공학석사 학위논문

Modeling Stock Prices using Textual Contents in 10-Q Reports

10-Q 분기별 재무보고서 원문을 이용한 장기적 주가예측모델

2018년 08월

서울대학교 대학원

융합과학부 디지털정보융합학과

이 원 호

ABSTRACT

Researchers in the field of finance have reviewed various adaptation of algorithms and data sources for modeling stock prices. In previous literatures, prediction models proved to be effective in modeling short-term prices. However, there has been relatively little application of these datasets on modeling long-term price trends.

Successful research in fundamental analysis, school of financial research that models stock prices using news and other company-related information, depended on text analysis of news media, social network services, and internet message boards. This study proposes that textual content of 10-Q form mandated by Securities and Exchange Commission is more useful in modeling long-term stock prices since 10-Q reports have relatively less variety in information content, more future-oriented set of topics, and accurate information regarding future company risks and performance.

In order to test informational applicability of 10-Q text, total of 18,237 10-Q reports of companies listed in January 2018 Standard & Poor's 500 index were collected from January 2004 to January 2018 from SEC website. The collected corpus has been split into train and test subsets for out-of-sample evaluation. Test subset consists of 3,000 observations that are published between

December 2015 to January 2018, whereas train subset consists of 11,132 documents that are published at least 90 days prior to the first publication date in test subset.

Due to inherent difficulties in modeling long-term prices, we use few adjustments to standard methods in our experiment. In order to utilize only the most current and relevant information, only the recent additions to each 10-Q document is used during the experiment. We then used 50 topic probabilities created using Latent Dirichlet Allocation on our corpus to mitigate curse of dimensionality and induce hidden topical structures. Moreover, information content in 10-Q corpus is known to be associated with possible future events, which may be more effectively modeled using abnormal gains or losses within a specified period. Thus, we also propose a binary classification of stock prices using modified return-on-investment to represent stocks that have had abnormal gains within 90 days and stocks that have not.

We then construct 10-Q corpus stock price prediction model using Stacked Ensemble of Generalized Linear Regression, Random Forest, Extremely Randomized Trees, Gradient Boosting Machine, and Deep Feed-Forward Neural Network. Models were evaluated through the entire test set, quarter subsets, and simulated investment portfolio. Notable findings of this study include: 1) Highest prediction area under the Receiver Operating

Characteristic curve (AUC) of 0.5878 on model using Latent Dirichlet Allocation, proposed performance measurement, and Stacked Ensemble. 2) Model performance increases by using Latent Dirichlet Allocation (highest AUC of 0.5878) compared to bag-of-words text representation (0.5727). 3) Prediction cannot be made by using traditional return-on-investment using publication price and price at 90 days after publication. 3) Portfolio earnings using stocks selected by our model had higher two-year compound earnings of 55.78% compared to S&P 500 average of 29.98%. 4) There is no observable difference in topics between false positive stocks and true positive stocks in investment simulation. 5) Earnings in simulated investment portfolio increased when our proposed performance measurement is closer to distribution of 90-day prices.

This study makes the following contributions to the growing body of research in finance and machine learning. First, we review application of previously overlooked textual contents in 10-Q reports for modeling long-term stock prices. Second, we propose a new method for representing future stock prices and review difference in performance of models that are built through conventional and proposed return on investment. Lastly, we review application of LDA in using 10-Q corpus to build price prediction models.

In sum, this study concludes that 10-Q corpus can be used for modeling

stock prices. 10-Q corpus is a unique, future-oriented dataset that can be used for analysis of future stocks and company value. This study only reviewed a handful number of approaches to test informational usability of 10-Q reports, but there are a number of measurements that may enhance performance and provide a better understanding of effects of 10-Q text on stock prices.

Key words: Machine Learning, 10-Q Reports, Latent Dirichlet Allocation, Natural Language Processing, Stacked Ensemble

Student Number: 2014-24825

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Modeling financial market and stock prices has always been a topic of interest and challenge for both academic and business researchers. There have been numerous attempts at integrating machine learning algorithms to build viable prediction models, but the general census has been that markets are too efficient in information dispersion and that prices already reflect all available public information for there to be a sustainable prediction model (Malkiel, 1985). Nonetheless, recent advancements in machine learning and integration of new data sources have allowed researches for in-depth analysis and modeling of future stock prices.

There are two types of research in finance: technical analysis and fundamental analysis. Technical analysis focuses on patterns established in historical prices on predictable principles such as emotional momentum (Murphy, 1999), whereas fundamental analysis utilizes relevant quantitative and qualitative information such as media, social network services (SNS) and financial statements to analyze disparity between current stock prices and the future outlook of the company.

While advancements in machine learning has also strengthened the

performance of technical analysis experiments, research in fundamental analysis has especially benefited from integration of machine learning and natural language processing algorithms. Machines can now process natural languages and investigate impacts of previously convoluted information sources such as media, social network services (SNS), and internet message boards on the stock market, as shown in Table 1.

Author	Method	Target Prediction Date
Hassan et. al (2007)	Technical Analysis	Next Day
Ince, et. al (2017)	Technical Analysis	Next Day
Leu, et. al (2008)	Technical Analysis	7 Days
Patel, et. al (2015)	Technical Analysis	1-10, 15, 30 Days
Ding, et. al (2015)	Fundamental Analysis	Next Day
Zhai et. al (2007)	Fundamental Analysis	Next Day
Schumaker et. al (2008)	Fundamental Analysis	20 Minutes
Retchenthin et. al(2013)	Fundamental Analysis	Next Day
Sehgal et. al (2007)	Fundamental Analysis	Next Day

Table 1. Examples of Researches on Stock Modeling

Most researches in finance had focused heavily on modeling next-day stock trends. However, day-trading has been always characterized by high risk and volatility. Before developments of more recent algorithms, about twice as

many day traders do not profit from day-trading (Jordan, 2001), and day-trading has frequently been compared to lottery (Statman, 2002). Especially since it is difficult to prove generalizability of financial research (Sullivan, 2001), diversification of short-term and long-term analysis may mitigate the inherent risks of algorithm-based portfolios.

There are few reasons why long-term stock price prediction is not heavily reviewed in fundamental research. First, there is lack of research in a feasible method to represent long-term stock prices. Stock markets are affected by changes in demand, and prices often move in wave-like Elliot pattern (Atsalakis, 2011). Strategy employed by short-term stock prediction models focus on arbitrage of difference between public opinion (i.e. Bollen, 2011) and current stock prices. Long-run models are more susceptible to demand shifts, unforeseen events, differences in company structures, and other exogenous surprises that make the rate of impact incomparable.

Second, informational contents of SNS, internet boards, and media agency are too diverse in nature. Since public information is mostly focused on the most pertinent and current information (Esser, 1999), short-term investment models suffer less from information diversity. On the other hand, long-term modeling of stock prices may be more difficult due to the noise introduced by overwhelming amounts of short-term relevant news.

Lastly, investors are not accurate about future company prospects. Most professional investment managers do not surpass passive index benchmark (Malkiel, 2005). While mining of public opinion may be useful for modeling changing demands, long-run change in stock prices depend on performance and decisions of companies.

This dissertation therefore reviews applicability of 10-Q reports mandated by United States Securities and Exchange Commission (SEC) for long-term stock analysis. Perhaps considered as one of the most useful channels of information for buy-and-hold investors (Lawrence, 2013) and yet overlooked in the field of machine learning, reports filed to the SEC are mandatory for all publicly traded companies to submit through SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database (EDGAR, 2018).

SEC reports have unique characteristics that are helpful for future stock price analysis. First, the pre-defined topics in the reports allow for better comparison of information for long-term modeling as SEC intends reports to convey as much forward-looking information (Schneider, 1972) such as risks and management discussions. Second, SEC reports are written and revised by the management, auditors, and lawyers hired by the respective firms, who have knowledge of information about the decisions and strategies that may otherwise be unavailable to the public.

Although direct implementation of SEC filings on stock market prediction has not been extensively researched, contents of the reports have shown to have relationships with certain events and activities like decision to go public (Helwege, 2003), financial restraints (Bodnaruk, 2015) and risk analysis (Uppsala, 2010; Huang, 2011).

There are currently 158 SEC forms, ranging from report of sales to risk assessment for brokers and dealers. While most of the forms are designed in case of a specific event, there are three forms that are designed to cover a wide array of topics: 8-K, 10-K, and 10-Q forms.

All three forms are designed to provide a unique channel of information from company to individual investors. SEC requires 8-K from companies in case of pre-specified events according to Securities Exchange Act of 1934, such as changes in management and bankruptcies. On the other hand, 10-Q and 10-K forms are more extensive documents that mainly focus on providing summary of past and future activities and potential risks. 10-K form, published at the end of every fiscal year, reviews and presents activities that occurred throughout the entire fiscal year, whereas 10-Q form is published at the end of all other financial quarters and are focused on content specific to respective period (SEC, 2018).

Table 2 presents required period of submission, contents, and example of contents using Apple's recent SEC reports:

Form	Type	Topics	Example
8-K	Current report	Bankruptcies; Change in Control; Significant Acquisitions; etc.	*A management proposal to ratify the appointment of Ernst & Young LLP as Apple's independent registered public accounting firm for 2018 was approved.*
10-K	Yearly report	Risk Factors; Unresolved Staff Comments; Legal Proceedings; Controls and Procedures; Financial Statements; etc.	*Net sales declined 8% or $18.1 billion during 2016 compared to 2015, primarily driven by a year-over-year decrease in iPhone net sales and the effect of weakness in most foreign currencies relative to the U.S. dollar, partially offset by an increase in Services*
10-Q	Quarterly report	Changes in Securities; Discussion and Analysis; Disclosures Regarding Market Risks; etc.	*The Company recognized $695 million and $509 million of interest expense on its term debt for the three months ended December 30, 2017 and December 31, 2016, respectively.*

Table 2. Examples and Contents of 8-K, 10-K, and 10-Q reports

(source: sec.gov/edgar.html; Apple)

In this dissertation, we assess the effects of textual contents of 10-Q forms on modeling future stock prices. Information in 8-K form is event-based and easily interpreted for the market to quickly react. Moreover, while 10-K filings are more comprehensive and may contain more information regarding future performance, 10-Q filings have less informational diversity for more accurate comparison between reports.

Textual contents of 10-Q form are forward-looking (Seligman, 1994) and review a number of items that may affect future stock prices. According to the SEC guidelines, 10-Q form consists of two main sections: financial information and other information (SEC, 2018). Financial information section of 10-Q reports contains summary of the firm's operations and associated risks. This section is further broken down into: financial statements; management's discussion and analysis of financial condition and result of operations; quantitative and qualitative disclosure about market risk; and controls and procedures. Other information section discloses miscellaneous topics not covered in financial information section and consists of subsections: Legal proceedings, unregistered sales of equity securities and use of proceeds; defaults upon senior securities; mine safety disclosures; other information; and exhibits.

The experiment setup in this dissertation is as follows: Total of 18,237 10-Q reports of companies listed in January 2018 Standard and Poor's 500 index (S&P 500) are gathered from EDGAR database from March 2004 to January 2018. Since markets converge according to public information and old news becomes obsolete for price prediction models (Fama, 1965), we use only the most recent additions to 10-Q reports by comparing each report to the respective firm's last published report in order to minimize informational diversity and maximize strength of information for prediction models. We then apply topic modeling algorithm known as Latent Dirichlet Allocation (LDA) to remaining corpus in order to reduce number of dimensions and infer hidden topical structures.

To ensure model validity on out-of-sample datasets, we further divide our database to two subsets. Test set, used exclusively in model evaluation, consists of 3,000 10-Q documents that were released from December 2015 to January 2018. Train subset, in which we build our models using 5-fold cross-validation, comprises of 11,132 documents that were released at least 90 days before the first observation in the test set and had abnormal gains or losses within 90-days after publication.

We then construct binomial classification model with stocks that are

represented as either UP (Stocks with top 25% abnormal price gains) or DOWN (Stocks with bottom 75% abnormal gains) according to a modified performance measurement designed to represent abnormal price changes within 90 days after 10-Q publication dates. We then build several models with random grid search on hyper parameters of five learning algorithms: Generalized Linear Regression, Random Forest, Extremely Randomized Trees, Gradient Boosting Machine, and Deep Feed-Forward Neural Network. Lastly, we aggregate them using Stacked Ensemble using Generalized Linear Regression as meta-learner and evaluate our models using area under the Receiver Operating Characteristic curve (AUC), precision, and simulation of investment using the test set.

The experiment results show that there is some relationship with generated corpus and future performance, with highest AUC of 0.5878. The analysis of results reports no distinctive relationship in topics or publication date between false-positive stocks, but investment simulation suggests that investment performance may improve by using a better performance measurement for long-term modeling.

The rest of this thesis is organized in the following order: Chapter 2 provides a review of related literatures in previous stock prediction models and 10-Q dataset. In Chapter 3, we define and experiment our research questions.

Before we further explain our experiment design, we review the algorithms that are used in this dissertation. The experiment design is then presented in Chapter 4, results in Chapter 5, and discussion of results in Chapter 6. Chapter 7 concludes this dissertation with summary of this paper and future direction of 10-Q stock prediction research.

2. PREVIOUS LITERATURE

We present previous literatures on stock market analysis and SEC filings in this chapter. Section 2.1 summarizes previous researches on stock prediction models, whereas section 2.2 discusses researches on impact and uses of SEC filings.

2.1 Stock Prediction

Although this dissertation does not integrate technical indicators, technical analysis is often used in conjunction with fundamental analysis to improve performance (Oberlechner, 2001; Weng 2017). In essence, technical analysis is aimed to compare and utilize investor response to adjusting prices (Murphy, 1999), and researchers have tried to enhance portfolio performance through adaptation of state-of-art learning algorithms such as Genetic Modeling (Leu,

2008) or ensemble of algorithms (Hassan, 2007; Ince, 2017).

Fundamental analysis on the other hand uses public information about firms and corresponding sub-markets. Most researches on algorithmic approach to stock analysis use consumer-oriented data sources such as news, SNS, and message boards.

Fundamental analysis models have two advantages that make them viable in actual stock markets. First, while markets have been known to converge rapidly in reaction to current events (i.e. Hamilton, 1995; Woolridge, 1990; Li, 2008; Campbell, 2003; Koh, 1991), there is still some, albeit very short, lag before prices adjust that investors can exploit. For example, Schumaker et. al (2009) estimates that there is at least a 20-minute time lag in price change after first publication of current events when building models using bag of words, noun phrases and named entities.

Second, machine learning is more accurate and apt at assessing complex information. Short-term stock prices can be attributed to change in demand, and sentiment-analysis on consumer-oriented data sources have had powerful results in predicting next-day stock prices. For example, models based on Yahoo! Finance message boards showed 75% accuracy (Rechenthin, 2013) and 81% accuracy (Sehgal, 2007) for predicting next day trends. Models built

on economic news produced 65% accuracy on next day stock prices (Kim, 2014), whereas Vu et. al (2012) found 75% accuracy for keyword-based sentiment analysis. The famous article by Bollen et. al (2011) achieved same-day Dow-Jones Industrial Average prediction accuracy up to 86.7% and expanded prediction periods with meaningful prediction rates up to 6 days after Twitter posts. Li et. al (2014) directly compared bag-of-words approach to sentiment analysis on same-day stock price prediction using news and found that sentiment analysis outperformed bag-of-words on both validation and independent testing sets.

While first advantage does not apply to long-term modeling of prices, research in sentiment analysis showed that inclusion of complex, intangible information not only improved model accuracy but extended duration of investment opportunity. Markets have already been reported to underreact to more complex 10-Q and 10-K filings (Haifeng, 2008), suggesting that information in 10-Q and 10-K filings are complex enough for long-term stock price modeling.

2.2 SEC Reports

SEC filings are important sources of information for investors. When 10-Q and 10-K are released, the absolute value of excess returns increases for a

short duration after each publication (Griffin, 2003), whereas delay in publication of SEC filings have been shown to induce negative market reactions (Cao, 2010).

There have been many previous attempts to integrate financial statement portion of SEC reports to predicting future stock prices. Accounting ratios provided in the financial statements have shown to be useful in producing positive excess returns in investment strategies (Holthausen, 1992; Ou, 1989) and financial statements in SEC reports have shown negative association between estimated unexpected discretionary accruals and stock returns (Balsam, 2002).

There is relatively little research on directly modeling stock prices using SEC documents. However, previous literatures have shown that SEC documents are effective in analyzing companies. Contents of 10-K reports have frequently been used to identify type of risk associated with a company (Uppsala, 2010), to automatically detect risk groups using K-nearest neighbors (Huang, 2011), or to detect financial restraints (Bodnaruk, 2015).

In general, SEC filings provide information relevant to events that have not occurred. Firms facing greater risks also tended to disclose more (Campbell,

2014). Textual contents of SEC filings have also been used to directly forecast future events. Analysis of SEC filings showed that some private firms are more likely to go public (Helwege, 2003). 10-K text has also been used to identify fraudulent firms with 75% accuracy and predict bankruptcy with 80% accuracy using vocabulary used in management discussion and analysis section (Cecchini, 2012).

There were few attempts at direct usage of text analysis on SEC reports. Lee et. al (2014) studied the usage of neural networks on text contents in 8-K form in building stock price prediction models up to three days after the publication date using Random Forest. Qui (2007) in his thesis explored the relationship between texts of 10-K reports and company performance indices one to two years after the initial publication of 10-K reports using support vector machines. However, although he found meaningful prediction results for predicting stock prices of documents published within the same year, his approach did not show significant performance for out-of-sample stock predictions.

3. RESEARCH QUESTIONS

The goal of this dissertation is to assess information value of 10-Q reports on stock evaluation and therefore focus on the following research questions:

RQ1. Do the textual contents of 10-Q reports have informational value?

SEC mandates a large amount of information from public companies, and on average companies experience about $1.5 million dollars as a recurring cost to become public and one-time cost of over $1 million dollars at initial public offering (Strategy&, 2012). Considering that there were 4,333 public companies in United States as of June 2016 (Ibrahim, 2016), publication and maintenance of SEC reports incur a large sum of money for the economy. The increasing costs associated with hiring financial analysts and auditors may even explain decreasing number of firms that go public (Doidge, 2017). Understanding that information content in SEC filings provide informational value for investors, however, may justify the associated costs.

RQ2. Does hidden topical structures of 10-Q corpus add value when modeling future stock prices?

Traditionally, texts have been represented with bag-of-words. However, while bag-of-words provide analysis on direct correlation between selected words and independent variable in prediction models, research using bag-of-words are often hindered by curse-of-dimensionality (Bellman, 1957) and by loss of semantic information between words. Previous literatures suggest that utilizing higher-level analysis of vocabulary improved general model accuracy and extend the duration of informational usability in modeling stock prices.

Moreover, semantic structure of human language may be important to representing information in large collection of text datasets that uses variety of jargons and vocabulary such as 10-Q reports. Topic modeling allows for inference of more complex information in a given corpus by uncovering hidden topical patterns in the corpus (Karthik, 2013).

There are many different methods of textual representation that may improve model quality. Improvement in model performance using topic modeling indicate that research in representation of 10-Q corpus may uncover more information relevant for analysis of long-term stock prices.

.

4. METHODOLOGICAL BACKGROUND

This chapter introduces and reviews application of LDA and machine learning algorithms used in this dissertation. Section 4.1 introduces the implementation of LDA, whereas section 4.2 describes the implementation of various machine learning algorithms that are used to build our prediction models.

4.1 Latent Dirichlet Allocation

We use topic modeling to represent 10-Q documents. Topic modeling has been known to be effective at mining hidden topical structures and reducing dimensionality. Texts that are rich in vocabulary such as 10-Q reports have problems of high dimensionality and places strain on both time it takes optimize an algorithm and generalization needed for out-of-sample predictions (Donoho, 2000).

Specifically, this paper uses unsupervised learning algorithm Latent Dirichlet Allocation (LDA) and topic probability created by LDA as features for supervised learning. More specifically, we use LDA on term-frequency inverse document frequency weights (TF-IDF) of collected corpus to

automatically derive topical structures.

LDA is a generative probabilistic linguistic model that assumes that documents may be explained by a set of topics and that each word may be attributed to one or more of the determined topics (Blei, 2003). TF-IDF on the other hand is an NLP weighting statistic technique aimed to adjust feature weights according to importance of words in a corpus (Ramos, 2003). The generated probabilities of LDA may be used to represent documents (Cai, 2016) and often increases accuracy for supervised machine learning algorithms that the traditional bag-of-words and simple TF-IDF models do not deliver (Hong, 2011).

There are many other algorithms to represent texts. In previous literatures on quantitative analysis, researchers ascribed words to sentiments as a method of representing public opinions. However, language structure in financial reports have been reported to be misleading (Bonsall, 2017) and Levels of jargons and length of document in SEC filings is known to be different according to company performance (Subramanian, 1993), which may reduce applicability of automatic sentiment analysis algorithms. Since the goal of this dissertation is to test usability of 10-Q reports on modeling stock prices, we do not use pre-defined set of sentiments that may produce different results per method of classification.

4.2 Supervised Learning

To maximize prediction accuracy and application of information in 10-Q reports, we adopt a number of individual learning methods and aggregate them together using Stacked Ensemble algorithm. Stacked Ensemble is a model that combines several base learners using a meta learner (Breiman, 1997) into one cumulative model using n-fold cross validation to evaluate algorithm weights (Laan, 2007), and has been proven to improve model performance.

This dissertation uses five independent learning algorithms: Generalized Linear Regression, Random Forest, Extremely Randomized Trees, Gradient Boosting Machine and deep Multilayered Feed-Forward Neural Network. Generalized Linear Regression (GLM) is a basic learning model developed from ordinary linear regression to incorporate error distribution models that do not follow normal distributions (John, 1972). We choose to use GLM since our main experiment uses binary classification to indicate stock performance. Moreover, GLM is used as a basic learner to compare accuracy with other learners, and as a meta learner for accumulation of individual models in SE in this dissertation.

Random Forest (RF), Extremely Randomized Trees (ET) and Gradient Boosting Machine (GbM) are three similar learning models based on ensemble of decision trees. RF is a common method based on bagging using decision trees created using random samples of data (Ho, 1995). Because RF is built on subsamples, larger model usually leads to better generalization and thus are not as susceptible to change in hyper parameters. Compared to the other two decision tree algorithms, RF models are less likely to over fit and thus may be more advantageous for out-of-sample predictions (Oshiro, 2012).

ET is a variant of RF. In RF, subsamples are selected using bootstrapping while optimal cut-off points for top-down splitting are calculated using a performance measurement (Ho, 1995). ET take a step further in randomization by creating decision trees using the entire dataset but randomly picking cut-off points. The additional random cut-off reduces variance in trade-off with increasing bias in prediction models and may outperform RF depending on how noisy the dataset is (Geurts, 2006). Since 10-Q documents contain many noise, such as outdated or superfluous information, ET may be more advantageous for 10-Q dataset compared to RF.

GBM, on the other hand, is an additive model where each decision tree is built on prediction residuals of previous trees (Ann, 2000). Every additional decision tree increases expressiveness of the model and increases model

performance. However, although in general GBM models outperform against RF models, GBM are more prone to over fitting as trees will continuously be added until all residuals are fitted if left unchecked (Elith, 2008), which can be countered by careful selection of appropriate hyper parameters such as learning rate and max depth.

Deep feed-forward neural network (DNN) is a different class of machine learning that is based on neural network. In essence, it is a supervised learning technique that utilizes at least three layers of nodes that are connected via back propagation, in which the errors are calculated at the output and fed backwards to the layers of nodes through a non-linear activation function (LeCun, 2015). DNN frequently outperform other learning problems as it reduces needs for feature engineering in complex problems. However, because DNN tends to over fit, this paper adopts dropout as a method of regularization (Srivastava, 2014) and adaptive learning rate for gradient descent to find appropriate learning rate (Zeiler, 2012)

Each machine learning algorithm employs a variety of hyper parameters, such as number of hidden nodes, dropout rate, and activation function for DNN; and number of trees and learning rate for GBM. In order to maximize generalization and prediction rates, we used adaptive random grid search of hyper parameters with n-fold cross-validation to find optimal set of hyper

parameters (Bergstra, 2012). We focus random grid search mostly on DNN and GBM since the two are more prone to overfitting and are more susceptible to changes in hyper parameters.

5. RESEARCH METHODS

This chapter presents the experimental design used in this dissertation in detail. Section 5.1 reviews collection of 10-Q reports and how LDA is implemented. Section 5.2 explains our proposed method of stock price performance measurement designed to represent price trends through abnormal price changes. Section 5.3 summarizes the experimental design and how train and test sets are divided to evaluate out-of-sample stock predictions.

5.1 10-Q Reports

5.1.1 Data

In order to test if stock prices can be modeled with 10-Q documents, we collected 18,737 10-Q reports of companies listed in January 2018 Standard and Poor's 500 index (S&P 500) from January 2004 to January 2018. As there is no API for downloading 10-Q reports at the time of research, the 10-Q reports used in this study were directly crawled from the EDGAR website using python script libraries such as *urllib2*, *BeautifulSoup4*, and *Selenium*. We then isolated textual contents of 10-Q reports using *BeautifulSoup4* as downloaded 10-Q documents were in *XML* format.

Many companies listed in S&P 500 have had their initial public offering during the experiment period, and thus each company produced different number of 10-Q reports. Table 3 presents descriptive statistics of 10-Q reports used in this experiment. Majority of the companies in this dataset have published 10-Q reports for the most of the experiment period but there are some companies that have issued considerably less reports due to late public offering. However, although a topic for investigation, we assume that it does not affect our model since the age of firms does not seem to affect at least the mortality rate of public firms (Daepp, 2015).

Lastly, S&P 500 companies are placed in various sectors of the economy and are known to be an efficient representation of the overall market. As also presented by the descriptive statistics in Table 3, there are sizeable amount of 10-Q reports per sub-industry sectors as defined by Standard and Poor's Global Industry Classification Standard (GICS). Although not all GICS subsectors are covered, the collected dataset covers 122 out of 157 GICS sub-industries.

At least 25% of the sub-industries are represented by less than 50 documents. Model proposed in this dissertation may be more effective for some sub-industries since overall language structure and jargons used in 10-Q reports may be different per industry. However, we do not cover this topic since there

are not enough documents to provide a meaningful analysis of industry effect.

	10-Q Reports	
Statistic	Per Company	Per GICS Sub-Industry
Mean	36.55	150.5
Min	2	15
25th Percentile	38	42
Median	42	120
75th Percentile	43	193
Max	49	582

Table 3. Descriptive Statistics of Collected 10-Q Documents

5.1.2 Text Representation

10-Q documents are designed to cover a wide array of historical, ongoing and new topics. However, efficient market hypothesis supports that stock prices eventually converges onto the market equilibrium onset by the release of public information (Fama, 1965). Thus, it is less likely for old content to be effective for analyzing future stock prices.

Even if information does not lose potency over time, using both old and current information greatly increases information diversity. It may not be

feasible to assume that duration of impact is same for old and new information. For example, information on market default risk that have been in 10-Q reports in 2008 may not be as applicable as additional information in reports published after 2018 for current price analysis. There may be other methods that weights words for machine learning according to their age, but we chose to remove old information altogether to simplify our model.

We therefore compared each 10-Q documents to the corresponding firm's previously published reports and utilized only the sentences that have been added since last publication. We also chose to omit deletions from our analysis since deleted words are less likely to hold homogeneity in variable effects. Sentences may have been promptly deleted according to changing market conditions, or overlooked until number of years later while cleaning extraneous information in quarterly reports.

The remaining words in the corpus is then filtered by removal of stop words and words that contain numbers or non-alphabet characters. The remaining tokens are lemmatized using *NLTK*'s *wordnet lemmatizer* (Porter, 2001). We chose to use lemmatization instead of stemming in order to minimize loss of semantic information that comes from part of speech. Since words in 10-Q reports are strictly reviewed by auditors and lawyers, even the smallest change in connotation that comes from differing part of speech may hold hidden

connotations about company status and future prospects.

An example of how words in 10-Q reports were processed for LDA is shown in Table 4 using Exelon 10-Q reports in August 2016 and October 2016.

Exelon 10-Q, August 2016	**Exelon 10-Q, October, 2016**	**Added Words**	**Added to Corpus**
Exelon estimates total commitments of approximately $444 million on a net present value basis will be provided	In the first quarter of 2016, Exelon estimated and recorded total nominal cost commitments of $508 million, excluding renewable generation commitments	In, the, first, of, quarter, recorded, total, nominal, commitments, excluding, renewable, generation, commitments	*Omitted* quarter record, total nominal, commitment, exclude, renewable, generation, commitment

Table 4. Example of Changes Made in 10-Q Reports

After removing infrequent words that appear in less than 10 documents, there are total of 32,798 word tokens that remain in our corpus. However,

using machine learning algorithms directly on the corpus is not only computationally heavy but also misleading due to curse of dimensionality (Bellman, 1961). We therefore use LDA to deduce hidden topical structures that may be useful for information retrieval and to reduce number of dimensions.

Moreover, vocabulary used in 10-Q documents are diverse and sometimes use language structure not congruous to other 10-Q documents. Some of the most common problems with SEC filings are weak or hidden verbs, superfluous words, legal and financial jargons (Bonsall, 2017) that may be mitigated by using LDA to categorize words according to possible topics.

In terms of implementation, the results of LDA may change according to number of topics (Blei, 2012). Large number of topics may decrease quality of topics, but small number of topics may lose too much information necessary for supervised learning. Although there are some algorithms designed to review quality of topic split (Arun, 2010), there is no standard method in finding the right number of topics to maximize model performance for supervised learning based on LDA word vectors. Thus, this paper uses arbitrary number of 50 topics, similar to 31 topics chosen by Qui (2007) who studied effects of textual contents in 10-K reports on stock prices.

Lastly, in order to answer RQ2, we also built prediction model using bag-of-words to evaluate applicability of LDA on 10-Q dataset. While text representation method is different between the two datasets, we used exactly same supervised learning algorithms and stock price performance representation to compare effects of LDA implementation for this paper's experiment.

5.2 Stock Performance

Stock prices were gathered from *Yahoo! Finance* using *NLTK* API for corresponding firms of 10-Q reports. This papers uses adjusted closing prices offered by *Yahoo! Finance*, which accounts for stock manipulation such as splits in order to provide better comparison of observations for stock price prediction models.

The task in this paper is to test the effectiveness of 10-Q reports on future stock performances. There are many conventional and newly developed methods, such as earnings per share and price to earnings ratio to growth ratio. While these measurements are often used to estimate company performances (Bradshaw, 2004), it is difficult to adopt these measurements in assessing long-term stock price prediction models due to increase in the effects of other stimuluses such as market conditions and catastrophic events.

One of the most commonly used is to use moving averages. While using only moving averages have been often used in previous researches on modeling short-term stock price (Kimoto, 1990), it also requires a specific date of saturation for information.

Instead, we model our corpus against abnormal changes to stock prices. Since information in SEC reports have shown to be effective in predicting future risks, decisions, and catastrophes, this paper uses abnormal changes in prices to represent events that had occurred according to information embedded in 10-Q reports.

In order to calculate abnormal changes, we use an adaptation of return on investment. Traditional return on investment uses the ratio between net gain on investment and the cost of the investment as a way to measure direct impact of stock price change for the investors:

$$\frac{Profit\ on\ Investment - Cost\ of\ Investment}{Cost\ of\ Investment}$$

There are few short comings to using traditional return on investment for long-term price modeling. First, it is difficult to set specific number of days

after publication as the date of information impact as it is not feasible to assume that information affects the market in the same manner over time. Second, stock market tends to react according to changes in demand (Bondt, 1985). Price is known to fluctuate (Atsalakis, 2011), making it even more problematic when trying to set a specific number of days after publication for price prediction models. By using a proposed return on investment, we aim to model effect of information on stock prices without restriction of homogeneity in duration of information impact.

To illustrate further, this paper defines the following vocabulary:

- Publication Date: Date on which 10-Q report was published.
- Moving Average: Average of stock prices three days prior to selected date
- Price at Publication: Moving average stock price at publication.
- Observation Period: Period between the publication date – purchase date - and 90 days after publication.
- Maximum/Minimum Stock Price: The maximum/minimum stock price during the observation period.
- Point of Maximum Distance: The point of which stock price deviates

furthest from the price at publication.

- Return on Investment: Return on investment if stock is considered to have been bought at publication date and sold at point of maximum distance using moving averages.
- Opposite Point of Maximum Distance: Maximum/minimum stock price in opposite direction from point of maximum distance.
- Minimal Difference in Distance: Minimal difference in distance between point of maximum distance and purchase price and distance between opposite point of maximum distance and purchase price.
- Stagnant Stocks: Stocks that do not meet the minimal difference in distance criteria and are considered to have either risen nor fallen
- UP Class: Observations with top 25% return on investment in a given dataset
- DOWN Class: Observations with bottom 75% return on investment in a given dataset.

The proposed method compares distances to maximum and minimum price within the observation period. Distances are calculated using traditional return on investment using publication price and maximum/minimum price. Only the

point that has deviated furthest from the initial price is used for final calculation of return on investment. In sum, the modified return on investment is calculated by using moving averages of stock prices at publication date as the purchase date, and point of maximum distance within the observation period as the sales date.

Since prices may simply be moving along the Elliot curve of an unchanging true value, we also add additional restriction in which point of maximum distance has to be at least twice as further from the distance in opposite direction. Stocks that do not meet the criteria are considered to not have experienced abnormal gains or losses, and modified return on investment used for supervised learning is set to zero.

Stocks that have neither risen nor fallen introduces bias in assessing impact of our variables toward zero. Thus they are omitted from model building. However, since it is impossible to distinguish these stocks, they are utilized for evaluation using test subset.

Figure 1 shows examples of modified return on investment:

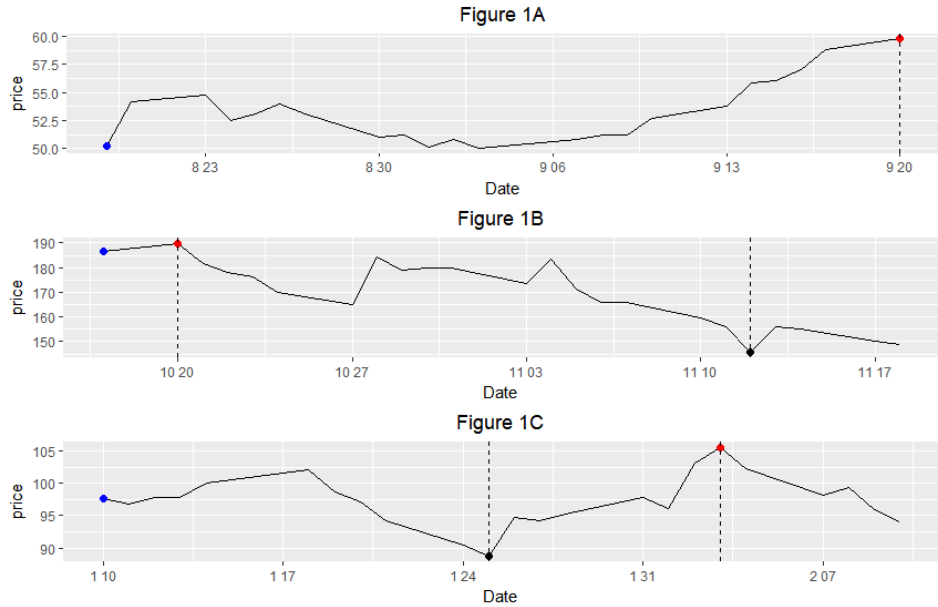

Figure 1. Example of Stock Prices

For simplicity, we assume that prices at publication in Figure 1 is three-day moving averages.

There are three cases presented in Figure 1. Figure 1A exemplifies easiest cases of modified return on investment. Stock price continuously rises, and never falls below price at publication, indicated by blue dot. Minimum price is then equal to publication price, and return on investment is calculated at the maximum price, indicated by red dot. Return on investment in this example therefore is (60-50) / 50, or 0.2.

On the other hand, in Figure 1B, the price initially increases to maximum

stock price but continues to decrease over the observation period. Minimum price, indicated by black dot, is sufficiently further away from price at publication with distance of (185-145) / 185, or 0.2162, compared to maximum price with distance of (190-185)/185, or 0.0270. Since the distance to minimum price is eight times more than the distance to maximum price, return on investment is calculated at the point of minimum price at (145-185)/185, or -0.2162.

Lastly, in Figure 1C, stock price does not noticeably move toward one direction. The distance from publication price to maximum price is 107-97, or 0.1031, whereas the distance from publication price to minimum price is 97-90, or 0.0722. Since neither distances are sufficiently larger than the other, stock in this example is considered to have neither risen nor fallen, and the return on investment is set to 0.

It is noteworthy to mention that this approach does not provide insight on optimal date of sales for selected stocks in portfolio. Prices may hit a maximum or minimum value within the given period and eventually revert in the other direction. Instead, the modified return on investment should be viewed as a representation of price changes due to significant future events, but not as a direct measure of price trends. Thus, we use our proposed method for model building and evaluation, but not for appraising simulation of

investment portfolio.

Lastly, we transform our return on investment to binary classification. It is also infeasible to assume that events impact prices at equal rates. Oil spillage, for example, impacts oil industry differently compared to automobile industry. We thus categorize stock prices as either UP or DOWN classes. Using proposed return on investments, UP class is defined as stocks that have increased significantly (top 25% of the given dataset) while DOWN class consists of all other observations not in the UP class. Since modified return on investment is intended to represent abnormal changes in stock prices, using a high cutoff point for class distribution ensures that UP class represents stocks with abnormal gains in stock prices within the observation period.

5.3 Experiment Design

The overview of this paper's experiment design is presented in Figure 2.

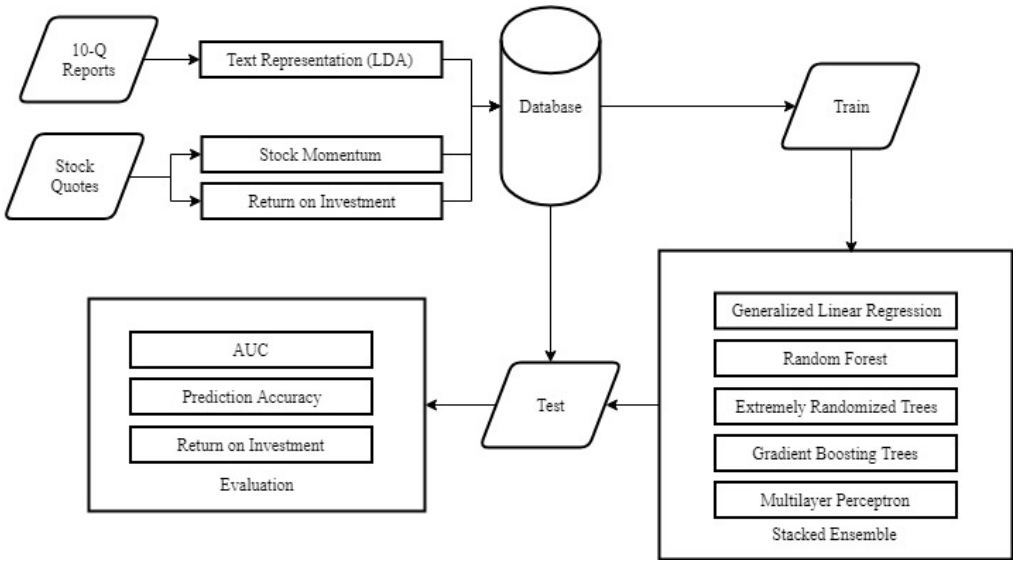

Figure 2. System Overview

In order to test generalizability of our models and out-of-sample prediction performance, we divide the corpus into train and test subsets to ensure model building only occurs separately from performance evaluation. Test subset consists of 3,000 observations published in between December 2015 to January 2018, whereas train subset consists of 14,792 documents that are at least 90 days prior to the earliest 10-Q report in the test subset. We further remove ambiguous documents that do not satisfy the two-fold criteria from train subset, leaving 11,132 documents for building prediction models.

Price distributions for our modified return on investment are almost

identical between train and test subsets prior to removal of ambiguous documents. Cutoff point for UP class is 14.25 for train subset and 13.92 for the test subset. For simplicity of performance evaluation, we use the test subset distribution for building our model. There are total of 750 UP and 2250 DOWN stocks in the test subset, and 3806 UP class and 7,326 DOWN class in train subset.

Prediction models are built strictly on the train subset. For generalizability, we adapted 5-fold cross-validation. After building a number of independent models using random grid search of hyper parameters, models are aggregated using SE with GLM as meta-learner. The models were then evaluated on the test dataset and reported in Chapter 6.

6. RESULTS

In this chapter, we present the results of our experiment. In section 5.1, we evaluate our models using standard evaluation metrics on different subsets of our test set. Section 5.2 presents an investment simulation by selecting stocks according to the predicted probabilities using the Stacked Ensemble model. Lastly, in section 5.3, we present top ten most important topics and their relative importance in percentages.

6.1 Model Evaluation

Models in this section are evaluated using out-of-sample performance measurements. However, we selected models with highest cross-validation (CV) AUC since there are more than one model per algorithm, especially for DNN and GBM that requires more stringent random grid search on hyper parameters. Model performance is reported using AUC, F-score, recall and precision. Predicted class cutoff point for recall and precision have been selected at the point that maximizes F-score.

Table 5 presents evaluation of bag-of-words, modified return on investment model:

Algorithm	Date Type	AUC	Recall	Precision	F-Score
GLM	CV	0.5050	1.0000	0.2644	0.4182
	Test	0.4934	0.9987	0.2502	0.4001
XRT	CV	0.5015	0.9987	0.2647	0.4186
	Test	0.4873	0.9987	0.2507	0.4007
RF	CV	0.5712	0.8817	0.2792	0.4241
	Test	0.5377	0.8947	0.2684	0.4128
GBM	CV	0.6144	0.8788	0.2946	0.4413
	Test	0. 5557	0.8440	0.2750	0.4148
DNN	CV	0.5920	0.8949	0.2830	0.4300
	Test	0.5436	0.8987	0.2635	0.4075
SE	CV	0.6268	0.7991	0.3099	0.4466
	Test	0.5727	0.7920	0.2809	0.4147

Table 5. Prediction Performance using Bag-of-Words Representation

Prediction failed in two out of six models. GBM and XRT both had lower than 0.50 AUC for out-of-sample testing. Moreover, performance barely passes the base-line in models using other three independent algorithms.

Table 6 presents model performances of models built using LDA-representation and return-on-investment calculated using price at publication

and price at 90 days after publication date.

Algorithm	Date Type	AUC	Recall	Precision	F-Score
GLM	CV	0.5683	0.8597	0.2492	0.3864
	Test	0.4891	0.9987	0.2503	0.4002
XRT	CV	0.5453	0.8831	0.2452	0.3838
	Test	0.4999	1.0000	0.2500	0.4000
RF	CV	0.5451	0.9813	0.2376	0.3825
	Test	0.5027	1.0000	0.2503	0.4003
GBM	CV	0.5622	0.9683	0.2400	0.3846
	Test	0.4958	1.0000	0.2500	0.4000
DNN	CV	0.5596	0.8920	0.2451	0.3845
	Test	0.4808	1.0000	0.2500	0.4000
SE	CV	0.5695	0.9259	0.2446	0.3869
	Test	0.4880	1.0000	0.2502	0.4002

Table 6. Prediction Performance using LDA, 90-Day Returns

Prediction performance failed to achieve performance greater than random guess in all 6 models for out-of-sample testing.

In Table 7, we present the performance of prediction model using LDA-representation and modified return-on-investment:

Algorithm	Date Type	AUC	Recall	Precision	F-Score
GLM	CV	0.5874	0.8142	0.2424	0.3736
	Test	0.5601	0.8827	0.2688	0.4121
XRT	CV	0.5614	0.8073	0.2363	0.3656
	Test	0.5560	0.8987	0.2687	0.4138
RF	CV	0.5773	0.8119	0.2402	0.3707
	Test	0.5604	0.9787	0.2547	0.4042
GBM	CV	0.5911	0.7240	0.2561	0.3783
	Test	0.5742	0.8613	0.2758	0.4179
DNN	CV	0.5819	0.8433	0.2380	0.3712
	Test	0.5610	0.9640	0.2571	0.4059
SE	CV	0.5967	0.8162	0.2476	0.3800
	Test	0.5878	0.9360	0.2687	0.4178

Table 7. Prediction Performance using LDA, modified Performance

Compared to the model using bag-of-words or 90-day returns, all six learners outperformed on out-of-sample testing. However, DNN, GBM, and SE models using bag-of-words corpus had higher performance during cross-

validation. The results are in concordance with the assumption that bag-of-words representation is usually more susceptible to overfitting due to curse of dimensionality.

All six representative models of supervised learning algorithms using LDA corpus indicate that there is a relationship between textual contents of 10-Q reports and future stock prices. Stacked Ensemble model reported highest performance AUC in both CV and test dataset and are used as the default evaluation model unless stated otherwise.

In order to test generalizability of our model outside the timeframe in our test set, we further divide our test set into subsets of 8 separate quarters, starting from 2016 Q1 (January - March) to 2017 Q4 (October – December).

Due to the fact that 10-Q reports are only published three times a year and some reports are delayed up to 90 days after due date, 10-Q reports are not equally distributed in all 8 periods. For Q2 to Q4, number of published 10-Q reports range from 449 to 476, but there are only 95 and 100 documents in Q1 for years 2016 and 2017. The distribution of modified return to investment per each period are presented in Figure 3:

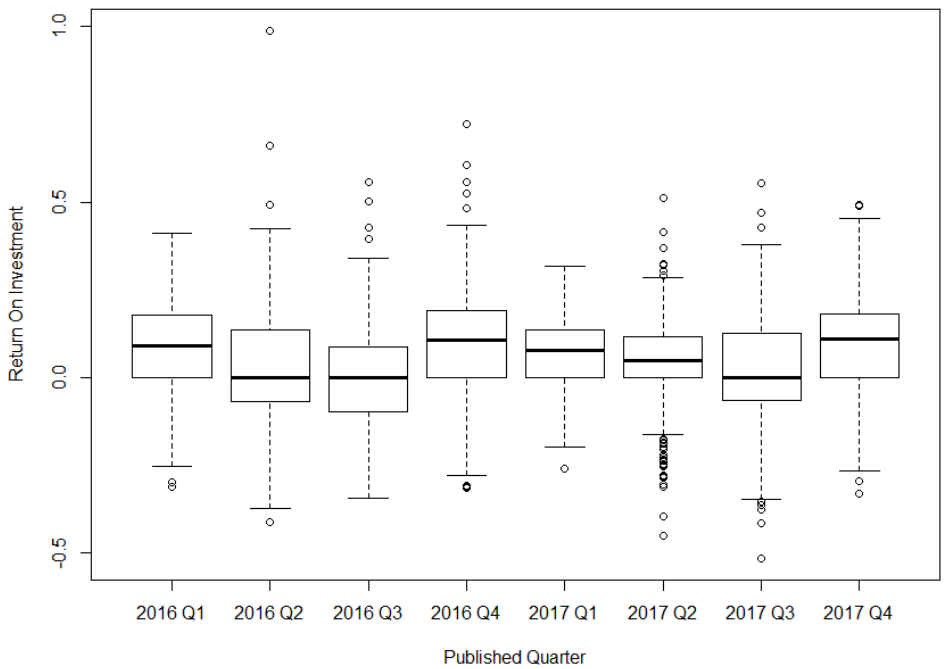

Figure 3. Modified Return on Investment per Published Quarter

While distribution of modified return on investment are relatively similar in most subsets, 2016 Q2, 2016 Q3, and 2017 Q3 saw marked lower average modified return on investment than in other publication quarters. However, the average on overall S&P 500 return during the three periods (-0.0148, 0.0414, and 0.0245, respectively) is not significantly different from the median return of 0.0373 during other quarters.

Table 8 summarizes the results of our experiment from 2016 Q1 to 2017 Q4.

Quarter	AUC	Recall	Precision	UP Class Percentage
2016 Q1	0.4652	1.0000	0.3298	0.3263
2016 Q2	0.5389	0.9730	0.2450	0.2362
2016 Q3	0.6347	0.7969	0.1855	0.1391
2016 Q4	0.6187	0.9167	0.4185	0.3741
2017 Q1	0.6217	0.6087	0.3500	0.2300
2017 Q2	0.5708	0.5441	0.1927	0.1429
2017 Q3	0.5410	0.9368	0.2259	0.2065
2017 Q4	0.6141	0.9840	0.4402	0.4026

Table 8. Evaluation per Quarter Subsets

In general, 7 out of 8 quarters reported positive AUC, with 2016 Q3 as the period of highest AUC with 0.6347 and 2016 Q1 as the period of lowest AUC with 0.4652.

The reported precision rates in Table 8 for individual quarters are not meaningfully higher than the actual distribution of UP class in each quarter. However, the metrics used in Table 8 are based on the cutoff point for predicted classes that maximizes the F-scores. For actual investment decisions, precision is more important in selecting and comparing stocks. As shown in

Figure 4, precision may be increased at the cost of recall by decreasing the number of predicted UP class stocks.

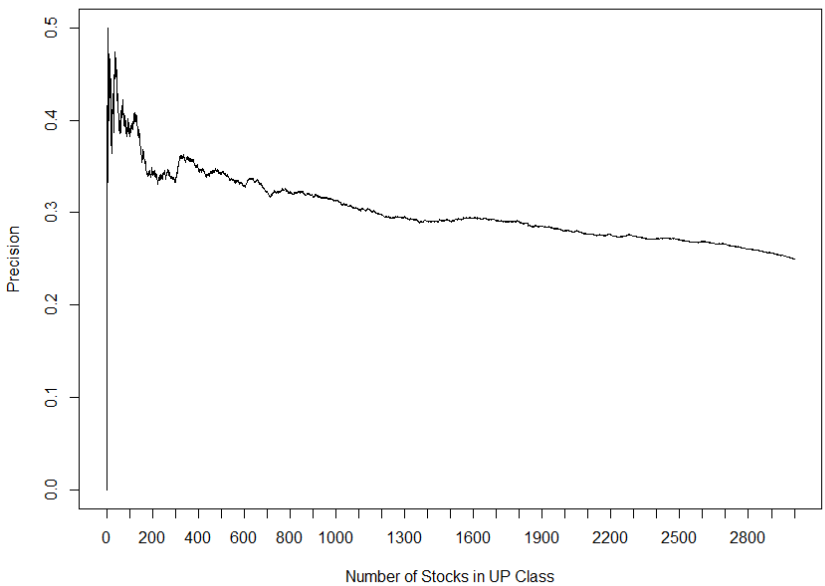

Figure 4 Precision by Number of Stocks Selected as UP Class

Precision continuously increases as we increase predicted probability cutoff point for UP class stocks. At all points except when only two stocks are classified falsely as UP class, precision is higher than the base rate of 25%.

6.2 Simulated Investment

In order to test practical applicability of our model in finance, we construct a simulation of investments with stocks selected according to predicted probabilities produced by SE model. Although precision rate shows to increase at higher cutoff points, the simulated portfolio invests equally in stocks that have predicted probabilities greater than the 75^{th} percentile of CV predicted probabilities in order to have sufficient amount of stocks per period (mean=92.625, min=23, max=145).

The models proposed in this dissertation do not give any information about optimal date of sales for selected stocks since date of sales for our modified return-on-invest vary per observation. While there are more sophisticated methods to calculate optimal sales date for stocks reviewed in section 2.1, profit maximization using other data sources or investment strategies is out of the scope of this paper. Therefore, the simulated portfolio buys stocks at the date of publication and sells exactly 90 days after purchase date.

In order to more clearly compare quarterly earnings of our constructed portfolio against average quarterly S&P 500 returns, we assume that 10-Q reports are reviewed and corresponding stocks are considered to have been purchased at the beginning and sold at the end of respective quarter. The

results of investment changes for compound returns to portfolio since purchase and sales date differ per each observation. Results of the simulation are shown in Table 9:

	Simulated Earnings	S&P 500 Average Earnings
2016 Q1	10.77%	02.81%
2016 Q2	07.64%	-01.48%
2016 Q3	00.28%	04.14%
2016 Q4	12.28%	05.78%
2017 Q1	02.35%	03.73%
2017 Q2	00.79%	03.43%
2017 Q3	02.79%	02.45%
2017 Q4	09.43%	05.97%
Compounded	55.78%	29.98%

Table 9. Results of Simulated Investment

The simulated portfolio earned positive average returns for all eight quarters. However, in 2016 Q3, 2017 Q1, and 2017 Q2, SE model failed to outperform the S&P 500 average earnings. The difference between earnings are minor in 2017 Q1 than in 2017 Q2 and 2016 Q3. In 2016 Q3 and 2017 Q2, S&P 500 average modified return on investment was considerably smaller than other

periods, with only 13.91% and 14.29% of the stocks classified as UP in respective subsets. However, there is no noticeable difference in S&P average earnings. Compounded return on portfolio outperformed compounded return on S&P 500 averages by two folds.

Table 10 reports the number of selected stocks and percentage of stocks that had more than 10% gains within the simulated portfolio and within S&P 500:

	Number of Stocks	Portfolio	S&P 500
2016 Q1	23	56.52%	27.35%
2016 Q2	79	30.38%	09.98%
2016 Q3	133	16.54%	26.15%
2016 Q4	145	51.72%	31.08%
2017 Q1	29	27.59%	29.06%
2017 Q2	72	18.06%	25.30%
2017 Q3	139	26.62%	21.12%
2017 Q4	121	46.28%	32.60%
Overall	741	34.21%	25.33%

Table 10. Percentage of Outperforming Stocks

Overall, percentage of selected stocks that had more than 10% gains within the portfolio was higher than the percentage of the outperforming stocks in for

all S&P 500. In all cases except 2016 Q3, 2017 Q1, and 2017 Q2, simulated portfolio outperformed S&P 500 average.

6.3 Top Ten Topics

This section presents top ten topics used in building the prediction models. Since SE model does not provide analysis of dependent variables, we instead use GBM model, which outperformed other models for both CV and test observations. Top ten topics used in GBM model are presented in Table 11.

Rank	Top words	Variable Importance (%)
1	Adversely, Affect, Law	0.0806
2	Energy, Power, Electric	0.0612
3	July, Goodwill, Intangible	0.0601
4	Network, Technology, Digital	0.0456
5	April, Incorporated, January	0.0410
6	Equipment, Receivables, Inventory	0.0403
7	Gas, Oil, Production	0.0394
8	Senior, Preferred, Loan	0.0359
9	Loan, Mortgage, Commercial	0.0331
10	Restructuring, Percent, Profit	0.0329

Table 11. Top Ten Topics Selected by Variable Importance

7. DISCUSSION

7.1 Model Performance

Overall, our results indicate that textual contents of 10-Q reports may be useful for modeling future stock prices. Combination of LDA and Stacked Ensemble recorded highest performance in out-of-sample AUC with 0.5878 out of all learning algorithms employed during the experiment. All representatives of the six algorithm with highest CV AUC recorded AUC over 0.5 (mean=0.5666, std=0.0121) and precision rate showed linear relationship with number of stocks selected according to predicted probabilities for UP class (Adjusted R^2=0.7851).

In Figure 5, we present LDA probabilities of top 10 topics for predicted UP class in our simulated investment portfolio. Since our model has been built using cross-validation for generalization, there should not be a considerable difference in topics used in false positive stocks and true positive stocks. As such, there is no topics distinguishably used more for false positive stocks. However, topic 6, with top words of equipment, receivables and inventory, performed exceptionally well for positively identifying UP Class within our

portfolio.

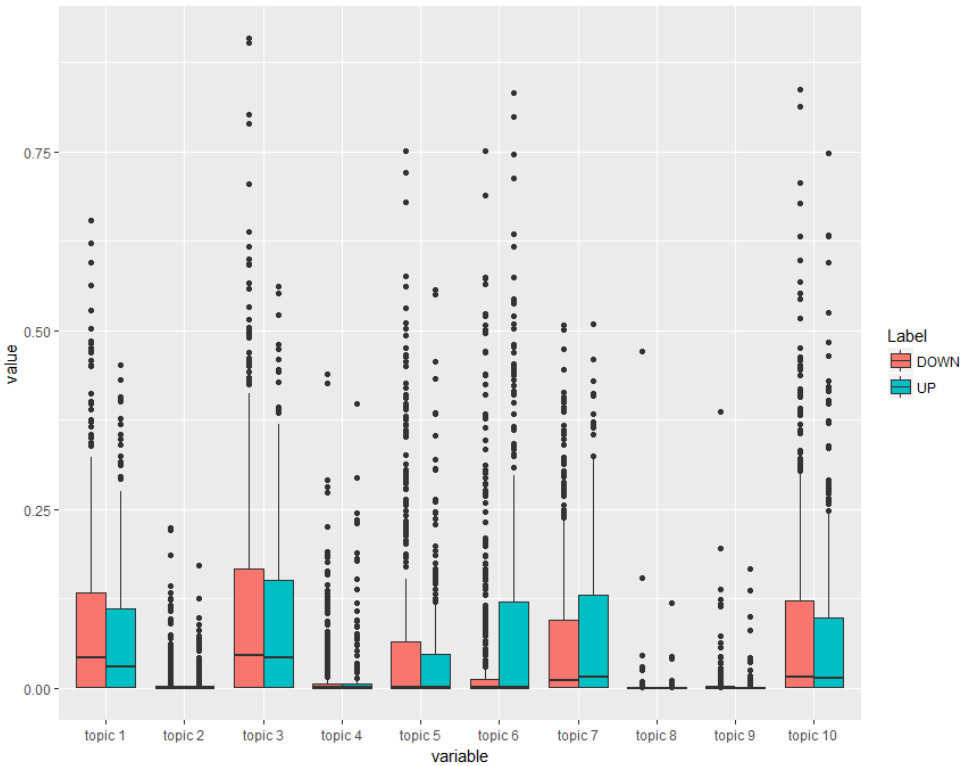

Figure 5. Topic Distribution of Predicted UP Class in Portfolio

In terms of publication date, our model recorded positive AUC in all eight subsets, but saw relative low simulated returns in 2016 Q3, 2017 Q1 and 2017 Q2. There are no direct, distinguishable patterns between simulated returns and number of selected stocks in portfolio, or between simulated portfolio earnings and average S&P 500 earnings (Table 10).

In all three time subsets when our simulated portfolio performed worse than S&P 500 average, there is a low ratio of UP class in proposed price performance classification (Table 9). Similarly, our simulation performed well in 2016 Q1, 2016 Q4 and 2017 Q4 when percentage of UP class is high within the subset. The relationship becomes clearer with respect to average S&P 500 earnings. Percentage of outperforming stocks increases when ratio between percentage of stocks classified as UP class and percentage of outperforming stocks (accuracy of performance measurement) is similar, as shown in Figure 6:

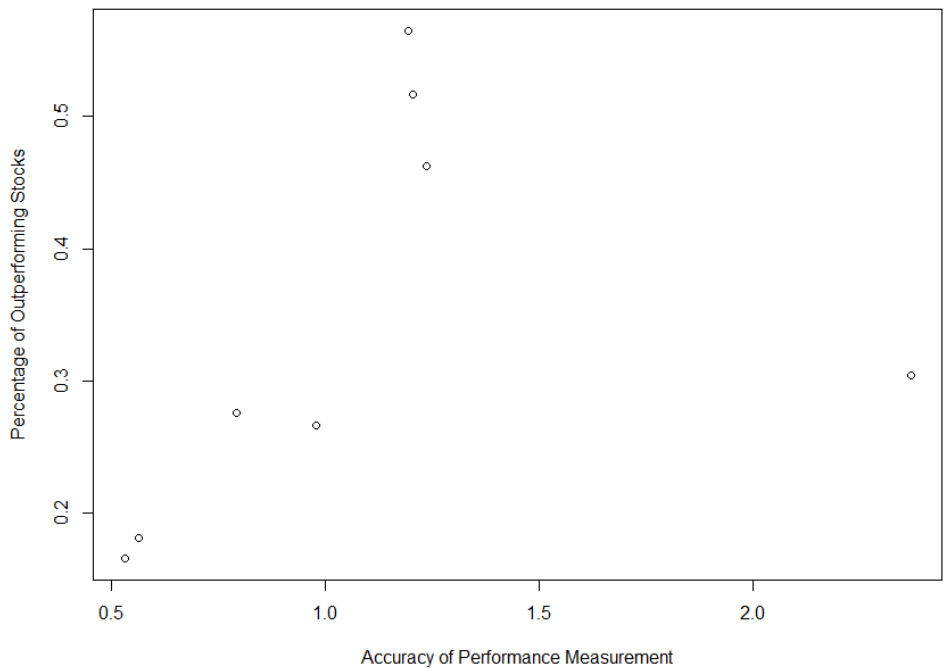

Figure 6. Outperforming Stock Percentage per Measurement Percentage

Since our model performance seems to be dependent on accuracy of our defined performance measurement on the actual price change within the observation period, this may indicate that our performance measurement does not always represent trend in stock prices accurately. Moreover, the results of simulation may improve by adopting methods to correctly identify optimal sales date for stocks that is more compatible with our proposed performance measurement. However, we found no relationship between direct representation of stock performance using price at publication and price after

90-days, with out-of-performance SE model AUC of 0.4880.

There may be other ways to better represent stock trends to improve model accuracy. For example, performance may increase by a more thorough assessment of the parameters arbitrarily set in the experiment. We chose to use an arbitrary span of 90-days as our observation period days but have not reviewed selecting different observation period for our proposed performance measurement. 10-Q corpus may be more effective in modeling short-term price trends or prices a year after publication. This dissertation chose to use 90-days as re-evaluating stocks every 90 days will be most natural for building real-life investment portfolio since 10-Q reports are published roughly every three months.

Our experiment also used an arbitrary cutoff point of 25% for differentiating between UP class and DOWN class. There may be ways to maximize model performance by adopting different cutoff point or representation methods such as multinomial classification. We have chosen binomial classification in this paper since the primary focus of this paper is to utilize textual contents in 10-Q reports for regular investment strategies. However, 10-Q dataset may be more potent for predicting catastrophes using price representation of abnormal losses or predicting stocks that have not faced a significant gain or losses within the observation period.

Since the main objective of the proposed performance was to differentiate stocks with abnormal gains within the observed period, this dissertation simply chose 25% to represent high gains for the purpose of reviewing application of 10-Q corpus. We may also have utilized any other cutoff point high enough to distinguish strongest stocks, but review of optimal cutoff point and classification methods may be beneficial to maximizing performance of our proposed models.

There are other potential methods that may also have increase overall model performance and simulated earnings. First, since this thesis deals with long-term stock price movements and does not provide information about optimal time of sales for stocks, integration of technical and fundamental approach reviewed in section 2.1 for predicting next day trend may improve investment performance greatly. Previous attempts at integrating multiple data sources or fundamental and technical analysis have indeed been known to increase performance.

Second, we only used a handful algorithms for building prediction models. There are more recent state-of-art supervised learning algorithms that may have improved model performance, such as Recurrent neural network that can process sequences of variables (Medsker, 2001) or Convolutional neural

network that suffers relatively less from curse of dimensionality (LeCun, 1999).

However, the main objective of this experiment is to review application of textual contents in 10-Q reports on modeling future stock prices. Although there may be better approaches in maximizing performance 10-Q text data for stock prediction, performance maximization is out of the scope of this dissertation.

7.2 Generalizability

We tested our model against 3,000 out-of-sample stocks listed in S&P 500 index. While S&P 500 index is regarded as representatives of publicly traded companies, our prediction models may act differently for companies not listed in the S&P 500 index. We have assumed that S&P 500 companies are diverse enough for initial investigation of 10-Q text application since S&P 500 are often used in financial research (i.e. Lam, 2004; Schumaker, 2009), but the results may change with other publicly traded companies or stocks sold outside of United States.

The test-subset only covered a handful of periods that were used for evaluating effects of time using quarter subsets. Because electronic format of

10-Q reports was consolidated under EDGAR system only after 2002 (SEC, 2002), there are not enough 10-Q reports to extend the timeline of our model without decreasing our train subset even further. Once there are more 10-Q reports under EDGAR system, our experiment may be tested on longer periods to ensure out-of-sample performance.

7.3 10-Q Representation

In response to RQ2, we compared prediction performance of models that are based on bag-of-words and LDA representation of 10-Q corpus. All 6 stock price prediction models saw increase in model performance using LDA representation of 10-Q text compared using bag-of-words representation in out-of-sample testing. However, CV set AUC of some models using algorithms such as SE, DNN and GBM was higher in bag-of-words representation than in LDA representation. The results are in concordance with the assumption that bag-of-words are more prone to curse of dimensionality with sparse, high dimensional dataset such as our 10-Q corpus.

Nonetheless, there may be other methods to better utilize information in 10-Q texts. We have only reviewed two possible representations methods in our experiment, LDA and bag-of-words. While LDA is often used in text representation and bag-of-words as a basis model for comparison, prediction

performance may increase by adopting more state-of-art text representation algorithms like Word2vec (Lilleberg, 2015) or adaptation of sentiment analysis (Feldman, 2013).

There may also be ways to use texts more directly by building prediction models that better handle natural language. There have been special adaptations of both Recurrent Neural Network (Mikolov, 2010) and Convolutional Neural Network (Hu, 2014) that are specifically designed to utilize topical and semantic information by using the corpus directly.

Moreover, we used an arbitrary number of 50 topics. While we based the number of topics based on previous research in 10-K reports in our prediction models, it may not be the most optimal number to accurately represent all information in the 10-Q dataset. Nonetheless, we have shown that using LDA representation improves the quality of models in our experiment.

8. CONCLUSION

Natural language processing and topic modeling have been applied to various domains like news and social network services for modeling future stock prices. While there are many viable information sources for company analysis, SEC's quarterly reports have been one of the most important direct channel of communication between firms and investors. However, it has largely been overlooked on assessing firm performances as they are not published as frequently as other domains. However, 10-Q reports have few advantages in predicting future stock prices. It has more homogeneity in information content, reviews future-oriented pre-defined set of topics, and is written directly by the company instead of speculators.

Therefore, the research goal of this paper is to review possibility of modeling stock prices using textual contents of 10-Q reports and application of topic modeling on evaluating 10-Q corpus. In order to test our research questions, we used LDA on recently added texts as our observations. Because 10-Q corpus has already been proved to be useful for predicting future events, we then modeled our dataset against a modified return on investment, directed to represent sudden changes in stock prices due to announcement of events. Then, we built prediction models using six algorithms using random grid

search of hyper parameters and cross-validation exclusively on the train subset. The resulting models were evaluated on test subset for measuring out-of-sample performance.

Stacked Ensemble model on LDA corpus achieved best performance compared to bag-of-words, modified return-on-investment dataset or LDA, 90-day return dataset. We then assessed potential investment returns by diving our test set into 8 different period subsets. In 5 out of 8 quarters, LDA-SE model outperformed overall S&P 500 average, and obtained positive returns in all 8 quarters. The compounded interest rate in our investment portfolio was significantly higher at 55.78% compared to 29.98% of S&P 500 average.

The results of our experiment indicated few characteristics of our approach. There was no conceivable explanation of errors within the prediction model when comparing distribution of topics with top ten variable importance in true positive observations and false positive observations or when comparing period subsets using number of selected stocks per period. However, we were able to attribute errors in simulated investment portfolio to difference in ratios of our proposed return-on-investment and the 90-day return on our stocks. Investment portfolio did well when our measurement closely reflected the true 90-day returns of selected stocks, but performed relatively poorly otherwise.

Adaptation of machine learning algorithms in finance is continuously being researched as state-of-art machine learning algorithms open up ways to incorporate new datasets and model changes in price trends. There are already a number of investment portfolios that strictly rely on artificial intelligence to pick optimal stocks (e.g., AIEQ, 2018; Robo Global, 2018.) and both academic and field researchers are continuously concocting new ways to analyze companies and stocks. This paper contributes to the growing body of research in finance and machine learning by reviewing the application of previously overlooked 10-Q reports as a possible source of information for stock market research through machine learning, proposing a different approach to representing future stock prices, and reviewing application of LDA on analyzing 10-Q text corpus.

BIBLIOGRAPHY

AIEQ. (2018). ABOUT AIEQ – The AI Powered Equity ETF. Retrieved from http://www.equbotetf.com/about-aieq/

Arun, R., Suresh, V., Madhavan, C. V., & Murthy, M. N. (2010, June). On finding the natural number of topics with latent dirichlet allocation: Some observations. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 391-402). Springer, Berlin, Heidelberg.

Atsalakis, G. S., Dimitrakakis, E. M., & Zopounidis, C. D. (2011). Elliott Wave Theory and neuro-fuzzy systems, in stock market prediction: The WASP system. Expert Systems with Applications, 38(8), 9196-9206.

Balsam, S., Bartov, E., & Marquardt, C. (2002). Accruals management, investor sophistication, and equity valuation: Evidence from 10–Q filings. Journal of Accounting Research, 40(4), 987-1012.

Bellman, R. E. (1957). Dynamic programming. Princeton, NJ: Princeton University Press.

Bellman, R. (1961). Curse of dimensionality. Adaptive control processes: a guided tour. Princeton, NJ.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305.

Bodnaruk, A., Loughran, T., & McDonald, B. (2015). Using 10-k text to gauge financial constraints. Journal of Financial and Quantitative Analysis, 50(4),

623-646.

Bondt, D., Werner, F. M., & Thaler, R. H. (1987). Further evidence on investor overreaction and stock market seasonality. The Journal of finance, 42(3), 557-581.

Bonsall IV, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. (2017). A plain English measure of financial reporting readability. Journal of Accounting and Economics, 63(2-3), 329-357.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal

of machine Learning research, 3(Jan), 993-1022.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.

Bradshaw, M. T. (2004). How do analysts use their earnings forecasts in generating stock recommendations?. The Accounting Review, 79(1), 25-50.

Breiman, L. (1996). Stacked regressions. Machine learning, 24(1), 49-64.

Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. Review of Accounting Studies, 19(1), 396-455.

Campbell, K., Gordon, L. A., Loeb, M. P., & Zhou, L. (2003). The economic cost

of publicly announced information security breaches: empirical evidence from

the stock market. Journal of Computer Security, 11(3), 431-448.

Cai, Z., Li, H., Hu, X., & Graesser, A. (2016). Can Word Probabilities from LDA

be Simply Added up to Represent Documents?. In EDM (pp. 577-578).

Cao, J., Calderon, T., Chandra, A., & Wang, L. (2010). Analyzing late SEC filings for differential impacts of IS and accounting issues. International Journal of Accounting Information Systems, 11(3), 189-207.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work:

Using financial text as a predictor of financial events. Decision Support Systems,

50(1), 164-175.

Daepp, M. I., Hamilton, M. J., West, G. B., & Bettencourt, L. M. (2015). The mortality of companies. Journal of The Royal Society Interface, 12(106), 20150120.

Donoho, D. L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. AMS Math Challenges Lecture, 1, 32.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015, July). Deep learning for event-driven stock prediction. In Ijcai (pp. 2327-2333).

Doidge, C., Karolyi, G. A., & Stulz, R. M. (2017). The US listing gap. Journal of

Financial Economics, 123(3), 464-487.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. Journal of Animal Ecology, 77(4), 802-813.

Esser, F. (1999). Tabloidization'of news: A comparative analysis of Anglo-American and German press journalism. European journal of communication, 14(3), 291-324.

Fama, E. F. (1965). The behavior of stock-market prices. The journal of Business,

38(1), 34-105.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.

Griffin, P. A. (2003). Got information? Investor response to Form 10-K and Form

10-Q EDGAR filings. Review of Accounting Studies, 8(4), 433-460.

Hamilton, J. T. (1995). Pollution as news: media and stock market reactions to the toxics release inventory data. Journal of environmental economics and management, 28(1), 98-113.

Helwege, J., & Packer, F. (2003). The decision to go public: evidence from mandatory SEC filings of private firms. Fisher College of Business, Ohio State University.

Hu, B., Lu, Z., Li, H., & Chen, Q. (2014). Convolutional neural network architectures for matching natural language sentences. In Advances in neural information processing systems (pp. 2042-2050).

Ince, H., & Trafalis, T. B. (2017). A HYBRID FORECASTING MODEL FOR STOCK MARKET PREDICTION. Economic Computation & Economic Cybernetics Studies & Research, 51(3).

Kim, Y., Jeong, S. R., & Ghani, I. (2014). Text opinion mining to analyze news for stock market prediction. Int. J. Advance. Soft Comput. Appl, 6(1), 2074-8523.

Koh, J., & Venkatraman, N. (1991). Joint venture formations and stock market reactions: An assessment in the information technology sector. Academy of Management Journal, 34(4), 869-892.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely Randomized Trees. Machine learning, 63(1), 3-42.

Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. Expert systems with Applications, 33(1), 171-180.

Ho, T. K. (1995, August). Random decision forests. In Document analysis and recognition, 1995., proceedings of the third international conference on (Vol. 1, pp. 278-282). IEEE.

Huang, K. W., & Li, Z. (2011). A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. ACM Transactions on Management Information Systems (TMIS), 2(3), 18.

Ibrahim, R. (2016). The number of publicly-traded US companies is down 46% in the past two decades. Retrieved from https://finance.yahoo.com/news/jp-

startup-public-companies-fewer-000000709.html

Jordan, D. J., & Diltz, J. D. (2003). The profitability of day traders. Financial Analysts Journal, 59(6), 85-94.

Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990, June). Stock market prediction system with modular neural networks. In Neural Networks, 1990., 1990 IJCNN International Joint Conference on (pp. 1-6). IEEE.

Lam, M. (2004). Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Decision support systems, 37(4), 567-581.

Lawrence, A. (2013). Individual investors and financial disclosure. Journal of Accounting and Economics, 56(1), 130-147.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999). Object recognition with

gradient-based learning. In Shape, contour and grouping in computer vision (pp.

319-345). Springer, Berlin, Heidelberg.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436.

Lee, H., Surdeanu, M., MacCartney, B., & Jurafsky, D. (2014, May). On the Importance of Text Analysis for Stock Price Prediction. In LREC (pp. 1170-1175).

Leu, Y., Lee, C. P., & Jou, Y. Z. (2009). A distance-based fuzzy time series

model

for exchange rates forecasting. Expert Systems with Applications, 36(4), 8107-

8114.

Li, H., Pincus, M., & Rego, S. O. (2008). Market reaction to events surrounding the Sarbanes-Oxley Act of 2002 and earnings management. The Journal of Law and Economics, 51(1), 111-134.

Lilleberg, J., Zhu, Y., & Zhang, Y. (2015, July). Support vector machines and word2vec for text classification with semantic features. In Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on (pp. 136-140). IEEE.

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. Financial Review, 40(1), 1-9.

Malkiel, B. G., & McCue, K. (1985). A random walk down Wall Street (Vol. 8).

New York: Norton.

Medsker, L. R., and L. C. Jain. "Recurrent neural networks." Design and Applications 5 (2001).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Murphy, J. J. (1999). Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. Penguin.

Oberlechner, T. (2001). Importance of technical and fundamental analysis in the European foreign exchange market. International Journal of Finance & Economics, 6(1), 81-93.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a Random Forest?. In International Workshop on Machine Learning and Data Mining in Pattern Recognition (pp. 154-168). Springer, Berlin, Heidelberg.

Pai, P. F., & Lin, C. S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497-505.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259-268.

Porter, M. (2001). Snowball: A language for stemming algorithms. Retrieved from http://snowball.tartarus.org/texts/introduction.html

Porter, M. (2001). English Stop Words list. Retrieved from http://snowball.tartarus.org/algorithms/english/stop.txt

Qiu, X. Y. (2007). On building predictive models with company annual reports. The University of Iowa.

Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, pp. 133-142).

Rechenthin, M., Street, W. N., & Srinivasan, P. (2013). Stock chatter: Using stock sentiment to predict price direction. Algorithmic Finance, 2(3-4), 169-196.

Robo Global. (2018). About Us - ROBO Global. Retrieved from https://www.roboglobal.com/about-us/

Schneider, C. W. (1972). Nits, Grits, and Soft Information in SEC Filings. U. Pa. L. Rev., 121, 254.

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems (TOIS), 27(2), 12.

Seligman, J. (1994). The SEC's Unfinished Soft Information Revolution. Fordham L. Rev., 63, 1953.

Securities and Exchange Committee. (2018). Form 10-Q. Retrieved from https://www.sec.gov/files/form10-q.pdf

Securities and Exchange Committee. (2002). SEC Announces Free, Real-Time Public Access to EDGAR Database at. www.sec.gov Retrieved from https://www.sec.gov/news/press/2002-75.htm

Securities and Exchange Committee. (2018). EDGAR Filer Manual (Volumes I - III). Retrieved from https://www.sec.gov/info/edgar/edmanuals.htm

Sehgal, V., & Song, C. (2007, October). Sops: stock prediction using web

sentiment. In Data Mining Workshops, 2007. ICDM Workshops 2007. Seventh IEEE International Conference on(pp. 21-26). IEEE.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.

Strategy&. (2018). Considering an IPO? Retrieved from https://www.strategyand.pwc.com/media/file/Strategyand_Considering-an-IPO.pdf

Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. The Journal of Business Communication (1973), 30(1), 49-61.

Sullivan, R., Timmermann, A., & White, H. (2001). Dangers of data mining: The case of calendar effects in stock returns. Journal of Econometrics, 105(1), 249-286.

Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. Statistical applications in genetics and molecular biology, 6(1).

Vu, T. T., Chang, S., Ha, Q. T., & Collier, N. (2012). An experiment in integrating sentiment features for tech stock prediction in twitter.

Weng, B., Ahmed, M. A., & Megahed, F. M. (2017). Stock market one-day ahead

movement prediction using disparate data sources. Expert Systems with Applications, 79, 153-163.

Woolridge, J. R., & Snow, C. C. (1990). Stock market reaction to strategic investment decisions. Strategic management journal, 11(5), 353-363.

You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. Review of Accounting Studies, 14(4), 559-
586.

Zhai, Y., Hsu, A., & Halgamuge, S. K. (2007, June). Combining news and technical indicators in daily stock price trends prediction. In International symposium on neural networks (pp. 1087-1096). Springer, Berlin, Heidelberg.

Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. arXiv preprint arXiv:1212.5701.

국 문 초 록

최근 인공 지능이 급격히 발달하면서 기계 학습을 통한 주가 예측에 대한 연구가 새로이 급부상하고 있다. 허나, 장기 주가 예측은 주가를 결정하는 여러 외부적 요인들과 효율적 시장 가설에 의해 상대적으로 적은 수의 연구가 진행되어 왔다. 기업분석을 통한 예측모델을 성공적으로 연구한 이전의 문헌들은 대부분 뉴스나 게시판 또는 소셜 네트워크 서비스의 본문을 이용해 연구를 진행해왔지만 본 연구는 기존에 사용된 사건 위주의 정보는 한계가 있다는 가정하에 미국 주식회사들이 증권거래위원회를 통해 기재하는 10-Q 분기별 보고서를 토대로 연구를 진행하였다. 분기별 보고서는 다른 문서들에 비해 주제가 유사하고, 미래지향적이며 정확하다는 장점이 있어 장기주가예측연구에 비교적 적합하다고 볼 수 있다.

따라서 본 연구는 10-Q 본문의 장기적 기업가치 분석에 대한 정보 유용성을 검토하기 위해 스탠더드 앤 푸어스 500 총 평균 주가지수(S&P 500)에 동원된 회사들의 보고서를 수집했다. 2004년 1월부터 2018년 1월까지 기재된 10-Q 보고서들을 전수 수집하고, 기계 학

습과 자연어 처리를 통해 정보유용성을 검토했다. 총 18,237건의 문서들이 수집되었으며, 표본 외 정확성 평가와 분석을 위해 2015년 12월과 2018년 1월 사이 공개된 3,000편의 문건과 모델 구축을 위한 11,132편의 문건으로 다시 나누었다.

또한, 장기적 주가에 영향을 미치는 방해 요인을 최소화하기 위해 본 연구는 두 가지 연구방법 체제를 제안하였다. 첫째, 정보의 효율과 균등성을 위해 각 10-Q 문헌에 새로이 기재된 단어들을 축출하였으며, 잠재 디리클레 할당(Latent Dirichlet Allocation) 문서 분석 방식을 통해 정보 표기 방식을 재정립하였다. 둘째, 기존 연구들은 미리 지정된 날짜를 통해 주가변동 수치를 계산하였으나, 본 연구는 정보의 정확한 변동 날짜를 유추하기 어렵다는 가정하에 정해진 기간 내에 가장 높은 주가변동 수치를 기반으로 높은 이익 잠재 가능성을 가진 주식과 아닌 주식들로 이항 분류법(binary classification)을 정립하였다.

마지막으로, 구축된 데이터에 Stacked Ensemble, Random Forest, Extremely Randomized Trees, Generalized Linear Regression, Gradient Boosting Machine, Deep Feed-Forward

Neural Network 등 6가지 기계학습 기법을 적용한 모델들을 구축하여 3,000편의 문건들을 통해 모델들을 평가한 결과를 요약하자면 다음과 같다.

1) 잠재 디리클레 할당, Stacked Ensemble, 본문에서 제안된 주가변동 수치 표기 방식 등을 사용한 모델이 가장 높은 AUC (area under the Receiver Operating Characteristic curve)인 0.5878을 기록하였다.
2) 잠재 디리클레 할당 문서 표기 방식(0.5878 AUC)이 bag-of-words 기법보다(0.5727) 살짝 높은 성과를 보였다.
3) 기존 연구 방식인 정해진 기간을 통한 주가변동 수치는 예측할 수 없었다.
4) 모의투자 포트폴리오는 2년이란 실험 기간 내에 S&P 500 평균치인 29.98%보다 높은 55.78%의 수익률을 기록하였다.
5) 모의투자 포트폴리오는 제안된 주가변동 표기 방법과 90-일 이후 변동 수치가 비슷할수록 높은 수익률을 보였다.

본 연구는 10-Q 재무보고서를 장기 주가 변동 예측의 새로운 데이터로 도입 해봄과 함과 동시에, 10-Q 본문의 접근방식과 새로운 주가 표기 방식을 제안함으로써 기계 학습을 통한 금융 연구에 기여한다. 또한 이를 통해 장기주가예측모델의 가능성을 새로이 검토하였다.

하지만 본 연구는 기존에 사용되지 않은 접근방식을 채용함에 따라 몇 가지 한계점이 존재한다. 우선 현재까지 기재된 문건들을 활용하였으나 좀 더 긴 기간을 두고 S&P 500에 속하지 않은 여러 공개기업의 문서들을 통해 범용성과 유용성을 더 검토해야 한다. 또한 문서분석방식이나 예측모델 구축에 있어 제한된 알고리즘만을 사용하였으므로 더 많은 학습방식을 통해 연구를 계속해야 할 것으로 보인다. 그러나 본 연구는 10-Q 보고서라는 새로운 데이터의 활용 가능성과 장기주가예측 모델의 가능성을 검토하여 기계 학습을 통한 장기주가예측 연구의 초석을 제공한 것으로 볼 수 있다.

주요어 : 기계학습, 10-Q 재무보고서, 잠재 디리클레 할당, 자연어처리, 스택 앙상블

학 번 : 2014-24825