Predicting the Direction of Movement of Abnormal Returns using Earnings Conference Calls and FinBERT

By

Rudolfsson Adam & Sandström Carl

Master's Thesis in Data Analytics and Business Economics

May 2022

Supervisor:

LUSEM: Simon Reese

Abstract

Earning conference calls is an important communication tool for companies to provide relevant information about the latest quarter based on the reported earnings. However, the research on the impact of the earning conference calls has for long been a relatively unexplored subject. This study was born out of the ambition to try and expand the research done on earnings conference calls and the possibility to analyse them with FinBERT. This was conducted using a sample of 1118 quarterly earning conference calls, comprising 74 firms from the Information Technology Sector of the S&P 500 in an event study outline. The results suggest that the earning conference calls can be used for predicting the direction of abnormal returns with the help of FinBERT and machine learning. Moreover, the Support Vector Machines achieved the highest accuracy of the tested models on the classification problem.

Acknowledgments

We wish to express appreciation for our supervisor Simon Reese's support and guidance during the process of writing this thesis. We are also grateful for having the opportunity to be the first students in the new Data Analytics and Business Economics Master Program. Even though it has been a challenging journey, it has also been extremely informative and interesting.

Table of Contents

1.Introduction

The purpose of earnings conference calls is that the company should provide relevant information based on the latest quarter to its stakeholders by discussing and clarifying reported earnings (Frankel, Johnson & Skinner, 1999). Previous literature suggests that earnings conference calls have become an increasingly important communication tool for companies and are not only favoured by their investors (Fu, Wu & Zhang, 2019). Generally, managers possess superior information compared to investors regarding a firm’s prospects (Healy & Palepu, 2001). In Bridging the Information Gap, Tasker (1998) concludes that managers provide additional disclosed information during earnings conference calls. Tasker's sample in this study includes small- and medium-sized firms in industries with many companies, but still, shows interesting patterns in how firms disclose information to their investors.

This paper seeks to determine if it is possible to predict the direction of movement of Abnormal Stock Returns with transcripts from quarterly earnings conference calls using FinBERT for Sentiment Analysis and four different Machine Learning methods for classification. This will be conducted using a sample of 1118 quarterly earnings conference calls from 2017 to 2021, comprising 74 firms from the Information Technology Sector of the S&P 500 in an Event Study framework.

Earnings conference calls have a key advantage relative to the more formal and structured earnings reports that they complement. More specifically, they leave room for other information, information that is not possible to display in balance sheets or cash flow statements. Information that might not fit in the very formally structured earnings reports or information that the company itself never thought would be interesting for its investors (Tasker, 1998). The argumentation of undisclosed information becomes even more relevant for the chosen sample of S&P 500 Information Technology as it includes a large share of companies, whose value is determined by far more parameters than the balance sheets. For example, the sample includes a couple of payment providers such as Paypal, Visa, Mastercard, Paycheck, and Paycom. These companies' value is affected by the development of cryptocurrencies and different legislations regarding that. Furthermore, there are a couple of companies, whose values could be affected by a political view. Cisco and Fortinet could for instance be positively impacted by events such as Huawei's connection to the Chinese government. Although they seem interesting and valuable for investors, the earnings conference calls are surprisingly unexplored.

Sentiment analysis is a part of Natural Language Processing that has seen rapid growth in the last few years. The main objective of sentiment analysis is to extract sentiment and opinions from large-sized textual data in a faster and more effective way with the help of machine learning. This makes it possible to analyse large textual data fast and effectively with almost no limitations (Cambria & White, 2014). Price, Doran, Peterson, and Bliss (2012) looked at earnings conference calls and stock returns and they conclude that the linguistic tone in the earnings conference calls is a significant predictor of abnormal returns and trading volume. Other literature about earnings conference calls shows that managers in some cases can provide undisclosed information and that sentiments in earnings conference calls have some effects on the stock market (Tasker, 1998; Amicis, Falconieri & Tastan, 2021). However, the main focus in research regarding stock prices, machine learning, and sentiments has so far been Twitter, news articles, and other information where the approach has proven some interesting results (Bollen, Mao & Zeng, 2011; Mohan, Mullapudi, Sammmmeta, Vijayvergia & Anastasiu, 2019; Tralaven & Aste, 2015; Tetlock, 2007; Valencia & Garcia, 2012).

Quarterly earnings conference calls are an opportunity for managers to freely speak about the company within the framework of the structure. The Q&A part of the call is particularly interesting since that leaves the managers even more space to freely express their thoughts about the company without a manuscript. The earnings conference calls became accessible to the public in 2000 as a response to Regulation Fair Disclosure which stated that when a public-traded company discloses any material nonpublic information, it must also make public disclosure of that information (The Securities and Exchange Commission, 2021). Therefore, there are nowadays transcripts available of almost all larger public companies. Despite that, there is an information gap when it comes to the impact of earnings conference calls. Especially when it comes to sentiment analysis where little research has been done. Therefore, the authors of this paper aim to contribute to this research area by empirically investigating if it is possible to predict the outcome of positive or negative abnormal returns using sentiment analysis on earnings conference calls.

The remainder of this thesis is structured in the following way: Section 2 provides a literature review covering previous theories and findings underpinning this study. Section 3 describes the collection, cleaning, and limitations of the data used in the study. Section 4 describes the research methodology, as well as the machine learning tools used for prediction. Section 5 presents the findings of the study. Section 6 reports the conclusion and recommendations for further research.

2. Literature review and theory

2.1 Earnings Conference Calls

Fu, Wu and Zhang (2019) find that US public firms with the less optimistic tone of the Q4 earnings conference call, experience a higher risk of declining stock prices in the following year. They also find evidence that an optimistic tone lowers the risk. The authors of the study use Loughran and McDonald’s word list to derive the scores to the different transcripts. Amicis, Falconieri, and Tastan (2021) compare the differences in sentiments between female and male CEOs and CFOs using 78 000 earnings conference calls using the same word list. Apart from showing that female executives are more positive and less ambiguous, they conclude that the stock market responds to the sentiment of the call.

Tasker (1998) concludes that managers try to find other sources of communication than classic financial statements. In addition, Tasker shows that managers provide additional disclosed information during earnings conference calls for the sample of small and mid-size companies.

Price et al. (2012) examines earnings conference calls and stock returns using a sample of 2800 earnings conference calls from companies in different industries during the period of 2004-2007. They compare Loughran and McDonald’s word list to Harvard IV-4 Psychosocial Dictionary and conclude that the Loughran and McDonald's dictionary, trained on financial corpora, shows better results. Moreover, the researchers find that earnings conference calls' linguistic tone is a significant predictor for predicting cumulative abnormal returns and trading volume.

Except for the above-mentioned articles, research within this area is relatively limited. There are however some similar articles examining other types of textual data. For example, Bollen, Mao and Zeng (2011) provide good predictions using Sentiment Analysis together with Twitter. Tetlock (2007) analyse the articles in the Wall Street Journal and finds that high media pessimism predicts declining stock prices. Mohan et al (2019) analyses more than 265 000 financial news and the corresponding stock prices on the S&P500. Their model provides good predictions, except in the cases where stock prices are low or highly volatile.

2.2 Text analytics and Text mining

Gaikwad, Chaugule, and Patil (2014) define the purpose of text mining as extracting valuable information from text. Textual data is usually unstructured data; a type of data that is generally stored in its native format. Unstructured data is typically text, audio, and video and it is referred to as unstructured data since the data is not structured in the form of a data table. This means that the data is more difficult to analyse compared to structured data. The main challenges of analysing textual data derive from the structure of the language, which is difficult for a machine to interpret without help structuring it.

In order to extract valuable information from text, one must first transform it into structured data. There are a few tools one can use in order to transform the data. Firstly, stop words can be removed. Stop words include the words that are most frequently used in the language, for example, “to”, “you”, “in”' and “is/are”. These words only serve as noise and do not have any significance to the results. Secondly, punctuations should be removed since it creates a lot of noise as well (Haddi, Liu & Shi, 2013). Furthermore, one can do stemming or lemmatization on the words in the textual data. The main purpose of stemming and lemmatization is to remove different grammatical forms that do not contribute to the meaning of the text. By stemming the words, the word is changed to its root form and the goal is to reduce the inflectional forms of the words (Chopra, Joshi & Mathur, 2016). Lemmatization is a similar method and the difference between the methods is that stemming does not always produce an actual word whereas lemmatization creates an actual word. Balakrishnan and Lloyd-Yemoh (2014) compare the different methods' performance and conclude that lemmatization is performing slightly better.

2.3 Sentiment Analysis

Bing and Lei (2012) define sentiment as the degree of positivity or negativity in text. In a financial context, the degree of positivity or negativity can be analysed in order to draw conclusions. To what extent earnings conference calls can help investors make decisions by only looking at the sentiment scores and the abnormal return is so far not narrowly investigated. Sentiment scores are usually done by a predefined word list. This wordlist would for example give “growth” a positive sentiment score and “disaster” a negative sentiment score. There are however different approaches to this.

One of the most popular wordlists for sentiment analysis in finance is the Loughran and McDonald (2011) dictionary; a dictionary tool based on a large number of financial texts. The financial vocabulary is very different in its sentiment. According to Loughran and McDonald (2011), 73.8% of the negative word counts in the Harvard list are not necessarily negative in a financial context. An example of a word that Loughran and McDonald bring up in the article is liability, which in most cases would be negative, in finance however, the word is neutral. Therefore, it is crucial to use a financial pre-trained dictionary when analysing sentiments in earnings calls.

A pre-trained financial language model that is increasing in popularity is FinBERT. FinBERT is based on the open-source Google's Bidirectional Encoder Representations from Transformers (BERT). BERT is a bidirectional model that is pre-trained on Wikipedia and book corpus to bring an understanding of the property representation of the language. The transformer includes two mechanisms, an encoder, and a decoder. The encoder reads the text and the decoder makes the prediction. Unlike directional models that read the text from left to right, the encoder in BERT reads the entire sequence at once (Gomez, Jones, Parmar, Polosukhin, Shazeer, Uszkoreit, & Vaswani. 2017). Similar to Loughran and McDonald, the modified version called FinBERT is trained on a substantial amount of financial corpus and is therefore better suited for the purpose of analysing financial texts. FinBERT reads the financial document and gives three types of scores that sum to one as output. The categories are positive, negative, and neutral (Araci, 2019).

Huang, Wang and Yang (2020) argue that other approaches such as the Loughran and McDonald dictionary and Word2Vec underestimate the information in earnings conference calls by at least 32% compared to FinBERT. The reason for this is primarily that FinBERT is able to uncover sentiment in sentences that the other approaches misclassify as neutral. Moreover, FinBERT was tested on Financial PhraseBank, a financial database with 5000 sentences reviewed by financial experts and it reached an accuracy of 97%. However, there are still some areas for improvement. FinBERT for example predicted negative on “Pre-tax loss totaled euro 0.3 million, compared to a loss of euro 2.2 million in the first quarter of 2005”, when it was tested on the PhraseBank (Araci, 2019). One of Araci’s (2019) suggestions for further research is to use FinBERT together with market return data. Even though FinBERT has shown good results, the number of researchers that have used FinBERT to predict stock returns or abnormal returns is relatively small.

2.4 Machine Learning Classification Algorithms

There are several different machine learning algorithms that can be used for classification problems. As the No-Free-Lunch-theorem suggests, there is not one algorithm that works for all types of data and problems. There are in other words no shortcut to success and therefore it is recommended to try different types of algorithms when classifying data (Wolpert & Macready, 1997).

Gupte, Joshi, Gadgul, and Kadam (2014) compare Naive Bayes, Max Entropy, Boosted Trees, and Random Forest Classifier for sentiment analysis. According to their study, the most accurate one is the Random Forest Classifier. Another machine learning algorithm that has been important for sentiment analysis research is Support Vector Machines (SVM). By transforming the data with kernels, SVMs can classify complex data. Amrani, Lazaar, and Kadiri (2018) compare SVMs and Random Forest (RF) accuracy on Amazon product reviews where SVM has higher accuracy. Yadav, Kudal, Rao, Gupta, and Shitole (2021) classifies Twitter data as positive and negative using Naive Bayes, Logistic Regression and SVMs and have the highest accuracy with SVMs as well.

K-Nearest Neighbor (KNN) is a relatively simple algorithm that has proven surprisingly good results in sentiment analysis. In Sentiment Analysis on Twitter Data using KNN and SVM, Huq and Rahman (2017) compares the reputable SVM to the KNN and the results were surprisingly in KNNs favour. Furthermore, the computational cost of KNN is also beneficial.

Extreme Gradient Boosting (XGB) has not been used to the same extent in previous research as the above-mentioned algorithms since it was introduced in 2015. However, the algorithm has been very efficient in other kinds of classification problems and has a very good reputation in the machine learning field (Huang, Liu, Qin, Shi, Wang & Zhao, 2019; Bansal & Kaur, 2018).

The number of researchers trying to predict abnormal returns using machine learning algorithms is limited, except for a number of attempts on historical data. However, the mentioned algorithms are not only regarded as successful in sentiment analysis but also in classification problems overall. In the article Supervised Machine Learning Algorithms: Classification and Comparison, RFs and SVMs are the best performing algorithms when the authors compare seven different classification algorithms with small data sets and few attributes (Osisanwo, Akinsola, Awodele, Hinmikaiye, Olakanmi & Akinjobi 2017).

2.5 Event Study

The event study methodology has been used in the literature of finance and economics for many years. One of the first published studies that is available today might be the event study by Dolley (1933) which dates all the way back to 1933. Dolley studied the effect of stock split on stock prices. Since then, Dolley's simple approach has made small improvements but great contributions to the research.

Conducting an event study on securities is generally divided into a few different steps. Firstly, the event of interest must be defined together with the period in which the security prices will be examined, which is defined as the event window. The event window is often longer than the event of interest itself. For instance, if the event of interest is the release of an earnings announcement, the event window will be longer than just the day of the event. This is so that the event study can look at non-normalities before and after the event of interest, which can be due to information leaks or inefficiencies in the market. Secondly, the selection criteria for the event study must be stated. This is often done in conjunction with the event of interest. The selection criteria involve the scope of securities as well as the time frame of the event if the event is repeated on a quarterly or yearly basis. To test the impact of the event, a model to estimate normal security return must be specified. The main object of estimating normal returns is to try and show the security return if no event would have occurred (MacKinlay, 1997). There are multiple ways of estimating the normal returns: naive benchmark, single factor, and CAPM for example.

The naive approach is effectively just the same as taking the market return and has no security unique parameter. The single factor model is almost identical to the widely used CAPM but built on statistical argumentation rather than the economic framework of CAPM which is based on theoretical arguments. The argumentation of CAPM, in comparison to the single-factor model, has proven to be a subject for critique whereas the single-factor model in comparison has shown fewer of these critiques (Kliger & Gurevich, 2014). There are examples of more complex factor models which include factors such as size and market to book. However, there is controversy if the increased complexity improves the expectations. Industry-specific effects are however effects that can be controlled and proven to be beneficial in comparison to the single-factor model, in cases where there are different industries in the sample (MacKinlay, 1997).

The last step in conducting an event study is to specify the estimation window. The estimation window is the time frame in which the estimation of parameters for the chosen model will be done. The estimation window is normally not overlapping with the event window (MacKinlay, 1997). Typically, the estimation window comprises somewhere between 100-300 days, where 100 days are sufficient for estimating the normal returns with most models (Peterson, 1989).

2.6 Efficient Markets

In Fama's (1970) article Efficient Capital Markets: A Review of Theory and Empirical Work, the market efficiency is defined as the degree to which current stock prices reflect available information about the value of financial assets. The efficient market hypothesis describes the highest level of market efficiency, where all stock prices fully reflect all available information and trade at their fair value. Sufficient conditions for the efficient market hypothesis are described as:

> (i)There are no transaction costs in trading securities, (ii) all available information is costlessly available to all market participants, and (iii) all agree on the implications of current information for the current price and distributions of future prices of each security (Fama, 1970, p.387).

Due to the complexity of conditions and the obvious problem of observing them in real markets, three levels of the efficient market hypothesis are derived: weak, semi-strong, and strong form of market efficiency.

The weak form states that prices reflect all past available information but may not reflect any new information that has yet not been made publicly available. Malkiel (2003) describes the weak form as similar to the famous random walk theory which states that stock prices reflect all available information of that day. Today's price reflects all information today, and tomorrow's price reflects all information tomorrow, and so on. The semi-strong form states that prices reflect both past and newly released information in a quick way so that no investor can benefit more in the market by trading on new information. The strong form states that the prices reflect all publicly available information and private information because all information is reflected in prices. No private information or fundamental analysis can help predict future prices (Fama, 1970). In reality, the efficient market hypothesis holds to a large extent, with some under and overreactions (Fama, 1998).

3. Data

The sample chosen in this event study comprises the 74 firms from the Information Technology sector of the S&P 500 listed on Wikipedia on the 20:th of April 2022. The time span of this event study comprises five years, 2017-2021. Mark that the calls are often taking place a while after the quarter, Q4 for 2021 is for example often held in Q1 of 2022. This paper covers all the Q4 for 2021, including the ones held in 2022. In total, the sample base comprises 1480 observations before cleaning.

The total S&P 500 consists of 500 companies that are divided into eleven non equally weighted sectors. Information technology is by far the most heavily weighted sector with approximately 28% of the weight. The S&P 500 Information Technology Index is specified as a market index, which is closely correlated to the firms specified, as the index comprises the same sample of firms. The index is a value-weighted index that is calculated based on the latest transaction price of each of the companies. It is in other words based on the index and not the 74 companies that the authors of this paper have chosen to investigate since there have been some changes during the time that is being observed. The companies in the sector are divided into eleven different sub-industries: Application Software, Communication Equipment, Data Processing & Outsourced Services, Electric Equipment & Instruments, Electronic Components, Internet Services & Infrastructure, IT Consulting & Other Services, Semiconductors Equipment, Semiconductors, Systems Software and Technology Hardware, Storage & Peripherals. The five biggest companies in the sector are Apple, Microsoft, Nvidia, Mastercard, and Visa.

3.1 Data Collection

Stock price data and earnings calls were scraped from *FinancialModelingPrep.com* covering all the 74 companies included in the sample. The data was gathered through an API together with Python code and contains 1480 observations before cleaning. The stock prices were collected for each of the 150 days before the event window and the 11-days around the event date, in total 161 days of stock prices for each observation. Index data was downloaded from the S&P website in the form of the S&P 500 Information Technology index. The index data was then matched to the stock data, so all observations had the same amount of 161 days of both stock and index data with correct starting and ending dates. If the earnings call was presented after the stock market was closed (4

PM), the following day was counted as the event day and if it was released during the time that the stock market was opened (9.30 AM to 4 PM), the same day was counted as the event day. The motivation for this procedure is that closing prices are used which are set at 4 PM, if the conference calls are not disclosed before 4 PM the conference call will not show any impact on the closing prices of that day, and the next day must be used as event-day instead.

3.2 Data Preprocessing

Stock data, index data, and earnings calls were all combined into a large document for preprocessing. If one of the observations was missing one measurement in either stock price, index price or earnings conference call, the whole observation was removed. This method was implemented instead of matrix completion or averaging because of the structure and quality of the collected data. If the earnings call was missing, the independent variables could not be calculated, and matrix computation would not be an option on text data. If stock data were missing, this was evidence that the stock was listed on the market at a later date or that the stock had been removed from the public stock market. The index data did not have any missing values and were therefore not a problem. There were in total 346 observations with some kind of missing data during the observed period. The explanation for the missing data is primarily that some companies were not listed during the whole time period but are included in the index as of 2022. There were also a few missing transcripts from companies that were listed during the whole time period. In other words, transcripts that for some unexplained reason were not available in our data source.

Earnings conference calls are manually transcripted from an oral presentation. Considering the construction of the earnings conference calls, one could argue that some parts could be removed such as the introduction, speakers that are not employed by the company, or the Q&A-part. However, there is no evidence that supports removing certain parts of the transcript, and the questions posed by non-company representatives could potentially be highly relevant to understanding the overall sentiments of the earnings conference calls.

The stop words and punctuation were removed from the earnings conference call transcripts using the list of stopwords in the Python nltk package, a list that contains 179 stopwords that can be found in List 3 in the Appendix. *Up* and *down* were removed from the list of stopwords since the words can have an important meaning in a financial context. The text was also lemmatized using the same

Python package. In figure 1 you can see an example of a part of a cleaned transcript. The reason that some of the words are unnaturally compounded is that they are written with a dash or something similar in the actual transcript.

Earnings conference calls usually has the following structure: It begins with company representatives delivering a speech about the past quarter and their thoughts about the upcoming quarter. Thereafter, the representatives respond to questions provided by the audience. The earnings conference calls usually take place within a few hours after the earnings announcement is released. In figure 1 down below, a part of a cleaned earnings conference call is presented.

Figure 1, Snippet of a cleaned Earnings Conference Call (Accenture Q1 2018)

Operator Ladies gentleman thank standing welcome Accenture ' First Quarter Fiscal 2018 Earnings Call time participant listenonly mode Later conduct questionandanswer session Operator Instructions reminder today ' call recorded hosting speaker today Managing Director Head Investor Relations Angie Park Please go aheadnAngie Park Thank Kevin thanks everyone joining u today first quarter fiscal 2018 earnings announcement operator mentioned ' Angie Park Managing Director Head Investor Relations today Pierre Nanterme Chairman Chief Executive Officer David Rowland Chief Financial Officer hope ' opportunity review news release issued short time ago Let quickly outline agenda today ' call Pierre begin overview result David take financial detail including income statement balance sheet first quarter Pierre provide brief update market positioning David provides outlook second quarter full fiscal year 2018 take question Pierre provides wrapup end call reminder discus revenue today ' call ' talking revenue reimbursement net revenue matter ' discus call including business outlook forwardlooking subject known unknown risk uncertainty including limited factor set forth today ' news release discussed annual report Form 10K quarterly report Form 10Q SEC filing risk uncertainty could cause actual result differ materially expressed call call today reference certain nonGAAP financial measure believe provide useful information investor include reconciliation nonGAAP financial measure appropriate GAAP news release Investor Relations section website accenturecom always Accenture assumes obligation update information presented conference call let turn call PierrenPierre Nanterme Thank Angie thanks everyone joining u today excellent first quarter extremely pleased result delivered strong broadbased revenue growth across dimension business gained significant market share strategy continues different Accenture marketplace seeing strong demand service particularly digital cloud security highlight quarter delivered strong new booking 10 billion generated revenue 95 billion 10 growth local currency delivered strong earnings per share 179 13 increase Operating margin 156 consistent first quarter last year generated strong free cash flow nearly 900 million returned 14 billion cash shareholder share repurchase dividend strong start fiscal year 2018 feel good continued momentum business let handover David review number greater detail David younDavid Rowland Thank Pierre Happy holiday thanks much joining u today ' call Building Pierre ' comment pleased quarter one result positioned u well achieve full year business outlook especially relates strong broadbased topline growth result demonstrate durability resiliency growth model high degree relevance differentiation capability marketplace get detail quarter let summarize important highlight Starting net revenue expanded business approximately 1 billion quarter 10 growth local currency diversity durability growth model evident strong extremely wellbalanced growth across five operating group three geographic area doubledigit growth four operating group Europe Growth Markets Strong doubledigit growth digital cloud security continued dominant driver growth pervasive across business estimate 10 growth significantly outpaced market continue gain share strengthen position leader new respect profitability operating margin 156 quarter consistent quarter one last year continues reflect significant level investment business delivered strong EPS 179 13 compared last year Looking cash generation capital allocation free

3.3 Descriptive Statistics

The table below shows the descriptive statistics of the abnormal returns on event-day together with the sentiment scores. The positive and negative outcomes of the earnings calls were relatively equally distributed. Of the 1118 calls, 582 had a negative outcome and 536 a positive outcome. The mean and median of the abnormal returns were -0.02699% respectively -0.01179% with a maximum positive value of 17.71% and a maximum negative value of -39.81%. In general, the quarterly earnings calls had a negative effect on the abnormal return, but the spread was wide between the negative and positive outcomes. The earnings call transcripts were assigned an overall sentiment score that sum to one. An average earnings call received a positive score of 0.122, a negative score of 0.046, and a neutral score of 0.831.

Table 1: Descriptive statistics

	Abnormal Return	Positive Score	Negative Score	Neutral Score
Values	[-100,\ infinity]	[0,1]	[0,1]	[0,1]
Min	-0,398144	0,021038	0,015247	0,149400
Max	0,177186	0,810618	0,799476	0,937341
Mean	-0,002700	0,122219	0,046354	0,831426
Median	-0,001180	0,088057	0,031556	0,872441

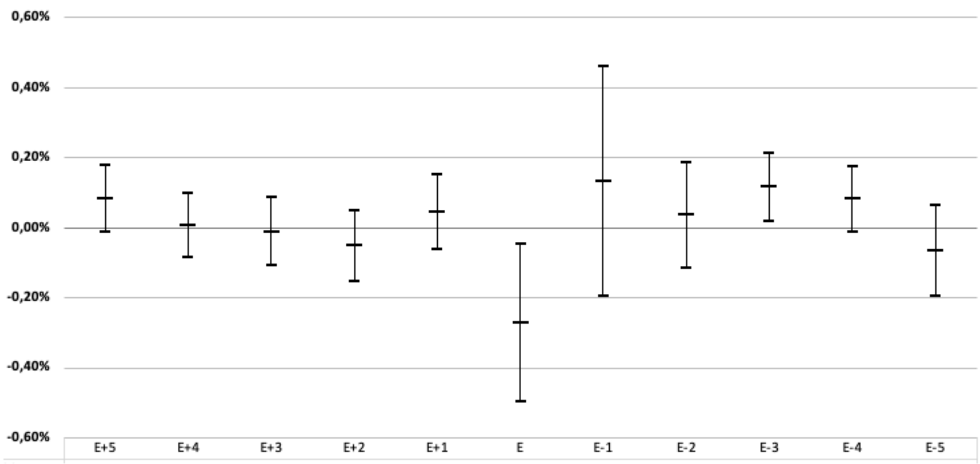
The table below shows the abnormal returns for each day in the event window, which includes five days before the event and five days after the event. It can be seen from the table that the day of the event and the day before the event show relatively large fluctuations from zero. All days after the event-date show small fluctuations from zero.

Table 2: Event Window Abnormal Returns

E+5	E+4	E+3	E+2	E+1	E	E-1	E-2	E-3	E-4	E-5
0,08%	0,01%	-0,01%	-0,05%	0,05%	-0,27%	0,13%	0,04%	0,12%	0,08%	-0,07%

The figure down below shows the abnormal return with a 95% confidence interval, all eleven days are included. It can be seen that the return of the event day deviates more from zero than the outcome of all other days. The day before the event day also shows large variance with a big confidence interval, and the days after the event day show no signs of extreme volatility or extreme return.

Figure 2: Event Window Abnormal Returns with 95% Confidence Intervals



3.4 Delimitations

Considering the limitation of time, some delimitations were unavoidable. Firstly, there were a couple of earnings call transcripts that were missing in the Financialmodelingprep database. These transcripts are in most cases available on the company’s websites or on other sources. However, the transcripts were left out of the study instead of being scraped manually.

Moreover, the sector has had some minor changes during the years 2017-2022. For example, Xilinix was acquired by Advanced Micro Device and Cadence was introduced. Furthermore, Teledyne was introduced as a company in the industrials sector mid 2020 but is according to some sources counted as part of the Information Technology Sector. In this paper, the authors have chosen to web scrape the 74 companies listed on Wikipedia's list of S&P 500 companies by the 20:th of April 2022. The 74 companies that were used in this study are listed in Table 6 in the Appendix. One could have looked at the specific changes that happened during these years and changed the collected transcripts according to that information. However, the authors have chosen not to do that since it is also interesting to follow the companies over time. This is also the reason why not more than five years were used in the model. Lastly, the sample size could preferably have been larger, since machine learning algorithms generally work better with more data.

4. Research Methodology

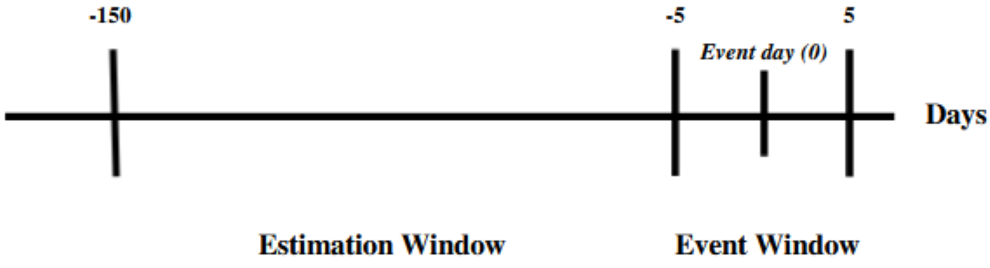
4.1 Outlining the Event Study

An event study methodology will be used to analyze the effect of earnings conference calls on stock market abnormal returns. The underlying idea of using the event study methodology is to track the prices of securities involved in the event to analyze any potential event-related reactions. In line with the Efficient Market Hypothesis presented in section 3.4, it will be assumed that the markets are efficient to a large extent. This assumption allows us to state that price adjustments are instantaneous and complete, any new information should be reflected in the price immediately after release.

The event study will be divided into two parts: the estimation window and the event window. The estimation window is where one formulates the expectations of the event window and the event window is the time where one expects that the event of interest will be reflected in the price and will therefore be the basis of this event study.

The event window will be a total of eleven days, including event-day, five days pre-event-day, and five days post-event-day. The five days pre-event-day will be used to look for non-normality or leakage of information for example. The five days post-event will be used for looking at long-time effects of the release of earnings calls. The event day will be our main goal of the investigation and is the day of the releasing of the earnings conference call. If the earnings conference call is released after the stock market is closed, the next day will be used as event day. To be able to predict the event windows and test for any type of hypothesis, 150 days pre-event window will be used to calculate the expectation of return during the Event Window with the chosen single factor-model.

Figure 3: Event study timeline



4.2 Positive or Negative Abnormal Return

With this baseline set, the authors of this study will in line with the event study methodology start by predicting the normal returns which will be the return expected if no event has occurred. The Single Factor Model is chosen to predict the normal returns. With the Single Factor Model, one can from the estimation window derive an expectation for the event window. This will be done in multiple steps, starting by rewriting the closing prices as log difference between the closing price at time *i* and the previous day at time *i-1* for each individual stock s:

$$R_{st} = ln\left(\frac{p_{st}}{p_{st-1}}\right)$$

Closing prices will be used as they reflect all information released during the day, the Natural Logarithm will be used because of its mathematical properties to standardize and decrease skewness (Fama, Fisher, Jensen & Roll, 1969). With the stock return, an expectation of normal returns can be formulated. The expectation of normal returns will be derived by using the Single Factor Model which assumes a linear relationship between stock *i* and market index *m* and it is derived with ordinary least squares:

$$R_{st} = \alpha_s + \beta_s R_{mt} + \varepsilon_{st} \qquad \text{Where: } \varepsilon_{st} \sim N(N, \sigma^2_{st})$$

R_{st} is the return of stock *s* at time *t*, R_{mt} is the return of market index *m* at time *t*. $\alpha_{s.}$ is the intercept of stock *s* and and β_s is the beta of stock *s*, ε_{st} is the error term with the expectancy 0. All 150 days will be used in the estimation window to derive the parameters of *alpha* and *beta*, they are derived with OLS. With the derived *alpha* and *beta* variables, the expected return of stock *s* at time *t* can be written as a function of the market index return given the zero expectation of error:

$$\widehat{E}\left(R_{st}|R_{mt}\right) = \widehat{\alpha_s} + \widehat{\beta_s} R_{mt}$$

With the expectation of normal return, the abnormal returns can be calculated which simply is the difference between actual observed return and expectation of return given the market index return:

$$AR = R_{st} - \widehat{E}\left(R_{st}|R_{mt}\right)$$

By construction, the expectation of abnormal return is zero, if no stock unique event has happened the expected normal return should be the same as the observed return and AR is therefore zero. This means that if AR≠0 is observed within the event window it is concluded that this price movement is derived from the event in question and not regular market fluctuations. In this paper, the abnormal returns will be rewritten to binary variables: Positive or Negative.

$$f(AR) = \left\{ \begin{matrix} Positive, & if\ AR > 0 \\ Negative, & otherwise \end{matrix} \right\}$$

This is done as the main goal is to test if it is possible to predict the market reaction with earnings conference calls and sentiment analysis, not to find a linear relationship between the independent variables of sentiment score and abnormal returns.

4.3 Sentiment Score

FinBERT is applied to analyze the sentiments of each earnings conference calls. FinBERT analyzes each and every earnings call transcript and gives an overall sentiment score for the analyzed transcript. The overall sentiment score is given in three categories: positive, negative, and neutral. The score is on a relative scale meaning that the total assigned score in all three categories sum to one.

4.4 Machine Learning Classification Algorithms

To test the hypothesis that earnings transcript calls can be used to predict the outcome of abnormal returns, four different categorization models are applied: SVM, K-Nearest Neighbor, Random Forest, and Extreme Gradient Boosting. The selection of models in this paper is based on previous research using classification algorithms. Four different models are applied since no model is superior in all cases, the performance of all models will vary depending on the problem and the data available for the models. All models are classification models and applied in similar ways. Abnormal return will be used as the dependent variable with the binary outcome: Positive/Negative. Sentiment scores will be used as explanatory variables formulated in three different categories: Positive, Negative, and Neutral.

$$f(AR) \sim Positive\ \beta_1 + Negative\ \beta_2 + Neutral\,\beta_3$$

Before the models are trained on the data, the data must be divided into training and a test set. The data is split 70/30, meaning that 70% of the data is used for training the models and 30% for testing and evaluating the models. After the data is split, the models can be fit on the training data. When fitting the models to the training data, each hyperparameter must be specified to its optimal value as this will have a large effect on how well the model can predict the data. All hyperparameters for each model with specified gird values are presented in List 1 in the Appendix. The specific hyperparameters after tuning are specified for each model in List 2 in the Appendix.

The hyperparameters In this step Random Search K-fold cross-validation is applied. Random Search allows one to search wide intervals with less computation cost than the grid search would.

The number of folds in the K-fold cross-validation is set to three which specifies that for each fit the data is divided into three new train/test splits and then trained. This allows one to fit the model on all of the training data. The second step of fitting the model is to use the Grid Search cross-validation within the smaller earlier specified hyperparameters values extracted from the Random Search. The best fit of the hyperparameters is then extracted from the best model and used to fit a new model on the same training data, this model can then be used to predict the test data and get an indication of how well the model can predict the potential relation between dependent and independent variables on new data. All four models return predictions in the form of an accuracy and a confusion matrix.

5. Results and Analysis

5.1 Empirical Results

The accuracy of all four classification models is presented in table 3 down below, the accuracy shows the percentage of how many of the predictions the model predicts in the correct class. The accuracy is represented in decimals, where 1 indicates 100% correct and 0,01 indicates 1% correct.

Table 3: Accuracy result from trained models

Model	Support Vector Machine	K-Nearest Neighbour	Random Forest	Extreme Gradient Boosting
Accuracy	0.5506	0.5327	0.5387	0.5357

The results of the accuracy show that all four models performed similarly. The best performing model was the SVM which achieved 55,06% accuracy on test data. All three of the remaining models performed in similar ways, but with slightly lower accuracy. The RF returned 53,87% accuracy, the XGB returned 53,57% accuracy and the least performing model was the KNN with an accuracy score of 53.27%. The results of the models should be interpreted as better than random guessing, this means that the models were usable for predicting the direction of abnormal returns.

The specificity and sensitivity of the four modes are presented in table 4 down below. Specificity and sensitivity indicate how many of the observations the model classifies as true/false in relation to how many true/false outcomes there are in the sample. The sensitivity and specificity are represented in decimals, where 1 indicates 100% correct and 0,01 indicates 1% correct.

Table 4: Specificity and Sensitivity for the trained models

Model	Specificity TN/(TN+FN)	Sensitivity TP/(TP+FP)
Support Vector Machine	0.5368	0.5810
K-Nearest Neighbour	0.5256	0.5455
Random Forest	0.5286	0.5596
Extreme Gradient Boosting	0.5316	0.5411

The specificity and sensitivity accuracy was above 50% for all four models. This result shows that all models were able to predict both positive and negative outcomes of abnormal return. This result should be interpreted as all four models are usable for predicting both classes and performs better than random guessing. The sensitivity was higher than the specificity for all four models, this should be interpreted as the models being slightly better at finding the outcomes of positive abnormal returns than negative abnormal returns. One thing that stands out is the SVM's sensitivity, the SVM was relatively much more accurate at predicting the true outcomes of positive abnormal return compared to the other models.

The table down below shows the confusion matrix of all four models, the confusion matrix shows the distribution of prediction in relation to if they were predicted correctly or not.

Table 5: Confusion Matrix for the Machine Learning Algorithms

Random Forest		
	PRED P	PRED N
Actual P	120	48
Actual N	107	61

Extreme Gradient Boosting		
	PRED P	PRED N
Actual P	101	67
Actual N	89	79

Support Vector Machine		
	PRED P	PRED N
Actual P	124	44
Actual N	107	61

K-Nearest Neighbour		
	PRED P	PRED N
Actual P	113	55
Actual N	102	66

The confusion matrix shows that all models predicted the positive outcome more often than the negative outcome, this is not an outcome of skewed distribution in the dependent variable. The dependent variables are distributed: 582 negatives and 536 positives, and the test data is distributed 186/186. So the result in conjunction with the specificity/sensitivity table should be interpreted as the model is better at predicting the positive outcomes than negative outcomes.

5.2 Analysis

The accuracy of the models in this paper was not more than a couple of percentage points higher than 50%. However, this was not expected due to two major reasons. The first reason is that this study implements and uses the Efficient Market Hypothesis to a large degree, which states that in efficient markets it should effectively be impossible to develop profit-making strategies. Secondly, the model was using only three inputs as explanatory variables which are few variables to explain the movements of the stock markets. These two factors make the results of this study more compelling as everything over 50% in finance is tradable and profit-making strategy that can be implemented.

The result of this study, finding that it is possible to predict the direction of movement of abnormal returns based on earnings conference calls, is in line with the findings of Price et al. (2012). In their article, they concluded that earnings conference calls linguistic tone is a significant predictor of abnormal returns and trading volume. However, the articles differ in execution. Price et al. (2012) compared Loughran and McDonald's world list to the Harvard IV-4 Psychosocial Dictionary, which both are old approaches in comparison to how FinBERT is providing sentiment analysis. The newer approach FinBERT is generally able to uncover sentiment in sentences that the Loughran and McDonald's world list would misclassify as neutral according to the findings of Huang, Wang, and Yang (2020). Furthermore, Price et al. (2012) use Cumulative Abnormal Return (CAR) instead of the abnormal return used in this article. This means that their study is aimed at medium to long term effects, whereas this study was built on the Efficient Market Hypothesis and is focusing on the immediate effects. Moreover, they use their own constructed ratios of measures derived from the sentiment analysis as input and not only the plain sentiment scores used in this article.

Similar to Price et al. (2012), Fu, Wu, and Zhang (2019) look at long-term effects and use Loughran and McDonald's which are preferred by Price et al. (2012). Their findings are that firms with less optimistic tone of the Q4 earnings conference call, experience a higher risk of declining stock prices in the following year. This is also in line with the result of this paper, even if the two studies differ in a couple of regards. For example, the study is done on long term effects, and also the main goal is to look for risks and the potential of large size stock price decline. Amicis, Falconieri and Tastan

(2021), also show that the sentiment has an effect on the stock market although their main objective was to look at the differences between men and women.

The results of this paper are also in line with the findings of Tasker (1998) who presented evidence for undisclosed information in earnings conference calls. However, they differ in some regards. Tasker is able to isolate the effect of undisclosed information and say that there is undisclosed information presented at the conference calls. In this study, it can merely be stated that there seems to be valuable information presented at the conference calls and in conjunction with studies from Tasker (1998) argue that this information most likely to some extent is undisclosed. In this discussion, it should be mentioned again that the sample was specifically constructed with the reasoning that these 74 firms within the Information Technology Sector from the S&P 500 could have properties that made them sensitive to the type of information that could be disclosed in Q&A seasons. Information that could be undisclosed because it is rarely included in earnings reports by traditional structurings of earnings reports or that the constructors of earnings reports are unaware of what type of information the investor wants to know. The Q&A format allows the investors to inquire about this type of information and is therefore contributing to the undisclosed information.

This information could, as presented in the introduction, be related to things that are hard to communicate in earnings reports. Such as cryptocurrencies or the relationship between the US and China. Another aspect could be the time frame 2017-2021 and undisclosed information regarding the Covid-19 pandemic. The time period specified for the sample, or at least a part of the time period, was heavily affected by the Covid-19 pandemic. Even though the researcher's objective was not to analyze the effects of the pandemic, it should be mentioned because this of course has had some impact on the study itself. The Q&A sessions are probably the place where the management has presented or discussed undisclosed information regarding the pandemic, which also means that this period has probably been a contributing factor to information available in the earnings conference calls.

In this study, the sentiment scores were the only explanatory variables. This was done to isolate the effect of the earnings conference calls, where no other complementary information would be included. Arguably, the predictive power could increase even more if these explanatory variables were combined with other finance terms such as price/earnings, market value, growth rates, or other information from the regular earnings reports, to derive a model that uses both sentiment analysis and finance multiples. However, this was not the objective of this study.

The focus of the sample made the sample smaller with fewer observations than if for example the whole S&P 500 would be used. Fewer observations are generally negative for machine learning algorithms. This could be one of the reasons why the simplest approach, KNN, almost had the same accuracy as the newer and more complex approach XGB. Even though simpler is sometimes better, it is likely that the XGB would have performed better on a larger dataset or with more attributes. The best performing model was the SVM which achieved the highest sensitivity score by far of the tested models. This is interesting as it was one of its key differentiators which resulted in overall better performance. High sensitivity is also beneficial in the sense that it is one of the easier trading strategies. Shorting is almost always possible in close to efficient markets, but going long will always be possible in close to efficient markets.

6. Conclusion

6.1 Conclusion

The intent behind this study was to look into a relatively unexplored area in sentiment analysis and finance. The results of this study imply, in line with previous research, that earnings conference calls seem to be a vital source of information for investors. The reason for this is that there seems to be undisclosed information presented in the quarterly earnings conference calls. Using FinBERT to analyze the sentiments of 1118 earnings conference calls from 74 firms in the Information Technology Sector of the S&P 500 between 2017 and 2021 together with four different classification algorithms, it can be concluded that the sentiments in earnings conference calls can be used to predict the direction of movement of abnormal returns. The Support Vector Machine was the best performing model, reaching an accuracy of 55.06%, followed by Random Forest, Extreme Gradient Boosting, and K Nearest Neighbor. In line with the Theory of Efficient Markets, an accuracy exceeding 50% should not be expected. An accuracy of 55.06% is a profit-making strategy in financial modeling and should therefore be seen as a relatively good predictivity.

6.2 Future Research

For future research, it would be interesting to look at small and midsize companies. There is no evidence that there would be any difference between smaller and bigger companies, but it is not unreasonable that it would be the case. Previous research has shown that managers in small and midsize companies provide undisclosed information during the earnings conference calls (Tasker, 1998). Furthermore, it could also be relevant to investigate if certain parts of the earnings conference calls provide more valuable information than others. One could for example remove all non-company representatives or focus only on the Q&A-part of the transcript. The Q&A-part could be particularly interesting since the information provided during this specific part is not a part of the manuscript and it is therefore probably more likely to contain undisclosed information.

Another interesting approach would be to combine the more regularly used models that use financial multiples together with quarterly reports and earnings conference calls with sentiment analysis to create a predictive model that has more inputs. This would possibly create a more accurate model that could be further explored.

References

Amicis, D, C., Falconieri, S., & Tastan, M. (2021). Sentiment Analysis and Gender Differences in Earnings Conference Calls. *Journal of Corporate Finance*, vol. 71

Amrani, Y., Lazaar, M., & Kadiri, K. E. (2018). Random Forest and Support Vector Machine Based Hybrid Approach to Sentiment Analysis, *Procedia Computer Science,* vol. 127, pp.511-520

Araci, D. (2019). Finbert: Financial Sentiment Analysis with Pre-trained Language Models

Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and Lemmatization: a Comparison of Retrieval Performances

Bansal, A., & Kaur, S. (2018). Advances in Computing and Data Sciences, *Springer,* pp.372-380

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, vol. 2, pp.1-8

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research, *IEEE,* vol. 9, pp.48-57

Chopra, D., Joshi, N., & Mathur, I. (2016). Mastering Natural Language Processing with Python, *Packt Publishing*

Dolley, C. j. (1933). Characteristics and Procedure of Common Stock Split-Ups, *Harvard Business Review,* vol.11, pp.316-26

Healy, P. M., & Palepu, K. G. (2001). Information Asymmetry, Corporate Disclosure, and the Capital Markets: A Review of the Empirical Disclosure Literature, *Journal of Accounting and Economics*, vol. 31, pp.405-440

Huq, M. R., Ali, A., & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM, *International Journal of Advanced Computer Science and Applications*, vol. 8, pp.19-25

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work, *The Journal of Finance*, vol. 25, pp.383-417

Fama, E.F. (1998). Market Efficiency, Long-term Returns, and Behavioral Finance. *Journal of Financial Economics*, vol. 49, pp.283-306

Fama, E,F.,Fisher, L., Jensen, M., & Roll, R. (1969). The Adjustment of Stock Prices to New Information, *International Economic Review*, vol. 10, pp.1-21

Frankel, R., Johnson, M., & Skinner, D. J. (1999). An Empirical Examination of Conference Calls as a Voluntary Disclosure Medium, *Journal of Accounting Research*, vol. 37, pp.133-150

Fu, X., Wu, X., & Zhang, Z. (2019). The Information Role of Earnings Conference Call Tone: Evidence from stock price crash risk, *Journal of Business Ethics*, vol. 173, pp.643-660

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text Mining Methods and Techniques, *International Journal of Computer Applications*, vol. 85.

Gomez, A. N., Jones, L., Parmar, N., Polosukhin, I., Shazeer, N., Uszkoreit, J., & Vaswani, A, (2017). Attention is all you need

Gupte, A., Joshi, S., Gadgul, P., Kadam, A., & Gupte, A. (2014). Comparative Study of Classification Algorithms used in Sentiment Analysis, *International Journal of Computer Science and Information Technologies*, vol. 5, pp. 6261-6264

Kliger, D., & Gurevich, G. (2014). Event Studies for Financial Research: A Comprehensive Guide, *springer,* pp.19-41

Haddi, E., Liu, X., & Shi, Y. (2013). The Role of Text Pre-processing in Sentiment Analysis, *Procedia Computer Science*, vol. 17, pp.26-32

Huang, A., Wang, H., & Yang, Y. (2020). FinBERT—A Deep Learning Approach to Extracting Textual Information

Hollander, S., Pronk, M., & Roelofsen, E. (2010). Does Silence Speak? An Empirical Analysis of Disclosure Choices During Conference Calls, *Journal of Accounting Research*, vol. 48, pp.531-563

Kearney, C., & Liu, S. (2014). Textual Sentiment in Finance: A Survey of Methods and Models, *International Review of Financial Analysis*, vol. 33, pp.171-185

Liu, B., & Zhang, L. (2012). Mining Text Data, A Survey of Opinion Mining and Sentiment Analysis, *Springer*, pp.415–463

Loughran, T., & McDonald, B. (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance*, vol. 66, pp.35-65

MacKinlay, A. C. (1997). Event Studies in Economics and Finance, *Journal of Economic Literature,* vol.35, pp.13-39

Malkiel, B. (2003). The Efficient Market Hypothesis and Its Critics, *The Journal of Economic Perspectives*, vol. 17, pp59-82

Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., & Anastasiu, D. C. (2019). Stock Price Prediction Using News Sentiment Analysis. *IEEE*, pp. 205-208

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised Machine Learning algorithms: Classification and Comparison, *International Journal of Computer Trends and Technology*, vol. 48, pp.128-138

Peterson, P. P. (1989). Event Studies: A Review of Issues and Methodology, *Quarterly Journal of Business and Economics*, vol. 28, pp.36-66

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings Conference Calls and Stock Returns: The Incremental Informativeness of Textual Tone*, Journal of Banking & Finance*, vol. 36, pp. 992-1011.

Shi, H., Wang, H., Huang, Y., Zhao, L., Qin, C., & Liu, C. (2019). A Hierarchical Method Based on Weighted Extreme Gradient Boosting in ECG Heartbeat Classification, *Computer Methods and Programs in Biomedicine*, vol. 171, pp.1-10

SP Global, S&P 500. Avalable: https://www.spglobal.com/spdji/en/indices/equity/sp-500/#data . Accessed: 18 May 2022

Tasker, S. C. (1998). Bridging the Information Gap: Quarterly Conference Calls as a Medium for Voluntary Disclosure, *Review of Accounting Studies*, vol. 3, pp.137-167

Wolpert, D. H., & Macready, W. G. (1997). No Free Lunch Theorems for Optimization. *IEEE*, vol. 1, pp.67-82

Yadav, N., Kudale, O., Rao, A., Gupta, S., & Shitole, A. (2021). Intelligent Data Communication Technologies and Internet of Things, *Springer,* pp.631-642

Appendix

List 1: The complete Hyper Parameter grid search

Support Vector Machine:

Kernel (type of kernel): Poly and RDF

Gamma (Increasing gamma means increasing the modelling of the training data such that the model tries to fit it exactly) : 0,001-10

Penalty (controls the trade-off between smooth decision boundary and classifying the training points correctly, similar to gamma): 0,001-100

Degree (Flexibility of kernel): 3-9

Random Forest:

Criterion (How to evaluate a split): Entropy and Gini
Maximum Depth (Max depth of single node): 2-20
Maximum Features (Max features to be evaluated in node splitting): 1-3
Minimum Samples Leaf (Minimum number of observations in end leaf): 2-10
Minimum Samples Split (Minimum number of observations that is needed for node splitting): 2-10
Number of Estimators (Number of trees to construct the random forest from): 70-100
Bootstrap (Should sampling with replacement be implemented or not): True/False

K-Nearest Neighbour:

Number of Neighbors: 2-15

Weights (Different neighbours can get different weight depending on how close they are): Uniform (No) and Distance (Yes)

Metric (Different measures on distance): Minkowski, Chebyshev, Euclidean and Manhattan

Extreme Gradient Boosting

Minimum Child Weight (Defines the minimum sum of weights of all observations required in a child): 0.5-7

Gamma (Regularization parameter, high gamma high regularization): 0-11

Colsample_bynode (Percentage of features to be used in splitting node): 0,3-1

Maximum Depth (Max depth of single node): 2-20

Learning Rate (How fast the models learn from improvements): 0,01-0,30

Number of Estimators (Number of trees to fit): 70-200

Reg Alpha (L1 regularization): 0,0001-1

Reg Lambda (L2 regularization) :0,0001-1

Early Stopping (Criterion to stop if the model test performance doesn't improve by x number fits): 10-40

List 2: Optimised Hyperparameter for each Model

Support Vector Machine:

Kernel: Poly

Gamma: 100

Penalty: 1

Degree: 6

Random Forest:

Criterion: Gini

Maximum Depth: 6

Maximum Features: 2

Minimum Samples Leaf: 2

Minimum Samples Split: 3

Number of Estimators: 90

Bootstrap: False

K-Nearest Neighbour:

Number of Neighbors: 10

Weights: Uniform

Metric: Minkowski

Extreme Gradient Boosting

Minimum Child Weight: 0.5

Gamma: 0.6

Colsample_bynode: 0.5

Maximum Depth: 14

Learning Rate: 0.15

Number of Estimators: 70

Reg Alpha: 0.001

Reg Lambda: 1

Early Stopping: 15

Table 6: Companies observed in this study

Accenture	Arista	Corning	Global Payments	KLA	Nvidia	Seagate	Verisign
Adobe	Autodesk	DXC Technology	Hewlett Packard Enterprise	Lam Research	NXP	ServiceNow	Visa
ADP	Broadcom	Enphase	HP	Mastercard	Oracle	Skyworks	Western Digital
Akamai	Broadridge	EPAM	Intel	Microchip	Paychex	SolarEdge	Zebra
AMD	Cadence	F5	IBM	Micron	Paycom	Synopsys	
Amphenol	CDW	FIS	Intuit	Microsoft	PayPal	TE Connectivity	
Analog Devices	Ceridian	Fiserv	IPG Photonics	Monolithic Power Systems	PTC	Teradyne	
Ansys	Cisco	Fleetcor	Jack Henry & Associates	Motorola Solutions	Qorvo	Texas Instruments	
Apple	Citrix	Fortinet	Juniper Networks	NetApp	Qualcomm	Trimble	
Applied Materials	Cognizant	Gartner	Keysight	NortonLifeLock	Salesforce	Tyler Technologies	

List 3: *List of stopwords in NLTK stopword list*

'then', 'the', 'do', 'all', 'some', 'hers', "won't", 'at', 'over', 's', 'does', 'how', 'have', 'both', 'hasn', "it's", 'been', 'again', 'during', 'o', 'those', 'don', 'itself', 'yourself', 'their', "didn't", "wouldn't", "isn't", 'than', 'needn', 'shan', 'off', 'wasn', 'are', 'this', 'about', 'until', 'now', "needn't", "haven't", "that'll", 'couldn', "don't", 'mightn', 'be', 'aren', 'very', 'for', "aren't", 're', 'we', 'down', "should've", 'own', 'ours', 'wouldn', 'you', 'isn', 'they', 'once', 'not', 'same', 'mustn', 'where', 'an', 'above', 'only', "you'd", 'didn', "hasn't", 'can', 'through', 'no', 'them', 'himself', 'such', 'weren', 'haven', 'while', 'under', 'between', 'nor', 'd', "couldn't", "doesn't", 'ourselves', 'by', 'me', 'in', 'from', 'hadn', 'below', 'it', 'my', 'her', 'i', 'doing', 'because', 'so', 'few', "mustn't", "weren't", 'what', 'has', 'if', 'will', 'm', 'had', 'to', 'too', 'our', 'here', "shouldn't", 'your', "you've", 'on', 'any', 'its', 'did', "wasn't", 'herself', 'against', 'out', 'into', 'there', 'ain', 'theirs', 'who', 'up', 'after', 'that', 'themselves', 'just', 'most', "shan't", 'should', 'these', 'were', 'being', 'his', 've', 'll', 'doesn', 'won', "you're", 'before', 'as', 't', 'whom', 'but', 'why', 'ma', 'each', 'was', 'further', "hadn't", 'y', 'and', 'yourselves', 'myself', 'which', "you'll", 'other', 'of', 'with', 'she', 'having', 'a', 'when', 'am', 'is', 'he', 'yours', 'shouldn', "she's", "mightn't", 'or', 'him', 'more'