Maja Z. Olsen & Sindre F. Fjellavli

# Digging for Returns: Can Text Mining Improve Equity Return Predictions?

Master's thesis in Industrial Economics and Technology Management
Supervisor: Stein-Erik Fleten & Peter Molnár
June 2020

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

NTNU
Norwegian University of
Science and Technology

Master's Thesis

# Digging for Returns: Can Text Mining Improve Equity Return Predictions?

Maja Z. Olsen        Sindre F. Fjellavli

**Norwegian University of Science and Technology, Department of Industrial Economics and Technology Management**

**Abstract** – **From examining context of news stories to looking at annual reports - their timing and content or investigating the effect of sentiment in text, text mining is gaining traction in financial applications. This study examines the joint effect of information gathered from news stories, company documents and financial factors on share returns. Using a sample of more than 13,000 public corporate documents, over 300,000 news stories and financial data across 21 years, we find that it is worthwhile including textual factors in combination with financial factors, and that it is possible to make abnormal return on the information contained in news stories and corporate documents. We train a neural network that obtains an out-of-sample prediction accuracy higher than the model using either textual or financial information, and a subsequent long/short portfolio that achieves returns of 9.4 times invested amount over a six-year period.**

**Keywords** : **Stock price prediction, decision support, textual analysis**

## 1 Introduction

Asset price prediction is a recurring topic in financial literature, with voluminous research seeking to explain asset price developments. However, asset price prediction is in no way limited to academia, it is a crucial topic in practice for both institutional and individual investors. Obtaining an accurate model for predicting future asset prices means being able to understand asset price fluctuations in terms of changes in relevant factors. An investor with an accurate model possesses a significant advantage in the stock market as compared to the multitudes of investors that flock to financial markets looking to earn an abnormal return. However, the large degree of uncertainty and volatility in the markets make predicting stock return a tedious task and many investors see their fortunes vanish into thin air.

A significant amount of literature emphasize the statistical behavior of time series [54], attempting to explain future return through historical behaviour. This branch of return forecasting is known as technical analysis. The field of asset price forecasting is, however, not confined to historical behaviour. Indeed, fundamental analysis has through several studies become an established part of return predictions in academia and in practice [12], and is arguably one of the most common ways of estimating share returns. Despite being widespread, intuitive and arguably easy to perform, there exists empirical evidence suggesting that large share price movements do not correspond with changes in fundamental factors [16] [58] [59]. Such findings suggest that researchers and investors will benefit from expanding their models beyond incorporating historical and fundamental factors. One way to do this is through including text mining and subsequent textual analysis. Textual analysis allows researchers to quantify informational sources including news stories and corporate documents, and estimate their effect on asset price developments. This method has made its inroads into stock price predictions and correspondingly it has made its way into literature within financial predictions [31]. Within this area, sentiment analysis allowing investors to tap into the market sentiment has grown increasingly prominent [11], with results mainly suggesting that negative words have a relationship with company value [51].

When valuing a company, the amount of information available is immense and this great pool of information will pose as a double edged sword for many investors. An extensive information environment reduces information asymmetry between investors, resulting in increased trading volumes [10] [30]. On the other hand, if the pool of information becomes too vast, some investors will find it challenging to access and understand all the information available [3]. Automated solutions have made successful inroads into financial research, enabling efficient collection and debunking of public information. Accordingly, automated solutions may allow for reduced information asymmetry when made available to several investors.

In terms of public information, as indicated above, a large degree of the information available is in the form of text as opposed to financial data. Despite being more challenging and costly to retrieve and analyze than financial data [47], textual analysis is gaining significant traction within financial applications. Recent literature supports this shift with several studies indicating that it is insufficient to explain stock prices only in terms of typical quantitative data [1] [16] [46]. Researchers including Abrahamson and Amir [1] and Cutler [16] argue that the solution is including non-conventional measures and textual data, whilst Henry [34] provide evidence that including verbal predictor variables improve prediction accuracy. As further explained below, it also allows for a more thorough examination of the Efficient Market Hypothesis.

In line with the Efficient Market Hypothesis (EMH), as new information is made available to investors an updated market equilibrium should emerge, eliminating any arbitrage opportunities. This hinges on the investors, who are thought to be fully rational therefore ensuring that stock prices correctly reflect all publicly available information. Assuming that this is not the case, and that the market participants instead struggle with limited attention and limited processing power, as Hirshleifer and Teoh argue [36], we can test the validity of the Efficient Market Hypothesis. It is worthwhile to include textual information when assessing the Efficient Market Hypothesis, as, according to the hypothesis, all information incorporated in news and corporate documents should be unbiasedly reflected by the stock price [35] immediately after publication. By the time the investor reads the report it should be too late to make abnormal returns on the information.

Aside from testing the Efficient Market Hypothesis, another motivation for this research relates to the fact that previous literature in this field is often limited to specific documents, such as earnings announcements, annual reports or news stories [59]. By including both news stories and various types of firm-related documents, we are able to analyze an extensive set of events and textual descriptions. Combining this with financial values allows us to look for patterns across a variety of different data sources. The technical difficulty of the study lies in the extensive and meticulous scraping work used to compose the data set, and the consequent extraction of sentiment from textual sources.

We chose the Oslo Stock Exchange for a number of reasons. Internationally, Oslo Stock Exchange is viewed as a leading exchange for segments including energy, shipping and seafood. However, we find that previous research is in large concentrated around the US market. In fact, to our knowledge, there has not been conducted any similar studies in Norway. By examining the effect of textual factors on share return on the Oslo Stock Exchange we are able investigate whether the results from similar international studies holds up in a Nordic exchange. It is furthermore likely that the usage of textual analysis for share price prediction in practice is less common in Norway. As a result, it should be easier to examine the informational value of textual sources.

In this study, the effect of including quantified textual information in stock predictions is investigated by creating a neural network running on news stories and corporate documents in conjunction with financial values. Our study is distinguished from similar research in many ways. Firstly, by scraping various sources in order to collect both news stories and a variety of corporate documents we produce an extensive data set. Secondly, by focusing on the Oslo Stock Exchange, we investigate the use textual information for stock return prediction on a market presumably less impacted by such analyses. From this, we interpret the importance of the different features included in the analysis. We furthermore interpret the explanatory power of the the textual information by constructing a portfolio of stocks to buy and sell. Collectively we find that the textual information do contain valuable information. We obtain a higher prediction accuracy when including financial and textual information, followed by a portfolio that returns 9.4 times invested amount.

The remainder of the article is organized as follows. In Section II, relevant background is introduced with a corresponding literature review. In Section III, data sources and variables are presented, followed by an explanation of the methods used in Section IV. We continue by presenting and discussing the empirical findings in Section V combined with a discussion of the findings. Finally, Section VI concludes the paper.

## 2 Theoretical Background and Literature Review

Textual analysis plays an important role for investors and analysts alike when valuing a company. In terms of previous literature, textual analysis of firm-related texts has shown promising results when it comes to predicting stock returns [31]. Within textual analysis, Kearney and Liu distinguish between three sources of information researchers rely on when performing sentiment analysis in finance; corporation-expressed sentiment, media-expressed sentiment and Internet-expressed sentiment [40].

Corporation-expressed sentiment originate from corporate disclosures, and are in large part centered around financial reports, particularly the MD&A section from US 10-Ks [1] [25] [47]. Collectively, the research suggests that including textual information from corporate documents is worthwhile. Abrahamson and Amir [1] found that the content of the president's letter contain information about the future of the company, and that the negativity holds particular explanatory power. Feldman *et al.* examined the MD&A section of corporate disclosures, indicating that the market reaction is largely associated with the overall tone for a short duration post publishing [25], whereas Li document a relationship with the risk sentiment in annual reports and future earnings [47]. On a whole, the literature suggest that negative sentiment have more explanatory power for future return [1] [21] [47] [59].

Media-expressed information is based on news stories, and a lot of the prominent research in this area is concentrated around US news outlets [31]. Within media-expressed information, Cutler, Poterba and Summer's work on economic news and their effect on stock return is viewed as pioneering [16]. They estimated the fraction of variance in aggregated stock returns that could be explained by macroeconomic news. Since then, there has been several studies on the impact of different media-expressed information. Tetlock *et al.* focused instead on firm-specific news from the Wall Street Journal and Dow Jones News Service attempting to explain individual firm's earnings and share performance. Their research provide evidence that the fraction of negative words can help forecast low firm earnings, and that the market is slow to react on the information reflected by the negative words [59]. Studying short sales and news releases from the Dow Jones, Engelberg *et al.* find evidence that short sellers are apt at processing information contained in news events, providing them with a trading advantage [21].

Lastly, Internet-expressed information is based on posts published by individuals online. Much of the prominent research within this area shows that sentiment extracted from stock message boards holds no effect [5] [60] or only a small effect on share returns [17]. Relying on Yahoo!, Das and Chen performed a sentiment analysis based on investor sentiment from stock message boards. Their research document a weak effect for individual stocks, but show that the aggregate sentiment for tech stocks holds a higher predictive power [17]. As opposed to Das and Chen, Antweiler and Frank find evidence that stock messages help predict market volatility, but the economic effect on stock returns is small [5]. Due to ambiguous results in related research in combination with a lack of historical data from Norwegian stock message boards, Internet-expressed information is not included in this study.

Research in the field of textual analysis is mainly centered around dictionary-based or machine learning approaches. The dictionary-based method has foundation in several studies, and is mainly concentrated around two major dictionaries; the Harvard IV Dictionary and Loughran and McDonald's Financial Sentiment Dictionary. Tetlock *et al.* [59] and Lillo *et al.* [49] document the usefulness of the Harvard IV Dictionary in language analysis for predicting stock market performance. Both studies investigate how firm-specific news stories affect stock return by looking at the constituents of the S&P 500 index and Nokia, respectively. Comparatively, Loughran and McDonald show that the Harvard IV Dictionary is somewhat unspecific in a financial context, with three-fourths of negative words in the Harvard IV Dictionary at risk of being misclassified [51]. Words such as *liability*, *capital* and *foreign* hold a negative value according to the Harvard IV Dictionary, but in a financial context these are frequently used words that, without any other context, should hardly be classified as negative. Based on the findings, Loughran and McDonald construct a financial sentiment dictionary aiming to better reflect the tone in financial texts. Using this dictionary on U.S. banks ' annual reports, Gandhi *et al.* find that the frequency of negative words used in annual reports can help predict bank distress [26].

Our research is most closely related to previous work by Davis *et al.* and Tetlock *et al.* Davis *et al.* investigate the relationship between qualitative information from earnings releases and share returns after controlling for various financial factors. By reviewing 24,000 quarterly earnings releases between 1998 and 2003, Davis *et al.* find evidence that the language used in earnings releases, either optimistic or pessimistic, do hold information about future stock performance. Our research differentiates from Davis *et al.*'s work in terms of the textual information included. We examine qualitative information from 14,800 corporate documents, including but not limited to earnings releases. In addition, we include over 300,000 news stories and look at quarterly returns over a time horizon of 21 years. Tetlock *et al.* use the Harvard IV Dictionary to quantify news stories, retrieving over 350,000 qualifying news stories from US news outlets. Controlling for financial variables such as earnings, size, book-to-market, trading volume and analyst forecasts, the researchers were able to document that the frequency of negative words can help forecast low firm earnings, and that the stock prices do not immediately reflect the information embedded in the news stories. Compared to Tetlock *et al.*, our study is more varied in terms of the textual sources included. In addition, we test explanatory power of positive sentiment, as opposed to only emphasizing the fraction of negative words present in news stories.

Our research adds to existing research in several ways. As mentioned, the study is based on an extensive set of textual sources and we test on the Oslo Stock Exchange. Comparatively, in including both qualitative and quantitative sources in order to predict the direction of share returns, this research differentiates itself from much of the literature within this field. By doing this, we increase the scope of the textual analysis as compared to Tetlock *et al.*, Hájek, and Lillo *et al.*, and we differentiate ourselves from the likes of Davis *et al.* by expanding beyond earnings press releases. We use a dictionary-based method to produce a sentiment score from annual reports, and a topic-based approach to understand news stories. An additional differentiating factor is that we gather several features from the annual reports; i.e. representation of women in the company and acquisitions. Lastly, we focus on quarterly returns. This is in line with Davis *et. al*, but is a longer time horizon than the studies conducted by Tetlock *et. al* and Hájek who focuses on daily and three-day returns, respectively.

## 3 Empirical Specifications and Data

The study encompasses the constituents of the Oslo Stock Exchange as of February 2020. This amounts to 246 companies. The study relies on various sources to produce the data sample;

- Thompson Reuters Eikon database for fundamental and technical factors
- Oslo Stock Exchange NewsWeb web page to retrieve news stories
- Company websites to access annual reports and other firm-related documents

Using the NewsWeb database managed by Oslo Stock Exchange we gathered news stories from March 18th 1998 to February 19th 2020. This amounts to more than 350,000 news stories. By excluding news stories that do not refer to active stocks as of February 2020, we have more than 300,000 qualifying news stories. These news stories consists of company announcements, and in line with the Securities Trading Act in Norway, all listed companies are obliged to publish notifiable inside information. Furthermore, according to Oslo Stock Exchange's web page, this information is published and stored in NewsWeb immediately upon publication. As a result, NewsWeb is viewed to be a comprehensive and viable source for news stories.

In retrieving company reports and presentations a three-layered scraping effort was performed. With a total of 246 firms, we have retrieved 14,814 documents resulting in 13,420 qualifying reports and presentations. Typically, these documents are pdfs that the firm has published under its investor relations pages. This is quite different to related research, which tends to focus on 10-Ks and then oftentimes extracting only the Management's Discussion and Analysis to be used in the analysis. When collecting firm-related documents from individual firms' web pages, we rely on companies to have an exhaustive repository of company documents since their initial public offering. By implementing a three-layered scraping solution, we are able to locate and obtain documents that are scattered across different URLs associated with each company's web page.

After having downloaded the corporate documents, we use Python to extract texts from the pdfs. This is a computationally complex task, and documents that are scanned as opposed exported to pdf are excluded as they cannot be read by the script. Reports and presentations consist of anything from one page to several hundred pages. Understandably, the textual analysis that ensues after extraction of text from pdf is lenghty and computationally intense. In performing the textual analysis, aside from gathering sentiment and other factors, we extract the quarter and year of publication. When the quarter and year is unattainable, the content of the pdf cannot be matched to financial factors and news factor. The pdf is rendered useless resulting in the exclusion of the report from the analysis. We furthermore test the language of the pdfs, finding that 11.2% of the pdfs are in other languages than English. These reports are however included as they will present reliable data for factors including readability, percentage women represented and acquisition focus.

The data is standardized before being included in the model, in addition to being adapted to the quarterly prediction horizon. The predictions are made 2 months into the next quarter, ensuring that the model does not predict using values that are not available at the time of prediction. Table 1 introduces and explains the variables used in the study.

**Table 1**: Variable definitions

| **Variable** | **Definition** | **Rationale** |
|---|---|---|
| | **Textual analysis** | |
| News sentiment | Sentiment score on news stories based on Loughran and McDonald Financial Sentiment Dictionary. We calculate positive and negative sentiment, as well as overall tone (sentiment polarity) | Many researchers have documented a correlation between sentiment in news stories and stock performance, including Tetlock *et al.* [59], Cutler *et al.* [16] and Engelberg *et al.* [21] |
| News context classification | Dummy variables indicating which company event has occurred. Events are shown in appendix | Allows us to capture context. Hájek show that capturing context in news stories result in a significant increase in prediction accuracy [31] |
| Sentiment on public corporate document | Sentiment score based on Loughran and McDonald Financial Sentiment Dictionary. We include the three sentiment factors: (i) positive sentiment, (ii) negative sentiment and (iii) sentiment polarity | Annual reports and other company documents is a way for firms to communicate corporate strategy to its stakeholders, and research indicates that the content of the reports have a correlation to stock performance [1] [41]. In line with related research, we utilize sentiment scores to quantify the qualitative textual information [26] [31] |
| Readability | We measure readability with the Gunning Fog Index. The index returns a number indicating the reading level as measured by an educational grade. $Gunning\ fog\ index = 0.4 * (words\ per\ sentence + percent\ of\ complex\ words)$ | As Bloomfield [9] describes in 2002, managers have incentive to obscure information that might adversely impact stock price, making it difficult for investors to uncover information. Li [48] tests this theorem by implementing the Fog index. By looking at MD&A of US companies, Li find that firms with annual reports with a higher readability have more persistent positive earnings. The Fog index is the most common measure for readability in financial research [52] |

| | | |
|---|---|---|
| % Female names | Using a proprietary database of Norwegian names we estimate the percentage representation of females in the company, as compared to males. We construct a variable that is calculated according to the formula below: $\%Females = \frac{name_{female}}{name_{female}+name_{male}}$ | Indicate how well the female gender is represented in the company. Several recent research is focused on the effect of having women in management and board positions, some research evidencing investor bias against women [19], whilst other research show that having women in boards or management positions improve corporate decisions [13] and/or stock return [13] [53]. As the names mentioned in corporate reports are mainly affiliated to management and board members, we expect this variable to, within reasonable bounds, reflect the amount of females in top positions. |
| Acquisition factor | Frequency analysis on acquisitions and related terms as a percentage of total number of words, excluding stop words | Acquisitions have been proven to have an effect on stock return. Kamaluddin *et al.* [39] documented a positive correlation between acquisitions and stock performance, whereas Alhenawi *et al.* [4] show that the effect is dependent on the type of acquisition, i.e. related or unrelated to the core business |
| | **Financial data** | |
| Dividend yield | Dividend as percentage of total revenues | Graham and Dodd stated in the 1930's that the stock market reacts favourably to high dividends, whilst Modigliani and Miller argued in the 1960's that a firm's dividend policy is irrelevant on its value [56]. Dividend policy is understandably a contentious topic in financial research, with many researchers attempting to understand the effect of dividend payments on firm value. Researchers including Erasmus [22], Henne *et al.* [33] and Lintner [50] argue that dividends have an impact on share returns |
| Dividend stability | Dividend stability is measured as point change in dividend yield for the past period to the current | Several studies provide evidence that dividend yield alone is not sufficient to explain stock return, but should be viewed in combination with dividend stability, indicating that dividend stability contain information about future performance [7] [22] [28] |
| EBITDA margin | Profitability ratio calculated as earnings before interest taxes, depreciation and amortization over total revenue. EBITDA is viewed as a proxy for cash flow | Frequently used by institutional and individual investors to measure profitability [44] |

| | | |
|---|---|---|
| EBITDA margin change | Point change in EBITDA margin compared to last period | Metric reflecting profitability development as measured through EBITDA |
| EPS 12 month growth | 12 month percentage change in the company earnings available per share issued | EPS is a popular ratio for investors and is viewed as a big determinant of share prices [8]. Haugen and Baker also show that past returns is correlated to future returns [32] |
| EPS 3 month growth | 3 month percentage change in the company earnings available per share issued | As mentioned above, EPS is a commonly used ratio amongst investors. We use both 3 and 12 month momentum |
| Revenue 12 month growth | 12 month change in total revenue as percentage | Past financial performance is an essential input into analysts' valuation model, and is used to substantiate stock performance. Within literature, Lakonishok *et al.* document a negative relation between returns and as sales growth, explaining this finding as investors overly emphasizing past performance [43] |
| Revenue 3 month growth | 3 month change in total revenue as percentage | As indicated above, revenue growth is an important input into many valuation models. Here we include the shorter time frame of 3 months to specify shorter-term developments |
| Market value | Absolute firm size as calculated by share price times number of shares | Fama and French is known to have documented that the size of a firm help explain returns [23] [24] |
| Market value / book value | Ratio as calculated by total market capitalization over total book value | Metric showing how much shareholders pay per unit of currency in assets. Used to indicate the value of the stock |
| P/E ratio | P/E, or the price earnings ratio, is a ratio reflecting the firm's capitalization over the firm's value | P/E is a well-known metric commonly used by investors to quickly grasp the pricing of a stock. Researchers including Basu [6], Haugen and Baker [32] and Miller *et al.* [55] have found that P/E ratio has an effect on share returns |
| Price / cash flow | Ratio comparing value of stocks to cash flow | Valuation multiple measuring operational cash flow relative to stock price |
| Price / sales | Ratio showing value of stocks over total sales | Valuation multiple comparing value of stocks to revenues |
| 3 month momentum | Technical indicator showing the share return over the past 3 months | Several studies indicate the effect of momentum on share returns, including Jegadeesh and Titman document correlation between 3-12 month momentum and future returns [37] |

| | | |
|---|---|---|
| 12 month momentum | Technical indicator showing the share return over the past 12 months | As mentioned above, this technical indicator is oftentimes used to estimate share returns. Here we use 12 month momentum to view long-term stock development in addition to the shorter three month frame explained above |
| ROIC | Return metric calculated as return as % of invested capital | Indicates a firm's ability to earn on capital invested |
| ROA | Return metric calculated as return as % of total assets | Indicates how well the company turns its assets into earnings |
| ROE | Return metric calculated as return as % of equity | Return metric that is widely used in practice. Indicates how well the firm translates its equity into returns. Haugen and Baker show that return on equity is correlated to future returns [32] |
| Sales / total assets | Total sales over total assets | Indicates the firm's ability to generate revenue from assets |
| Total asset growth | % growth in assets | Cooper *et al.* indicate a correlation between growth in a firm's assets and future returns [15] |
| Current ratio | Liquidity ratio as shown by current assets divided by current liabilities | Indicate the firm's ability to cover short-term debt |
| Debt / Total capital | Financial leverage measure | Metric showing the firm's leverage, used to indicate the risk associated with the firm |
| Share turnover as % of market value | Liquidity proxy calculated as total volume traded over market value | Glosten *et al.* investigated liquidity in terms of share turnover and found evidence that share turnover is negatively correlated to returns [27]. This is also evidenced by Haugen and Baker [32] |
| Volatility | Measure of dispersion of returns calculated as standard deviation | Commonly used risk metric. Haugen and Baker found that volatility does not affect share returns [32] |

## 4 Research design

In this section, we formulate three research questions intended to be answered throughout the study. The methods used in order to answer each topic are subsequently introduced.

The research questions are as follows:

**Question 1:** *Does information obtained from news stories and corporate documents improve prediction accuracy ?*

Despite an increasing number of research focused on textual analysis, few studies incorporate both news stories and corporate documents for stock return prediction. If the model has a higher prediction accuracy by including textual analysis, we find evidence that corporate documents and news stories contain information that investors can act upon to make abnormal returns. On the contrary, if the prediction accuracy does not benefit from including qualitative factors, we find evidence that the content of firm-specific news stories and corporate documents is inconsequential for how the stock price develops.

**Question 2:** *Which features hold importance for prediction of stock performance?*

This is a worthwhile questions as it allows us to determine which factors are of high importance to explaining stock price movement.

**Question 3:** *Can the findings be used to substantiate a profitable investing strategy*

Constructing a portfolio of long and short positions allow us to interpret how the results translate into returns for investors. We test whether the results may be exploited in practice by constructing a long/short portfolio with 5%, 10% and 20% of stocks for each prediction period. We include liquidity restrictions and transaction costs, meaning the investor has a limited amount of money to invest, and that transaction costs incur when the portfolio is rebalanced.

### 4.1 Textual Analysis

Over 300,000 news stories and over 13,000 company documents were collected to provide the basis for a sentiment analysis and context classification.

#### 4.1.1 Sentiment analysis

Following related research [20] [21] [31] we construct quantitative variables from textual features by structuring the content into singular words populating a matrix. Unnecessary words and characters are excluded from the analysis, followed by tokenization and stemming using the Porter method allows us to group related words. Based on this, we construct a matrix of sentiment-bearing words. By using a dictionary-based method, the matrix is translated into a quantitative variable by examining the occurrences of words and aggregating their sentiment value. The sentiment value is based on a classification into positive and negative words. This holds true when using both the Harvard IV Dictionary and Loughran and McDonald's Financial Sentiment Dictionary. An advantage of using the dictionary-based approach is that human judgement is not involved, as would have been necessary in a supervised machine learning context. In fact, if we were to use a machine learning approach for sentiment analysis, we would have to pre-classify parts of the data sample into positive, neutral and negative sentiment in order to test and train the model. The reliance on human judgement and the tedious work associated with the pre-classification, makes it unsuited for the premise of this study.

We use raw-term frequency of word categories in calculating the sentiment. Using raw term frequency means that all sentiment-bearing words are weighted equally. After understanding the count for positive and negative words, the values

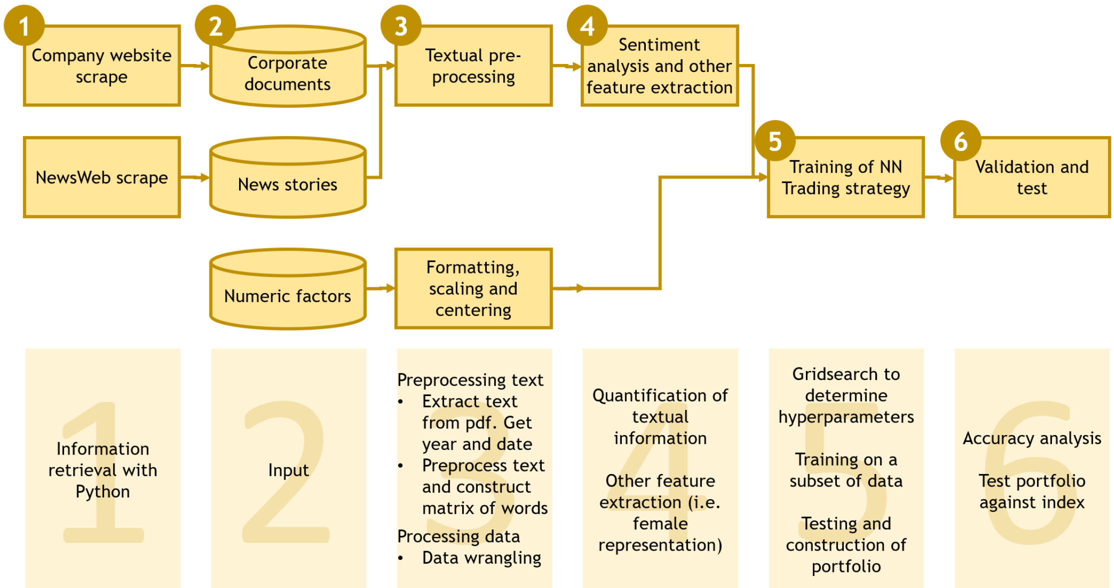



**Figure 1** Overview of methodology used in this study

are normalized over the length of the text. We apply this approach when calculating both positive and negative sentiment for corporate documents and news stories. In addition, we investigate the tone of the news stories and corporate documents by calculating the ratio of positive minus negative words over total sentiment-bearing words. Lastly, we calculate readability to indicate how difficult the text is. Complex words are defined as words containing three syllables or more, and again, each word is weighted equally in order to calculate the factor.

#### 4.1.2 News Context Classification

We extract the context of each news story by grouping the stories into 29 different classes. Using automatic classification, the news story is classified according to the appearance of given words. These words, or key words, are the most frequent unique words associated with each class and are defined using a bag-of-words approach on a subset of stories. The method is applicable for both Norwegian and English news stories as the classes are defined using key words from both languages - an advantage as this news outlet publishes stories in both languages. This is a reliable approach for grouping of news stories retrieved from Oslo Stock Exchange NewsWeb as the content is in a relatively standardized format.

### 4.2 Financial Factors

For each period, we obtain several financial factors used in the analysis. In line with similar research, we calculate logarithmic stock returns. The predictors are scaled so that the magnitudes of each coefficient are of equal unit, centered so that the we achieve a mean of zero and standard deviation of one. This computation is shown in Equation 1. These two computations enable us to compare predictors that initially have varying scales.

$$x_{i,standardized} = \frac{x_i - \mu}{\sigma} \tag{1}$$

### 4.3 Neural Network

In this study, we apply a neural network to train and test a model based on three data sets: (i) financial factors gathered from Eikon, (ii) quantified textual factors from news and corporate documents and (iii) the aggregate data set. The principal advantage to using a neural network as opposed to a linear regression is that the model is unperturbed by multicollinearity and nonlinearity. In terms of textual analysis within financial literature, neural networks are not as prominent as support vector machines and naive Bayes [31]. However, neural networks have shown promising results in several financial applications including share price prediction with textual analyses, as shown by Khadjeh *et al.* [38] and Hájek [31].

The model is apt at discovering seemingly hidden patterns in the data by passing the data through several layers. Between each layer the input is weighed and summed together with a bias, and passed through an activation function. This allows for interactions and nonlinearity. As for the activation functions, ReLU, or a rectified linear unit, is one of the most popular ones. ReLU also happens to be the activation function implemented in this study. This is a piecewise linear function that outputs the value if positive or zero if negative. Behind the curtain, an optimization algorithm helps determine the weights and bias in order to minimize errors. As is common, the gradient descent algorithm will be used as the optimization algorithm, supported by backpropagation that determines the gradient.

Naturally, the neural network requires several hyperparameters. Hyperparameters are predetermined variables that set the rules for how the model learns and predicts. These hyperparameters are tuned using a cross validation grid search approach. This allows us to test on the training data which values of hyperparameters are most suited for the task at hand.

### 4.4 Feature Importance

Feature importance is a valuable part of this study allowing us to understand how the features, or factors, contribute to the prediction in the model. Determining feature importance is simple for the linear regression case, here one looks at the size of the coefficient and the p-values to determine the impact of a one-unit change and the statistical significance of the factor. In the case of the neural network, understanding feature importance is not quite as simple. The model is less transparent oftentimes described as a black box.

Fortunately, there are ways to calculate feature importance for the more complex models, one such way is to implement permutation feature importance. Using this method, we randomize the values for each feature one at a time, and calculate the prediction accuracy associated with each model. Mean directional accuracy is used as the performance measure. In the case of a significant drop in accuracy for one model, the factor that is randomized in that model contributes significantly to the prediction.

### 4.5 Trading Strategy

We interpret the usefulness of the model through constructing a portfolio of long and short positions for the test period. The portfolio consists of a total of 20 companies, divided across long and short positions. The long position comprises of the 10 stocks with the highest predicted returns, the short position of the 10 stocks with the lowest predicted returns. We introduce liquidity constraints and transaction costs in the portfolio optimization problem. The transaction costs are calculated using the Nordic bank Nordnet's brokerage fee's for long and short transactions. Using this, we paint a realistic picture for returns obtainable for individual investors investing in line with the recommendations of the model.

### 4.6 Prediction Accuracy

The validity of the features included and the model implemented in the study are investigated through comparative analysis. On a whole, directional accuracy is used as the main performance metric, complemented by root mean squared error (RMSE). Using directional accuracy, we compare the neural network with a simple linear regression. Directional accuracy forecasts the direction of change, and it has support within recent literature. Leitch and Tanner argue that directional accuracy is a valid evaluation criterion when predicting firm profits [45]. Other researchers including Abhyankar *et al.* [2], Cheung *et al.* [14] and Moosa *et al.* [57] argue that direction of change correlates more strongly with profitability as compared to other statistical measures including root mean squared error. The formula for mean directional accuracy (MDA) is shown below:

$$MDA = \frac{1}{N} \sum 1_{sign(A_t - A_{t-1}) == sign(F_t - A_{t-1})} \quad (2)$$

$A_t$: actual value
$F_t$: forecast value

## 5 Results and Discussion

In this section, we present the results from our analysis with a subsequent discussion. Using 13,420 corporate documents, over 300,000 news stories and supplementing with financial values over a period of 21 years, our results suggest that the information contained in textual sources are of value. We arrive at such a conclusion by comparing two different statistical models, a deep neural network and a simple linear regression. Both models are implemented on three different data sets:

- Financial factors (technical and fundamental)
- Factors obtained by textual analysis of news and corporate documents
- Both data sets (aggregate data set)

We find the feature importance of the various factors by implementing the method explained above. In addition, we compare the results of the neural network prediction with different data sets by constructing long/short portfolios. All of the results are shown and discussed in the consecutive subsections.

### 5.1 Value of Including Qualitative Factors

We present the prediction accuracies in Table 2, showing both mean directional accuracy (MDA) and root mean squared error (RMSE) for the neural network (NN) and ordinary least squares regression (OLS). It is evident that the neural network achieves a higher directional accuracy than linear regression, and this goes for all data sets. The highest directional accuracy is associated with the neural network run on the data set containing all data, with a directional accuracy of 56.4%. This is a clear improvement over 53.2% that is obtained using the linear regression. Existence of multicollinearity or hidden patterns in the data set are possible reasons. The neural network uses interactions and nonlinearity to expose these hidden patterns, and successfully handles multicollinearity. Simultaneously, the results point out that the linear regression might be too simple for the prediction task at hand.

**Table 2**: Prediction accuracies

| | **OLS** | | **NN** | |
|---|---|---|---|---|
| **Model** | **MDA** | **RMSE** | **MDA** | **RMSE** |
| **Financial** | 48.6% | 0.597 | 53.8% | 0.602 |
| **Textual** | 52.0% | 0.603 | 51.1% | 0.596 |
| **Aggregate** | 53.2% | 0.606 | 56.4% | 0.596 |

From Table 2 it is clear that the data set containing both financial and textual values obtains the highest prediction accuracy in this study. This above the financial case at 53.8% and the textual case at 51.1%. The results suggest that the factors obtained from textual sources have a beneficial presence when combined with financial factors. Excluding one or the other results in omitted variable bias. This is supported by the root mean squared error, that show the lowest error for the aggregate model and the textual model at 0.596, followed by the financial model at 0.602.

Table 2 reveals one interesting observation regarding the distinction between financial and textual information. For textual information, the neural network does not produce higher prediction acccuracies than ordinary least squares regression. On the other hand, for financial information, neural networks works better than the OLS regression. This could indicate that in this market, investors analyze financial data using regressions. As a result, regressions might not be useful in this context anymore, while neural networks are. However, it seems that textual information is not utilized by investors in this market, as linear regression based on textual data makes better predictions.

For comparison, the directional accuracies across different prediction ranges are shown in table 3. According to the table, the extreme cases of very high or very low predicted returns show a higher directional accuracy for the neural network. This result reveals that the neural network model is comparatively bad at distinguishing the direction of stocks that are expected to have moderate return over the period. At the same time, the model is better at predicting the direction of stocks whose prices are expected to move more. The model is, in other words, bad at predicting direction of share returns in general, but provide reliable results for stocks that it finds likely to have significantly high or low returns. It furthermore becomes evident that the aggregate model outperforms the textual and financial model in terms of directional accuracy in the high and low percentiles. The lowest percentiles of the aggregate model, with a MDA of 68%, obtains the highest directional accuracy in this study. The pattern of strong prediction accuracies is not quite as present for the OLS regression. The model generally has a higher prediction accuracy for the bottom percentiles on all data sets, but performs worse in the top percentile. In fact, the OLS regression performs worst in the 90-100 percentile for the aggregate data set.

### 5.2 Feature Importance

We calculate permutation feature importance of the aggregate model in order to understand how the factors contribute to the predictive power of the model. The results are shown in table 4. As mentioned, a lower permutation feature importance is associated with a higher importance to the model, as we measure the importance with mean directional accuracy. Table 4 is ordered after ascending feature importance, meaning that the higher up the feature, the more important it is.

It is clear that news stories and corporate documents contain information that is valuable in a return prediction model. According to the results, 5 of the 10 most important factors are extracted from news stories, 2 from corporate documents. The textual features, despite having a lower prediction accuracy standalone as shown in table 2, is evidenced to be significant when used in conjunction with financial factors. Looking at the factor documented to hold most importance, the directional accuracy decreases 4.3% to 95.7% of its original value when permuting the news class associated with new projects. This is not a significant decrease, but it reflects that the model is not excessively reliant on any one feature. In fact, the results indicate that the model uses a vast amount of information to find signals, and that the strength of the model is in combining the different factors. As a side note, table 4 shows that seven features are associated with an increase in directional accuracy when permuted. This does not necessarily imply that the features do not have any significance for share returns, rather it may exemplify how permutation importance is not able to capture the effects of multicollinearity, a typical issue with permutation feature importance.

**Table 3**: Directional accuracy for percentiles of predicted return, NN

| | | **0-10%** | **10-20%** | **20-80%** | **80-90%** | **90-100%** |
|---|---|---|---|---|---|---|
| **NN** | Aggregate | 68.0% | 55.0% | 54.9% | 53.9% | 58.1% |
| | Financial | 60.9% | 47.3% | 52.7% | 57.0% | 56.6% |
| | Textual | 53.9% | 55.0% | 50.6% | 50.8% | 47.3% |
| **OLS** | Aggregate | 59.4% | 54.3% | 53.2% | 57.8% | 51.2% |
| | Financial | 56.3% | 45.0% | 49.3% | 57.0% | 55.0% |
| | Textual | 59.4% | 58.1% | 50.7% | 53.9% | 51.9% |

**Table 4**: Permutation feature importance

| **Feature** | **Importance** |
|---|---|
| News context classification: New project | 0.957 |
| Dividend stability | 0.963 |
| News context classification: Dividend announcement | 0.965 |
| Corporate document: % female names | 0.966 |
| 12 month trailing ROA | 0.966 |
| News context classification: Other news | 0.967 |
| News context classification: Dividend is paid out | 0.968 |
| News context classification: Information about company presentations | 0.971 |
| 3 month momentum | 0.971 |
| Corporate document: Sentiment polarity | 0.972 |
| Dividend yield | 0.972 |
| EPS 3 month growth | 0.974 |
| Price to sales | 0.975 |
| News context classification: Insider buys shares | 0.975 |
| EBITDA margin | 0.976 |
| EPS 12 month growth | 0.977 |
| Debt % Total capital | 0.977 |
| News context classification: Change in interest rate | 0.978 |
| 12 month momentum | 0.980 |
| Market value / book value | 0.981 |
| News: Sentiment polarity | 0.981 |
| Volatility | 0.982 |
| ROE | 0.984 |
| News context classification: Acquisition of other firm | 0.984 |
| Corporate documents: Sentiment negative | 0.985 |
| ROIC | 0.986 |
| Sales / total assets | 0.987 |
| News: Negative sentiment | 0.988 |
| Share turnover as % of market value | 0.990 |
| News context classification: Reduced debt | 0.990 |
| News context classification: General meeting | 0.990 |
| News context classification: Cancellation/delay in project | 0.990 |
| News context classification: Large shareholder buys shares | 0.991 |
| News context classification: Company is IPO'd | 0.991 |
| News context classification: Large shareholder sell shares | 0.993 |
| Price / cash flow | 0.994 |
| News context classification: Emission/increase share capital | 0.994 |
| Revenue 12 month | 0.994 |
| News context classification: Challenging market conditions | 0.995 |

| | |
|---|---|
| Total asset growth | 0.995 |
| News context classification: Insider sells shares | 0.995 |
| News context classification: Company may be acquired | 0.996 |
| News context classification: Increase debt | 0.996 |
| News context classification: Company is excluded from observation | 0.996 |
| News context classification: Judicial or regulatory enforcement | 0.997 |
| News context classification: Changes in leadership / board | 0.997 |
| News context classification: Company put on observation | 0.998 |
| News: positive sentiment | 0.998 |
| News context classification: dividend reduction or cancellation | 0.998 |
| News context classification: Cost reduction programme | 0.999 |
| Current ratio | 0.999 |
| P/E ratio | 0.999 |
| News context classification: Favourable market conditions | 0.999 |
| EBITDA margin expansion (contraction) | 0.999 |
| News context classification: Share buyback | 0.999 |
| Market value | 0.999 |
| News context classification: Divestment of company assets | 1.001 |
| News context classification: Decrease in share capital | 1.003 |
| Corporate document: Readability | 1.003 |
| Corporate document: Acquisition factor | 1.005 |
| News context classification: Implementation of employee incentive program | 1.006 |
| Corporate document: Sentiment positive | 1.009 |
| Revenue 3 month growth | 1.023 |

### 5.3 Investment Strategy

Using the neural network model, we construct a portfolio of the shares predicted to outperform and underperform in each quarter. We take into account transaction costs and liquidity constraints. For comparison, we examine the portfolio return across two models: (i) the neural network and (ii) linear regression, based on three different data sets: (i) financial (ii) textual and (iii) the aggregate data set. The portfolio performance of the neural network portfolios are shown in figure 2, and the performance of the OLS regression portfolios are shown in figure 3.

As evident from figure 2, the neural network obtains returns of 9.4x, 6.3x and 3.5x invested amount using the data sets aggregate, financial and textual, respectively. In other words, the aggregate model performs well above the financial model, which in turn returns close to twice that of the textual model. The return achieved by the aggregate model is significant and can in part be explained by the prediction accuracy split we show in table 3. The table shows that the stocks that are included in the aggregate portfolio are more likely to move in the favourable direction as the model is better at predicting in the top and bottom percentiles. This holds true for the financial model as well, but this trend is not as evident for the textual model, which has a directional accuracy worse than a coin flip in the 90th percentile. As mentioned, transaction costs are included, but will not greatly impact the returns of the portfolios as we operate with quarterly rebalancing.

For comparative reasons, we also look at the returns of the linear regression shown in figure 3. Notably, the portfolios based on the linear regression performs worse than the neural network. There is one exception, and that is for the textual model where OLS regression yields a higher portfolio return than the neural network. It seems the linear model is seemingly unable to find profitable patterns in the data, this is particularly true for the case relying on financial data. A possible explanation is that share price prediction with financial values is widely used in practice, and that the underlying data is readily accessible. This is further supported by the fact that the portfolio relying on textual data outperforms the financial portfolio by a large margin, as text mining is more complex and the data is subsequently less accessible. Furthermore, apart from the model based on textual data, the neural network gives significantly better results than the linear regression, supporting the findings in table 2.

In theory, returns of this portfolio should be only alpha driven. By including an equal weighting of long and short positions and assuming that the long and short stocks have similar relationship to the market developments, we should be able to eliminate market risk. Theoretically, the result is a portfolio with an insignificant beta and high alpha. We test this by running a CAPM regression on the portfolio versus an index constructed of the stocks included in the analysis. The results are shown in table 5. We get a quarterly alpha of 4.7% with a p-value of 0.092. This corresponds to a yearly alpha of 20.2% and a confidence level of 90.8%. The beta is at 0.056 with a p-value of 0.750. This is a low beta, with a confidence level of only 25%. In other words, the portfolio is not exposed to market risk, and the returns are based on alpha. This provides the backdrop of a possible trading strategy consisting of a combination of our portfolio and index, yielding a stable portfolio returning more than only investing in the market portfolio.

**Table 5**: CAPM regression of long-short strategy

| | Coefficient | t-statistic | p-value |
|---|---|---|---|
| Alpha | 0.047 | 1.762 | 0.092 |
| Beta | 0.056 | 0.323 | 0.750 |

**Table 6**: Fama-French Three-Factor Model regression of long-short strategy

| | Coefficient | t-statistic | p-value |
|---|---|---|---|
| Alpha | 0.051 | 1.821 | 0.084 |
| MKT | 0.079 | 0.421 | 0.679 |
| HML | -0.352 | -0.467 | 0.644 |
| SMB | -0.345 | -0.785 | 0.442 |

By including the two additional factors HML and SMB, we test whether the portfolio return can be explained by the factors in the Fama-French Three-Factor model. Norwegian values for HML and SMB is gathered from Bernt Ødegaard's webpage [61]. When adding HML and SMB we are able to test whether the return can be explained by systemic risk or if the portfolio obtains a true alpha. The results are shown in table 6. This time around we obtain a quarterly alpha of 5.1% with a p-value reaching 0.083, meaning that this alpha is statistically significant at the 10% confidence level. The market beta is calculated to be 0.079 with a p-value of 0.697. In other words, the market beta is not statistically significant at the 10% confidence level, meaning we cannot with certainty say that it is different from zero. We find the same for the factors HML and SMB, both have p-values well above 0.10, and yet again we cannot with certainty say the factors are unequal to zero. As both CAPM and Fama-French Three-Factor model yield positive alphas that are statistically significant at the 10% confidence level, we find evidence that the portfolio return is in fact driven by alpha and not by taking on systematic risk.

We, furthermore, test the sensitivity of the model by including a varying number of stocks and examining the consequent portfolio return. Figure 4 shows portfolio returns for portfolios consisting of 3, 5, 6, 7, 10 and 12 stocks. The results show that the number of stocks in the portfolio is inversely correlated with return. There are two reasons as to why this happens. First, each additional share included has a lower predicted return than the previous, therefore, assuming that the prediction is correct, the overall return diminishes. Secondly, as shown in table 3, the model predicts accurately for extreme cases of high and low predicted returns. Increasing the portfolio size would, according to table 3, cause a higher probability of including stocks with erroneous predictions. The returns of the portfolio with 10 and 12 long and short positions makes this painstakingly clear. On the other side, portfolios containing few stocks are more volatile and are associated with a higher idiosyncratic risk. This is suggested by the return of the portfolio with three short and three long positions - which had a relatively bad performance prior to 2015, reaching a higher overall return than the main portfolio (five stocks per position) from 2017 onward. Increasing the portfolio size to five stocks per position yields a comparatively lower idiosyncratic risk while maintaining the high return profile.

The return of the portfolio consisting of five stocks per position is compelling, and to further understand what is driving the strong results we segment the return across long and short positions. Doing this allows us to understand how the two positions contribute to the overall return. The result is illustrated in figure 5. From the figure, it is evident that both positions provide high returns, but is affected by a higher volatility than the portfolio containing both positions. Both positions have an overarching positive drift, but develops quite differently over the period. The long portfolio quickly achieves high return multiples, achieving at best nearly 12 times the invested amount before dropping heavily to 7.8 times invested amount. The short portfolio on the other side is slower to progress, returning below invested amount in 2013. This turns rapidly and the short portfolio ends up with returns of 10.8 times the original amount over the six year period. Overall, the gradient is close to opposite of that of the long portfolio, and has quite extreme slopes.

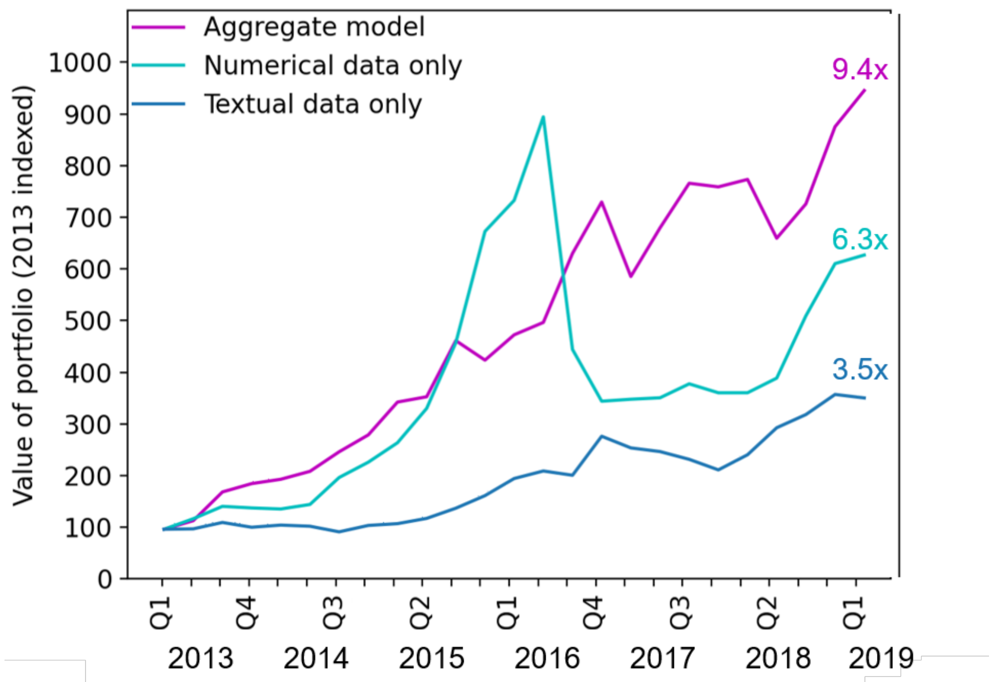



**Figure 2** Portfolio performance neural network

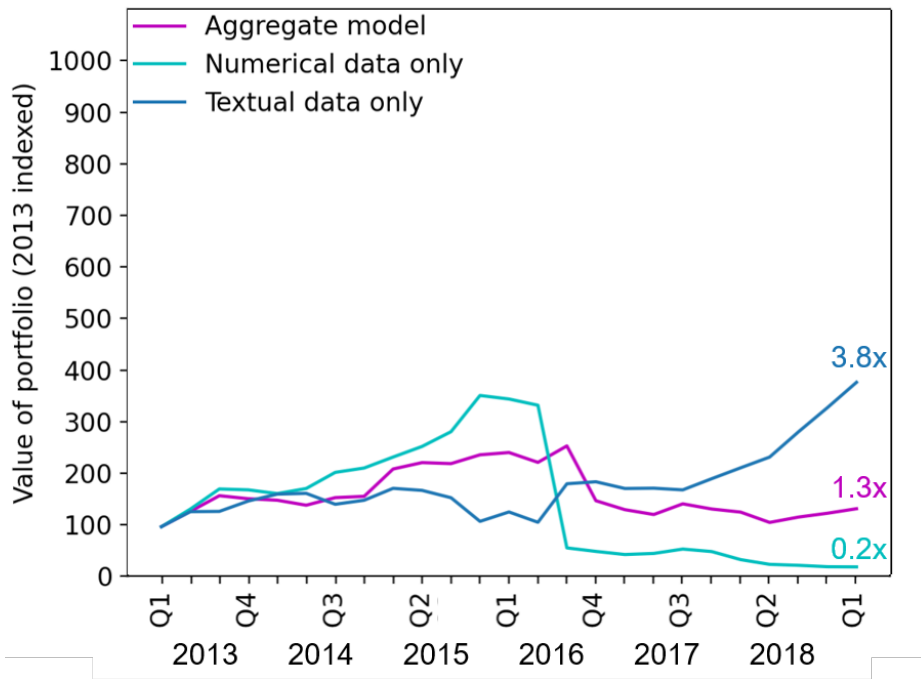



**Figure 3** Portfolio performance linear regression

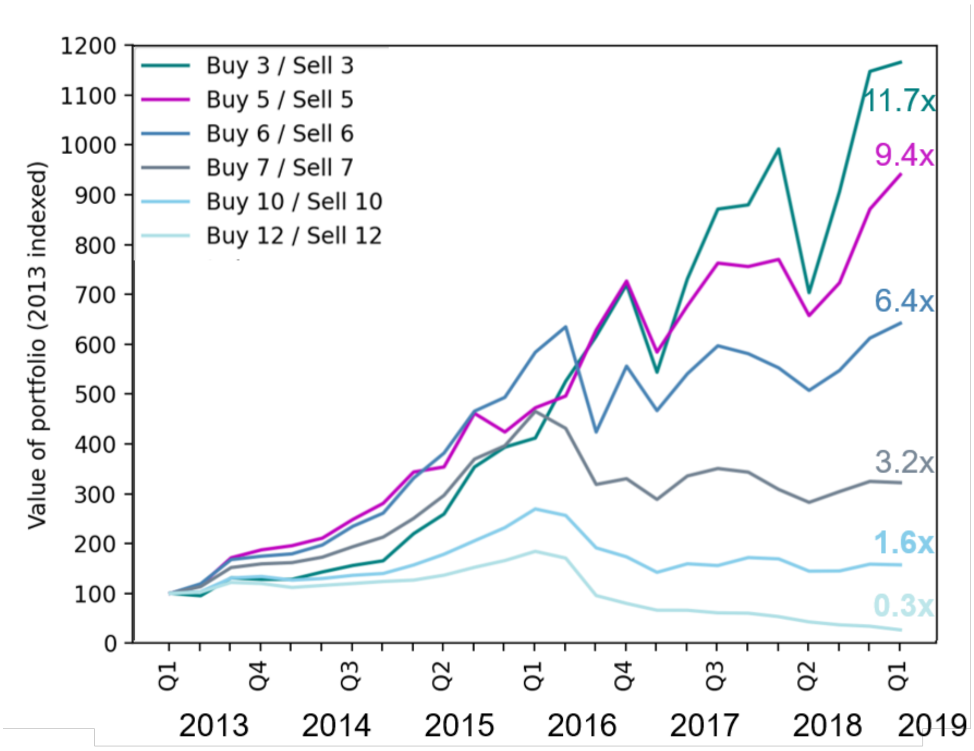



**Figure 4** Portfolio return for differing number of stocks in portfolio

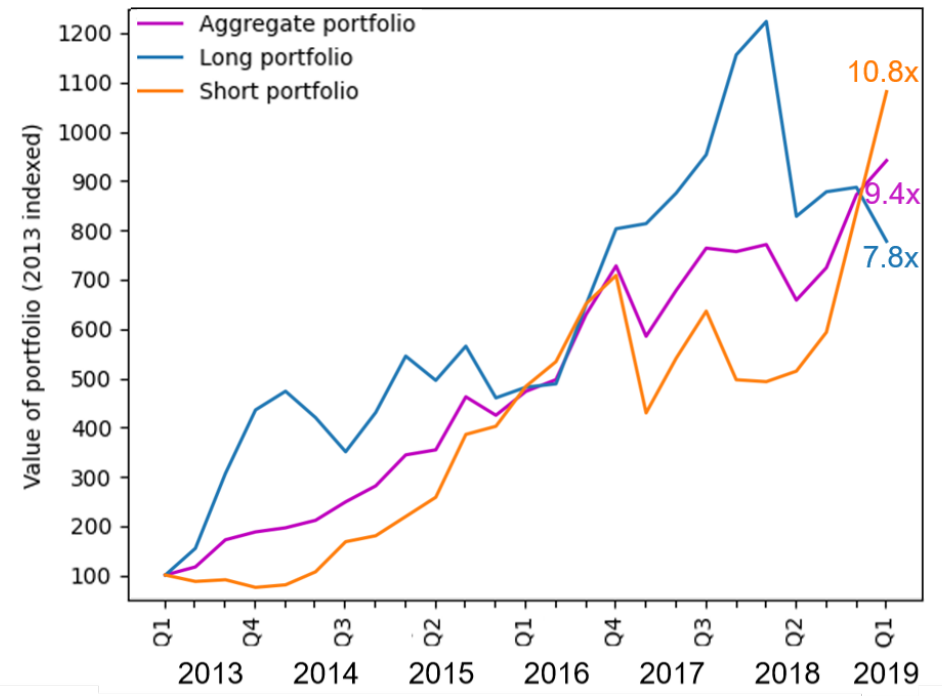



**Figure 5** Portfolio return split across long and short portfolio for portfolio with five long and five short positions

## 6 Conclusion

This study attempts to measure the information contained in corporate documents and news reports by retrieving vast amounts of data from company websites and news outlets. We combine this data with classical technical and accounting data and test the predictive power of a neural network. Our results suggest that textual sources contain useful information that can help indicate future performance of a company, showing that it is worthwhile to scrape and quantify public textual sources. We find that a model containing financial and textual information outperforms a model with either financial or textual with a directional accuracy of 56% in a neural network. In fact, we find that the directional accuracy is even higher for the highest and lowest percentiles of predicted returns, reaching 68% at most. As a direct result of this, a portfolio constructed from the predictions based on the aggregate model yields a return reaching 9.4 times invested amount over a six-year period, and outperforms the portfolio constructed on either financial or textual information. We also look at how much each factor contribute to prediction accuracy, finding that no specific factor is uniquely important to the model, rather that the predictive power is result from the combination of features.

Recent literature reports a relationship between textual information from corporate documents and news stories on share return, and our study corroborates this research. By looking at the Oslo Stock Exchange, we find evidence from the Norwegian market supporting the research findings of several studies, including Davis *et al.* [18] and Tetlock *et al.* [59], on the importance of qualitative factors for stock prediction. Furthermore, our analysis indicates that it is possible to earn an abnormal return by assessing available public information, thus contributing to the literature supporting market inefficiency. Retrieving and quantifying textual information from annual reports and news stories is time-consuming and computationally complex, meaning that although the data is accessible it is not necessarily *easily* accessible. It is likely that the investor is unable to process and correctly act upon every bit of information, giving rise to arbitrage opportunities. This is consistent with the Inefficient Market Hypothesis.

There are ways to further build upon this study. This study is limited by access to information, which has been prevalent in the earlier parts of the historical period included. Firms with non-readable pdf's or Norwegian report language was more common, and this results in less viable data from the earlier periods. Implementing a more extensive text mining approach to investigate syntax, and extract content from corporate documents could yield valuable information, but correspondingly requires increased processing power. In gathering information, we have attempted to be exhaustive, but it could be of interest to increase the scope further by including Internet-expressed information. A major drawback is that much of the content on stock message boards is written in Norwegian, whereas the dictionary-based approach relies on English as the input language.

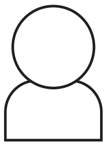
### Acknowledgement

This thesis is the concluding chapter of our Master's degree in Industrial Economics and Technology Management at the Norwegian University of Science and Technology.

We extend our sincerest gratitude to our supervisors Professor Stein-Erik Fleten and Associate Professor Peter Molnár for their enduring guidance, sound advice and helpful discussions throughout the process.

## Authors

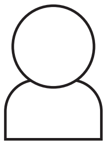
**Maja Z. Olsen** (Student)



**2015-2020 NTNU**: MSc Industrial Economics and Technology Management, specialization in Finance

**Previous employments: Goldman Sachs, Boston Consulting Group, FSN Capital**

**Sindre F. Fjellavli** (Student)



**2015-2020 NTNU**: MSc Industrial Economics and Technology Management, specialization in Finance

**Previous employments: Boston Consulting Group, Summa Equity, Arkwright Consulting**

## References

[1] Abrahamson, Eric and Eli Amir (1996), "The Information Content of the President's Letter to Shareholders", Journal of Business Finance Accounting 23:8, pp 1157-1182

[2] Abhyankar, Abhay, Lucio Sarno and Giorgio Valente (2005) "Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability", Journal of International Economics 66:2, pp 325 - 348

[3] Agnew, Julie R. and Lisa R. Szykman (2005) "Asset Allocation and Information Overload: The Influence of Information Display", Asset Choice, and Investor Experience, The Journal of Behavioral Finance, 6:2, pp 57-70

[4] Alhenawi, Yasser and Sudha Krishaswami (2015) "Long-Term Impact of Merger Synergies on Performance and Value", The Quarterly Review of Economics and Finance 58, pp 93-118

[5] Antweiler, Werner and Frank Z. Murray (2004) "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards", The Journal of Finance 59:3, pp 1259-1294

[6] Basu, S. (1977) "Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis", Journal of Finance 32:3, pp 663-682

[7] Bhattacharya, Sudipto (1980) "Imperfect Information, Dividend Policy and the 'Bird in the Hand' Fallacy", Bell Journal of Economics 10, pp 259-270

[8] Black, John, Nigar Hashimzade and Gareth Myles (2012) "A Dictionary of Economics", Oxford University Press

[9] Bloomfield, Robert J. (2002) "The 'Incomplete Revelation Hypothesis' and Financial Reporting", Accounting Horizons 16:2, pp 233-243

[10] Botosan, Christine A. and Marlene A. Plumlee, (2000) "A Re-Examination of Disclosure Level and the Expected Cost of Equity Capital", Journal of Accounting Research, 40:1, pp 21-40

[11] Bollen, Johan, Huina Mao and Xiao-Jun Zeng (2011) "Twitter Mood Predicts the Stock Market", Journal of Computational Science, 2:1, pp 1-8

[12] Campbell, Sean D. and Francis X. Diebold (2009), "Stock Returns and Expected Business Conditions: Half a Century of Direct Evidence", Journal of Business Economic Statistics 27:2, pp 266-278

[13] Chen, Jie, Woon Sau Leung, Wei Song and Marc Goergen (2019) "Why Female Board Representation matters: The Role of Female Directors in Reducing Male CEO Overconfidence", Journal of Empirical Finance 53, pp 70-90

[14] Cheung, Yin-Wong, Menzie D. Chinn and Antonio G. Pascual (2005) "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive", Journal of International Money and Finance 24:7, pp 1150-1175

[15] Cooper, Michael J., Huseyin Gulen and Michael J. Schill (2008) "Asset Growth and the Cross-Section of Stock Returns", Journal of Finance 63:4, pp 1609-1651

[16] Cutler, David M., James M. Poterba and Lawrence H. Summers (1989) "What Moves Stock Prices?", Journal of Portfolio Management 15:3, pp 4-12

[17] Das, Sanjiv R. and Mike Y. Chen (2007), "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web", Management Science 53:9, pp 1375 -1388

[18] Davis, Angela K., Jeremy M. Piger, and Lisa M. Sedor, 2006, "Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases", Working paper, Federal Reserve Bank of St. Louis

[19] Dobbin, Frank and Jiwook Jung (2011) "Corporate Board Gender Diversity and Stock Performance: The Competence Gap or Institutional Investor Bias", North Carolina Law Review 89:3, pp 809-838

[20] Doran, James, David Peterson and S. Price (2012) "Earnings Conference Call Content and Stock Price: The Case of REITS", The Journal of Real Estate Finance and Economics 45:2, pp 402-434

[21] Engelberg, Joseph, Adam Reed and Matthew Ringenberg (2012) "How are Shorts Informed? Short Sellers, News and Information Processing", Journal of Financial Economics 105:2, pp 260-278

[22] Erasmus, Pierre (2013) "The Influence of Dividend Yield and Dividend Stability on Share Returns: Implications for Dividend Policy Formulation", Journal of Economic and Financial Sciences 6:1, pp 13-32

[23] Fama, Eugene F. and Kenneth R. French (1992) "The Cross-Section of Expected Stock Returns", Journal of Finance 47:2, pp 427-466

[24] Fama, Eugene F. and Kenneth. R. French (1993) "Common Risk Factors in the Returns

on Stock and Bonds", Journal of Financial Economics 33:1, pp 3-56

[25] Feldman, Ronen, Suresh Govindaraj, Joshua Livnat and Benjamin Segal (2008) "The Incremental Information Content of Tone Change in Management Discussion and Analysis", Available at SSRN: https://ssrn.com/abstract=1126962

[26] Gandhi, Priyank, Tim Loughran and Bill McDonald (2019) "Using Annual Report Sentiment as a Proxy for Financial Distress in U.S. Banks", Journal of Behavioral Finance 20:4, pp 424-436

[27] Glosten, Lawrence and Lawrence E. Harris (1988) "Estimating the Components of the Bid-Ask Spread", Journal of Finance and Economics 21:1, pp 123-142

[28] Gwilym, Owain A., Gareth Morgan and Stephen Thomas (2000), "Dividend Stability, Dividend Yield and Stock Returns: UK Evidence", Journal of Business Finance Accounting 27:3, pp 261-281

[29] Hagenau, Michael, Michael Liebmann and Dirk Neumann (2013) "Automated News Reading: Stock Price Prediction Based on Financial News Using Context Capturing Features", Decision Support Systems 55, pp 685 - 697

[30] Hail, Luzi (2002) "The Impact of Voluntary Corporate Disclosures on the Ex-Ante Cost of Capital for Swiss firms", European Accounting Review, 11:4, pp 741-773

[31] Hájek, Peter (2018) "Combining Bag-of-Words and Sentiment Features of Annual Reports to Predict Abnormal Stock Returns", Neural Computing Applications 29, pp 343-358

[32] Haugen, Robert A. and Nardin L. Baker (1996) "Commonality in the Determinants of Expected Stock Returns", Journal of Financial Economics 41:3, pp. 401-439

[33] Henne, Antje, Sebastian Ostrowski and Peter Reichling (2007)" Dividend Yield and Stability versus Performance at the German Stock Market", IDEAS Working Paper Series

[34] Henry, Elaine (2006) "Market Reaction to Verbal Components of Earnings Press Releases: Event Study Using a Predictive Algorithm", Journal of Emerging Technologies in Accounting 3:1, pp 1-19

[35] Hines, R. D. (1982) "The Usefulness of Annual Reports: the Anomaly Between the Efficient Markets Hypothesis and Shareholder Surveys", Accounting and Business Research 12:48, pp 296-309

[36] Hirshleifer, David and Siew H. Teoh (2002), "Limited Attention, Information Disclosure and Financial Reporting", Journal of Accounting and Economics 36:1, pp 337-386

[37] Jegadeesh, Narasimhan and Sheridan Titman (1993) "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", Journal of Finance 48:1, pp 65-92

[38] Khadjeh Nassirtoussi, Arman, Saeed Aghabozorgi, Teh Ying Wah and David Chek Ling Ngo (2014) "Text Mining for Market Prediction: A Systematic Review", Expert Systems With Applications 41:16, pp 7653-7670

[39] Kamaluddin, Kamilah and Nadisah Zakaria (2018) "Long-Term Post-Acquisition Share Performance: The Case of Saudi Stock Exchange", Journal of Economic Management Perspectives 12:2, pp 473-481

[40] Kearney, Colm and Sha Liu (2014) "Textual Sentiment in Finance: a Survey of Methods and Models", International Review of Financial Analysis 33, pp 171-185

[41] Kohut, Gary F. and Albert H. Segars (1992) "The President's Letter to Stock-Holders: An Examination of Corporate Communication Strategy", Journal of Business Communication 29:1, pp 719

[42] Krinsky, Itzhak, and Jason Lee (1996) "Earnings Announcements and the Components of the Bid-Ask Spread", The Journal of Finance 51, pp 1523-1535

[43] Lakonishok, Josef, Andrei Shleifer and Robert W. Vishny (1994) "Contrarian Investment, Extrapolation and Risk", Journal of Finance 49:5, pp 1541-1578

[44] Law, Jonathan (2016) "A Dictionary of Accounting 5th Edition", Oxford University Press

[45] Leitch, Gordon and J. Ernest Tanner (1991) "Economic Forecast Evaluation: Profits Versus the Conventional Error Measures", American Economic Review 81:3, pp. 580-590

[46] Lev, Baruch and S. Ramu Thiagarajan (1993), "Fundamental Information Analysis", Journal of Accounting Research 31:2, pp 190-215

[47] Li, Feng (2006) "Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?", Working paper, University of Michigan.

[48] Li, Feng (2008) "Annual Report Readability, Current Earnings, and Earnings Persistence", Journal of Accounting and Economics 45:2, pp 221-247

[49] Lillo, Fabrizio, Salvatore Miccich, Michele Tumminello, Jyrki Piilo and Rosario N. Mantegna, (2014) "How News Affects the Trading Behaviour of Different Categories of Investors in a Financial Market", Journal of Quantitative Finance 15:2, pp 213-229

[50] Lintner, John (1962) "Dividends, Earnings, Leverage, Stock Prices and the Supply of Capital to Corporations", Review of Economics and Statistics 44:3, pp 243-269

[51] Loughran, Tim, Bill McDonald (2011) "When is a Liability not a Liability? Textual Analysis, Dictionaries and 10-Ks", Journal of Finance 66:1, pp 35-65

[52] Loughran, Tim and Bill McDonald (2014) "Measuring Readability in Financial Disclousures", Journal of Finance 69:4, pp 1643-1671

[53] Lückerath-Rovers, Mijntje (2010) "Women on Boards and Firm Performance", Journal of Management and Governance 17:2, pp 1-19

[54] Mandelbrot, Benoit (1966) "Forecasts of Future Prices. Unbiased Markets, and Martingale

Models", Journal of Business 39, pp 242-255

[55] Miller, Paul F. and Ernest R. Widmann (1966) "Price Performance Outlook for High Low P/E Stocks", Stock and Bond Issue, Commercial Financial Chronicle, pp 26-28

[56] Modigliani, Franco and Miller Merton (1961) "Dividend Policy, Growth, and the Valuation of Shares", Journal of Business 34:4, pp 411-433

[57] Moosa, Imad and John Vas (2015) "Directional Accuracy, Forecasting Error and the Profitability of Currency Trading: Model-Based Evidence", Applied Economics 47:57, pp 6191-6199

[58] Shiller, Robert J. (1981) "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends", American Economic Review 71:3, pp 421:436

[59] Tetlock, Paul, Maytal Saar-Tsechansky and Sofus Macskassy (2008) "More Than Words: Quantifying Language to Measure Firms' Fundamentals", Journal of Finance 63:3, pp 1437-1466

[60] Tumarkin Robert, Robert F. Whitelaw (2001) "News or Noise? Internet Postings and Stock Prices", Financial Analysts Journal 57:3, pp 41-51

[61] Ødegaard, Bernt (2020) "Empirics of the Oslo Stock Exchange. Basic, descriptive, results 1980-2019", http://finance.bi.no/ bernt/financial_data/ose_asset_pricing_data/index.html, Accessed June 2020

## Appendix

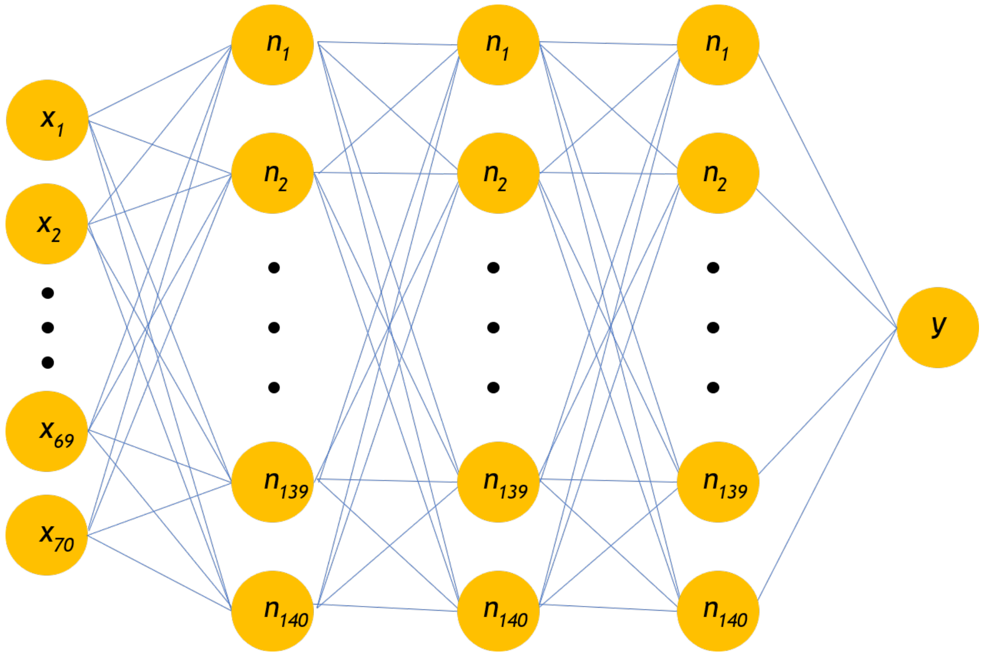
**Table A1: Context classification**

| News event |
|---|
| New project |
| Dividend announcement |
| Other news |
| Ex Dividend |
| Information about company presentations |
| Primary insider buys shares |
| Change in interest rate |
| Acquisition of other firm |
| Reduced debt |
| General meeting |
| Cancellation/delay in project |
| Large shareholder buys shares |
| Company is IPO'd |
| Large shareholder sell shares |
| Change in share capital |
| Challenging market conditions |
| Primary insider sells shares |
| Company may be acquired |
| Increase debt |
| Company is excluded from observation |
| Judicial or regulatory enforcement |
| Changes in leadership / board |
| Company put on observation |
| Dividend reduction or cancellation |
| Cost reduction programme |
| Favourable market conditions |
| Share buyback |
| Divestment of company assets |
| Decrease in share capital |
| Implementation of employee incentive program |

**Table A2: Directional accuracy for all percentiles of predicted returns**

| | **0-10%** | **10-20%** | **20-30%** | **30-40%** | **40-50%** | **50-60%** | **60-70%** | **70-80%** | **80-90%** | **90-100%** |
|---|---|---|---|---|---|---|---|---|---|---|
| Aggregate | 0.680 | 0.550 | 0.555 | 0.496 | 0.578 | 0.496 | 0.531 | 0.636 | 0.539 | 0.581 |
| Financial | 0.609 | 0.473 | 0.578 | 0.457 | 0.578 | 0.481 | 0.563 | 0.504 | 0.570 | 0.566 |
| Textual | 0.539 | 0.550 | 0.516 | 0.488 | 0.500 | 0.512 | 0.492 | 0.527 | 0.508 | 0.473 |

**Figure A1: Neural network used in the analysis**




The neural network included in the analysis consist of three layers with 140 nodes each. Using the ReLu activation function we obtain predicted share return. A simple gridsearch has been used to tune epochs and number of nodes. Testing 50, 100 and 200 epochs and 50, 100, 140 and 200 nodes we obtained the best prediction accuracy using 100 epochs and 140 nodes in each layers.

NTNU

Norwegian University of
Science and Technology