Individual Stock Price Prediction by Using KAP and Twitter Sentiments with Machine Learning for BIST30

Muhlis Sarıyer
Computer Engineering Department
Marmara University
Istanbul, Turkey
muhlissariyer@marun.edu.tr

Ahmet Akıl
Computer Engineering Department
Marmara University
Istanbul, Turkey
ahmetakil@marun.edu.tr

Feyza Nur Bulgurcu
Computer Engineering Department
Marmara University
Istanbul, Turkey
feyzabulgurcu@marun.edu.tr

Fatih Emin Öge
Computer Engineering Department
Marmara University
Istanbul, Turkey
fatihoge@marun.edu.tr

Murat Can Ganiz
Computer Engineering Department
Marmara University
Istanbul, Turkey
murat.ganiz@marmara.edu.tr

2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA) | 978-1-6654-9810-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/INISTA55318.2022.9894172

***Abstract*—In this study we used machine learning models for predicting individual stock price and volume changes using sentiments from public disclosures and tweets. Public Disclosure Platform (KAP) is the mandated regulatory platform for disclosing news about companies listed in Borsa Istanbul. Investors in Borsa Istanbul use Twitter to express their sentiments about stocks. By combining people's sentiment on Twitter and companies' disclosures, our prediction model predicts the volume and price changes of individual company stocks listed in BIST30. Financial data regarding market conditions consisting of daily price changes of BIST30, DJI, USD, and Gold per Ounce are also added to enhance the prediction accuracy of the model. Our model achieves an maximum of 80% individual stock price prediction accuracy for companies with high social media presence and public disclosure count. We also achieve 74.7% mean volume prediction accuracy across all BIST30 companies.**

***Index Terms*—Borsa Istanbul (BIST), Public Disclosure Platform (KAP), sentiment analysis, stock market price prediction, individual stock prediction, stock volume prediction**

I. Introduction

The study of stock markets in general is a study of human behavior [1] and their emotions in the markets. This assumes that people who participate in the market behave emotionally and their collective movements are impacted by the current news about individual companies [2]. Efficient stock market theory assumes that the participants in an efficient market make their decisions based on information available to the public [2]. Borsa Istanbul provides investors with KAP (Public Disclosure Platform) for companies to disclose information publicly to provide equal opportunity of information to all investors [3].

Stock price prediction using financial information is a rigorously studied area. This is due to it's potential applications on building profitable trading strategies in the market. Most of the studies in the field are done to predict market indices such as BIST30 or BIST100 [4].

BIST30 and BIST100 are the indices that represent the stock performance of the 30 and 100 most valuable companies on the Borsa Istanbul Stock Exchange respectively. Public Disclosure Platform (KAP) is the mandated regulatory platform for disclosing news about companies.

In conjunction with our research focus, we trained different ML models for each company to observe the impact of public disclosures on individual stocks. This allowed us to achieve higher confidence in a trading strategy as well as provided us the ability to distinguish the impact of news sentiments on companies. Some companies are more affected by disclosures than others.

We build a predictive model on individual stocks for a trading behavior that would maximize profit for trading on companies listed in BIST30. We focus on creating models for predicting individual stock price changes.

The literature on sentiment analysis driven stock price prediction, mostly consists of financial news sentiment analysis [5], [6]. In this study, we focused on public disclosures as they are the initial source of information [7] and Twitter so we can determine the outreach of information regarding the stock. We create a novel individual stock price and volume change prediction model to foresee volatility and price changes using public disclosures and Twitter sentiments. We improved our model by including data about market conditions. We tested and implemented imputation methods to assess news relevance across time.

978-1-6654-9810-4/22/$31.00 ©2022 IEEE

II. Related Work

Individual stock prediction can be seen in these works with lower accuracy with respect to index predictions [3], [8]. Although financial news [9] used for sentiment analysis are more widespread than Twitter, there are many studies with Twitter data cases, whose approaches produce significant results [10], [4]. KAP disclosure use for prediction, is rarely seen in the literature with approaches focusing on indices [11]. KAP data comes in an unstructured format and it's challenging to structure the data for Machine Learning (ML) projects.This and some other reasons, makes it being avoided in many studies in the field. Twitter data used for stock prediction falls under two categories, the first is hashtag use [12], and the second is a broader keyword detection implementation [7], we found hashtags to be more reliable because they provide a more precise impression for the companies. Keyword detection implementations are comprehensive in planning and provide bigger data that can be used for Deep Learning applications. BIST100 companies represent the majority of Borsa Istanbul's trading volume and BIST30 companies are the biggest and most prestigious companies in BIST100. BIST30 Index and BIST30 companies are widely used for news sentiment analysis by studies conducted on Borsa Istanbul [13], [14]. Companies in the BIST30 index use KAP to disclose information much more frequently as well as attract much more attention on social media, especially Twitter. Using BIST100 companies would produce a much more sparse dataset that would yield inconclusive results, therefore we decided to use BIST30 as our target. Gunduz et al. [15] [16] pioneered news sentiment analysis using ML models for Borsa Istanbul in 2013 and achieved greater prediction accuracy with their study in 2018 [17]. Literature on finance and machine learning overlap in papers regarding the stock market analysis. Therefore, there are many papers with data included from financial analyses. Compared to other studies in the field from the perspective of price prediction accuracy financial models can improve models [18] but require a more manual approach. General sentiment analysis is done for assessing the predictability of stock returns is done by You et al [19]. The combination of sentiments and market information in our method can be compared to aggregate/ensemble learning done by Pasupulety et al [20]. Since their research also uses twitter data to find the public sentiment about a stock and also uses technical indicators that is calculated from price information. In [21] with stock prices Deep Learning techniques used with Systematic Literature Review(SLR). Experiments performed with different combinations with deep learning models and different datasets; they gave their dataset as stock market, stock market indices, European stock market indices etc. With SLR, best performing deep learning model was LSTM according to this studies' finding.

III. Approach

For financial analysis of big data, deep learning methods are used as machine learning models [3], [5], [22]. In our case, a total of 38.864 data points exist for training the final ML model. Due to the scarcity of labeled data, deep learning methods performed poorly in our experiments. Therefore, we conducted our experiments with ML models that are being used in literature for similar cases.

BERT [23], [24] is a pre-trained machine learning model for general purpose natural language processing developed by Google. BERTurk [25] is a community driven language model developed to calculate sentiments in Turkish sentences. BERTurk is the state-of-the-art sentiment analysis model for Turkish text and is used in financial news sentiment analysis in many studies [26], [27]. The literature on news sentiment analysis for stock market prediction is overwhelmingly done for predicting Market or Sector Indices such as BIST100 or the banking sector. BIST100 is the index that represents the stock performance of the 100 most valuable companies in Borsa Istanbul (BIST). Banking sector index in Borsa Istanbul is the index for all banks that are listed in BIST. As a novel approach, we choose the prediction of individual stocks rather than a market or sector index, these predictions can form trading strategies to maximize profit in a systemic and reliable way. Our literature analysis shows that indices and sectors are preferred in this area because of their predictable nature. And information about individual companies is too fragmented and sparse to form into a machine learning model. Our study overcomes this fragmentation and sparsity by enriching public disclosure data and combining it with Twitter and wider market data

A. Twitter Data Collection

Stock symbols/tickers are unique abbreviations for each company's stock in a stock exchange such as THYAO for Turkish Airlines or MSFT for Microsoft. For Twitter data "snscrape" Python module is used and all the tweets that mentioned a stock symbol in BIST30 are collected with additional data that will be used to give weights such as like count of the tweet, quote count of the tweet, reply count of the tweet, retweet count of the tweet, followers count of the account at the time of tweeting.

In total, we gathered **1.494.108** tweets regarding BIST30 companies and their interactions on Twitter between the 1st of January 2016 and the 1st of January 2021

B. KAP Data Collection

For KAP data python requests library is used to gather all disclosures about companies listed under BIST30 index. In total, we gathered 20.667 public disclosures from KAP's website between the 1st of January 2016 and the 1st of January 2021. 1629 days of trading occurred inside the timeframe.

C. Sentiment Score Calculation

For sentiment score calculations for both KAP and Twitter datasets, the sentiment BERT model introduced in [25] [6] is used, the model has a character limit of 512. For longer KAP contents we divided up the contents into chunks of size 512 characters and used the average of the calculated sentiments. Table I shows the number of positive and negative sentiments derived from Twitter and KAP disclosures.

TABLE I

SENTIMENTS OF DATA POINTS OF DIFFERENT PLATFORMS

Sentiment	Twitter	KAP
Positive	6364	3914
Positive	6364	3914

D. Twitter Data Preparation

While collecting the twitter data we also gathered extra metadata belonging to the tweets including like count (lc), quote count (qc), retweet count (rc), and follower count(fc). Using these values, we generate an interaction score that will be used as weight as seen in equation (1). Each count is normalized using a logarithm with a 1 added to avoid log(0).

$$IS(lc,qc,rc,fc) = (\sum_{x}^{[lc,qc,rc,fc]} log(x+1)) * log(fc+1) \quad (1)$$

$$twitter\_score = IS * twitter\_sentiment\_score \quad (2)$$

After the calculation of the interaction score, we calculated the Twitter score by multiplying the interaction score and Twitter sentiment score. As seen in formula (2). And twitter_score is normalized using a Z-score normalization method.

E. Public Disclosure Text Preparation

Our time interval for data collection is between January 2016 and 2021. This period is chosen because it presents a time of high volatility and increased informational awareness in public both present on Twitter and in public disclosures. Disclosures are in text format without any length limit therefore their length varies greatly, information regarding a disclosure's length can provide an indication of its relevance towards the change in price. To enrich public disclosures data, we formed the length of disclosures into a feature for our ML model to capture readability, we also added the ratio of numeric to alphabetic characters in each disclosure to capture the interpretability. Higher ratios are expected to have higher interpretable results. To obtain sentiment scores using information relevant to the stock we preprocessed public disclosures by removing all the HTML tags from the texts as well as English versions that exist in 5% of the disclosures.

F. Market Data Preparation

We chose prices of Gold, USD, DJI, and BIST30 indices. The price of Gold per ounce is generally seen as an indicator of perceived risk in financial markets. The ratio of USD/TRY is used for benchmarking against inflation in Turkish Lira. We included the DJI index which represents the global trend in stock market investments. For the BIST30 index and the price change of individual stocks, we calculated 1, 3, 5, 8, 10, 20, 30, 60, 90 days change ahead of the trading day to understand the expected descent of news' relevance and our ability to predict the price changes going into the future. Trading volume is the number of shares bought and sold in a stock exchange for a given stock each day, it represents the activity and attention on a company's stock by investors[1]. We calculated volume changes of stocks and included them to predict volatility. Volume changes vary between -100% to 500%, which is grouped into equal parts of 6, 12, 24 bins respectively. -100% means no trading is done for that stock on that day. 500% means 4 times more trading compared to its all-time average trading volume is done on that stock for that day. Table II shows the 6 bins with respect to their intervals:

TABLE II

BIN IDS RANGE OF BINS FOR VOLUME CHANGE PREDICTION

Bin Id	Volume Change Range
0	(-1.0, 0.0)
1	(0.0, 1.0)
2	(1.0, 2.0)
3	(2.0, 3.0)
4	(3.0, 4.0)
5	(4.0, 5.0)

G. Data Consolidation

Trading days refer to days that a company's stock can be traded within a stock exchange. We consolidated all the gathered data with respect to trading days. Our analysis is conducted for both a combined model where all stocks are joined together, and predictions are made on a single model, and we also trained separate ML models for each stock. To combine different stocks, we used one hot encoding method where each stock is turned into a column with a one or a zero value. We combined stocks into a single ML model to provide a unified system and be able to build more complex models. In conjunction with our research focus, we trained different ML models for each company to observe the impact of disclosures on single stocks. This allowed us to achieve higher confidence in a trading strategy and provided us the ability to distinguish information efficiency on news sentiments of companies. Some companies are more affected by disclosures than others.

Distinguishing companies and building multiple ML models proved to be more successful in predicting stock price changes therefore further study is conducted on building 30 ML models for each company and combining results by taking maximum, minimum, and mean of predictions made on single stocks. Companies do not publish public disclosures every trading day therefore to make use of data on those days, imputation is made. Imputation [26] is the practice of determining what values should be used when there are gaps in a dataset. Zero fill method involves filling the gaps with neutral values. Forward filling method refers to using previous data points. Decay filling means filling values by using several data points before and assigning decaying weight with respect to their distance. Decay fill is intuitively more appropriate for news because news does not lose its informational value in a single day but that value decays over time.

H. Model training

Training is done by testing 9 different machine learning models from 5 different model families:

- Linear Models
- Support Vector Machines
- Cluster Models
- Ensemble Models
- Bayesian Models

Each model is trained to predict for 2, 4, 6, 8, 10 bin versions with 3 different imputation methods and different targets such as ”TodayChange”, ”Next5DayChange”, ”Next10DayChange”, vs. For model hyper-parameters we used the defaults for respective models since according to the results of our observations, it didn’t change accuracy significantly while we used hyper-parameter tuning with grid search and random search.

IV. Experiment Results and Discussion

The accuracy of the models is calculated using F1 score and 80% of the data is used for training and 20% is used for testing. Stocks price changes are time dependent therefore random sampling is not feasible for our case. Some of the results for scores can be seen in Table III Division for training and test is done by splitting data time-wise into before and after a certain day. Table IV shows the comparison of F1 scores of 9 different Machine Learning models:

- Logistic Regression (LR)
- Bagging Classifier (BagC) Random Forest (RandF)
- AdaBoost (AdaB)
- K Nearest Neighbors (KNN)
- Decision Trees (DecisionT)
- ExtraTrees Classifier (ExtraTC)
- Support Vector Classifier (SVC)
- XGBoost (XGB)

TABLE III
F1-score examples for 2 bins

Stock Name	Max	Mean
AKBNK	76.4	55.7
ARCLK	61.5	49.3
GUBRF	68.3	53.2
FROTO	65.5	53.2
EREGL	61.0	50.0

These algorithms are applied to get maximum, minimum, and mean predictions of BIST30 individual stocks’ price changes. Due to their faster processing times, novelty, and interpretability of results, four of the models (LR, RandF, AdaB, DecisionT ,and ExtraTC) are chosen to be used in our studies.

Table V shows the results of predictions using 2 bin predictions with 3 imputation methods and the best results are achieved by using decay and zero fill to achieve more precise results further research is done by only using zero fill method.

A. Stock Trading Volume Change Prediction

In literature, daily stock volume prediction is an understudied topic in terms of sentiment analysis. Our results portrayed in Table VI shows a strong correlation between sentiment scores and volume changes. Volume changes are calculated by taking the average daily volume of stock and dividing the current day’s volume by the average value. As resolution of prediction by means of bin counts increases radical drop in accuracy is observed.

TABLE IV
Prediction Accuracy Change With Different ML Models

Model	Max	Mean	Min
LR	76.4	47.5	52.3
BagC	62.7	56.9	56.9
RandF	60.1	51.9	56.2
AdaB	63.6	51	56.6
KNN	62.5	52	57.3
DecisionT	64.0	52.2	56.6
ExtraTC	59.6	52.7	56.4
SVC	66.9	48.3	54
XGB	62	51.3	57.3

TABLE V
Prediction Accuracy Change With Different Imputation Methods

Imputation	LR	RandF	AdaB	DecisionT	ExtraTC
Forward Fill	38.3	50.1	49.2	49.9	50.1
Decay Fill	52.3	56.2	56.6	56.5	56.3
Zero Fill	52.3	56.2	56.6	56.6	56.1

B. Stock Price Change Prediction

In Table VII results of predictions using Twitter and KAP sentiment scores are shown using different Machine Learning models. Maximum, minimum, and mean of different stocks’ predictions show that different companies have different susceptibility to sentiment across KAP and Twitter.

Table VIII shows the accuracy decreasing as the resolution of prediction increases. This is the expected behavior of ML models and provides confidence in the connection of the findings to the actuality of data.

Inclusion of disclosure length and alphabetic/numeric character ratios improves the prediction accuracy of models by 15%. Figure 1 shows that the first day of prediction yields higher accuracy as KAP and Twitter sentiment directly impact

TABLE VI
Prediction Accuracy Of Volume Prediction On Different Bin Sizes

F1 Score	LR	RandF	AdaB	DecisionT	ExtraTC
6 bins	74.7	62.4	70.2	61.8	62.8
12 bins	34.6	34.4	35.7	34.2	34.1
24 bins	17.8	20.2	18.9	20.1	20.1

TABLE VII
Accuracy Of Stock Price Change Predictions Of Different Stocks For 2 Bins

Stock	LR	RandF	AdaB	DecisionT	ExtraTC
MAX	81.76	81.31	81.58	77.00	76.68
MEAN	45.90	49.15	5.82	45.81	48.22
MIN	67.82	67.22	59.82	62.85	62.82

TABLE VIII
ACCURACY OF MEAN STOCK PRICE CHANGE PREDICTION FOR DIFFERENT BIN LEVELS

Bin Level	LR	RandF	AdaB	DecisionT	ExtraTC
2 Bin	52.7	60.4	62.2	62.7	62.4
4 Bin	51.9	58.9	65.6	51.1	51.1
6 Bin	49.1	54.8	51.9	56.3	56.7
8 Bin	44	48.9	45.8	50.7	50.7
10 Bin	44	48.7	45.6	50.1	50.4

the price and after the first day a sharp decrease in accuracy occurs. In 30 trading days' time, we can see the information in the sentiment integrating into the price of the stock and stabilizing afterward. This means that the accuracy of predictions gets higher for longer periods because the noise of daily price fluctuations can make predictions harder in the short term. But after 30 days the impact of disclosures and tweets is stabilized.

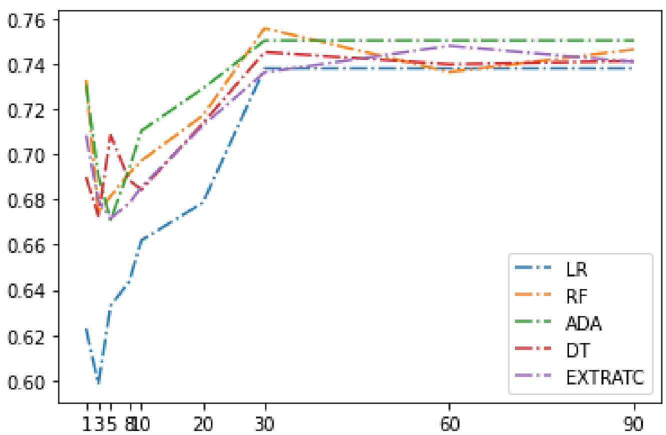

Fig. 1. Price Prediction Accuracy For Different Periods Of Time (Day)

V. CONCLUSION AND FUTURE WORK

In this study, we focus on building ML models for predicting individual stocks that are compatible with forming a trading strategy based on public disclosure published on KAP and people's sentiments on Twitter. We use state-of-the-art Turkish BERT sentiment model for sentiment analysis of tweets and disclosures. To build our dataset and make predictions for days without tweets or disclosures we test several imputation methods to assess their value, days after their publishment. We append financial data regarding market conditions consisting of daily price changes of BIST30, DJI, USD and Gold per Ounce to improve prediction accuracy and integrate markets overall sentiments and risk appetite. Our study also consists of predicting daily volume changes of stocks as a novel study field. We can predict the volume changes with high accuracy using sentiments from public dis- closures and tweets. We analyze the volatility of stocks by predicting daily volume changes of stocks and this provides a reliable correlation between the sentiments and volume changes for the upcoming trading day with 74.7% prediction accuracy. But because volume changes are generally considered to be reactionary the prediction falls quickly for further days Building a trading strategy using an ML model with accuracy levels higher than 60% can yield higher than market average profits and our results can support such levels of accuracy of our predictions. Our study shows prediction accuracies forBIST30 companies' individual stocks of 67%. Our study suggests that predicting the stock price changes for longer periods of time provides higher accuracy levels therefore by using larger timeframes and more advanced sentiment models better trading strategies can be formed for long term investment. In the future we would like to apply the following methods to obtain better predictions: Compressing data inside of disclosure into a single sentiment score omits a lot of information; therefore, more information regarding KAP disclosures will be extracted, and combined word vectors will be tested for higher accuracy. Research on dividend and stock split data to predict and explain the behavior of changes in stock prices goes back to Eugene Fama's 1970 paper on Efficient markets [2] and is proven to be reliable information, compared to the disclosure and tweet data on dividend and stock splits occur less frequently but can be integrated into the model for a study that covers a longer period. Imputation methods in financial news —-[21]— is an unexplored area of research that can provide fruitful results towards assessing the impact of the relevance of financial news over a period. In our research we made progress in this area and future studies imputation methods can append greater value towards NLP in financial news. NLP models specific to finance in Turkish are not publicly available therefore a clear improvement in this area will be the development and implementation of financial NLP models for Turkish as well as using established financial NLP models in English for KAP disclosures that are published in English.

REFERENCES

[1] R. J. Shiller, "Human behavior and the efficiency of the financial system," *Handbook of macroeconomics*, vol. 1, pp. 1305–1340, 1999.
[2] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.
[3] B. Aslan and R. C. Erdur, "Stock market prediction with deep learning using public disclosure platform data," in *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, pp. 1–5, IEEE, 2020.
[4] E. Ates and A. Güran, "Pearson correlation and granger causality analysis of twitter sentiments and the daily changes in bist30 index returns," *J. Fac. Eng. Archit. Gazi Univ*, vol. 36, pp. 1687–1701, 2021.
[5] Z. H. Kilimci and R. Duvar, "An efficient word embedding and deep learning based model to forecast the direction of stock exchange market using twitter and financial news sites: a case of istanbul stock exchange (bist 100)," *IEEE Access*, vol. 8, pp. 188186–188198, 2020.
[6] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong," *Information Processing & Management*, vol. 57, no. 5, p. 102212, 2020.
[7] O. Çörtük and M. Erten, "Türkiye'de kamuyu aydınlatmanın sermaye piyasasına etkisi," *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, vol. 45, no. 1, pp. 65–77, 2016.
[8] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert systems with applications*, vol. 42, no. 1, pp. 259–268, 2015.
[9] H. Gunduz and Z. Cataltepe, "Borsa istanbul (bist) daily prediction using financial news and balanced feature selection," *Expert Systems with Applications*, vol. 42, no. 22, pp. 9001–9011, 2015.

[10] U. U. Acikalin, B. Bardak, and M. Kutlu, "Turkish sentiment analysis using bert," in *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2020.

[11] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[12] G. Bozma and S. Kul, "Can twitter forecast uncertainty of stocks?," *SOSYOEKONOMI*, vol. 28, no. 45, 2020.

[13] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[14] İ. Sabuncu, "Forecasting stock value based on data from social media and investment instruments," *Acta Infologica*, vol. 5, no. 2, pp. 267–285, 2021.

[15] H. Gunduz, Y. Yaslan, and Z. Cataltepe, "Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations," *Knowledge-Based Systems*, vol. 137, pp. 138–148, 2017.

[16] H. Gunduz and Z. Cataltepe, "Prediction of istanbul stock exchange (ise) direction based on news articles," in *Conference: The Third International Conference on Digital Information Processing and Communications (ICDIPC2013)*, pp. 320–330, 2013.

[17] H. Gündüz, Y. Yaslan, and Z. Çataltepe, "Stock market prediction with deep learning using financial news," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2018.

[18] X. Guo and J. Li, "A novel twitter sentiment analysis model with baseline correlation for financial market prediction with improved efficiency," in *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 472–477, IEEE, 2019.

[19] W. You, Y. Guo, and C. Peng, "Twitter's daily happiness sentiment and the predictability of stock returns," *Finance Research Letters*, vol. 23, pp. 58–64, 2017.

[20] U. Pasupulety, A. A. Anees, S. Anmol, and B. R. Mohan, "Predicting stock prices using ensemble learning and sentiment analysis," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 215–222, IEEE, 2019.

[21] S. G. P. V. C. T. T. Ketsetsis, Kourounis, "Deep learning techniques for stock market prediction in the european union: A systematic review," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 605–610, IEEE, 2020.

[22] S. Yildirim, "Comparing deep neural networks to traditional models for sentiment analysis in turkish language," in *Deep learning-based approaches for sentiment analysis*, pp. 311–319, Springer, 2020.

[23] Z. A. Guven, "Comparison of bert models and machine learning methods for sentiment analysis on turkish tweets," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*, pp. 98–101, IEEE, 2021.

[24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[25] J. Han, M. Kamber, and J. Pei, "Data mining: Concepts 790 and techniques," *Morgan Kaufmann, Waltham. URL: http://myweb. sabanciuniv. edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mi-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011. pdf% 0Ahttp://scholar. google. com/schol, doi*, vol. 10, p. 795, 2012.

[26] L. A. Hunt, "Missing data imputation and its effect on the accuracy of classification," in *Data Science*, pp. 3–14, Springer, 2017.