Boosting the Accuracy of Stock Market Prediction using XGBoost and Long Short-Term Memory

Publisher: IEEE
Cite This PDF

Agustinus Bimo Gumelar ; Haryati Setyorini ; Derry Pramono Adi ; Sengguruh Niloward... All Authors

Abstract:
Stock exchange is one of the famous economical strategy that finally find its way to be experimented with ever-growing Machine Learning (ML) algorithm. With ML, many aspects regarding stock is learnable, to the point where one can predict stock prices. Although tempting, stock price prediction is still a challenging task due to its natural dynamic and real-time movement. Thus, predicting stock prices are deemed unseemingly. On the other hand, different patterns of stock prices are capable of represent a whole lot of detailed data, which is in favor for Deep Learning. In this study, we conducted an experiment of predicting the close stock price for 25 companies. To ensure data reliability and regional notion, these selected companies are officially enlisted in the Indonesia Stock Exchange (IDX). The two ML algorithms used for this experiment are the Long Short-Term Memory (LSTM) and Extreme Gradient Boosting (XGBoost), both known for its high accuracy of prediction from various representative data. By setting two thresholds, we were able to present a trading approach: when to buy or when to sell. This prediction result from the ML algorithm using in the ensuing trading approach leads to distinct aspects of benefit. In this experiment, XGBoost shown best performance by 99% prediction accuracy result.

Published in: 2020 International Seminar on Application for Technology of Information and Communication (iSemantic)

Date of Conference: 19-20 September 2020

INSPEC Accession Number: 20130727

Date Added to IEEE *Xplore*: 26 October 2020

DOI: 10.1109/iSemantic50169.2020.9234256

▸ **ISBN Information:**

Publisher: IEEE

Conference Location: Semarang, Indonesia

I. Introduction

The stock market is an ever-growing investment instrument in not-a-game-of-chances business consisted with many unpredictable moments [1]. The stock exchange has always tended to be a risky arena for those outside banking and numbers. Many are perplexed and find it difficult to comprehent its inherent yet comprehensive data. The data stream itself is ever-changing in real-time, makes it harder for traders to bu[...] f time series forecasting data is deeme[...] k [2], [3]. However, a prediction strategy is doable by Machine Learning (ML), as in fact, the stock market generates more data overtime. The ability of ML algorithm to predict with high precision for many field is astonishing [4]–[6], including of providing insight

Authors

Agustinus Bimo Gumelar
Fakultas Ilmu Komputer, Universitas Narotama, Surabaya, Indonesia

Haryati Setyorini
Dept of Management, STIE PERBANAS Surabaya, Surabaya, Indonesia

Derry Pramono Adi
Fakultas Ilmu Komputer, Universitas Narotama, Surabaya, Indonesia

Sengguruh Nilowardono
Economic and Business Faculty, Narotama University, Surabaya, Indonesia

Latipah
Faculty of Computer Science, Narotama University, Surabaya, Indonesia

Agung Widodo
Fakultas Ilmu Komputer, Universitas Narotama, Surabaya, Indonesia

Achmad Teguh Wibowo
Fakultas Ilmu Komputer, UIN Sunan Ampel, Surabaya, Indonesia

MY Teguh Sulistyono
Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

Evy Christine
Faculty of Business, Widya Mandala Catholic University, Surabaya, Indonesia

Figures

References

Citations

Keywords

Metrics

International Seminar on Application for Technology of Information and Communication (iSemantic)

Copy Persistent Link | Browse Title List | Sign up for Conference Alerts
Proceedings | All Proceedings | Popular

2020 International Seminar on Application for Technology of Information and Communication (iSemantic)
19-20 Sept. 2020
DOI: 10.1109/iSemantic50169.2020

Search within results | Items Per Page | Export | Email Selected Results

Showing 101-116 of 116

Refine
Author
Affiliation

Quick Links
Search for Upcoming Conferences
IEEE Publication Recommender
IEEE Author Center

Proceedings

The proceedings of this conference will be available for purchase through Curran Associates.

Application for Technology of Information and Communication (iSemantic), 2020 International Seminar on

Print on Demand **Purchase at Partner**

External Hard-drive **Purchase at Partner**

Select All on Page | Sort By Sequence

< 1 2 3 4 5

Exploring Indonesian Netizen's Emotional Behavior Through Investment Sentiment Analysis Using TextBlob-NLTK (Natural Language Toolkit)

Esther Talahaturuson
Fakultas Ilmu Komputer
Universitas Narotama
Surabaya, Indonesia
estheryola99@gmail.com

Agustinus Bimo Gumelar
Dept. of Electrical Engineering,
The Faculty of Intelligent Electrical and Informatics Technology (F-ELECTICS),
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
bimogumelar@ieee.org

Adri Gabriel Sooai
Dept. of Computer Science,
University Centre of Excellence in Artificial Intelligence for Tourism & Agriculture,
Universitas Katolik Widya Mandira.
Kupang NTT, Indonesia
adrigabriel@unwira.ac.id

Sueb
Fakultas Bahasa dan Sains
Universitas Wijaya Kusuma Surabaya
Surabaya, Indonesia
suebhadi_fbs@uwks.ac.id

Suprihatien
Fakultas Bahasa dan Sains
Universitas Wijaya Kusuma Surabaya
Surabaya, Indonesia
titien.suprihatien_fbs@uwks.ac.id

Hikmah Ali Altway
STIE AMM Mataram
Mataram NTB, Indonesia
hikmah@stieamm.ac.id

Chatarini Septi Ngudi Lestari
Dept of. English Literature
School of Foreign Language and Literature (STIBA) Satya Widya Surabaya
Surabaya, Indonesia
chatarinisnl@gmail.com

Perwi Darmajanti
Shipbuilding Institute Of Polytechnic Surabaya (PPNS)
Surabaya, Indonesia
perwi@ppns.ac.id

Urip Zaenal Fanani
Dept of. Japanese Language And Literature
Faculty of Languages
State University of Surabaya
Surabaya, Indonesia
zaenalfanani@unesa.ac.id

Tuty Hariyanti
Indonesia Naval Academy
Surabaya, Indonesia
tutyhariyanti1978@gmail.com

Sengguruh Nilowardono
Dept of Management
Fakultas Ekonomi dan Bisnis
Universitas Narotama
Surabaya, Indonesia
sengguruh@narotama.ac.id

Sulistiyani
Dept of. English Education
STKIP Bina Insan Mandiri
Surabaya Indonesia
sulistiyani@stkipbim.ac.id

***Abstract*— The investment industry has recently continued to provide the greatest experience and has grown the number of investors to this day. It also increases the quantity of traded investment assets because of its varied central and decentralized operating processes. Yet another aspect that can't be isolated from the investment process is price volatility and monetary policy. This means that the present price movement is influenced by every market mood. According to this study, a three-month period of tweets from Indonesian citizens was used to analyze attitude towards investment patterns in Indonesia. Because of the large number of people throughout the world who use Twitter to voice their opinions on investments, Twitter was selected as the primary source for this study. Twint, an open-source Python library, is used to retrieve tweet data. To process and analyze each tweet's data, TextBlob will be used, which values subjectivity and polarity. There were 92% favorable feelings and 42% positive sentiments on Indonesian tweets after a succession of research stages. These results were obtained by the limitation of data preprocessing and data labeling has been used.**

Keywords*— *Netizen's Emotional Behavior; Sentiment Analysis; Investment; TextBlob; NLTK

I. Introduction

The behaviour of investors has been the subject of a significant number of studies, and several theories have been developed in an effort to explain the remorse or overreaction that buyers and sellers frequently suffer in relation to monetary matters. When an investor is under a lot of stress, whether that stress is brought on by exhilaration or terror, it is possible for their emotions to take precedence over their ability to think rationally [1]. During what may seem like a limited period of time for capitalising on euphoric market developments or frightening market developments, it is crucial to take a sensible and realistic attitude to investing.

Financial independence can be achieved by direct and indirect investment in the stock market and other financial instruments [2][3]. From the same perspective, investment is one of the methods needed by Indonesia to improve the welfare of its people. Until now, there were many investment instruments and products that could be chosen by individuals or business entities to invest some of their funds [4][3]. In recent years, financial products have emerged as one of the most preferred vehicles for the people of Indonesia to use for the purpose of investing. This is illustrated by the rise in the number of Single Investor Identification (SID) cards, which has continued to see an annual increase of over 50 percentage points on average [5][3]. On the other hand, according to information provided by the Financial Services Authority or Otoritas Jasa Keuangan (OJK) Indonesia, there are already 143 Peer-to-Peer Landing Fintech platforms in Indonesia, in addition to 14 illegal investment enterprises that are thriving

2022 International Seminar on Application for Technology of Information and Communication (iSemantic) | 978-1-6654-8839-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/iSemantic55962.2022.9920431

[6]. When it comes to picking investing strategies and making judgments, one of the benefits that technology now offers to investors is the ability to make use of it. Because of its growing popularity and the fact that it makes it simpler for users to spread information, social media can be a useful tool for determining the benefits and drawbacks of an outstanding investment dilemma [7][8]. In social media sentiment analysis, consumers' opinions on a product, service, or brand are considered. Sentiment analysis of social media user comments helps effectively understand netizen psychology and decrease the threats posed by online public opinion[9]. According to information provided by the Ministry of Communication and Information of the Republic of Indonesia, there are approximately 19 million people living in the Republic of Indonesia who use Twitter as a platform to express a variety of opinions, including those that are positive, neutral, or negative [10].

Sentiment Analysis has been recognised as a valuable technique for assisting psychological evaluations and predicting mental health disorders, as well as the varied nature of valence expression, particularly with regard to emotion, thought, affect, decision, and decisiveness[11]. In addition to body language and facial expressions, human language is a means of expressing emotions. Whether symbolic or auditory representations of semantically significant and syntactically arranged units of a phrase or sentence are conveyed through it remains to be seen[12]. Though they are two independent systems, language and emotion are connected and influence one another's performance. Both aid in the exchange of information between individuals. Although facial emotions are more culture-specific than words such as ecstasy and astonishment, they may convey the objective message more effectively. Fridlund claims that Paul Ekman's universal facial expressions do not adequately explain cultural variances or aberrations in diverse social circumstances[13].

Previous research on analysing the public's perspective on investment applications was carried out by Amalida and Nuryana [10]. This study makes use of the data found on Twitter under the heading "Investment Application," which is then analysed using the Lexicon-Based and Naive Bayes Classifier methods. To provide a comparative value between the two approaches that were employed is the end goal of the research that is being done [10]. Since sentiment analysis is one of trend issue in the field of natural language processing (NLP), there are various models available to detect tha state of sentiment within a text, paragraph, or entire document [14]. It shows the impact of social media such as Twitter to understand user tweets and their behavior [15]. In this study, an exploration of the data on Twitter was carried out to find out and process every public opinion in regards to investment goods and instruments in Indonesia. This exploration was based on earlier research that was carried out. Negative information regarding publicly traded firms should be disclosed in the media for three reasons: to help promote a healthy market environment, to protect small and medium-sized investors' interests; and to motivate publicly traded companies to improve and flourish. Using the hashtag (#) on social media [16], various forms of digital activism have evolved in cyberspace. A hashtag (#) is a way for social media users to express their opinion on a specific topic during a specific event, such as on Twitter.

This paper is composed of the following sections: The first section discusses the backdrop of the challenges related with investment in Indonesia; the second section describes the methods utilised in this research; the third section contains the results gained from each stage of the methodology; and the fourth section concludes the study.

II. Research Method

Each process that will be applied in this research is illustrated through the flow chart in Figure 1 and is explained in detail in each sub-chapter of this section.

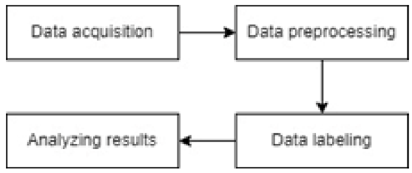

Fig. 1. Research Method

A. Data acquisition

The primary source of the data that is being managed for this research comes from tweets written in Indonesian that contain the word "invest." A technique called "scraping," which is a method for automatically extracting information from a website, is used to obtain each tweet. Scraping is a sort of data mining [17]. The process of web scraping enables us to collect data from a variety of websites directly onto our local system from the internet. It gathers information from a variety of internet portals by utilising Hypertext Transfer Protocols, and then processes this information in accordance with our specifications. This is utilised by many businesses in the process of data harvesting as well as the production of search engine bots. The Twitter Intelligence Tool, known as Twint, is an open-source Python library that was developed specifically to address issues that might arise while scraping data from Twitter. This library is what is used to implement the scraping process. The scraping tool that was used for this research was chosen after taking into account the number and period of tweet data that is not limited by Twint. As a result, it is possible to have tweet data that varies over a significant amount of time. This was one of the deciding factors in the scraping tool's selection. [18].

B. Data preprocessing

Data preprocessing is one of the crucial steps in analyzing sentiment, it will help to understand each tweet which is known as unstructured data in a better way to obtain a more optimal data format. Text preprocessing is the process of combining multiple word forms into one. Preprocessing's primary objective is to extract essential document characteristics from the obtained data set in order to improve the significance between phrases and documents and between words and classes[19]. Viewpoint mining features are the words, terms, or phrases that clearly reflect the opinion in a favourable or bad light. Thus, their influence on the text's direction is more significant than that of any other words within the same text [20]. At this stage, tweet data cleaning is carried out in several stages, including:

- *Cleaning*: to take off the extra spaces and punctuation, as well as any characters that aren't required
- *Case folding*: to change all sentences to no capital letters.

Before the data could be included in the model, contractions had to be expanded, links, hashtags, capitalization, and punctuation had to be removed. Negatives had to be addressed. A dictionary of negations had to be created so that negated terms could be handled efficiently. Whitespace and hyperlinks had to be eliminated. In order to make the model more robust, stop words other than the normal NLTK stop words were eliminated. The terms "investment" and "Indonesia" appeared prominently in the word clouds, as did the days of the week and their acronyms, as well as the names of the months. Following this step, the tweets were tokenized and PorterStemmer was used to extract the tweet's content.

C. Data labeling

After the data have been processed through the preprocessing stage, the labelling stage is carried out to analyse each and every sentiment that is contained in each tweet. This is done in an automated fashion with the help of TextBlob [21], which is a Python package that offers a straightforward Application Programming Interface (API) for gaining access to Natural Language Processing (NLP) operations [22]. TextBlob plays a part in this scenario by supplying the polarity and subjectivity values, which will be used as the fundamental values for assessing sentiment [23]. It would be very simplistic to infer that a negative sentiment score signifies depression for a binary classification model or a binary labelling system. Sentiment scores on their own do does not indicate depression, and it would also be incorrect to conclude that a negative score denotes depression.

III. Results and Discussion

A. Scraping Tweet Data

There are many libraries available, such as Twint, that have the ability to automate the process of web scraping. All of these libraries make use of a variety of application programming interfaces (APIs), allowing us to scrape data and save it locally in a data frame. Twint is a python tool that is open-source and used for scraping Twitter. This means that we can use Twint to extract data from Twitter even when we are not using the Twitter API. Twint provides a number of features that set it apart from other Twitter scraping APIs and make it more helpful than those APIs. It is possible to scrape tweets by using several parameters with the help of Twint. These parameters include hashtags, usernames, subjects, and more. It is even able to glean information like email addresses and phone numbers from the tweets themselves.

TABLE I. Twint Scraping Process

```
import twint

#Configure
c = twint.Config()
c.Search = “investasi”
c.Lang = “id”
c.Since = “2022-01-01”
c.Until = “2022-04-30”
c.Output = “./invest-id.csv”
c.Store_csv = True

#Run
twint.Run.Search(c)
```

The procedure for retrieving the data is carried out on May 1, 2022 within the time span beginning on January 1, 2022 and ending on April 30, 2022. There is no requirement for establishing a connection to a specific Twitter account in order to configure Twint. The procedure of installation just needs to be carried out in the same way as the process of installing Python libraries in general. We involved in this investigation opted to perform the scraping process is carried out by running the code as shown in Table I.

The scraping process's output will be saved in CSV format in the directory specified in the "c.Output" section. At this stage, a total of 143,079 tweets were gathered, as shown in Table II as a sample of twitter data that will be managed in the next step.

TABLE II. Tweet Data Examples

@adnanunique yep, properti sebagai instrumen investasi is wrong tapi ya itu yang paling aman, selama gak dilarang ya gas2 aja
Sales Investasi Cari Kesehatan Disingkat menjadi SICK
Yeayyy, gaji pertama as social media strategist!!! Investasi leher keatas gaakan pernah rugi!! https://t.co/7qo2SvN47y
Mencoba menjalani hidup dengan idealis, konsisten itu super sulit. Apalagi lihat manusia lain yg labil mirip bunglon, lebih sejahtera, lebih nanjak karirnya. Ya gpp, anggep aja investasi jangka panjang. Walaupun sering dilabeli oleh yg labil. Munafik emang, tapi itulah kehidupan.
Orang-orang pintar juga banyak yang ketipu soal bisnis dan investasi lho https://t.co/nLKX7GdeBf

B. Data preprocessing

To prepare the data for classification, preprocessing is necessary. HTML elements, scripts, and ads are common in online writings, which are often riddled with noise and clutter. In addition, many of the text's words have little bearing on the overall direction of the tweet. As a result, following the initial stage of preprocessing, the categorization process is aided by using similar word forms. This stage is implemented in Python by combining each cleaning procedure into a single function in accordance with Python standards. Table III depicts the code in detail.

TABLE III. Cleaning and Case Folding Process

```
def cleanTxt(text):
    #removing @mentions
    text = re.sub('@[A-Za-z0–9]+', '', text)
    #removing '#' hash tag
    text = re.sub('#', '', text)
    #lowercases all words
    text = text.lower()
    #removes all symbols
    text = re.sub(r'[^\w]', ' ', text)
    #removing RT
    text = re.sub('RT[\s]+', '', text)
    #removing hyperlink
    text = re.sub('https?:\/\/\S+', '', text)
    #removing tab, new line, ans back slice
    text = text.replace('\\t'," ").replace('\\n'," ").replace('\\u',"
").replace('\\',"")
    #removing non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')

    return text

df['tweet'] = df['tweet'].apply(cleanTxt)
```

We have done some basic cleaning of tweet that is denoising, because some of the symbols and words have no impact on the orientation of the text. All the characters that aren't required such as non-ASCII characters, mention symbol, hashtag symbol, any link, and additional white spaces are removed from tweet. While case folding converts all characters to lowercase, it is used to create the same letterform in the dataset. The data produced after running the code in Table III has been cleaned and case folded, as shown in Table IV.

TABLE IV. SAMPLE OF DATA THAT HAS BEEN THROUGH PREPROCESSING

yep properti sebagai instrumen investasi is wrong tapi ya itu yang paling aman selama gak dilarang ya gas2 aja
sales investasi cari kesehatan disingkat menjadi sick
yeayyy gaji pertama as social media strategist investasi leher keatas gaakan pernah rugi https t co 7qo2svn47y
mencoba menjalani hidup dengan idealis konsisten itu super sulit apalagi lihat manusia lain yg labil mirip bunglon lebih sejahtera lebih nanjak karirnya ya gpp anggep aja investasi jangka panjang walaupun sering dilabeli oleh yg labil munafik emang tapi itulah kehidupan
orang orang pintar juga banyak yang ketipu soal bisnis dan investasi lho https t co nlkx7gdebf

C. Subjectivity and Polarity labeling

Subjectivity represent the objectivity of a text provided, it relates to how much a person is personally interested in an issue. When presented as a personal experience, an opinion can have a high degree of subjectivity, whereas a low degree may represent someone else's viewpoint on something else. On the other hand, polarity reflects the probability a certain text has been classified as subjective, it relates to the strength of an opinion, which can be positive or negative. Sentiments can also have various levels of polarity depending on type of communication. This task is carried out using TextBlob's features. TextBlob performs sentence-level analysis, it will return two parameters called polarity and subjectivity as arranged in Table V.

TABLE V. APPLICATION OF TEXTBLOB

```
def getSubjectivity(text):
   return TextBlob(text).sentiment.subjectivity

def getPolarity(text):
   return  TextBlob(text).sentiment.polarity

df['Subjectivity'] = df['tweet'].apply(getSubjectivity)
df['Polarity'] = df['tweet'].apply(getPolarity)
```

The polarity score goes from -1 to 1, while subjectivity ranges from 0 to 1, with 0 being the most objective and 1 being the most subjective. Polarity values will determine the positive, negative, or neutral sentiment for the given tweet. The details of the polarity values are shown in Table VI.

TABLE VI. RANGE VALUE OF POLARITY

Polarity Value	Sentiment
(0) – (1)	Positive
0	Neutral
(-1) – (0)	Negative

The underlying programming logic for sentiment tagging is subject to this value limitation. Table VII shows the code in action, and Table VIII shows some instances of tweets that have been processed this way.

TABLE VII. TWEET LABELING PROCESS

```
def getAnalysis(score):
   if score < 0:
      return 'Negative'
   elif score == 0:
      return 'Neutral'
   else:
      return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)
```

TABLE VIII. EXAMPLES OF DATA THAT HAVE BEEN LABELED

Tweet	Subjectivity	Polarity	Sentiment
yep properti sebagai instrumen investasi is wrong tapi ya itu yang paling aman selama gak dilarang ya gas2 aja	0.9	-0.5	Negative
sales investasi cari kesehatan disingkat menjadi sick	0.857143	-0.71429	Negative
yeayyy gaji pertama as social media strategist investasi leher keatas gaakan pernah rugi https t co 7qo2svn47y	0.066667	0.033333	Positive
mencoba menjalani hidup dengan idealis konsisten itu super sulit apalagi lihat manusia lain yg labil mirip bunglon lebih sejahtera lebih nanjak karirnya ya gpp anggep aja investasi jangka panjang walaupun sering dilabeli oleh yg labil munafik emang tapi itulah kehidupan	0.066667	0.033333	Positive
orang orang pintar juga banyak yang ketipu soal bisnis dan investasi lho https t co nlkx7gdebf	0	0	Neutral

D. Analyzing results

Based on the findings of the research presented in this paper, the following are some investment recommendations: Because the unfavourable news events that were analysed in this research are just a tiny portion of the influencing variables of investment fluctuation and because they cannot reflect the entire scenario, the suggestions that were offered in this paper were only meant to serve as a reference. The way emotions are expressed in a language varies from culture to culture. People in the eastern country use more emotionally charged terms than those in the western country. Emotions can be expressed in both verbal and nonverbal ways, but verbal communication is easier to understand than nonverbal communication. Language enhances human intellect and speeds up problem-solving and decision-making by allowing us to express ourselves.

TABLE IX. FINAL RESULT OF PROCESSED DATA

Month	Total Tweet		
	Positive	Negative	Neutral
January	1784	655	29408
February	1987	823	29473
March	2057	1024	33511
April	3158	1196	38003

IV. Conclusion

An automated web scraping technique was utilised to compile 143,079 tweets representing public sentiment regarding the investment. After using the text processing library, there were a total of 130,395 opinions that were either favourable or negative, with 8986 opinions having a positive sentiment and 3698 opinions having a negative feeling. When taken into further consideration, negative sentiment witnessed a continuous growth of around 25 percent in both February and March, while positive sentiment experienced a significant spike of 53 percent in April.

Netizens are paying more attention to online news and public opinion because of the fast expansion of online media. Every netizen is both a consumer of online media and a developer and disseminator of that media as well. Internet users are more likely to be exposed to conflicting, irritating and ambiguous negative information in an increasingly diverse media environment because of the increased timeliness and convenience of the mobile Internet and the increased attention given to negative news on the Internet, according to an analysis of public psychological factors. This can be one of the things that people look at to symbolise how they feel about the growing importance of investing. supported by a monthly growth in either negative or positive sentiment, which indicates that a growing number of individuals are thinking about this investment and having an opinion about it. In later research, it may be able to implement a greater variety of approaches at the pre-processing techniques and feature selection process.

References

[1] T. K. Purnamaningrum, "Investor Sentiment and Stock Price Movement Property Sector in Indonesia Stock Exchange," *Proc. Int. Conf. Manag. Accounting, Econ.*, vol. 151, pp. 366–368, 2020, doi: 10.2991/aebmr.k.200915.083.

[2] J. Manajemen, V. Nomor, J. Desember, and M. Berinvestasi, "Value," vol. 15, pp. 60–73, 2020.

[3] U. Safryani, A. Aziz, and N. Triwahyuningtyas, "Analisis Literasi Keuangan, Perilaku Keuangan, Dan Pendapatan Terhadap Keputusan Investasi," *J. Ilm. Akunt. Kesatuan*, vol. 8, no. 3, pp. 319–332, 2020, doi: 10.37641/jiakes.v8i3.384.

[4] F. Firdhausa and R. Apriani, "Pengaruh Platform Media Sosial Terhadap Minat Generasi Milenial Dalam Berinvestasi di Pasar Modal," *J. Supremasi Huk.*, vol. 17, pp. 96–103, 2021.

[5] M. H. Purnomo and C. A. Kartika, "Dampak Pandemi Covid-19 terhadap Pasar Saham di PT Bursa Efek Indonesia (BEI)," *J. Kaji. Ekon. dan Bisnis Islam*, vol. 5, no. 2, pp. 245–253, 2022.

[6] M. Finthariasari, E. Febriansyah, and K. Pramadeka, "Pemberdayaan Masyarakat Desa Pelangkian Melalui Edukasi Dan Literasi Keuangan Pasar Modal Menuju Masyarakat Cerdas Berinvestasi," *J. Pengabdi. Masy. Bumi Raflesia*, vol. 3, no. 1, 2020, doi: 10.36085/jpmbr.v3i1.763.

[7] E. A. Sukma, A. N. Hidayanto, A. I. Pandesenda, A. N. Yahya, P. Widharto, and U. Rahardja, "Sentiment Analysis of the New Indonesian Government Policy (Omnibus Law) on Social Media Twitter," *Proc. - 2nd Int. Conf. Informatics, Multimedia, Cyber, Inf. Syst. ICIMCIS 2020*, no. July 2017, pp. 153–158, 2020, doi: 10.1109/ICIMCIS51567.2020.9354287.

[8] Y. Ruan, A. Durresi, and L. Alfantoukh, "Using Twitter trust network for stock market analysis," *Knowledge-Based Syst.*, vol. 145, pp. 207–218, 2018, doi: 10.1016/j.knosys.2018.01.016.

[9] X. Jiang, C. Song, Y. Xu, Y. Li, and Y. Peng, "Research on sentiment classification for netizens based on the BERT-BiLSTM-TextCNN model," *PeerJ Comput. Sci.*, vol. 8, p. e1005, 2022, doi: 10.7717/peerj-cs.1005.

[10] F. Amaliah and I. K. D. Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter," *J. Informatics Comput. Sci.*, vol. 03, no. 03, pp. 384–393, 2022.

[11] S. Saraff, R. Taraban, R. Rishipal, R. Biswal, S. Kedas, and S. Gupta, "Application of Sentiment Analysis in Understanding Human Emotions and Behaviour," *EAI Endorsed Trans. Smart Cities*, vol. 5, no. 13, p. 166547, 2018, doi: 10.4108/eai.6-10-2020.166547.

[12] M. Bamberg, "Language,Concept and Emotion," *Lang. Sci.*, vol. 19, no. 4, pp. 309–340, 1997.

[13] A. J. Fridlund, "The new ethology of human facial expressions," *Psychol. Facial Expr.*, no. May, pp. 103–130, 2010, doi: 10.1017/cbo9780511659911.007.

[14] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks," *Kurdistan J. Appl. Res.*, pp. 54–65, 2020, doi: 10.24017/covid.8.

[15] V. Bonta, N. Kumaresh, and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," *Asian J. Comput. Sci. Technol.*, vol. 8, pp. 1–6, 2019, doi: 10.51983/ajcst-2019.8.s2.2037.

[16] A. B. G. A. Y. Indar Sugiarto, "Social Network Analysis of Citizen Initiated Vaccination Campaigns on Twitter," in *International Conference on Community Empowerment and Engagement (ICCEE 2021)*, 2022, no. ICCEE 2021, pp. 122–132, doi: 10.1108/S2043-905920210000015014.

[17] A. Hernandez-Suarez, G. Sánchez-Pérez, K. Toscano-Medina, V. Martínez-Hernández, V. Sanchez, and H. Meana, "A Web Scraping Methodology for Bypassing Twitter API Restrictions," *ArXiv J.*, vol. abs/1803.0, 2018.

[18] V. Santoni and S. Rufat, "How fast is fast enough? Twitter usability during emergencies," *Geoforum J.*, vol. 124, no. May 2018, pp. 20–35, 2021, doi: 10.1016/j.geoforum.2021.05.007.

[19] W. A. Prabowo and F. Azizah, "RESTI journal," no. 10, pp. 11–12, 2021.

[20] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.

[21] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.

[22] D. Hazarika, G. Konwar, S. Deb, and D. J. Bora, "Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing," *Proc. Int. Conf. Res. Manag. Technovation 2020*, vol. 24, pp. 63–67, 2020, doi: 10.15439/2020km20.

[23] S. Ahuja and G. Dubey, "Clustering and sentiment analysis on Twitter data," *2017 2nd Int. Conf. Telecommun. Networks*, pp. 1–5, 2017, doi: 10.1109/TEL-NET.2017.8343568.