This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Building a Calendar of Events Database by Analyzing Financial Spikes

PRAKASH K. AITHAL[1], (Member, IEEE), DINESH ACHARYA U.[2], GEETHA M.[3], AND PARTHIV MENON[4]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: prakash.aithal@manipal.edu)
[2]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: dinesh.acharya@manipal.edu)
[3]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: geetha.maiya@manipal.edu)
[4]Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India (e-mail: parthiv.menon@learner.manipal.edu)

Corresponding author: Geetha M. (e-mail: geetha.maiya@ manipal.edu).

ABSTRACT
An event is a piece of news that triggers a change in stock prices. Here, an event study is undertaken to capture the effect of abnormal returns due to an event. The event can affect the stock market in the long term or short term. Event research is relevant to both the efficient market hypothesis and behavioral finance.
In this study, we collected data from websites that manage financial and economic data, performed a sentiment analysis, and correlated news article data with changes in a particular company's stock prices in the stock market. Data were collected from two well-known financial news websites. We observed a correlation between stock prices and news items. An event period of one day was considered for the study. The regression equation determined the relationship between stock returns and polarity and subjectivity. Bayesian model averaging was performed to identify the effects of polarity and subjectivity on stock returns. Time-series data were decomposed into components and detrended via regression. Prominent keywords and their polarity values for a particular day were plotted. An event enhanced some stock returns while adversely affecting other stocks. We found a variation range of -400 to 200 for different company stocks for the selected period.

INDEX TERMS Event Database, Social Media Data, Sentiment Analysis, Data Analytics, Event Studies

I. INTRODUCTION

DATA Science involves extracting information from raw data, converting information into knowledge and knowledge into an actionable plan, which has broad applications for fields ranging from agriculture and space science to arts and finance and which applies to all fields in which data are used.

Finance is essential to a nation's growth. Primarily two kinds of financial data are used: quantitative data on variables such as stock prices, the volume of stock, and the PE ratio and qualitative data drawn from material such as director reports, auditor reports, financial statements, and financial news. Quantitative data reveal a stock's movement, while qualitative data unearth investor sentiment, which is the driving force behind stock movement.

Stock prices fluctuate based on financial events captured by qualitative data. Events can include management changes, declarations of dividends, festivals, natural calamities, pandemics, recessions, budget and earnings releases, and quarterly and yearly statements. Some events, such as festivals and yearly statements, occur once a year. Changes in management have either positive or adverse effects. Dividends always decrease the value of a stock. Events have either short-term consequences or affect stocks over the long run. Events that affect the stock market for less than a week are considered to be short-term events and that rest are long-term events. An event can affect either a particular stock or the entire stock market. The present event study aims to determine the abnormal stock return attributable to new information carried by an event.

Financial events provide both quantitative and qualitative data. Qualitative data such as financial news must be quanti-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

fied. One can apply a sentiment analysis to quantify financial news. Sentiment can be measured based on tone, also known as polarity, or subjectivity. Polarity values vary from -1 to 1 where a value of -1 denotes extremely negative sentiment and +1 denotes total positive sentiment. Subjectivity levels vary from 0 to 1. The lower the value of subjectivity is, the more meaning can be extracted from the news.

There is a correlation between financial news and stock prices. If the correlation between news and the stock difference is greater, news influences the stock market. Event studies assume the semistrong form of market efficiency. The semistrong form of market efficiency applies two assumptions[1].

- All publicly available information is reflected in stock prices
- Stock prices vary when new information arrives

The latter assumption is utilized to predict the future value of a stock in advance. An event changes a stock price, and an event study can then be utilized to analyze future cashflows.

A. BAYESIAN MODEL AVERAGING

An analysis is conducted by first selecting the best model based on a particular criterion and then learning about the parameters of the chosen model. Any uncertainties of the model selection process are ignored. Contrary to this approach, Bayesian model averaging (BMA) learns parameters for all candidate models and allows one to select a model with specific posterior probabilities. BMA offers the following advantages over other models:

1) BMA captures the uncertainty of a model.
2) BMA reduces error.
3) BMA retains all model uncertainty until the final stage.
4) BMA continuously adjusts model weights.
5) BMA uses a set of models, which is preferable to using a single model.

B. TIME SERIES COMPONENTS

It is helpful to decompose the time series into systematic and nonsystematic components. Systematic components can be modeled as either consistent or recurring. Nonsystematic elements are random and cannot be modeled. The time series consists of the following components:

1) Level
2) Trend
3) Seasonality
4) Noise

The level is the mean value of the time series. The trend is either increasing or decreasing in the time series. The repeating short-term cycle is defined as seasonality. Randomness present in the time series that cannot be modeled is defined as noise.

The rest of the paper is organized as follows. Section 2 reviews the related literature, section 3 elaborates on the methodology used, and section 4 presents the results. Conclusions and avenues for future research are given in section 5.

II. LITERATURE SURVEY

Event studies determine excess returns attributable to a corporate event. Events can range from an earnings release by a company or budget announcement to a new product launch, announcement from a competing company, regular financial statement or announcement made by a regulatory body. Researchers look forward in time and predict the expected value a firm will derive from a corporate action announced to the public. This forward-looking focus of event studies makes them more potent than other metrics such as return on investment (ROI) and profits. Investors respond positively to product launches, sponsorships, and mergers, while they respond negatively to events such as product recalls. [2]. It has been observed that historical stock returns are affected by calendar events. When event studies have been conducted in developed countries, little related literature is available from developing nations. The stock market declines on Mondays and strengthens on Friday. The start of a new month or year also affects the stock market[3]. The presence of long-term abnormal returns contradicts the efficient market hypothesis. Different testing strategies are applicable to asset pricing and buy-and-hold methodologies[4].

To identify the abnormal returns of the portfolio of stocks affected by an event, a control portfolio of stocks not affected by the event is considered as a benchmark, and the results of both portfolios are compared[5]. Event studies can evaluate a government's policies, as markets react to new news immediately and adjust their value. If a policy has a positive effect on the market, it will eventually succeed. Speculation on an event also has a major impact on the stock market. The stock market reacts to speculation and is affected long before an event is announced[6]. Event studies require the use of a window period under which an event is studied. An event study evaluation is done based on the methodology utilized and can involve a regression with a t-test or the use of a capital asset pricing model (CAPM) with a t-test[7].

An election is a major event that significantly influences the stock market. A set of stocks are affected by an election. Elections have an effect for approximately five months before and after an event. The announcement of an election and the election itself greatly impact the stock market [8]. Event studies have become the defacto standard for measuring stock prices during and after an event. Abnormal returns are determined either as the residuals from standard normalized benchmarks or as a dummy variable in a regression equation[9]. News can be classified as repeating or surprising. A repeating event can be easily predicted, and markets adjust to such an event. A surprising event, on the other hand, leads to abnormal returns. This is the case because such an event is unknown, and no estimate has been made, which results in risk that must be addressed[10].

In [11], the authors conclude that intraday abnormal returns are normally distributed, and Pattel test statistics function better than other statistical measures in measuring abnormal returns. The market responds to both good and bad news. Good news reduces risk, while a piece of bad news increases it. The efficient market hypothesis and behavioral finance apply in some cases and fail in others. Both have explanatory power and must be explored. According to event studies, stocks will not alter prices unless and until a piece of news arrives[12]. Events do not have structure and cannot be stored in relational databases. Therefore, Not Only Structured Query Language (NoSQL) databases are utilized to store information on such events. NoSQL databases require no downtime and are different from regular relational database management systems (RDBMSs), as they can scale well at a lower cost. A RDBMS has strict atomicity consistency isolation durability (ACID) properties, while NoSQL databases adhere to the consistency availability and partition tolerance (CAP) theorem. Complex event processing cannot be achieved using a regular RDBMS, so one must in this case use NoSQL databases for faster processing. Complex event processing requires low latency, high throughput, and temporal and spatial event processing capabilities[13]. Stock prices move with a correlated random walk rather than with an uncorrelated random walk. Correlations can be found through event studies[14].

One comment posted on the Yahoo financial website almost crashed the Nevada company, after which the company had to call a press conference to prove that its fundamentals are strong. This incident demonstrates that event studies are critical. The simplest measures used include the number of times a stock name is mentioned, the frequency of certain keywords, and the sentiment of news about a given company[15]. There are two types of events, namely, simple events and complex events. A simple event occurs instantaneously and independent of other events. A complex event is a set of events that are interrelated[16]. The traditional investment strategy takes months or years to arrive at a decision. The advancement of information technology in the realm of finance has made the decision-making process faster. Events can be analyzed in milliseconds using information technology[17].

The efficient market hypothesis states that market prices reflect current news and that as new news arrives, prices change. Efficient market theory considers investors to be unemotional. Behavioral finance, on the other hand, treats investors as an emotional entity and assumes news events to affect investments. Both the efficient market hypothesis and behavioral finance agree that news events are responsible for abnormal returns[18]. Text mining is used in event studies, and it requires words to be segmented, and spaces are used to separate words. Stop words need to be removed. The sentiment of a sentence and frequency of words measure the quantum of the jump of stock prices[19].

Sentiment analysis is widely used in the field of finance. The method is also known as opinion mining. Sentiment analysis is broadly classified into emotion recognition and tone analysis or polarity analysis. Opinion mining can be classified based on whether a dictionary or machine learning is used to arrive at an opinion. Machine learning-based approaches require domain-specific content to work correctly, while dictionaries are costly[20]. Event analysis must be conducted by batch processing of historical events. The event pattern must support mathematical, logical, and statistical patterns among events[21]. An event study involves several steps and many choices must be made at each step. Thus, there is no standard means of conducting an event study. The timing of an event study and the event window size need to be defined by the researcher. Surprise events need to be evaluated through event studies[22].

A historical analysis of 145100 news articles has shown a significant relationship between tone and companies' performance. However, a single instance of news does not have an impact on stock movement. If a company calls a conference call and the tone is positive, the stock of that company will move multifold in a positive direction, resulting in an abnormal positive return[23]. The global financial crisis resulted in governments intervening and taking regulatory action. Global regulatory action is an event that affects the stock market worldwide[24].

The coronavirus pandemic has affected US asset prices. The pandemic has resulted in high levels of volatility in small and unstable markets. Stock markets are interrelated and sensitive to news. Media coverage on the pandemic has resulted in an adverse crash in some highly volatile stocks and unstable industries. The global stock markets are becoming interdependent. Even when a single stock market is affected, this impact spreads to other stock markets[25]. [26] studied corporate social responsibility with the help of a dictionary. The authors showed that when a firm is not responsible and this results in a negative event, its stock price will fall drastically. Event and firm performance are directly proportional. Narrative economics has developed into a field of its own. A narrative event influences a stock, and numerous examples of this phenomenon prove this fact[27]. The pandemic has caused a loss to the economy. A historical influenza revealed that a pandemic has effects similar to those of influenza but on a larger scale. When a pandemic affects key persons, this will have a tremendous impact on the stock market[28].

Event studies cover two time periods, pre-event and postevent periods, and examine two groups, control and treatment groups. If the event period varies, its average can be taken as defined in [29]. Event information can be soft or hard. Hard information can be quantified and quickly converted into a number. Soft information requires context. When one separates soft information from its source, it becomes useless. Hard information is known as quantitative, while soft information is known as qualitative[30]. Big data technologies can store qualitative data on events rather than quantitative data along. The field of finance has concentrated on the use of quantitative data for analyses. With the emergence of big data technologies, the storage and processing of qualitative data have become more accessible. The focus of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

the researcher has thus switched from the use of quantitative data to the use of qualitative data[31].

[32] calculate economic uncertainty based on newspaper coverage frequency and focus on monetary policy making. On the other hand, our work deals with event studies on how much abnormal return can be obtained from an event.

The above literature review that event studies require the examination of an event period and that there are many means of performing them. An event can be studied with two groups, control and treatment groups, or the standard index value. There are a variety of ways to conduct an event study. An event study must determine the period of research and adopt a regression equation.

III. METHODOLOGY

Our methodology is presented in subsections III-A and III-B, respectively, titled Tools and Process.

A. TOOLS

A thorough review of the two websites used, namely, moneycontrol.com and economictimes.com, revealed which routes gave relevant data. Then, the type of database and database schema were selected.

We selected Python as our language of choice for data collection and its related mathematical analysis and we selected MongoDB as are database. MongoDB is a scalable and fast NoSQL database and is known for its efficient querying capabilities. Initially, PyMongo, a native Python driver for MongoDB, was used to query the database. Subsequently, the code was modified to use mongoengine—a Python library that acts as an object document mapper for MongoDB. It then was easier to specify schemas and manipulate data obtained from the database in the form of objects with each field of a known type.

A request library was used to retrieve content from websites. BeautifulSoup, a Python package for parsing Hyper Text Markup Language (HTML) and eXtensible Markup Language (XML) documents, was used to analyze the data, scrape webpages, and parse the data into useful data. In addition, the pandas library was used to manipulate .csv files and other data frames to efficiently structure the pre-existing data.

TextBlob, a Python library built from the Natural Language Tool Kit (NLTK), provides user-friendly interfaces for performing natural language processing (NLP) tasks such as sentiment analyses of text. The NLTK was also used later in the project to process text and extract relevant information.

Git (and GitHub) was used for version control of the project files and folders. To make the code portable, environmental variables, along with the dotenv module, were used to dynamically load variables during execution.

B. PROCESS

Events were collected from moneycontrol.com and economictimes.com. The data are stored in the MongoDB database. Two collections are stored in the database—detailed stock values and events from the websites. The data fields of these two collections are as shown below.

NewsItem(News,
Date,
Time,
CompanyName,
Polarity,
Subjectivity)

Stocks(CompanySymbol,
OpeningValue,
ClosingValue,
Difference)

Field news stores scrapped news from the two websites. When information is published, the date and time are stored in the date and time fields, respectively. The company name is considered optional since news may affect multiple companies without citing a company name. Polarity values vary from -1 to 1. If TP represents the total number of positive words and TN represents the total number of negative words, polarity is defined as follows (1).

$$Polarity = (TP - TN)/(TP + TN) \tag{1}$$

Subjectivity is a type of probability that determines how objective a sentence is. Subjectivity, as it is a probability value, varies from 0 to 1. Further details on computing subjectivity can be found in [33].

The company symbol includes a ticker symbol for the Nifty 50 companies. The opening price is the stock price at the beginning of the day. The closing price is the price of the stock at the end of the day. The difference field contains the difference between closing and opening prices.

We collected and processed our data over the following steps.

1) The date and time of the event and the actual news article were extracted from the web.
2) The opening and closing prices of various companies' stock prices were collected, and the difference between these prices was computed. Data regarding companies and their symbols were obtained from NSE . These data were then used to dynamically obtain the stock values for every required instance from the Yahoo Finance website.
3) The correlation between the difference in opening and closing stock values and the various news articles of a company is found using equations (2) and (3), where C_P and C_S denote the correlations of polarity and subjectivity with returns, respectively. P denotes polarity, R denotes the return, and S denote subjectivity, relating

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

the news to stock fluctuations.

$$C_P = \frac{\sum(P_i - mean(P)) * (R_i - mean(R))}{\sqrt{\sum(P_i - mean(P))^2 * \sum(R_i - mean(R))^2}} \tag{2}$$

$$C_S = \frac{\sum(S_i - mean(S)) * (R_i - mean(R))}{\sqrt{\sum(S_i - mean(S))^2 * \sum(R_i - mean(R))^2}} \tag{3}$$

4) The following graphs are plotted
   - Keywords and their impact on investor sentiment for mean polarity
   - Keywords and their impact on investor sentiment for mean subjectivity
   - Mean polarity on a particular day
   - Variation in Nifty 50 stocks within a day due to a budget event
   - Correlation between stock returns and polarity
   - Correlation between stock returns and subjectivity

As the script is executed, the news item collection is updated without duplicates, and stock collection is refreshed.

BMA utilizes a linear regression model for predictions, considers all parameters and applies linear regression models for all possible combinations of the parameters. Two independent variables are used in the current paper, namely, polarity and subjectivity, and one dependent variable, return, is also used. BMA applies linear regression in the following parameter combinations:

1) Return
2) Polarity
3) Subjectivity
4) Return and Polarity
5) Return and Subjectivity
6) Polarity and Subjectivity
7) Return, Polarity, and Subjectivity

BMA finds the likelihood of each combination and the posterior probability of each parameter. A decision is made based on the posterior probability of the variable.

IV. RESULTS

Events were collected in the MongoDB database as two collections, namely, news items and stocks. News articles were collected during Indian budgeting time, and the articles have a positive tone. The sampled database collections of news items and stocks are described in Fig.1 and Fig.2 respectively. We collected information on the date and time of each news item, on the article itself, and on degrees of article polarization and subjectivity. For stocks, we collected stock symbols, the dates of closing and opening prices, the opening prices of particular days, the closing prices of particular days, and the differences between closing and opening prices.

The keywords and their impacts on the given investor for the given time period with respect to tone are presented in Fig. 3, and the same results with respect to subjectivity are shown in Fig. 4 From the figures, it is clear that news on budgets has a positive tone. The heat maps depicted in Figures 1 and 2 show that the article keywords have a positive tone, and words such as "high" are more popular than word such as "loss" during the period. The word "loss" is not used on February 1st, 2021, India's budget day. The word "high" is often used in the news articles.

The mean polarity within a day is shown in Fig. 5 During the Indian annual budget event held in Feb. 2021 , the mean polarity value is positive for the entire week of the event.

The variation in the stock prices of Nifty 50 companies is shown in Fig. 6. From Fig.6, it is clear that due to budget events, some individual stocks are positively affected, while others are negatively affected. The variation ranges from -400 to 200.

Fig.7 and Fig.8 demonstrate the effect of polarity and subjectivity on the stock returns of Nifty 50 companies. For a given polarity and subjectivity value, the companies react differently. From Figures 7 and 8, the return is not approximately zero for all of the companies. We also find variation ranging from -400 to 200 due to the event. Certain companies react to specific events.

The time components of return, polarity, and subjectivity are given in Fig. 9 to Fig. 11, respectively. The figures show that the seasonality component is not present in the data, and the observed component reflects a trend.

The regression equation of degree one with returns as the dependent variable and polarity as an independent variable is illustrated by equation 4. The regression equation with returns as a dependent variable and subjectivity as an independent variable is presented as equation 5. The polynomial regression equation of degree five with returns as a dependent variable and polarity as an independent variable is given by equation 6. The polynomial regression equation of degree five with returns as a dependent variable and subjectivity as an independent variable is written as equation 7. The top 15 keywords of the news items for the budget week were identified. Table3 lists the top 15 keywords in descending order of frequency. The table shows that the development of coronavirus vaccines has led to positive sentiment among investors. Pandemic-related words account for 5 of the 15 most frequent words.

$$Returns = 0.003 + 0.009 * Polarity \tag{4}$$

$$Returns = 0.0002 + 0.007 * Subjectivity \tag{5}$$

$$\begin{aligned} Returns = 0.001 - 0.122 * Polarity \\ -0.025 * (Polarity)^2 \\ -11.81 * (Polarity)^3 \\ +0.46 * (Polarity)^4 \\ +241.89 * (Polarity)^5 \end{aligned} \tag{6}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

$$\begin{aligned} Returns = -1.02 - 12.68 * Subjectivity \\ -60.48 * (Subjectivity)^2 \\ +138.82 * (Subjectivity)^3 \\ -152.97 * (Subjectivity)^4 \\ +64.47 * (Subjectivity)^5 \end{aligned} \quad (7)$$

The BMA results are tabulated in tables 1 and 2 and show that polarity and subjectivity improve the model's performance. Polarity outperforms subjectivity. It is better to consider polarity and subjectivity as model parameters. We also find that certain parameters influence the model, but they are not included in the model.

TABLE 1: Variables and their Likelihood Values

Variable	Likelihood
Const	4.58e+81
Polarity	6.58e+81
Subjectivity	5.85e+81
Const and Polarity	1.29e+81
Const and Subjectivity	8.46e+80
Polarity and Subjectivity	1.44e+81
Const, Polarity, and Subjectivity	2.10e+80

TABLE 2: Variables and their Posterior Probability Values

Variable	Probability	Average Coefficient
Const	0.333	-0.00033
Polarity	0.458	-0.00567
Subjectivity	0.402	-0.00178

TABLE 3: Keywords and their frequencies of use during budget week Feb. 1 to Feb. 7, 2021

Keyword	Frequency
Government	186
Vaccine	178
Budget	170
Economy	72
Pandemic	70
COVID-19	68
Tax	68
Growth	67
Virus	64
Farmers	56
Coronavirus	42
Infrastructure	41
Capital	40
Inflation	40
GDP	23

Spearman rho and Kendall tau rank correlations are found as well as a slightly positive correlation between polarity and returns. Additionally, we find a slightly positive correlation between subjectivity and returns.

V. CONCLUSIONS AND FUTURE WORK

Two prominent financial websites, namely, moneycontrol.com and economictimes.com, were scraped to build a calendar of events database. The Python language was utilized to scrape the websites. Time components were decomposed from data, and we found the only trend is present in the data, which was detrended via regression. BMA was applied with linear regression, and we found polarity to be a more significant parameter than subjectivity. We also show that the discovery of coronavirus vaccines is positively affecting the Indian stock market. To summarize, this article contributes to the existing body of knowledge in the following ways:

1) A calendar of events database is created.
2) Insights are derived from the events database with quantitative and qualitative data.
3) We find that the invention of coronavirus vaccines has impacted the stock market.

Event studies are a useful means of analyzing the stock market, as stock markets are sensitive to events. Portfolio optimization via event studies can be considered in future enhancements. Further work can also categorize events into the following categories: global, national, sectoral, and company-specific events.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

FIGURE 1: News Item Collection

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access


```
_id: ObjectId("601f5990e3f45f20e6e089d2")
symbol: "INFY"
date: "05-Feb-2021"
opening: 1286.4
closing: 1272.1
difference: -14.300000000000182
```

```
_id: ObjectId("601f5996e3f45f20e6e089d3")
symbol: "JSWSTEEL"
date: "05-Feb-2021"
opening: 403
closing: 402.2
difference: -0.8000000000000114
```

```
_id: ObjectId("601f599fe3f45f20e6e089d4")
symbol: "KOTAKBANK"
date: "05-Feb-2021"
opening: 1915
closing: 1982.7
difference: 67.70000000000005
```


FIGURE 2: Stock Collection

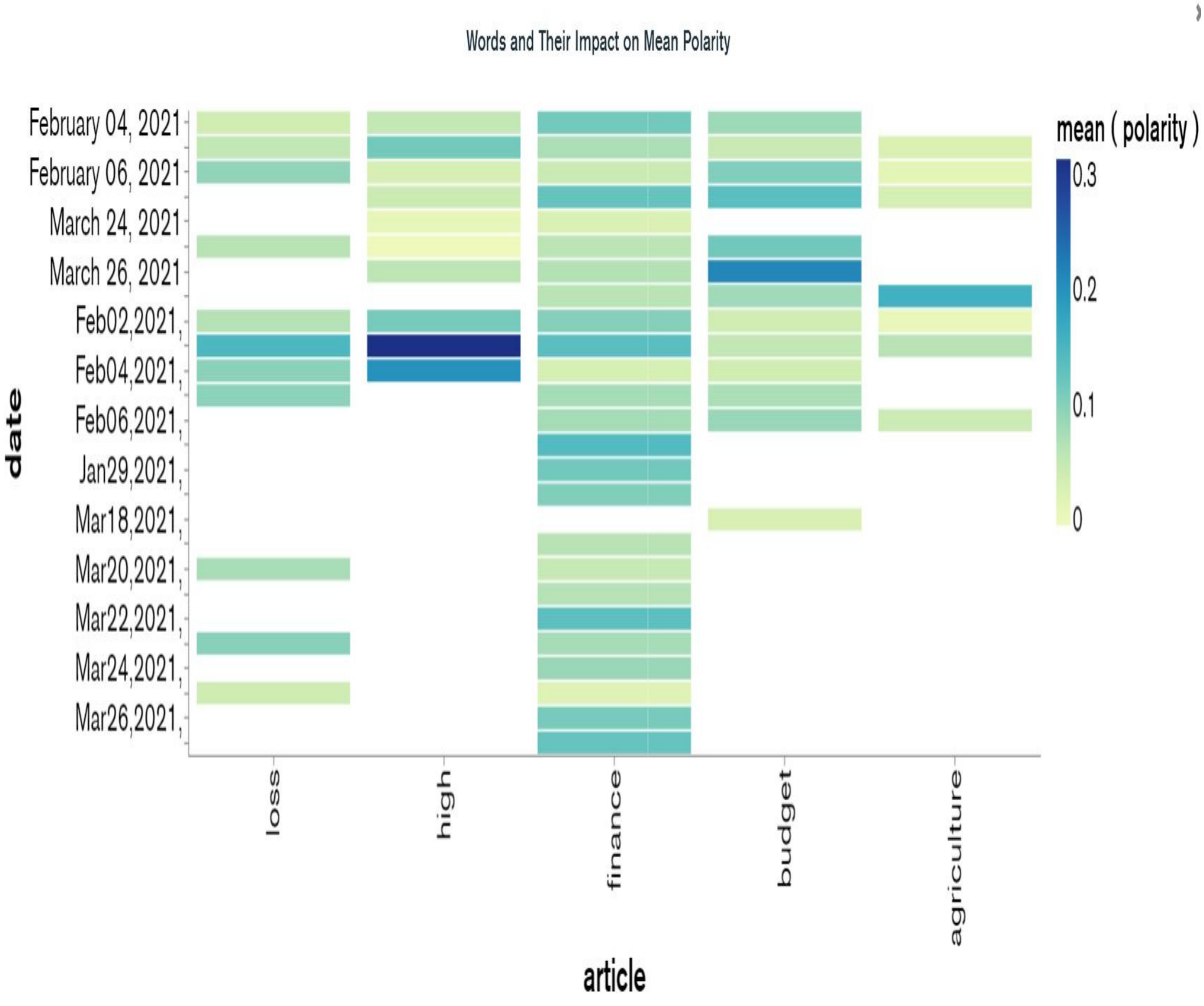

FIGURE 3: Keywords and their impact on investor sentiment for mean polarity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

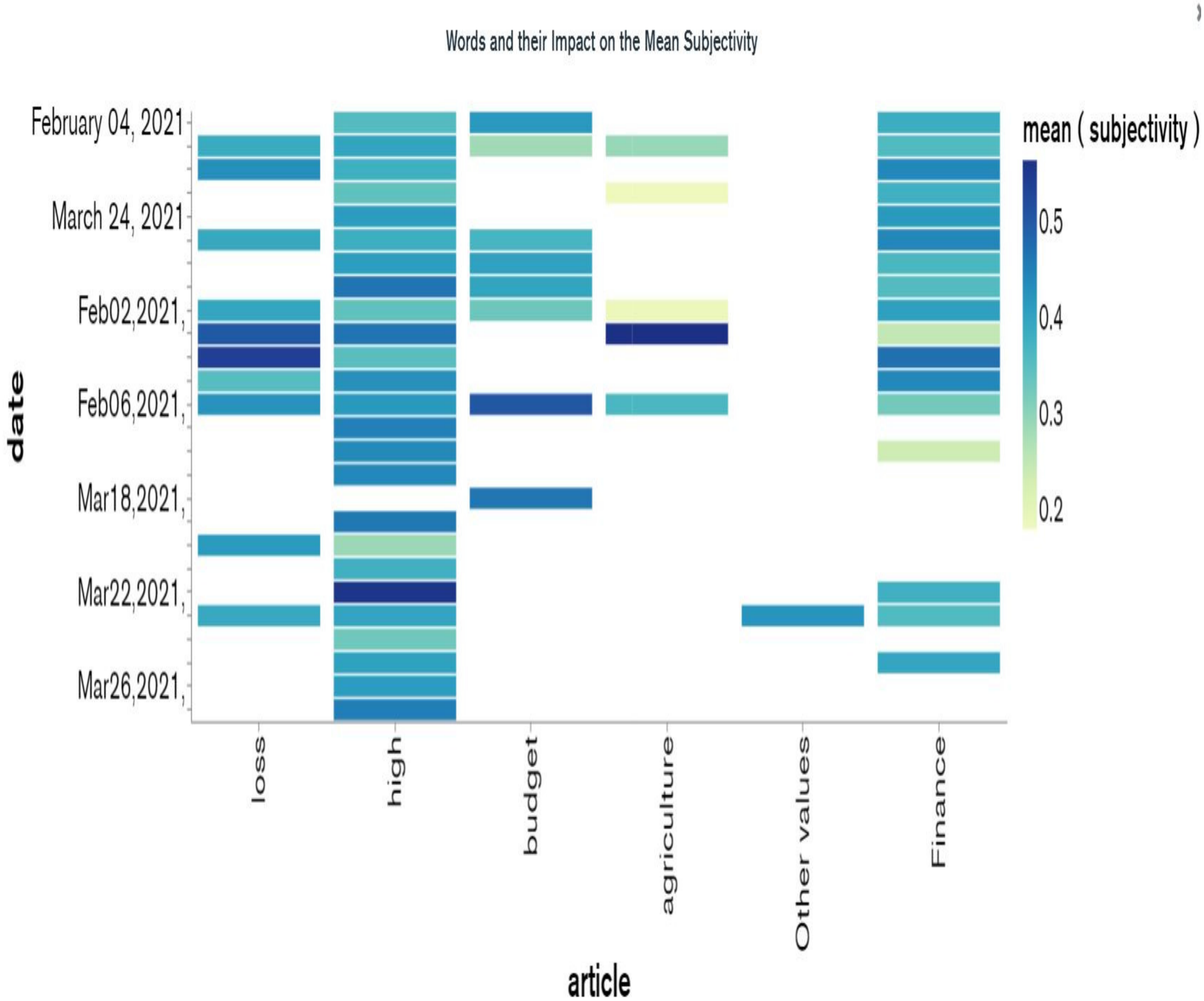

FIGURE 4: Keywords and their impact on investor sentiment for mean subjectivity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

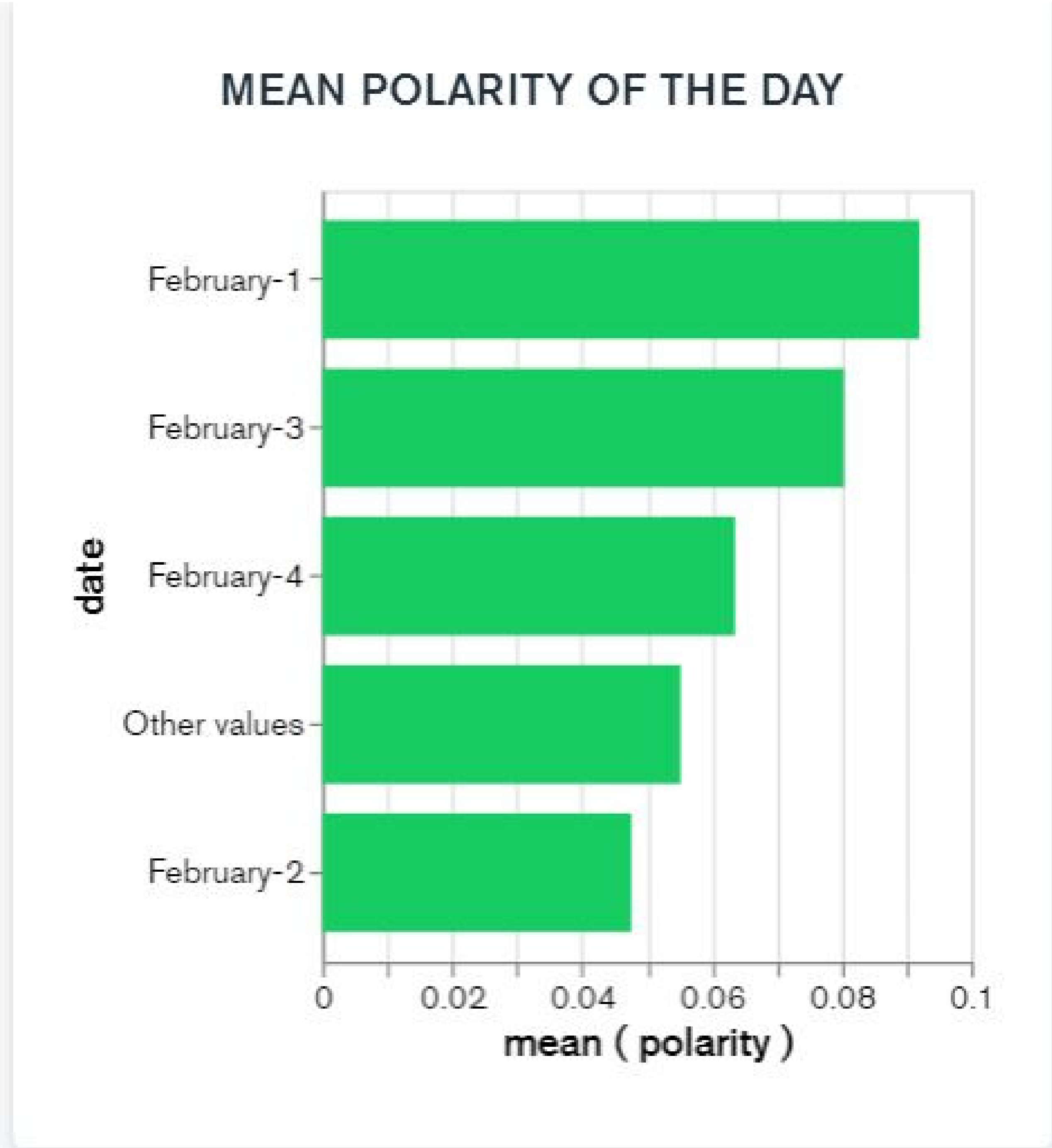

FIGURE 5: Mean Polarity in a Day

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

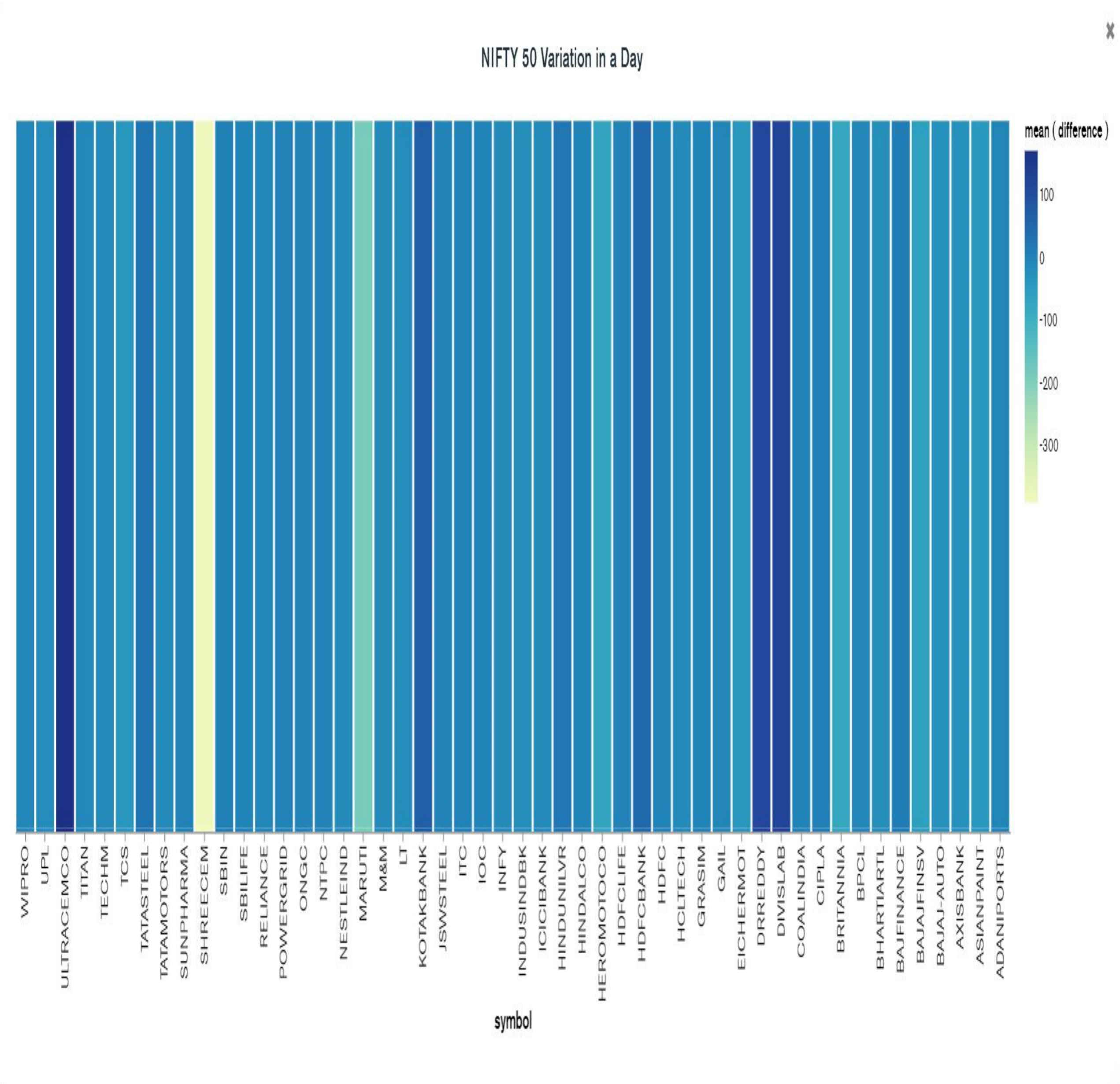

FIGURE 6: Nifty 50 Variation in a Day

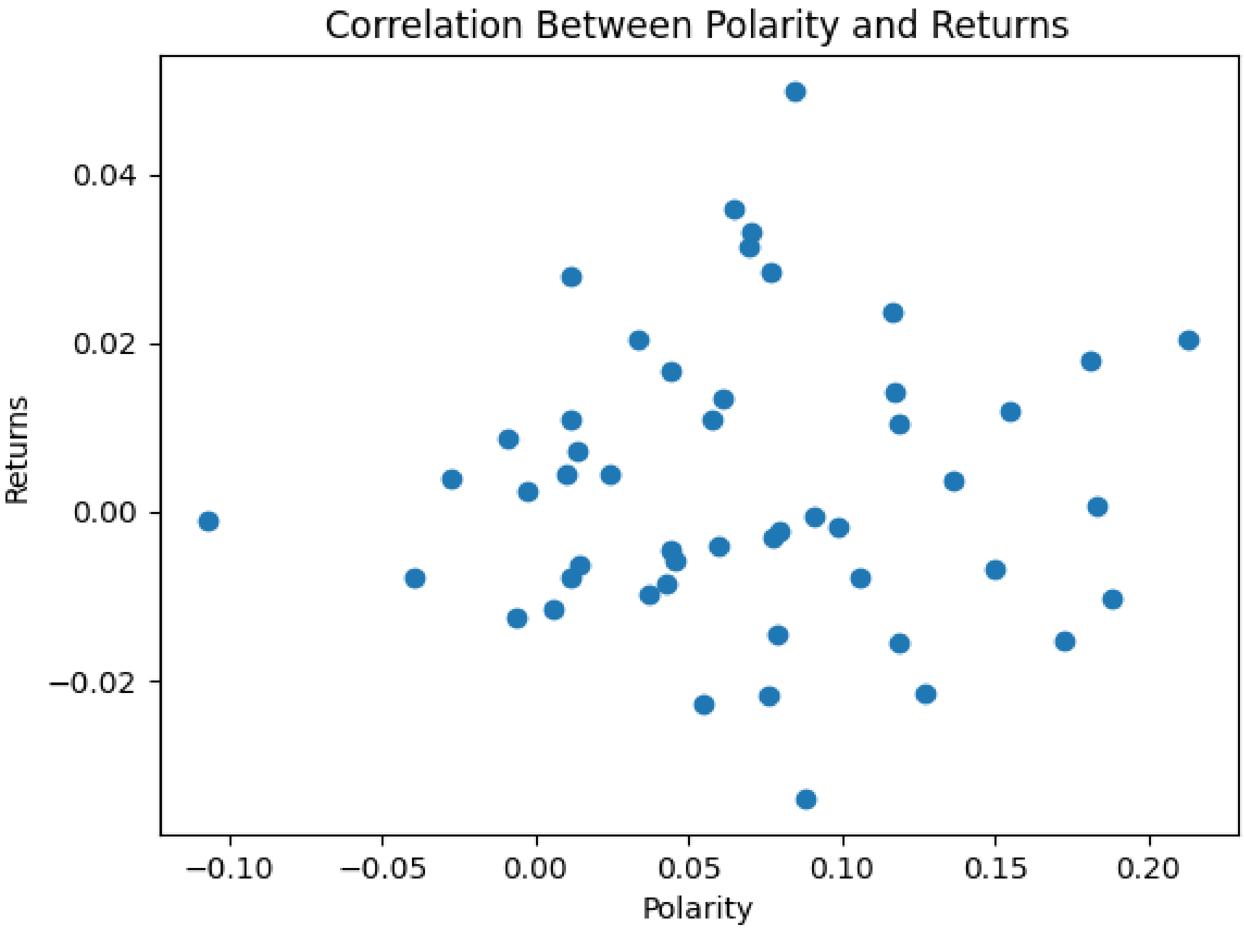

FIGURE 7: Correlation Between Stock Returns and Polarity

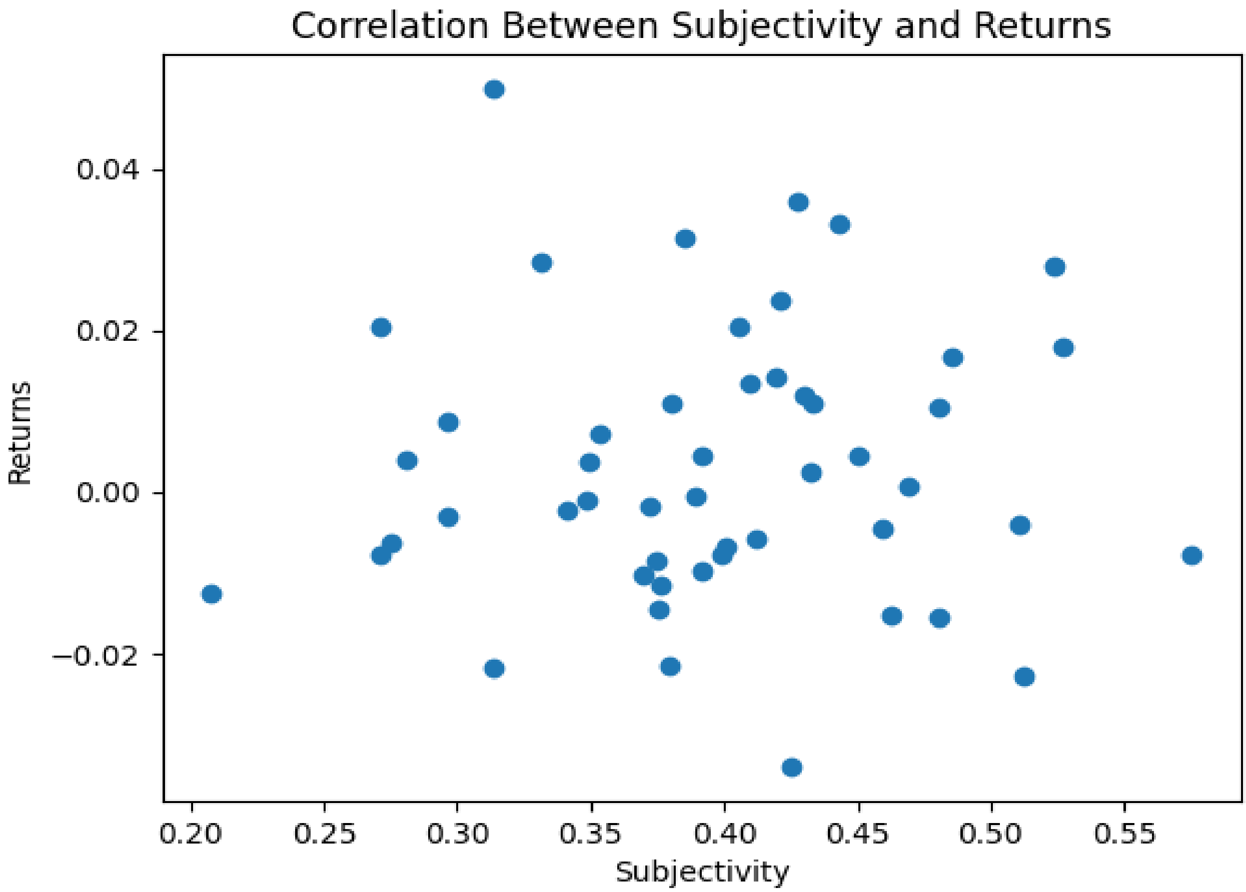

FIGURE 8: Correlation Between Stock Returns and Subjectivity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

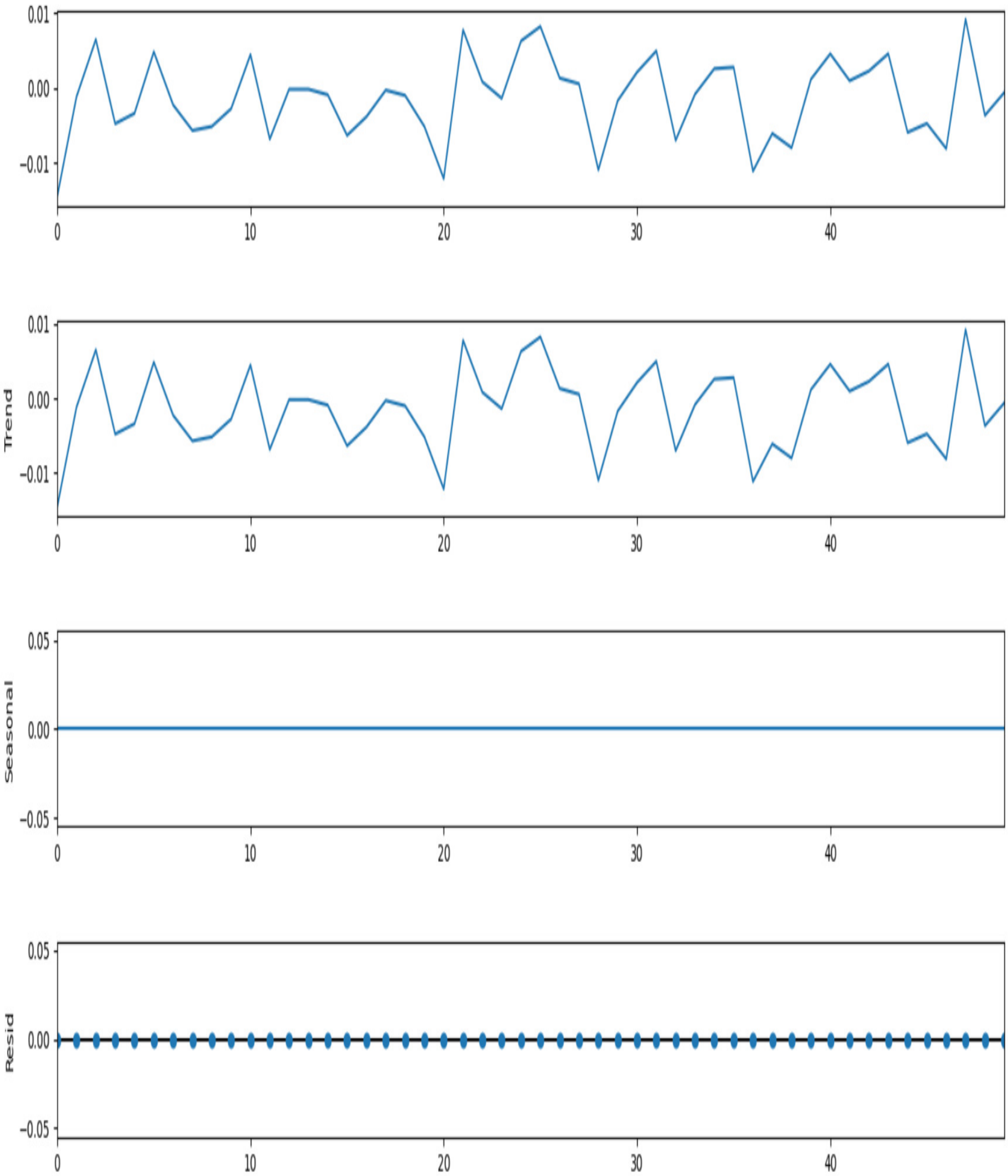

FIGURE 9: Time Components of Return

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3102495, IEEE Access

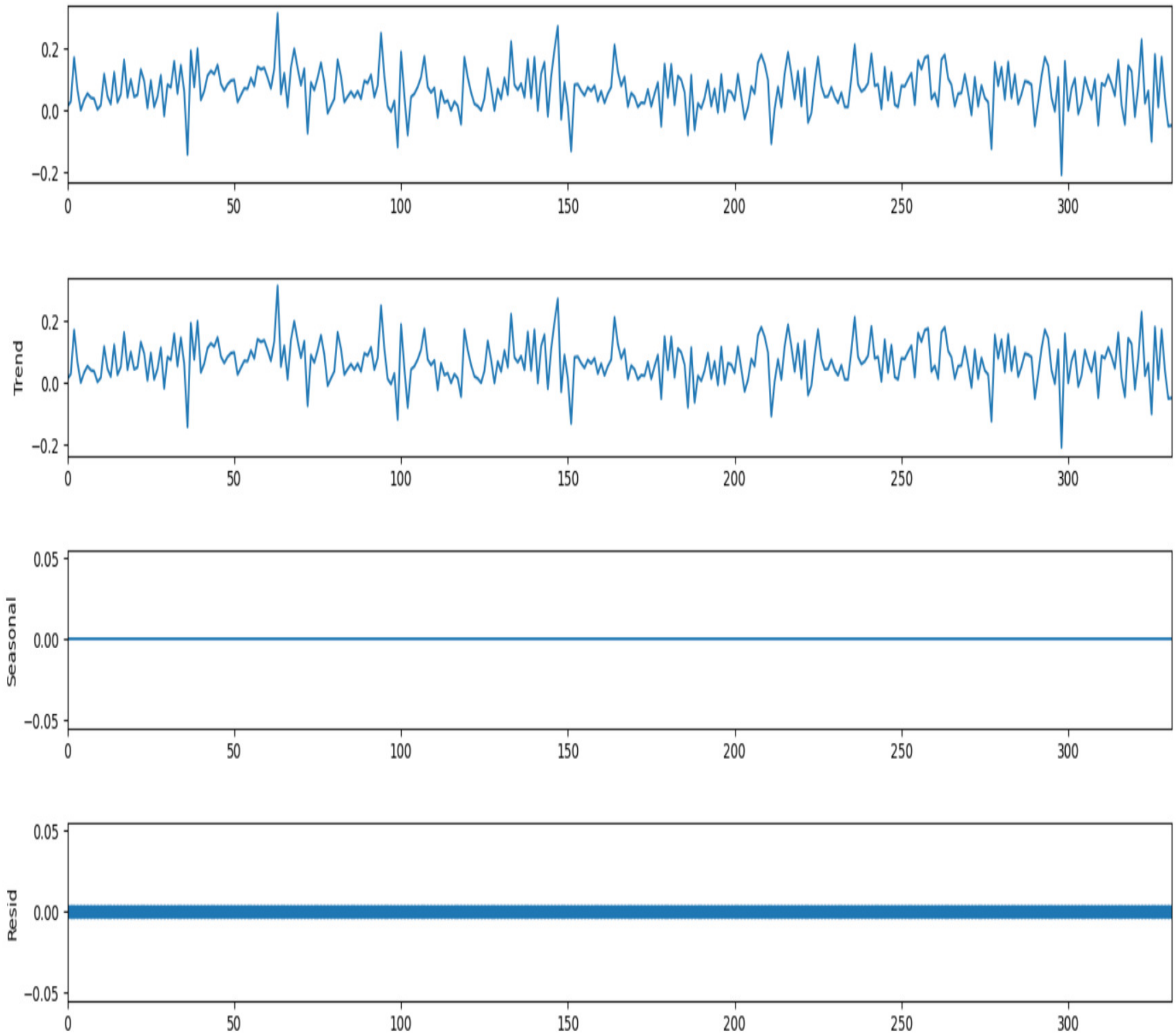

FIGURE 10: Time Components of Polarity

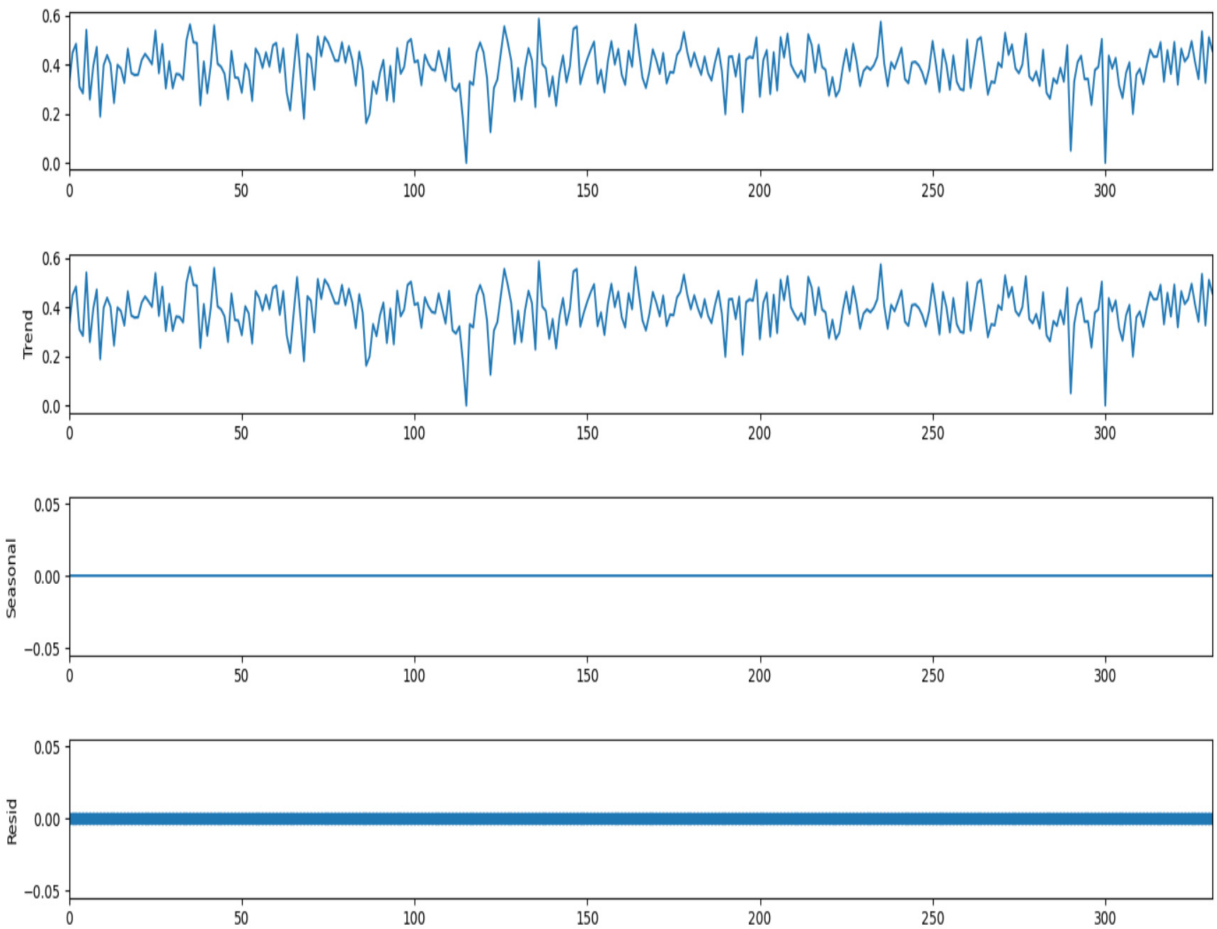

FIGURE 11: Time Components of Subjectivity

REFERENCES

[1] Burton G Malkiel. The efficient market hypothesis and its critics. Journal of economic perspectives, 17(1):59–82, 2003.

[2] Alina Sorescu, Nooshin L Warren, and Larisa Ertekin. Event study methodology in the marketing literature: an overview. Journal of the Academy of Marketing Science, 45(2):186–207, 2017.

[3] Diego Winkelried and Luis A Iberico. Calendar effects in latin american stock markets. Empirical Economics, 54(3):1215–1235, 2018.

[4] James S Ang and Shaojun Zhang. Evaluating long-horizon event study methodology. Available at SSRN 1865625, 2011.

[5] Carlos Castro-Iragorri. Does the market model provide a good counterfactual for event studies in finance? Financial Markets and Portfolio Management, 33(1):71–91, 2019.

[6] Lukas Raynaud. Evaluating industrial policies using event-studies: Evidence from stock market reactions to obama's stimulus package. Available at Skidmore.edu, 2017.

[7] Kurt Schmidheiny and Sebastian Siegloch. On event study designs and distributed-lag models: Equivalence, generalization and practical implications. CESifo Working Paper, 2019.

[8] I Putu Arie Argantha and I Made Surya Negara Sudirman. Stock market reaction to the event of indonesia's general election events in 2019. American Journal of Humanities and Social Sciences Research, 2019.

[9] John Binder. The event study methodology since 1969. Review of quantitative Finance and Accounting, 11(2):111–137, 1998.

[10] Nicolai Bohn, Fethi A Rabhi, Dennis Kundisch, Lawrence Yao, and Tobias Mutter. Towards automated event studies using high frequency news and trading data. In International Workshop on Enterprise Applications and Services in the Finance Industry, pages 20–41. Springer, 2012.

[11] John Affleck-Graves, Carolyn M Callahan, and Ramachandran Ramanan. Detecting abnormal bid-ask spread: a comparison of event study methods. Review of Quantitative Finance and Accounting, 14(1):45–65, 2000.

[12] Graham Bird, Wenti Du, and Thomas Willett. Behavioral finance and efficient markets: What does the euro crisis tell us? Open Economies Review, 28(2):273–295, 2017.

[13] Bin Fang and Peng Zhang. Big data in finance. In Big data concepts, theories, and applications, pages 391–412. Springer, 2016.

[14] HE Stanley, Xavier Gabaix, Parameswaran Gopikrishnan, and Vasiliki Plerou. Correlated randomness: Rare and not-so-rare events in finance. In Practical Fruits of Econophysics, pages 2–18. Springer, 2006.

[15] Dehua Shen and Shu-Heng Chen. Big data finance and financial markets. In Big Data in Computational Social Science and Humanities, pages 235–248. Springer, 2018.

[16] Z. Milosevic, A. Berry, W. Chen, and F. A. Rabhi. An event-based model to support distributed real-time analytics: Finance case study. In 2015 IEEE 19th International Enterprise Distributed Object Computing Conference, pages 122–127, 2015.

[17] C. Huang, H. Li, and B. Chang. An intelligent computational finance model for microstructure-based trading system. In 2016 International Conference on Applied System Innovation (ICASI), pages 1–4, 2016.

[18] Q. Li, J. Tan, J. Wang, and H. Chen. A multimodal event-driven lstm model for stock prediction using online news. IEEE Transactions on Knowledge and Data Engineering, pages 1–1, 2020.

[19] Z. Li, Y. Cai, and S. Hu. Research on systemic financial risk measurement based on hmm and text mining: A case of china financial market. IEEE Access, 9:22171–22185, 2021.

[20] Aaryan Gupta, Vinya Dengre, Hamza Abubakar Kheruwala, and Manan Shah. Comprehensive review of text-mining applications in finance. Financial Innovation, 6(1):1–25, 2020.

[21] Zoran Milosevic, Weisi Chen, Andrew Berry, and Fethi A Rabhi. An open architecture for event-based analytics. International Journal of Data science and analytics, 2(1):13–27, 2016.

[22] Johannes Distler. The event study as a research method for measuring m&a performance. In Acquisitions by Emerging Multinational Corporations, pages 221–268. Springer, 2018.

[23] Bikesh Raj Upreti, Philipp Martin Back, Pekka Malo, Oskar Ahlgren, and Ankur Sinha. Knowledge-driven approaches for financial news analytics. In Network Theory and Agent-Based Modeling in Economics and Finance, pages 375–404. Springer, 2019.

[24] Martin Hoesli, Stanimira Milcheva, and Alex Moss. Is financial regulation good or bad for real estate companies?–an event study. The journal of real estate finance and economics, pages 1–39, 2017.

[25] HaiYue Liu, Aqsa Manzoor, CangYu Wang, Lei Zhang, and Zaira Manzoor. The covid-19 outbreak and affected countries stock markets response. International Journal of Environmental Research and Public Health, 17(8):2800, 2020.

[26] Gunther Capelle-Blancard and Aurélien Petit. Every little helps? esg news and stock market reaction. Journal of Business Ethics, 157(2):543–565, 2019.

[27] Robert J Shiller. Narrative economics: How stories go viral and drive major economic events. 2020.

[28] Victoria Y Fan, Dean T Jamison, and Lawrence H Summers. Pandemic risk: how large are the expected losses? Bulletin of the World Health Organization, 96(2):129, 2018.

[29] Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research, 2018.

[30] José María Liberti and Mitchell A Petersen. Information: Hard and soft. Review of Corporate Finance Studies, 8(1):1–41, 2019.

[31] Mirjana Pejić Bach, Živko Krstić, Sanja Seljan, and Lejla Turulja. Text mining for big data analysis in financial sector: A literature review. Sustainability, 11(5):1277, 2019.

[32] Steven J. Davis Scott R. Baker, Nicholas Bloom. Measuring economic policy uncertainity. National Bureau of Economic Research, pages 1–75, 2015.

[33] Ahmad Kamal. Subjectivity classification using machine learning techniques for mining feature-opinion pairs from web opinion sources. arXiv preprint arXiv:1312.6962, 2013.