はたまた「買う」、「買わない」、「売る」というような離散的なデータに落とし込むのかという、金融市場の分析に使用する上で適切な感情データの特徴に関する検証が不足している。こうした背景から私たちは、GameStop 騒動の発端となったソーシャルメディアである「Reddit」及びそのプラットフォームにおける投資家コミュニティである「wallstreetbets」に注目し、投資家コミュニティのみを対象とした 2 年間という比較的長期間にわたるデータの感情分析や、異なる特徴を持つ感情データを用いた予測モデルの構築を行うという、これまでの研究とは違う観点からの検証を行う意義があると考えた。本研究では、Reddit のコミュニティである「wallstreetbets」の投稿に対して感情分析を行い、抽出した感情データを用いてダウ平均株価の予測モデルを構築し、その精度を検証した。

2 先行研究

2.1 先行研究の紹介, 及び課題の提示

Wang(2018)は, インターネット上のニュースサイト(New York Times)の記事に対して感情分析を行い, そこで得られた感情データ及びダウ平均株価を組み合わせ, 複数のニューラルネットワークモデルを用いた株価予測モデルを作成することで, ニュースの感情が株式市場と関係していること, また, ニューラルネットワークによる感情データの学習が株価予測の精度を向上させることを示した. また, Adekoya(2021)は, Twitter のツイート, インターネット上のニュース, フォーラムディスカッションなどの複数のデータを用いて感情分析を行い, そこで得られた感情データと株価データを独自のニューラルネットワークモデル(IKN-ConvLSTM)に学習させることで, ガーナ株式市場の 2 社の株価予測を行った. この研究も同様に, 感情データによって株価予測の精度が向上することを示した. 他にも, Bollen(2010)は Twitter のテキストデータの感情分析を通して複数の感情データの抽出を行い, 特定の感情とダウ平均株価との間に相関関係, 及び Granger 因果性があること, また, 感情データを学習させたニューラルネットワークを用いたダウ平均株価予測モデルは, 予測精度が向上することを示した. 以上の先行研究は, ソーシャルメディアの感情分析を用いた, 人々の株式市場に対する感情データの抽出の妥当性や, 非線形関係を捉えることが可能なニューラルネットワーク, 及び感情データの株価予測における有用性を明確に示している. しかしながら, これらの研究においては, 分析対象期間が数週間, もしくは数か月であることから, 長期的なトレンドを持つと考えられる株価・感情データの十分な分析が出来ていないという点や, 感情分析に用いたソーシャルメディアの特性上, 株式市場に関連しないノイズの情報が感情データに反映されやすいという点, また, 株式市場に対する人々の感情はある一つの尺度による連続的な特徴を持つのか, はたまた「買う」「買わない」「売る」というような離散的な特徴を持つのかという, 株価予測に用いる上で適切な感情データの特徴に関する検証が不足しているなどの点において, 未だ課題が残されている.

2.2 今回の研究目的

これら先行研究を踏まえ, 感情データは株価予測において有用であるのか, また, どのような感情データを用いることが適切であるのかを検証するという目的のもと, 海外のソーシャルメディアである「Reddit」上に存在する投資家コミュニティの「wallstreetbets」における 2 年分のデータを利用し, 複数の種類の感情データの抽出, 及びそれら感情データと株価データを特徴量として持つ LSTM を用いた複数の予測モデルの構築, 及び予測精度の比較を行った.

3 分析

3.1 使用したデータ

3.1.1 Reddit のデータ収集

今回は、Institute for Reproducible Research が運営するプラットフォームである Academic Torrents から Reddit のコメントデータセットを取得した。このプラットフォームには、2005 年 6 月から 2024 年 1 月までの Reddit のコメントおよび投稿が ZST 圧縮ファイル形式でアーカイブされている。本研究では株式取引関連の Subreddit である「wallstreetbets」のデータをダウンロードし、Python スクリプトを用いて 2022 年 1 月から 2024 年 1 月までのデータを抽出した後、ダウ平均株価を構成する企業に関連するコメントを特定した。この期間の設定は、2021 年初頭に発生した GameStop 事件による投稿数の急増や、2024 年 2 月のダウ平均構成銘柄の変更の影響を避けるためである。抽出したデータは、ZST 形式から CSV 形式に変換し、CSV ファイルにはコメントの upvote 数と downvote 数の差 (fav)、コメントの投稿日時、ユーザー名、コメントへのリンク、コメントの全文を含んでいる。また、抽出したデータのうち fav が 0 未満のコメントのデータにおける fav を 0 とする。これは、fav が低い不人気なコメントは不正確な情報や感情に基づくコメントやスパムであることが多く、後の処理でこれらのコメントの影響を取り除くためである。

3.1.2 株価のデータ収集

ダウ平均株価および構成企業の株価データの取得には、Excel の関数である STOCKHISTORY 関数を用いた。この関数によって得られるデータは、ロンドン証券取引所グループ傘下の企業である Refinitiv が提供している。

3.2 データ処理

3.2.1 センチメント分析

Reddit のコメントのセンチメント分析を行うにあたって、ソーシャルメディアの感情分析を目的とした辞書及びルールベースの感情分析ツールである VADER (Valence Aware Dictionary and sEntiment Reasoner) を利用した。VADER の辞書 (vader_lexicon) は株取引に関する単語が十分羅列されていないため、Long(2022)の研究で制作された、「wallstreetbets」上で頻出する 130 単語のスコアを追加し辞書の補完を行った。コメントの感情スコア (score) は辞書に従って付けられた各単語の感情スコアを合計して、-1 から +1 の間で正規化した値とする。VADER による詳細なスコア付けの手法は Hutto(2014)の研究を参照されたい。また、感情スコアを単純化したデータとして score が 0.5 以下のコメントに -1、0.5 超で 0.5 未満のコメントに 0、0.5 以上のコメントに 1 を与える type、コメントごとの影響力の違いを反映させるため各コメントで score と fav の積を取る weighed を追加した。

3.2.2 日次データ化

企業および日付ごとに異なるコメント数を統一するため、同一日付の `weighed` および `type` の平均を取得し、日次データとした。また、同一日付における `type` の最頻値を取った `represent`, ダウ平均株価終値を各株価終値で割った `ratio` を新たに追加し, `weighed` と `ratio` の積を取った `sentiment`, `type` と `ratio` の積を取った `type_score`, `represent` と `ratio` の積を取った `represent_score` という日次感情スコアを新たに作成した。ここで、市場が開いていない日の感情スコアについては、直前の開場日のスコアと合計し平均を取ることで 1 日分のデータに集約した。以下の表 1 が、今回用意したデータセットの一例である。

表 1 処理前後のデータセット

処理前のデータセット（企業：スリーエム）

	fav	date	score
0	0	2022-01-01	-0.7269
1	1	2022-01-01	0.6597
2	1	2022-01-01	-0.6562
3	2	2022-01-01	0.0000
4	3	2022-01-01	0.7656

処理後のデータセット

	date	fav	score	weighed	ratio	sentiment	type	represent	type score	represent score
0	2022/1/1	1.375000	-0.085587	0.105813	0.004062	0.000430	-0.125000	-1	-0.000508	-0.004062
1	2022/1/2	2.300000	0.172670	0.521640	0.004062	0.002119	0.100000	0	0.000406	0.000000
2	2022/1/3	3.700000	0.014205	0.111535	0.004062	0.000453	0.100000	0	0.000406	0.000000
3	2022/1/4	2.172414	0.130710	0.281031	0.004119	0.001158	0.103448	0	0.000426	0.000000
4	2022/1/5	2.115385	0.058785	-0.114112	0.004102	-0.000468	0.038462	0	0.000158	0.000000

ここで、`sentiment`, `type_score`, `represent_score` はそれぞれ、「連続的」「連続的・離散的の両方」「離散的」の特徴をもつ感情スコアとして抽出している。これらの異なる特徴を持つ感情スコアを用いて分析を行うことで、株価予測に適切な感情データを検証した。

3.2.3 ダウ平均株価への反映

ratio で重み付けした日次感情スコア (sentiment) の 30 社分の合計値を用い、ダウ平均株価に対する日次感情スコアを-1 から 1 で表現した。type_score および represent_score についても同様に、30 社分合計した値をダウ平均株価に対する日次感情スコアとして用いた。

3.3 分析手法

3.3.1 モデルの構築

感情データが株価予測に有用であるかを検証するため、上述した 3 種類の感情スコア、及びそれらを組み合わせた 4 種類の感情データに加えて、ダウ平均株価の終値を特徴量として用いた株価予測モデルを構築し、合計 7 種類のモデルの予測精度を、ダウ平均株価の終値のみを特徴量として用いた予測モデルと比較した。また、株価予測における感情データの適切な扱い方についても検証するために、感情スコアを特徴量として用いたモデル間でも、予測精度の比較を行った。モデルの構築には、株価や感情データの非線形な関係性を捉える上で有用であり、多くの先行研究でも採用されていたニューラルネットワークを用い、中でも時系列データを扱うことが可能な「回帰型ニューラルネットワーク (RNN)」を用いた。RNN は、従来のニューラルネットワークとは異なり、中間層に「回帰結合」を持つことで、系列データの取り扱いを可能にしている。回帰結合とは、中間層における出力を再び中間層に入力することで、系列の前後間の関係をニューラルネットワークに学習させる仕組みである。例えば、入力にベクトル系列 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ 、出力にベクトル系列 $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$ が対応する場合、RNN の中間層における n 番目の出力 $\mathbf{h}_n$ および出力層における n 番目の出力 $\mathbf{y}_n$ は、以下のように表される。

$$\mathbf{h}_n = \tanh(\mathbf{W}_h \mathbf{h}_{n-1} + \mathbf{W}_x \mathbf{x}_n)$$

$$\mathbf{y}_n = \tanh(\mathbf{W}^y \mathbf{h}_n)$$

ここで、 $\mathbf{h}_{n-1}$ は $n-1$ 番目の中間層の出力であり、 $\mathbf{W}_h, \mathbf{W}_x, \mathbf{W}^y$ はそれぞれに対応する重みベクトルとなっている。また、上式では出力を決定づける活性化関数に $\tanh$ 関数 (双曲線正接関数) を用いており、これは入力値を $[-1, 1]$ の値に変換する関数である。しかしながら、RNN には「勾配消失問題」、もしくは「勾配爆発問題」という致命的な欠点が存在し、特に今回のような長期に渡る時系列データの分析には、単純な RNN を用いることは難しい。そこで、今回のモデルの構築に際して、それらの問題を解決した RNN の一種である、「LSTM (Long Short-Term Memory)」を用いた。LSTM では、中間層に新たに「LSTM Block」という処理層を加えることで、一般的な RNN では扱えなかった、系列の長期の関係を学習することが可能になっている。LSTM に関するより詳細な解説については、Hochreiter, S., & Schmidhuber, J. (1997)の研究を参照されたい。そして、今回の研究では、Python のライブラリである Keras を用いて LSTM を構築し、60 日分の時系列データから 61 日目のダウ

平均株価の終値を予測するモデルを構築した。

3.3.2 モデルの予測精度の検定

合計 4 種類のモデルの予測精度の評価に際し、まず、「二乗平均平方根誤差 (RMSE)」「決定係数 (R^2)」「平均絶対パーセント誤差 (MAPE)」の 3 つの値を利用した。この内、「二乗平均平方根誤差」は、以下の数式で表される。

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

ここで、 $\hat{y}_i$ はモデルの予測値、 y_i は実測値、 n はサンプル数を表している。また、「決定係数」は以下の数式で表される。

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ここで、 $\bar{y}$ は実測値の平均値を表している。次に、「平均絶対パーセント誤差」は以下の数式で表される。

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

これらの値を用いて、それぞれのモデルの予測精度の検証を行っていくが、これらの値の差は、単に標本の偏りによって生じている可能性があるため、単純に値そのものを比較してモデルの精度を評価することはできない。そこで、二つのモデルの予測誤差の差を利用し、それぞれのモデルの予測精度の違いを検定する「Diebold-Mariano 検定」を用いた比較を行った。Diebold-Mariano 検定では、2 つの予測誤差の系列 $\{e_t\}_{t=j}^T$ と $\{e_t\}_{t=j}^T$ が与えられる際、損失関数 $L(e)$ を用いて表される損失差を

$$d_t = L(e_t) - L(e'_t)$$

とし、「損失差の期待値が 0 であること」を帰無仮説、「損失差の期待値が 0 でないこと」を対立仮説に設定する。すなわち、

$$H_0: E(d_t) = 0 \quad H_1: E(d_t) \neq 0$$

である。この帰無仮説に対して、以下の検定統計量を用いて検定を行う。

$$DM_T = \frac{\bar{d}_T}{se(\bar{d}_T)}$$

ここで、 $\bar{d}_T$ は損失差の平均、 $se(\bar{d}_T)$ は損失差の標準誤差である。この検定統計量に基づき p 値を計算し、2 つのモデルの予測精度に統計的に有意な差が存在するかを検定した。

4 結果

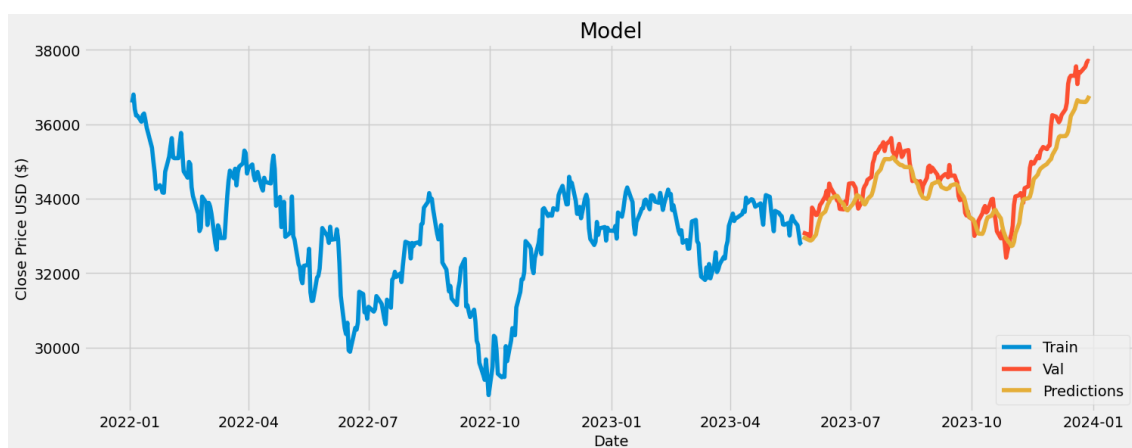
4.1 それぞれのモデルの予測結果及び予測精度

以下の表 2 がそれぞれのモデルの予測結果及び予測精度である。用意した 4 種類の時系

列データの内、前半の 70%を学習データとして用いそれぞれの LSTM を学習させ、残りの 30%をテストデータとして予測精度の計測を行う、という一連の処理を 30 回繰り返し、それらの平均を取った値を最終的な予測精度の結果とした。グラフの横軸には日時、縦軸には株価の値を取り、青線は学習期間のデータ、赤線・橙色はそれぞれ実測値・予測値を表している。

表 2 各 LSTM モデルの予測結果のグラフ及び予測精度

モデル 1. 入力データ：ダウ平均株価

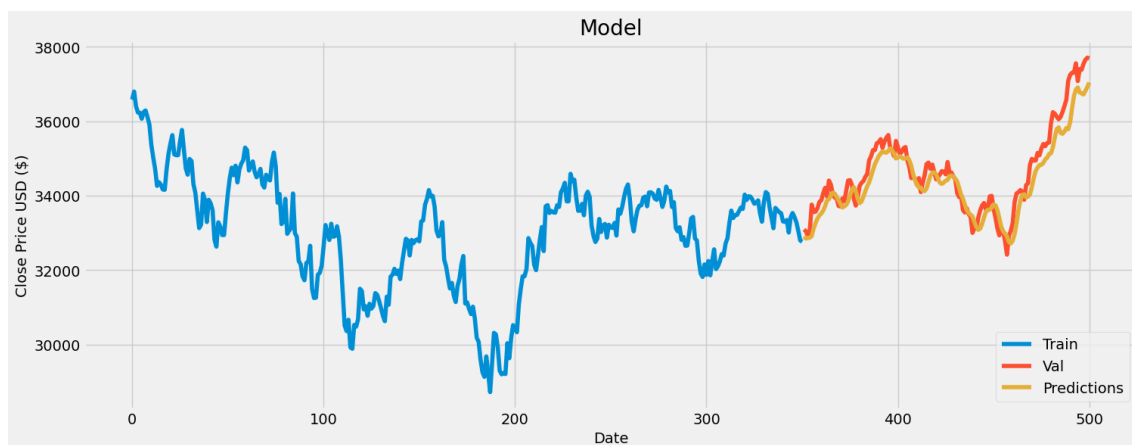


RMSE : 525.7298

決定係数 : 0.7854

MAPE : 1.2572

モデル 2. 入力データ：ダウ平均株価, sentiment

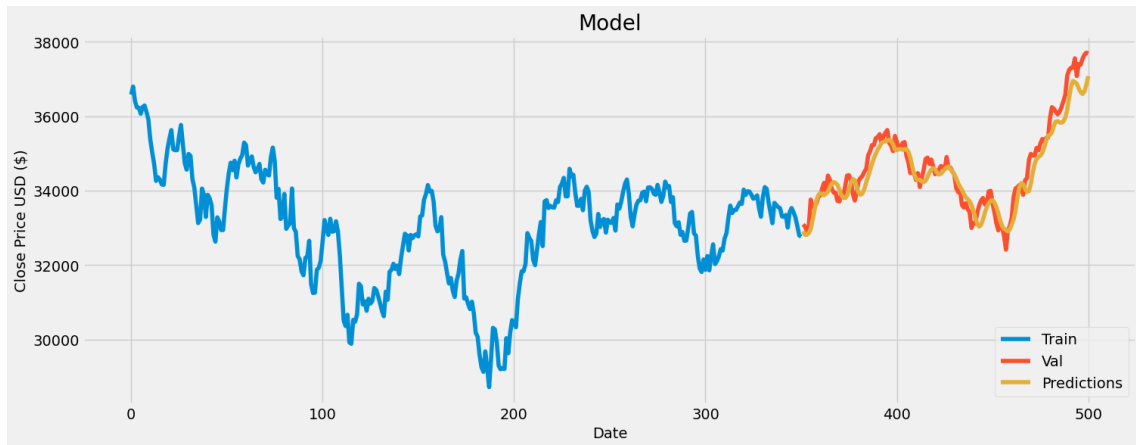


RMSE : 441.2590

決定係数 : 0.8533

MAPE : 1.0280

モデル 3. 入力データ：ダウ平均株価, type_score

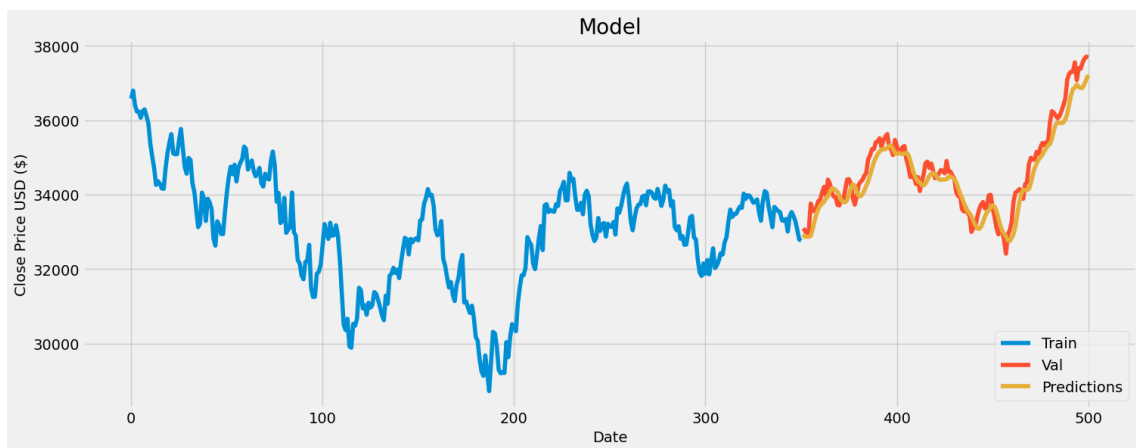


RMSE : 406.0711

決定係数 : 0.8759

MAPE : 0.9474

モデル 4. 入力データ：ダウ平均株価, represent_score

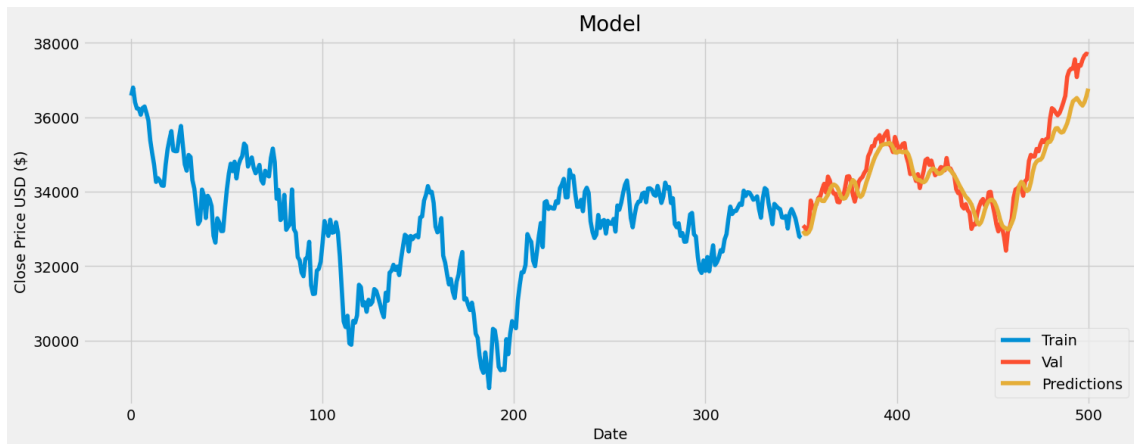


RMSE : 395.6401

決定係数 : 0.8802

MAPE : 0.9387

モデル 5. 入力データ：ダウ平均株価, sentiment, type_score

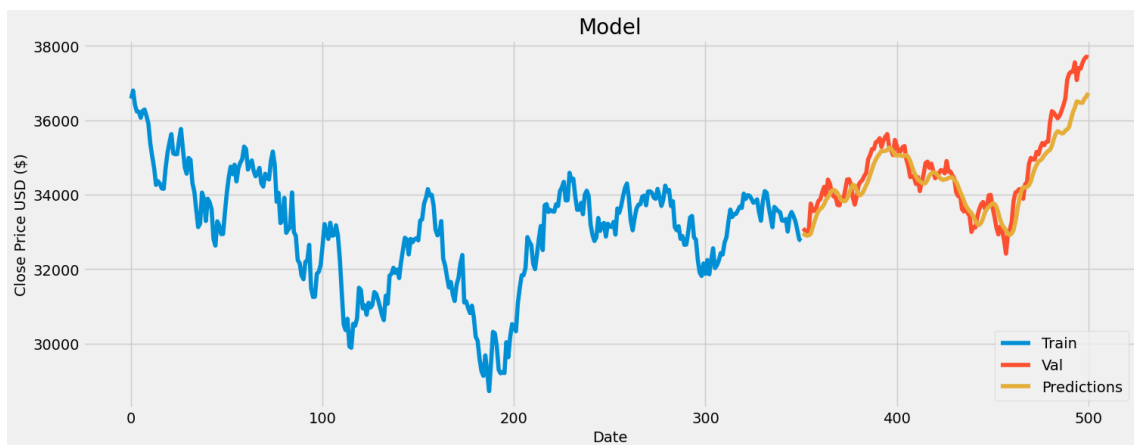


RMSE : 480.4314

決定係数 : 0.8240

MAPE : 1.0761

モデル 6. 入力データ : ダウ平均株価, sentiment, represent_score

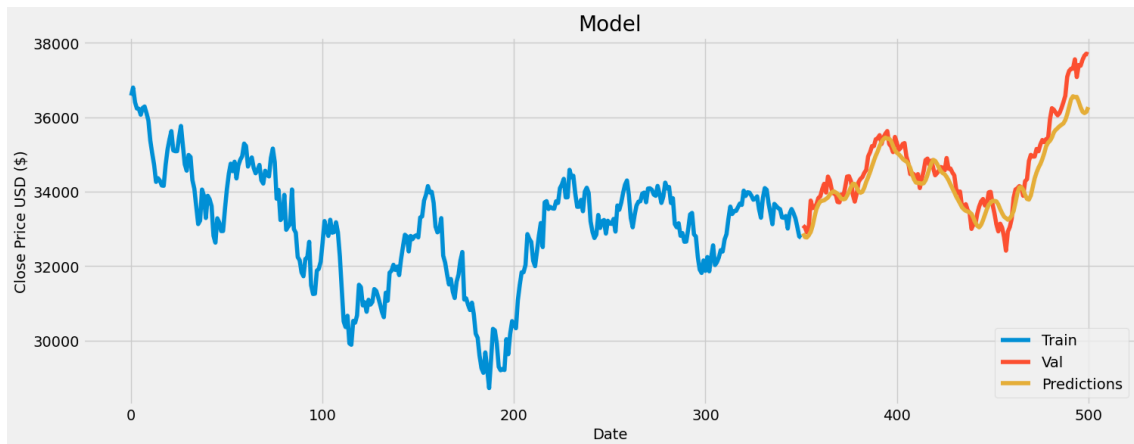


RMSE : 490.6780

決定係数 : 0.8198

MAPE : 1.0985

モデル 7. 入力データ : ダウ平均株価, type_score, represent_score

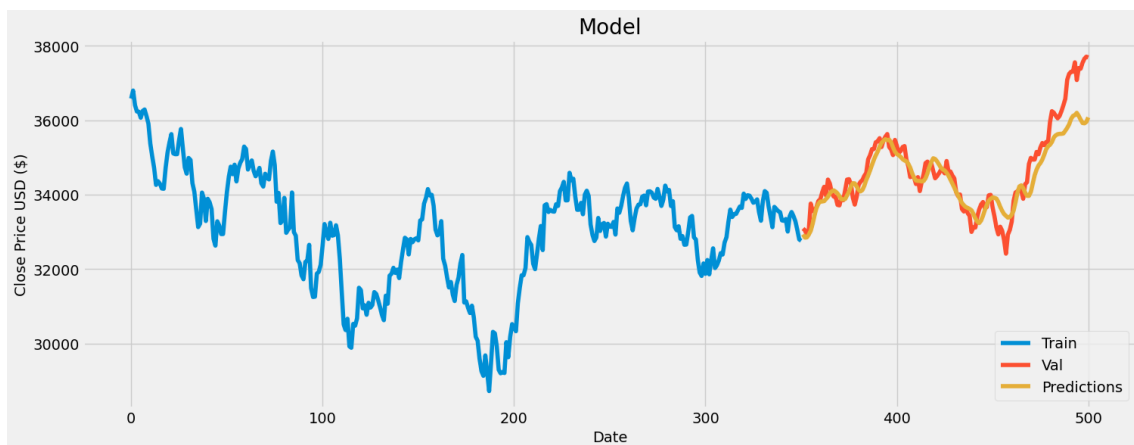


RMSE : 519.3818

決定係数 : 0.7941

MAPE : 1.1715

モデル 8. 入力データ : ダウ平均株価, sentiment, type_score, represent_score



RMSE : 553.1154

決定係数 : 0.7713

MAPE : 1.1415

※重みの最適化アルゴリズムには「Adam」, 損失関数には「最小二乗誤差 (MSE)」を用いた.

※バッチサイズは 32, エポック数は 100 に設定した.

4.2 モデルの予測精度の検定結果

以下の表 3 がそれぞれのモデルの予測精度の Diebold-Mariano 検定の結果である. ここで, 損失差の計算において, 誤差関数には平均絶対誤差(MAE), 平均二乗誤差(MSE), Huber

損失を用いた．それぞれの関数は以下の通りに表される．

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

$$Huber\ Loss = \begin{cases} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} (\hat{y}_i - y_i)^2 & \text{for } |\hat{y}_i - y_i| \leq \delta \\ \frac{1}{n} \sum_{i=1}^n \delta \left(|\hat{y}_i - y_i| - \frac{1}{2} \delta \right) & \text{otherwise} \end{cases}$$

ここで， $\hat{y}_i$ はモデルの予測値， y_i は実測値， n はサンプル数を表している．また，今回の計算において， $\delta = \frac{1}{n} \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2}$ と設定した．

表 3 モデルの予測精度の検定結果

		誤差系列2						
		モデル2	モデル3	モデル4	モデル5	モデル6	モデル7	モデル8
モデル1	Diebold-Mariano統計量	16.5303	12.0840	16.2105	11.5671	17.1841	9.3277	13.7801
	p値	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***
	検定結果	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり
	Diebold-Mariano統計量	13.1726	10.9713	12.4124	11.3126	14.4111	7.7631	12.8601
	p値	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***
	検定結果	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり
モデル2	Diebold-Mariano統計量	16.6534	12.4899	16.5039	12.0500	17.4566	9.4854	14.265
	p値	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***	0.0000 ***
	検定結果	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり
	Diebold-Mariano統計量		1.7698	4.7349	3.2075	-6.4547	-2.0134	2.6040
	p値		0.0788	0.0000 ***	0.0016 ***	0.0000 ***	0.0459 *	0.0101 *
	検定結果		有意差なし	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり
モデル3	Diebold-Mariano統計量		1.6333	4.3189	3.5006	-5.2571	-2.2692	0.8225
	p値		0.1045	0.0000 ***	0.0006 ***	0.0000 ***	0.0247 *	0.4121
	検定結果		有意差なし	有意差あり	有意差あり	有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量		1.8644	5.0534	3.6021	-6.3741	-1.9805	2.7591
	p値		0.0642	0.0000 ***	0.0004 ***	0.0006 ***	0.0495 *	0.0065 **
	検定結果		有意差なし	有意差あり	有意差あり	有意差あり	有意差あり	有意差あり
モデル4	Diebold-Mariano統計量			0.8390	2.7542	-5.3692	-3.9235	0.9522
	p値			0.4028	0.0066 **	0.0000 ***	0.0001 ***	0.3426
	検定結果			有意差なし	有意差あり	有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量			1.5669	2.4472	-5.3820	-4.2173	-0.9881
	p値			0.1193	0.0156 *	0.0000 ***	0.0000 ***	0.3247
	検定結果			有意差なし	有意差あり	有意差あり	有意差あり	有意差なし
モデル5	Diebold-Mariano統計量			1.1818	3.0618	-5.5051	-3.9988	1.0473
	p値			0.2392	0.0026 **	0.0000 ***	0.0001 ***	0.2967
	検定結果			有意差なし	有意差あり	有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量				1.3060	-7.9629	-3.9149	0.0865
	p値				0.1936	0.0000 ***	0.0001 ***	0.9312
	検定結果				有意差なし	有意差あり	有意差あり	有意差なし
モデル6	Diebold-Mariano統計量				0.2999	-6.4292	-3.8047	-1.8579
	p値				0.7647	0.0000 ***	0.0002 ***	0.0652
	検定結果				有意差なし	有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量				1.1963	-8.2049	-4.1771	-0.1702
	p値				0.2335	0.0000 ***	0.0001 ***	0.8651
	検定結果				有意差なし	有意差あり	有意差あり	有意差なし
モデル7	Diebold-Mariano統計量					-6.1257	-5.0800	-1.7112
	p値					0.0000 ***	0.0000 ***	0.0891
	検定結果					有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量					-6.8244	-4.5648	-3.0765
	p値					0.0000 ***	0.0000 ***	0.0025 **
	検定結果					有意差あり	有意差あり	有意差あり
モデル8	Diebold-Mariano統計量					-6.5232	-5.2360	-1.8434
	p値					0.0000 ***	0.0000 ***	0.0673
	検定結果					有意差あり	有意差あり	有意差なし
	Diebold-Mariano統計量						1.5194	7.8439
	p値						0.1308	0.0000 ***
	検定結果						有意差なし	有意差あり
モデル9	Diebold-Mariano統計量						0.5325	8.3345
	p値						0.5952	0.0000 ***
	検定結果						有意差なし	有意差あり
	Diebold-Mariano統計量						1.5293	8.2169
	p値						0.1283	0.0000 ***
	検定結果						有意差なし	有意差あり
モデル10	Diebold-Mariano統計量							5.3913
	p値							0.0000 ***
	検定結果							有意差あり
	Diebold-Mariano統計量							4.1554
	p値							0.0001 ***
	検定結果							有意差あり
モデル11	Diebold-Mariano統計量							5.4836
	p値							0.0000 ***
	検定結果							有意差あり
	Diebold-Mariano統計量							
	p値							
	検定結果							

※誤差系列 1 および 2 は、損失差を $d_t = L(e_t) - L(e'_t)$ としたときの $\{e_t\}_{t=1}^T$ と $\{e'_t\}_{t=1}^T$ にそれぞれ対応する。

※有意性の説明：p<0.05 ”*” p<0.01 ”***” p<0.001 “***”

5 考察

5.1 モデル 2 について

モデル 1 とモデル 2 の RMSE, 決定係数, MAPE を比較すると, 全ての値においてモデル 2 の予測精度が優れていることを示す結果となったことが分かる. また, Diebold-Mariano 検定の結果についても同様であり, 予測精度には統計的に有意な差があるという結果になった. 以上の結果は, 株価の予測における感情データの有用性を示唆している. 一方, モデル 3, 4 とモデル 2 を比較すると, RMSE, 決定係数, MAPE ではいずれのモデルに対してもモデル 2 の精度が劣っていることを示しているが, Diebold-Mariano 検定の結果を見ると, モデル 3 との間には予測精度に有意な差がないこと, モデル 4 との間には予測精度に統計的な差があることを示す結果となった. また, モデル 5~8 とモデル 2 を比較すると, RMSE, 決定係数, MAPE ではいずれのモデルもモデル 2 に対して精度が劣っていることを示す結果となったが, Diebold-Mariano 検定の結果を見ると, おおむね全ての検定でモデル 2 の予測精度に有意な差があること, つまりはモデル 2 の予測精度が劣後していることを示す結果となった. この結果は, 株価の予想において感情スコアを用いる際には, モデル 2 で用いた感情スコアである `sentiment` のような連続的な感情スコアのみを用いるよりも, モデル 4 で用いた `represent_score` のような離散的な性質を持つ感情スコアや, 連続的・離散的両方の性質を持つ `type_score` のような感情スコアを組み合わせる用いるのが適切であることを示唆している.

5.2 モデル 3 について

モデル 1 とモデル 3 の RMSE, 決定係数, MAPE を比較すると, 全ての値においてモデル 3 の予測精度が優れていることを示す結果となったことが分かる. また, Diebold-Mariano 検定の結果についても同様であり, 予測精度には統計的に有意な差があるという結果になった. 以上の結果は, 株価の予測における感情データの有用性を示唆している. 一方, モデル 2, 4~8 との比較については, RMSE, 決定係数, MAPE では, モデル 2, 5~8 に対してモデル 3 の精度が優れていることを示す結果となった一方, モデル 4 に対しては精度が劣っていることを示す結果となった. また, Diebold-Mariano 検定については, モデル 2, 4, 8 で精度に有意差なし, モデル 5~7 においては有意差ありで精度が劣後しているという結果となった. この結果は, 株価の予測に用いる感情スコアについて, 連続的, 離散的, もしくはその両方の性質を持つ感情スコアの間で, その適切さに差はないことを示唆しているほか, 複数の感情スコアを適切に組み合わせることで予測精度が向上することを示唆している.

5.3 モデル 4 について

モデル 1 とモデル 4 の RMSE, 決定係数, MAPE を比較すると, 全ての値においてモデル 4 の予測精度が優れていることを示す結果となったことが分かる. また, Diebold-

Mariano 検定の結果についても同様であり、予測精度には統計的に有意な差があるという結果になった。以上の結果は、株価の予測における感情データの有用性を示唆している。一方、モデル 2, 3, 5~8 との比較については、RMSE, 決定係数, MAPE では、モデル 2, 3, 5~8 の全てに対してモデル 4 の精度が優れていることを示す結果となった一方で、Diebold-Mariano 検定では、モデル 2, 6, 7, との間には予測精度に有意差があり、2 に対しては精度が優れている、6, 7 に対しては精度が劣後しているという結果となり、モデル 3, 5, 8 との間には有意差がないという結果となった。この結果は、株価の予測に用いる感情スコアには、sentiment のような 連続的な感情スコアよりも、represent_score のような 離散的な感情スコアを用いるのが適切であることを示唆する一方、連続的、離散的両方の性質を持つ type_score のような感情スコアと、represent_score のような離散的な感情スコアとの間で、その適切さに差はないことを示唆するほか、複数の感情スコアを適切に組み合わせることで、予測精度が向上することを示唆している。

5.4 モデル 5 について

モデル 1 とモデル 5 の RMSE, 決定係数, MAPE を比較すると、全ての値においてモデル 5 の予測精度が優れていることを示す結果となったことが分かる。また、Diebold-Mariano 検定の結果についても同様であり、予測精度には統計的に有意な差があるという結果になった。以上の結果は、株価の予測における感情データの有用性を示唆している。一方、モデル 2~4, 6~8 との比較については、RMSE, 決定係数, MAPE では、モデル 2~4 の全てに対してモデル 5 の精度が劣後していることを示す結果となった一方で、モデル 6~8 に対しては精度が優れていることを示す結果となった。Diebold-Mariano 検定では、モデル 2, 3, 6, 7 との間には予測精度に有意差があり、2, 3 に対しては精度が優れている、6, 7 に対しては精度が劣後しているという結果となり、モデル 4, 8 との間には有意差がないという結果となった。この結果は、複数の異なる性質を持つ感情スコアを、適切に組み合わせて使用することで株価予測の精度が上がることを示唆している。

5.5 モデル 6 について

モデル 1 とモデル 6 の RMSE, 決定係数, MAPE を比較すると、全ての値においてモデル 6 の予測精度が優れていることを示す結果となったことが分かる。また、Diebold-Mariano 検定の結果についても同様であり、予測精度には統計的に有意な差があるという結果になった。以上の結果は、株価の予測における感情データの有用性を示唆している。一方、モデル 2~5, 7, 8 との比較については、RMSE, 決定係数, MAPE では、モデル 2~5 の全てに対してモデル 6 の精度が劣後していることを示す結果となった一方で、モデル 7, 8 に対しては精度が優れていることを示す結果となった。Diebold-Mariano 検定では、モデル 2~5, 8 との間には予測精度に有意差があり、2~5 に対しては精度が優れている、8 に対しては精度が劣後しているという結果となり、モデル 7 との間には有意差がないとい

う結果となった。この結果は、複数の異なる性質を持つ感情スコアを、適切に組み合わせて使用することで株価予測の精度が上がることを示唆している。

5.6 モデル 7 について

モデル 1 とモデル 7 の RMSE, 決定係数, MAPE を比較すると、全ての値においてモデル 7 の予測精度が優れていることを示す結果となったことが分かる。また、Diebold-Mariano 検定の結果についても同様であり、予測精度には統計的に有意な差があるという結果になった。以上の結果は、株価の予測における感情データの有用性を示唆している。一方、モデル 2~6, 8 との比較については、RMSE, 決定係数, MAPE では、モデル 2~6 の全てに対してモデル 7 の精度が劣後していることを示す結果となった一方で、モデル 8 に対しては精度が優れていることを示す結果となった。Diebold-Mariano 検定では、モデル 2~5, 8 との間には予測精度に有意差があり、2~5 に対しては精度が優れている、8 に対しては精度が劣後しているという結果となり、モデル 6 との間には有意差がないという結果となった。この結果は、複数の異なる性質を持つ感情スコアを、適切に組み合わせて使用することで株価予測の精度が上がることを示唆している。

5.7 モデル 8 について

モデル 1 とモデル 8 の RMSE, 決定係数, MAPE を比較すると、全ての値においてモデル 1 の予測精度が優れていることを示す結果となったことが分かる。一方、Diebold-Mariano 検定の結果については、予測精度に統計的に有意な差があり、モデル 8 の予測精度が優れているという結果になった。以上の結果は、株価の予測における感情データの有用性を示唆している。一方、モデル 2~7 との比較については、RMSE, 決定係数, MAPE では、モデル 2~7 の全てに対してモデル 8 の精度が劣後していることを示す結果となった一方で、Diebold-Mariano 検定では、モデル 2, 6, 7 との間には予測精度に有意差があり、モデル 8 の精度が優れているという結果になり、モデル 3~5 との間には有意差がないという結果となった。この結果は、感情スコアの組み合わせ方によって、かえって予測精度が悪化してしまうことを示唆している。

6 終わりに

6.1 今回の研究から得られる示唆

本研究では、Reddit の投稿を感情分析で評価し、そこから得た感情データ及び株価データを用いて学習を行った LSTM によりダウ平均株価の予測を行い、異なる評価方法を用いたモデル同士で比較と考察を行った。その結果、株価データのみを用いた予測モデルと比較して、感情データを追加することで予測精度が有意に向上することを示した。これは、感情データが株価予測において有用であることに加え、ソーシャルメディア上での感情が金融

市場と深く結びついていることを示唆している。また、複数の異なる特徴を持つ感情データ、特に `type_score` や `represent_score` などの、離散的な特徴を持つ感情データを適切に組み合わせることで、株価予測の精度が有意に向上することも示すことができ、これは株価予測における感情データの適切な扱い方に示唆を与える結果となった。

6.2 今後の課題

今回提案した複数のモデル間で予測精度に違いが見られたため、感情データの利用方法についてさらなる検討が必要であるほか、Diebold-Mariano 検定において、一部矛盾するような結果が得られたため、これをより妥当な結果に落とし込むためにも、よりサンプル数を増やした検証を行う必要がある。また、本研究はダウ平均株価を対象とした分析であり、その性質上、個別銘柄同士の影響が平均化されていると予想されるが、同様の手法を用いて個別銘柄の予測精度向上にどの程度貢献できるかは、今後の課題である。さらに、金融市場は常に一貫した動きをすることは限らないため、学習期間や市場の休場日でのソーシャルメディアの投稿データの扱い方、対象とする年度の選定などが結果に与える影響も考慮する必要がある。今後、普遍的に本手法を用いて高い精度の予測が実現できるかどうかを検証することが求められる。

6.3 謝辞

本論文の作成にあたり、指導教官として終始多大なご指導を賜った、慶應義塾大学経済学部長倉大輔教授に深謝いたします。また、長倉大輔研究会の先輩方もメンターとして様々なご助言をして頂きました。ここに感謝の意を表します。

7 参考文献

湯原心一. (2021). ゲームストップ株価の乱高下. 法律時報, 93(5), 1-3.

Adekoya, A. F., Nti, I. K., & Weyori, B. A. (2022). Long Short-Term Memory Network for Predicting Exchange Rate of the Ghanaian Cedi. *FinTech*, 1(1), 25–43. <https://doi.org/10.3390/fintech1010002>

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/https://doi.org/10.1016/j.jocs.2010.12.007>

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>

Long, S. (Cheng), Lucey, B., Xie, Y., & Yarovaya, L. (2023). “I just like the stock”: The role of Reddit sentiment in the GameStop share rally. *Financial Review*, 58(1), 19–37. <https://doi.org/https://doi.org/10.1111/fire.12328>

Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, 7, e476. <https://doi.org/10.7717/peerj-cs.476>

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, 42(24), 9603–9611. <https://doi.org/https://doi.org/10.1016/j.eswa.2015.07.052>

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2021). A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *Journal of Big Data*, 8(1), 17. <https://doi.org/10.1186/s40537-020-00400-y>

Wang, Z., Ho, S.-B., & Lin, Z. (2018). Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1375–1380. <https://doi.org/10.1109/ICDMW.2018.00195>

Institute for Reproducible Research. n.d. “Academic Torrents.” Accessed November 6, 2024. <https://academictorrents.com/>.

8 付録

表 1 分析対象である企業と抽出する際に用いた単語の対応

	企業名	抽出時に用いた単語	抽出されたデータ数
ダ ウ 平 均 株 価 を 構 成 す る 3 0 企 業	アップル	"Apple","AAPL"	156381
	ナイキ	"Nike","NKE"	73479
	プロクター・アンド・ギャンブル	"Procter&Gamble","PG"	64804
	マイクロソフト	"Microsoft","MSFT"	55721
	インテル	"Intel","INTC"	49212
	ディズニー	"Disney"	23554
	スリーエム	"MMM"	23357
	マクドナルド	"McDonald","MCD"	18838
	JPモルガン	"JPMorgan","JPM"	11894
	ユナイテッド・ヘルス	"UnitedHealth","UNH"	8529
	ゴールドマン・サックス	"Goldman"	6335
	ベライゾン	"Verizon","VZ"	5784
	ウォルマート	"Wal-Mart","WMT"	4960
	アメリカン・エクスプレス	"American Express","AXP"	4910
	ウォルグリーン	"Walgreen","WBA"	4676
	セールスフォース	"Salesforce","CRM"	4577
	ビザ	"Visa"	4454
	シスコ	"Cisco","CSCO"	4452
	ボーイング	"Boeing"	4404
	アイビーエム	"IBM"	3079
	ホーム・デポ	"Home Depot"	2857
	シェブロン	"Chevron","CVX"	2598
	ジョンソン&ジョンソン	"Johnson","JPM"	991
	メルク	"Merck","MRK"	801
	コカ・コーラ	"Coca-cola"	723
	ダウ	"DJIA"	674
	キャタピラー	"Caterpillar"	457
	アムジェン	"Amgen","Amgen"	288
	ハネウェル	"Honeywell"	203
	トラベラーズ	"Travelers Company","TRV"	133