Norwegian School of Economics

Bergen, Spring 2022

Sentiment Analysis in the Norwegian Stock Market

Predicting Stock Price Movements Using Media Sentiment

Herdís Birta Jónsdóttir, Live Wold Thorsø

Supervisor: Christian Langerfeld

Master thesis, MSc in Economics and Business Administration,
Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Preface

This thesis is a part of a master's degree in Economics and Business Administration from NHH Norwegian School of Economics, with a specialization in Business Analytics. It was written during the spring semester of 2022 under the supervision of assistant professor Christian Langerfeld. We would like to thank him for his valuable input and guidance throughout the semester. Furthermore, we would like to thank our partners for their support during the writing process.

Abstract

A historical belief in financial economics states that stock prices react immediately to available information, making it impossible to predict changes in stock prices. However, research in the last few decades contradicts this theory to a degree. Some research has been done on whether the sentiment or "tone" of financial media can be used to predict stock price movements, although it has often been focused on U.S. newspapers and markets. This thesis will analyse Norwegian news articles from the online newspaper Dagens Næringsliv and stock prices from companies listed on the Oslo Stock Exchange, or Oslo Børs, to explore whether a relationship can be found between the sentiment of news articles and stock price changes. Sentiment analysis will be used to identify the level of positivity or negativity in the news articles and four statistical methods will be performed to attempt to predict stock prices using media sentiment. The methods are logistic regression, K-nearest neighbors, gradient boosted trees and support vector classifier. The accuracy of using media sentiment to predict stock price changes varies from 53.69% to 57.38% using the four methods. These results are in line with the accuracy that previous research on U.S. newspapers and companies has found, suggesting that there is a weak but significant relationship between the sentiment of Norwegian news articles and the stock price movements of Oslo Børs-listed companies.

Contents

List of Figures

List of Tables

1. Introduction

With the vast amount of data available today, researchers have raced to identify ways in which this data can be used to profit. It is a common assumption in finance that predicting the stock market is an extremely difficult task due to the market's complexity (Angadi and Kulkarni, 2015). Supporters of the efficient market hypothesis argue that the stock market reacts instantly to available information, making it impossible to use any type of information analysis to predict changes in the market. However, previous literature contradicts this hypothesis since researchers have found that it is possible to predict stock price movements to a degree by using data such as historical prices or investors' market outlook (Malkiel, 2005). In this thesis, an attempt will be made to predict stock price movements of companies listed on the Oslo Stock Exchange, or Oslo Børs, using a type of content analysis called sentiment analysis. Sentiment analysis is a method that identifies the polarity of a text based on certain conditions (Hardeniya and Borikar, 2016). Financial news articles from one of the most popular newspapers in Norway, Dagens Næringsliv, will be analysed to identify the relationship between the sentiment of news articles covering companies listed on Oslo Børs and the stock price movements of said companies. This is the foundation of the subsequent research question:

> *Can a relationship be found between media sentiment and stock price changes of Oslo Børs-listed companies?*

Previous literature has identified a weak but significant relationship between media sentiment of popular U.S. newspapers and stock price changes of companies listed on the New York Stock Exchange. For example, researchers in 2012 found that subjective news articles could predict the direction of a stock better than chance alone (Schumaker et al., 2012). The hypothesis of this analysis is that a similar relationship exists between the sentiment of articles written in a popular Norwegian newspaper and price changes of companies listed on Oslo Børs. It is suggested that the accuracy of predicting stock price movements using media sentiment is in line with the accuracy that previous literature has found. The hypothesis is presented as follows:

> *A weak but significant relationship exists between media sentiment and stock price changes of Oslo Børs-listed companies, with a prediction accuracy consistent with what previous literature has found.*

In order to explore the relationship between the two variables, predictions of stock price changes will be made using four statistical methods: logistic regression, K-nearest neighbors, gradient boosted trees and support vector classifier. Several methods are attempted since it eliminates the possibility of rejecting the hypothesis that stock price movements can be predicted using media sentiment because of a poor choice of prediction method. The results of the four methods implemented will be used to conclude whether or not media sentiment is a good predictor for stock price movements.

This thesis starts with a literature review, presenting two contradictory financial theories and relevant sentiment analysis research. The following chapter, methodology, is split into four parts. The first part covers the collecting and pre-processing of both the textual and the numeric data used in the analysis. The second part explains the theory and technicalities behind sentiment analysis and how the sentiment score is calculated. The third part discusses classification and the methods implemented, and finally, the fourth part explains the performance metrics chosen to evaluate the results of the analysis. The results chapter covers the performance of the analysis, first for each classification method implemented and then a comparison of the four methods. When the results of the four methods have been compared, an exploration of the nature of the relationship between the two variables will be done, relating to the findings of previous researchers. This is followed by a discussion of the limitations of the analysis and examples of further research that can be conducted. Lastly, the final chapter presents the conclusion of the thesis.

2. Literature Review

For investors, the idea of being able to accurately predict the financial market's movements to maximize their own profit is an enticing one. Researchers have attempted to find analytical methods to accurately predict the market, based on the assumption that a predictive relationship exists between public stock market data and future stock returns. Due to the stock market's complexity, trying to predict future stock returns is considered to be one of the most challenging tasks when it comes to financial computations (Angadi and Kulkarni, 2015). A dominating theory until the turn of the century was the efficient market hypothesis, arguing that no type of technical analysis could identify securities that were undervalued since the market was efficient enough to instantly incorporate any new information that came forward. Supporters of the efficient market hypothesis therefore believed that it was impossible to predict future stock returns. However, some criticism of the hypothesis arose as many experts started to consider the behavioural aspects of markets, suggesting that future stock price patterns could be at least partially predicted. Thus, the field of behavioural finance was born (Gupta et al., 2014). The following sections will discuss the two conflicting ideas and the theoretical background of the problem at hand.

2.1 Efficient Market Hypothesis

The efficient market hypothesis was developed independently by economists Paul Samuelson and Eugene Fama. In 1965, Samuelson defined an informationally efficient market as one whose prices always include the market participants' information and expectations. Since prices already reflect the relevant information, they cannot be properly forecasted. This summarizes the main idea behind efficient market hypothesis. However, the idea was mostly based on Samuelson's intuition that he had gained while researching stock prices over many years and the mathematics behind the theory were not particularly sophisticated. In 1970, it was then Fama who first used the term "efficient markets": markets where all available information is reflected in the prices. Whereas Samuelson's original idea of efficient markets was based on complicated and inaccessible calculations, Fama based his ideas on simple and comprehensible concepts which are widely known in the world of finance today. Fama defined three different forms of market efficiency: strong, semi-strong and weak (Read, 2013).

The market efficiency is considered to be strong if all available information is reflected in the price of a publicly traded security, whether the information is public or private. In a more practical sense, this would mean that if even a few individuals within a company have insider information about its strategies, they will already have acted on the information by buying or selling the security, and the price has already adjusted to incorporate the insider information. This is however unlikely to be a realistic case since individuals who could potentially have access to sensitive information are often banned from trading with the stock. It is more likely that a semi-strong form of market efficiency exists, which is the case if only public information is reflected in market prices. Public information includes financial statements and other knowledge that the company or regulators have made accessible, which investors can then use to help them determine whether the stock price will go up or down. The weak form of market efficiency assumes that the price reflects the most basic technical information about a stock, such as trading patterns and volume. This form assumes that no type of technical analysis of the stock's historical information will result in a profit for the investor since this type of information will already be included in the price (Read, 2013).

However, economists have started to question this theory that dominated the field just decades ago, arguing that evidence exists that is strong enough to completely reject the efficient market hypothesis. Experts have pointed out quite a few examples of events where public information did not seem to be reflected in market prices, such as the case of the internet "bubble" around the year 2000 that showed irrational behaviour of markets on a significant scale. Additionally, the idea that future stock returns cannot in any way be predicted by using historical returns or other relevant information has been rejected by econometricians. Research has shown that the returns of actively managed mutual funds outperform the Standard & Poor 500 stock index in over 80% of cases over a period of 10 years or longer (Malkiel, 2005). These findings support the idea that outperforming the market is not only possible, but actively carried out by some market participants. A conflicting field to the efficient market hypothesis, behavioural finance, will be discussed in the next chapter as well as the field's explanations for the efficient market theory's weaknesses.

2.2 Behavioural Finance

Behavioural finance introduced a new approach of analysing finance and investor behaviour based on psychology and its role in financial markets. Some of the building blocks of standard finance suggest that people are rational and markets are efficient. One researcher described behavioural finance as “finance with normal people in it”, arguing that people cannot be considered rational since they are prone to cognitive errors and emotions (Statman, 2014). The forefathers of behavioural finance, psychologists Tversky and Kahneman, suggested in their work from 1979 that cognitive errors, or heuristics, could explain the rationale behind investors’ judgment under uncertainty. Examples of those heuristics include herding which is the act of following others’ lead instead of analysing available information, loss aversion which describes people’s tendencies of being risk-averse for losses rather than gains, and regret aversion involving an investor’s aversion to possibly regretting making a decision that results in a loss (Gupta et al., 2014).

Behavioural finance rejects some of the foundations of the efficient market hypothesis, for example the assumption that historical returns and other relevant information cannot be used to predict stock returns as they follow a random walk model. Experts have found that the movements of financial markets are not completely random and have some predictable components. Therefore, certain metrics of valuation as well as historical returns can be used to predict the movements of financial markets to some extent (Malkiel, 2005).

Additionally, the psychological focus of behavioural finance has encouraged experts to explore other potential market predictors beyond technical aspects. Price movements in financial markets reflect investor behaviour which is based on their perception and beliefs about the market. Behavioural finance argues that this behaviour is not random, even though it can be irrational at times. The behaviour of financial markets can for example be explained by investors’ feelings and subjective judgments, as well as how they perceive the economy’s condition at a certain point in time (Fung et al., 2005). Analysing what lies behind investors’ behaviour can therefore provide important insights that can potentially be used to predict market movements. The following section will discuss researchers’ previous attempts at analysing certain information in order to draw conclusions about future market movements.

2.3 Sentiment Research

Using the large amount of data available today, some behavioural scientists have turned to content analysis, a technique that identifies certain text characteristics in a systematic way to draw conclusions. One type of content analysis is sentiment analysis. Experts in behavioural finance have mostly studied two types of sentiment, investor sentiment and textual sentiment.

Researching investor sentiment includes identifying certain beliefs about a stock's future beyond the facts and analysing the potential effects on the financial market. Analysing textual sentiment however involves identifying whether the "tone" of a text is positive or negative, strong or weak, or active or passive. Beyond simply looking at investors' subjective judgments, textual sentiment can also identify market conditions in a more objective manner (Kearney and Liu, 2014).

Some research has been done specifically on media sentiment and its role in financial markets. Tetlock believed he was one of the first researchers to use news media to predict stock market movements. In 2007, he analysed the sentiment of a daily market column in the Wall Street Journal and its relationship to the stock market. He analysed the daily variation in the market column over a 16-year period by using a quantitative content analysis program, constructing what he called a "pessimism factor". Tetlock used a dictionary-based approach, a method that will be explained in more detail in chapter 3.2.1, to sort the words in each market column into certain categories within the Harvard IV-4 psychosocial dictionary. He then counted the number of words that the dictionary defined as pessimistic to construct the pessimism factor. Basic vector autoregressions between the pessimism values and stock prices led him to draw the following conclusions: Firstly, a high level of media pessimism could forecast falling market prices. Secondly, unusually low or high values of media pessimism predicted high trading volume on the market. Thirdly, a high level of media pessimism followed low market returns (Tetlock, 2007). While Tetlock's approach was fairly rudimentary, his results suggest that the media might play a role in the stock market.

In 2012, Schumaker et al. analysed whether there was a correlation between financial news sentiment and stock price movements. Using a prediction system revolving around financial news paired with sentiment analysis, they identified articles as subjective (positive or negative) and neutral. The analysis found that subjective news articles could predict the direction of a stock better than chance alone. However, in contradiction to Tetlock's previous work, the analysis found that positive articles were better predictors of downswings in stock prices, and negative articles were better predictors of upswings in stock prices. The researchers believed that this was the results of investors' tendency to respond to positive news about a stock by selling it and responding to negative news about a stock by buying it. Schumaker et al. suggested that additional research could offer more thorough explanations of these effects (Schumaker et al., 2012).

Although the forementioned literature identifies a relationship between news article sentiment and stock price movements, it has the common feature of not having a high accuracy. Due to the complexities of financial markets, it is common for forecasting accuracy to range between 55% and 65% (Zhou et al., 2020). Accuracy and other metrics of performance will be covered in more detail in chapter 3.4 but simply put, previous literature has suggested a weak but significant relationship between the sentiment of financial news and stock price movements. The basis of this thesis is the idea that a similar relationship might exist between the financial news in Norwegian media and the stock price changes of companies listed on Oslo Børs. The next section will discuss the data and methodology, including text mining, sentiment analysis and the statistical methods that will be used to attempt to identify said relationship.

3. Methodology

This section will cover the collecting and pre-processing of the data at hand, the technicalities and implementation of the sentiment analysis, the classification methods used, and the performance metrics chosen to evaluate the results.

3.1 Data

The data in this analysis consists of textual and numeric data. The textual data is composed of Norwegian news articles from the popular newspaper Dagens Næringsliv, or DN, as well as a list of companies registered on Oslo Børs for all years in a chosen period of time. The time period of interest in this analysis is the years 2014-2019, chosen to capture observations during a relatively normal time period where no large events, such as the financial crisis of 2008 or the Covid-19 pandemic of 2020, could skew the results. Oslo Børs stocks are divided into five separate market segments: OBX, OB Match, OB Standard, OB Nye and OB Egenkapitalbevis (Finansleksikon, 2022). The analysis will consist of companies with shares in the first four segments, excluding the category of OB Egenkapitalbevis or equity certificates. A description of each segment can be seen in Table 1:

Market segments of Oslo Børs	
OBX	25 most liquid shares that make up the OBX index
OB Match	Shares with a minimum of 10 official daily trades on average (excluding OBX shares)
OB Standard	Shares that do not fulfil the conditions of OBX and OB Match
OB Nye	Newly listed shares
OB Egenkapitalbevis	Issuers of equity certificates

Table 1. Market segments of Oslo Børs (Finansleksikon, 2022).

The numeric data consists of stock prices from the companies registered on Oslo Børs between 2014 and 2019. The stock prices are retrieved from Yahoo! Finance based on the tickers of the companies with shares in the market segments of interest. The analysis will attempt to identify a relationship between the sentiment of news articles written on a given day, *t=0*, and the stock price changes between that day and the following business day, *t=1*. Therefore, the dependent variable is the stock price direction, and the independent variable is the sentiment score that will be constructed.

3.1.1 News Article Retrieval

To identify the sentiment of news articles and analyse its potential effect on stock prices, the news articles are extracted from DN's online news article collection. DN is one of the five largest online newspapers for most of the time period between 2014 and 2019. The main reason that DN was chosen over other newspapers is that it is one of the most read newspapers for financial news in Norway (Medienorge, 2022).

The analysis is built on financial news that appear when certain conditions are put into DN's search page. It is possible to choose between different *publications, contents, categories,* and *time periods* on the search page. The *publication* selected, DN, contains news articles while other *publications* such as D2, Smak, Magasinet and DNX contain articles that are more focused on lifestyle. It is also possible to choose a type of *content* with options like articles, advertising, recipes, and videos. In the case of this analysis, only articles will be chosen. There are also several different *categories* available for choosing. The *categories* that contain news articles that discuss the stock market are finance (finans), stock market (børs), Oslo stock market (Oslo Børs), trade (handel) and working life (arbeidsliv). There are 18.359 articles within these choices of *publication*, *content,* and *categories* during the chosen *time period* of 2014-2019.

Web scraping is used to scrape data from websites by directly accessing information within a web browser. DN's website is built using HTML code which can be read directly by using R functions that extract the information needed. The first step is extracting the dates and URLs for each article. The dates are later used to match articles to specific dates and companies, a process covered in more detail in chapter 3.1.4. The second step is extracting the text from each article, including DN+ articles which are only accessible to those who use DN's subscription-based service. A preview of the extracted data from DN is presented in Table 2 for further analysis.

Text	Date	URL
Jeg tror aldri vi har handlet fyrverkeri for så mye før. Vanligv...	2019-12-30	https://www.dn.no/handel/nyttarsaften/handel/fyrverkeri/tr...
Forvalter Jan Petter Sissener har gjennom SEB Prime Solutio...	2019-12-30	https://www.dn.no/marked/kjoper-seg-opp-i-bank/2-1-730...
Veidekke skal bygge kunstsenter i Trondheim. Kontrakten ha...	2019-12-30	https://www.dn.no/marked/bygger-kunstsenter/2-1-730575
I rekrutterernes floskelpregede verden skinner det av dyp o...	2019-12-30	https://www.dn.no/innlegg/rekruttering/sprak/jobbsoknad/i...
Idet og tiåret nærmer seg slutten kan forvaltere og investor...	2019-12-30	https://www.dn.no/bors/oslo-bors/norge/europa/sikre-pa-a...
Den siste tiden har sentralbanksjef, Mark Carney i Bank of En...	2019-12-30	https://www.dn.no/marked/storbritannia/mark-carney/fn/br...

Table 2. Data frame with news articles, retrieved from dn.no.

3.1.2 Pre-Processing Text

Pre-processing is the task of preparing the data for further analysis. This is an important step which might affect the final results. For DN’s news articles, the pre-processing is mostly revolved around removal and tokenization. Removal includes removing duplicate articles and URLs, broken URLs, HTML code, numbers and special characters. Removing duplicated articles and URLs is done by searching for duplicates both in the URLs and the text in each article after the extraction. While extracting text from each URL, there are some URLs that do not retrieve any text and are therefore no longer functioning. This includes articles that DN has removed and articles with only pictures and no text. The rows with duplicated URLs or articles and rows containing broken URLs are removed. After completing these pre-processing steps, 18.014 articles remain. When extracting articles from DN’s website, HTML code, numbers and special characters are included in the text. The HTML code makes the text difficult to read and includes words and phrases that should not be included in the sentiment analysis. Numbers and special characters also contain no value for the sentiment analysis. Therefore, removing HTML code, numbers and special characters is necessary before further analysis is conducted. Additionally, upper-case letters in the articles are changed to their lower-case forms. This is done since stopwords and sentiment dictionary words consist of lower-case letters only.

Removing stopwords from a text is often an important step when working with textual data. Stopwords are words that do not add any meaning to a text, for example: “and”, “but” or “that”. Words that do not carry any specific sentiment do not add value to the text when performing a sentiment analysis and can be considered noise. Stopwords are therefore commonly removed when a text is pre-processed (Saif et al., 2014). In this instance, stopwords are removed by utilizing a list of Norwegian stopwords provided by the R package *stopwords*. It is important to note that in some cases, stopwords do add meaning and removing them can sometimes result in an unreliable sentiment score. An example of this would be the sentiment of the following sentence: “I didn’t like the movie”. The sentence is clearly a negative statement but when English stopwords within the *stopwords* R package are excluded, it becomes “like movie” which then has a positive sentiment. This can however be avoided by checking if an overlap exists between the stopwords dictionary and the sentiment dictionary. When the list of Norwegian stopwords has been retrieved, it is compared to the translated LM sentiment dictionary, explained in more detail in chapter 3.2.1. The word lists are compared to identify any overlapping words. Since no overlap is found, the Norwegian stopwords are removed from the articles.

To perform a sentiment analysis on the news articles, the articles must be changed from a character format to a corpus format. Corpus is often also called dataset in machine learning, but corpus is preferred when referring to a collection of data that consists of text. The corpus is then tokenized, which is the process of going from a full text to a list of words. This will create character strings called tokens. An example of this is the sentence: “mens hovedindeksen falt”, which will turn into a list of separate words: “mens”, “hovedindeksen”, “falt”.

3.1.3 Stock Data

Working with the stock data is split into a few parts. Firstly, files with the names and tickers of companies listed on Oslo Børs from 2014 to 2019 are retrieved and read into R. Secondly, the names and tickers are manually adjusted to possible changes that have occurred in the time since the files were published. Next, the company names are changed to the versions they are more commonly referred to. Then, the tickers are used to retrieve historical stock prices from Yahoo! Finance. Lastly, the stock prices are used to calculate the daily average price for each company, as well as creating a column in the data that indicates whether the daily average price has gone up or down from the previous business day.

The webpage of Oslo Børs contains annually updated statistics, including files with each year's main figures for companies listed on Oslo Børs. Each file contains a column with the names of Oslo Børs-listed companies in a given year, and a column with each company's respective ticker. The files for the years 2014-2019 are downloaded and read into R. As explained in chapter 3.1, only the names of companies listed under OBX, OB Match, OB Standard and OB Nye are of interest. Therefore, only the names and tickers of companies with shares in those market segments are retrieved. Since many companies are listed in more than one year, the unique company names and tickers are retrieved. This process leaves 258 company names and tickers.

Next, a manual search for each ticker and company names is done since some might have changed since the files were created. This is the case with *Aker BP*'s ticker for example, which has been changed from *AKERBP* to *AKRBP*. When a company's ticker cannot be found, a search is conducted to attempt to find an updated ticker. Examples of the ticker alterations made are shown in Table 3.

Old ticker	**New ticker**
AKERBP	AKRBP
AKA	AKAST
ARCHER	ARCH
ASC	ABG
ASETEK	ASTK
AVANCE	AGAS

Table 3. Examples of ticker changes made.

Since only the stock prices of companies with active tickers can be retrieved, companies with tickers that are inactive today are dropped. Therefore, the analysis is subject to some survivorship bias; a problem that a researcher faces when firms that no longer exist are excluded from an analysis. This is often seen in performance research where the performance of smaller firms cannot be accurately assessed since they are more likely to have folded before the analysis is performed (Rohleder et al., 2011). In this case, folding is not necessarily the reason for an inactive ticker. Tickers can also become inactive if a company is deregistered from Oslo Børs or if the company has been acquired by another company which has an active ticker. Some survivorship bias is however inevitable since other methods of retrieving stock prices are infeasible and would result in even less data.

After searching for and altering the Oslo Børs tickers, a similar search is conducted for each company name. Some companies have had different names during the relevant time period and therefore need to be altered. This is the case with a company listed in all files before 2018, *Statoil*. A transition from an oil company towards a broader energy company prompted the company's name change in 2018 to its current name, *Equinor* (Equinor, 2018). In Oslo Børs files published before 2018, the company is referred to as *Statoil* but as *Equinor* in 2018 and onwards. It is therefore necessary to include both *Statoil* and *Equinor* in the list of companies that will be searched for, and combine the search results of those companies later. Similar cases are identified by researching each company's history to identify any name alterations that might have happened between 2014 and 2019. Additionally, certain companies are known under two names, such as *Norwegian Energy Company* which is also known as *Noreco*. In this case, the company is listed in the Oslo Børs under *Norwegian Energy Company*. The company's alternative name *Noreco*, along with alternative names for 13 other companies, will later be added to the list of companies that will be searched for in DN's articles.

It is also important to alter some company names to maximize the results obtained. This is done by manually searching for the names on DN's search page. Between the years 2014 and 2019, searching for the name *Lerøy Seafood Group* on DN's website in all categories gives 286 articles. However, when the name is shortened to *Lerøy*, the number of articles increases to 758. Reading through a few articles that come up in the search results for *Lerøy* shows that the articles are indeed about the company *Lerøy Seafood Group*, but the shorter version of the name seems to be more commonly used. A search for each company name is conducted on DN's website to identify how DN's journalists most commonly refer to the companies. This method is applied to companies with names that can be shortened, and each company's website is checked to identify how the company refers to itself. Examples of the company names before and after the alterations are shown in Table 4.

Old company names	**New company names**
Norwegian Air Shuttle	Norwegian Air
Yara International	Yara
ABG Sundal Collier Holding	ABG Sundal Collier
AKVA Group	AKVA
BW Offshore Limited	BW Offshore
SAS AB	SAS

Table 4. Examples of company name changes made.

Companies that are likely to cause confusion when searched for in DN's articles are then removed. For example, both *Aker* and *Aker BP* were registered on Oslo Børs between 2014 and 2019. Because the names are so similar, all articles that include the name *Aker BP* will also include the name *Aker*. This will result in search result inaccuracy and *Aker* is therefore removed. The same is done with companies like *SpareBank 1* and *B2Holding* since all numbers have been removed from the articles. Lastly, names of companies listed with two share classes with different rights are removed. For example, *Odfjell ser A* and *Odfjell ser B* are both listed between 2014 and 2019, with the former providing the shareholder with more rights than the latter. To eliminate any confusion with another listed company, *Odfjell Drilling*, the companies *Odfjell ser A* and *Odfjell ser B* are removed and the process is repeated for similar cases. This process leaves 245 companies and tickers.

When the alterations of company names and tickers have been completed for all companies listed on Oslo Børs between 2014 and 2019, the R package *BatchGetSymbols* is used to retrieve the daily stock prices for the obtained tickers from all business days between 01.01.2014 and 31.12.2019. Note that stock price observations on days when trading is closed, such as weekends and holidays, do not exist. A feature of this retrieval method is a threshold for bad data between 0 and 1, meaning that the user can define how much data needs to be available for the prices of each company to be downloaded. If the threshold is 0.75, companies that have price information for less than 75% of the defined time period are skipped. In this case, the threshold will be set to 0 to retrieve the largest possible number of observations. As previously explained, prices of companies with inactive tickers cannot be retrieved which results in some companies from the list being dropped. This leaves 194.826 price observations for 155 companies in total. At this point, a list of companies is constructed including the 155 companies with stock price information, as well as the alternative company names that 14 of the companies are known under.

When the stock prices have been retrieved, a column is created in the data that shows the average price for each company on each date. This is done by calculating the mean of two columns: *Price.open* which is the opening price of a stock on a given date, and *Price.close* which is the closing price of a stock on a given date. This column is called *Av.price*. Then, a *Difference*-column is created that subtracts the previous business day's average stock price from today's average stock price. Lastly, a *Direction*-column is created to indicate whether the average price is lower or higher today than it was on the previous business day. The value "up" is entered into the *Direction*-column if the value in the *Difference*-column is higher than

0, meaning that today’s price is higher than the previous business day’s price, and the value “down” is entered into the *Direction*-column if the value in the *Difference*-column is lower than 0, meaning that today’s price is lower than the previous business day’s price. Since only changes in stock prices are of interest, observations where the value in the *Difference*-column is exactly 0, meaning that today’s price is exactly the same as the price of the previous business day, are removed. Additionally, NA-values are added for the first *Direction*-observation of each company. This is done to make sure that the last price observation of one company and the first price observation of another company are not calculated together in the *Direction*-column. After removing NA-values, 176.214 observations remain. A preview of the data can be seen in Table 5:

Company	**Ticker**	**Date**	**Price.close**	**Price.open**	**Av.price**	**Diff**	**Direction**
akastor	AKAST	2014-01-03	28.977	29.278	29.128	-0.314	down
akastor	AKAST	2014-01-06	27.854	28.813	28.333	-0.794	down
akastor	AKAST	2014-01-07	27.799	28.073	27.936	-0.397	down
akastor	AKAST	2014-01-09	28.429	27.964	28.196	0.260	up
akastor	AKAST	2014-01-10	28.758	28.429	28.594	0.397	up
akastor	AKAST	2014-01-13	29.141	28.977	29.059	0.465	up

Table 5. Data frame with stock prices.

3.1.4 Matching Articles to Stock Prices

To analyse the effects and prediction power of news article sentiment on the following business day’s stock prices, each article needs to be matched to the company that it discusses and the relevant company’s stock prices. The company names chosen in the process described in chapter 3.1.3 are searched for in each article. If company names are found in the article, they are pasted into a new column. Note that since upper-case letters in the articles have been changed to their lower-case forms, the same is done to the company names when conducting the search. The company names are searched for with a word boundary on each side, meaning that the company name must be a separate word in the text. For example, searching for the company name *SAS* in lower-case letters without word boundaries would identify all instances of the letters *sas* in the text, such as in the lower-case version of the word *USAs* which occurs frequently in the articles. Searching for the company names with a word boundary on each side eliminates this issue.

Next, rows with articles where no company name is mentioned are removed. Then, articles where a company is only mentioned once are removed. This is done since articles that only mention a company once are unlikely to have any true coverage of the company in question. For example, an article that mentions the company *Norwegian Air* once is not likely to contain any relevant information that could indicate whether the stock price of *Norwegian Air* will rise or fall on the following business day. An assumption is made that an article with relevant information about a company will mention the company at least twice, therefore articles that mention a single company only once are removed.

The process continues with finding the most mentioned company for each article. This is done by counting the number of times each company is mentioned in each article and then allocating each article to the company mentioned most often in the article. The articles with the company names and dates are then merged with the stock prices of matching companies and dates, going from 18.014 to 4.127 observations of articles and stock prices. In some instances, more than one article has been written about a company on the same date. If a company has been allocated multiple articles on the same date, all those articles are combined into one text. Note that the search results for companies known under two names are combined at this point. For example, the articles that discuss *Noreco* are matched to the stock price information of *Norwegian Energy Company*. Because not all companies have articles written about them in the relevant time period, the number of companies has now dropped to 126.

A final tweak of the data needs to be made as the articles and stock prices are merged. Since the goal is to see if there is a relationship between the sentiment of an article published on one day, *t=0*, and the stock price movement of the company discussed in the article between that day and the next business day, *t=1*, the articles must be moved one business day ahead. For example, the sentiment of an article published on 01.01.2019 discussing *Equinor* is thought to perhaps have a relationship with *Equinor's* stock price change between 01.01.2019 and 02.01.2019. The stock price change between these two dates can be found in an observation labelled 02.01.2019. To facilitate the date difference and compare the article from 01.01.2019 to a stock price change observation on 02.01.2019, the date of each article is moved one business day forward. To account for articles landing outside of business days, the articles written on Fridays are moved three days ahead, to Monday, and articles written on Saturday are moved two days ahead, to Monday. This step is necessary to make sure that articles are not moved to Saturdays and Sundays, which would then cause them to be lost when the articles are combined with the stock prices since no stock observations exist on weekends. Table 6 shows examples of the date alterations made on the articles in order to move the date to the following business day.

Written on weekday	Written on date	Moved to weekday	Moved to date
Sunday	2019-12-15	Monday	2019-12-16
Monday	2019-12-16	Tuesday	2019-12-17
Tuesday	2019-12-17	Wednesday	2019-12-18
Wednesday	2019-12-18	Thursday	2019-12-19
Thursday	2019-12-19	Friday	2019-12-20
Friday	2019-12-20	Monday	2019-12-23
Saturday	2019-12-21	Monday	2019-12-23

Table 6. Date alterations of articles.

After the alterations have been made and the data has been merged, a data frame has been created that contains four relevant columns. An example of the data is displayed in Table 7 below.

Date	Company	Text	Direction
2014-02-06	storebrand	kollapsen aksjemarkedet usa asia natt ...	up
2014-02-12	dnb	synes virket komplisert først kommet ...	up
2014-02-26	bw lpg	dnb markets artic fund management bwg ...	down
2014-03-04	aker solutions	bussen redd nærkontakt sier camilla ...	up
2014-03-05	telenor	uroen ukraina tok løpet helgen russiske ...	up
2014-03-11	dnb	forrige uke tok articforvalter albert collett ...	down

Table 7. Example of articles after pre-processing, retrieved from dn.no.

Firstly, the *Date*-column represents the business day following the day that the article was written on, as described in Table 6. Recall that since stock price information is only available on business days, articles that have been moved to a non-trading day, such as an official holiday, are dropped when the data is merged. The *Company*-column shows which company the article has been allocated to, based on how many times it is mentioned in the article. For example, the first article has *Storebrand* in the *Company*-column. This means that *Storebrand* was the company name mentioned most often in that article out of all Oslo Børs-listed companies. The *Text*-column shows the text in the article after it has been pre-processed like chapter 3.1.2 describes. Finally, the *Direction*-column shows whether the stock price of the company in the *Company*-column went up or down from the previous business day from the date in the *Date*-column. The following chapters will discuss sentiment analysis and how it is implemented in this analysis.

3.2 Sentiment Analysis

To assign score to textual data in a way that describes the author's attitude is referred to as sentiment analysis. By using a collection of words that have been assigned a numeric value based on their sentiment, for example by assigning negative values to negative words and positive values to positive words, one can create an output of a score that represents the text's overall tone or sentiment on a numeric scale (Kudyba, 2014). This analysis is based on a dictionary-based sentiment score. The technical details and implementation of this type of analysis are explained in the following chapters.

3.2.1 Sentiment Dictionary

When working with a significant amount of data, for example news articles, it is much easier to implement methods that automatically determine the sentiment of a text instead of manually appointing certain words a positive or negative word score to calculate a sentiment value. One way to do this is using a dictionary-based method. This method determines polarity in data by matching words in a text to words in a word list known as a sentiment lexicon; a dictionary that provides a score for each word based on its sentiment (Hardeniya and Borikar, 2016). Sentiment lexicons can be split into negative and positive words but can for example also include word categories such as strong modal, weak modal, uncertain and litigious. Some of the commonly used sentiment lexicons are the Harvard-IV-4, SentiWordNet and MPQA Subjectivity Lexicon (Loughran and McDonald, 2011). However, most sentiment lexicons do not take financial context into consideration.

Loughran and McDonald conducted some research on the use of non-financial sentiment dictionaries in financial text and created an alternative sentiment lexicon with a focus on business. They used the Harvard dictionary, which was developed for psychology and sociology, as the non-financial part of the sentiment dictionary. In their research, they found that almost three-fourths of negative words in the Harvard dictionary are not usually considered negative in a financial context. Their research conclusion includes a caution against using non-financial sentiment dictionaries in a financial context. Based on these results, Loughran and McDonald constructed a financial sentiment dictionary, commonly referred to as the *Loughran-McDonald sentiment lexicon* or LM dictionary (Loughran and McDonald, 2011). Because of the financial context of the news articles from Dagens Næringsliv, the LM dictionary will be used in this analysis.

The LM dictionary is originally in English and consists of 2.702 words, 2.350 negative and 352 positive words. It is therefore necessary to translate the dictionary to Norwegian for the purposes of this analysis, which is done by using Google's translate API. This method of translating has the drawback of creating translations that are more than one-word, and duplicated words. For example, for the word "able", the translated Norwegian version actually consists of two words: "i stand". Additionally, both "abandon" and "abandoning" are translated to the Norwegian word "forlate". The first observation of the duplicated word will be the one that is not removed from the dictionary. As a result of the removal of more than one-word translations and duplicated words, the translated LM dictionary is left with 1.849 words, 1.570 negative and 279 positive words. A sample of the translated lexicon is shown in Table 8 where the words are assigned a score, -1 or 1, to indicate whether the word is negative or positive.

Word	Score
påstand	-1
fravær	-1
overflod	1
rikelig	1
misbruke	-1
forsikre	1

Table 8. Sample of the translated Loughran-McDonald sentiment lexicon.

Since the LM dictionary both before and after translating contains a much larger number of negative words than positive, the probability of finding negative words in a text will be larger than the probability of finding positive words. However, along with being widely used, this dictionary was developed specifically to include negative words that are generally not considered to be negative outside of the world of finance (Loughran and McDonald, 2011). It is therefore preferred over other dictionaries that do not take financial context into account and the uneven ratio between the number of positive and negative words is accepted as an unpreventable limitation. Additionally, even though many words are lost from the dictionary when it is translated, the ratio between positive and negative words is almost the same for the English and Norwegian versions. The translated version of the LM dictionary is therefore deemed an acceptable sentiment lexicon for this analysis.

When performing a sentiment analysis with a dictionary-based approach, there are a few possibilities on how to compute the actual sentiment score for a text. In the following section, examples of methods to calculate sentiment scores are discussed as well as their advantages and disadvantages.

3.2.2 Implementation of Sentiment Analysis

Previously assigned to a company and tokenized, the articles are now assigned a sentiment score. To create the sentiment score, a loop is used to check each token in each article to see if it matches any of the positive and negative words in the LM dictionary. For example, assume that the headline of an article is “TGS Nopec aksjen falt”, or “TGS Nopec stock fell”. This article has previously been assigned to the company TGS Nopec. Each word is now compared to the LM dictionary to see if there are any matches. The only word in the headline that has a match in the LM dictionary is “falt”, with a score of –1. The rest of the words get scores of 0. The headline is scored as shown in Table 9:

TGS	**Nopec**	**aksjen**	**falt**
0	0	0	-1

Table 9. Words scored by sentiment.

A few formulations were considered to construct a sentiment score, for example dividing the number of positive words by the number of negative words as seen in sentiment score (3.1):

$$sentiment\ score = \frac{number\ of\ positive\ words}{number\ of\ negative\ words} \tag{3.1}$$

This method however has the drawback of overestimating the power of positive words. In an extreme case, such as if the number of positive words in a text is 1 and the number of negative words in a text is 10, the sentiment score would be $\frac{1}{10} = 0.1$. However, one extra positive word would double the sentiment score to $\frac{2}{10} = 0.2$. One extra negative word would not have the same effect as it would result in a sentiment score of $\frac{1}{11} = 0.09$. The same would occur if the position of positive and negative words were switched, except negative words would then be more influential than positive words. Therefore, this method is regarded as insufficient for the purposes of this analysis.

Another method would include calculating the sum of the word scores and divide it by the total word length as is done in sentiment score (3.2) below:

$$sentiment\ score = \frac{\sum word\ \ scores}{text\ length} \tag{3.2}$$

This method has the advantage of calculating sentiment score as a ratio between –1 and 1, with a text only consisting of negative words having the score of –1 and a text consisting only of positive words having the score of 1. These extremes are however unlikely to happen so it is likely that the range of sentiment scores calculated with this method will be very small, which could affect the results when stock price changes are predicted. This method also has the drawback of penalizing longer articles by lowering their total sentiment score. For example, imagine a text with 20 positive words and 10 negative words. The total word length of the text is 100. The sentiment score calculated with this method would be $\frac{20-10}{100} = 0.1$. However, if equally as many positive and negative words are present in another text that is 200 words long, the sentiment score would be much lower, or $\frac{20-10}{200} = 0.05$. Therefore, a text with 20 positive and 10 negative words with a total word length of 100 will have double the sentiment score of an equally positive and negative text of 200 words, even though the reader would perhaps observe the same sentiment in both texts due to their equal number of positive and negative words.

As the examples above show, different methods of calculating a sentiment score have strengths and weaknesses. Keeping in mind that there is no correct or incorrect method to calculate the sentiment score of a text, the sentiment score in this analysis is calculated by simply calculating the sum of the word scores as shown in sentiment score (3.3) below:

$$sentiment\ score = \sum word\ \ scores \tag{3.3}$$

The calculation method in sentiment score (3.3) is deemed to accurately represent the text's sentiment while avoiding the drawbacks of the methods previously explained. With this method, positive words do not have more value than negative words or vice versa, which was a drawback of the method shown in sentiment score (3.1) above. Additionally, this method does not penalize longer articles by lowering their total sentiment score, which was a drawback of the method shown in sentiment score (3.2) above.

Although the method shown in sentiment score (3.3) avoids the drawbacks of the methods previously discussed, it is not without fault. For example, since only positive and negative words are identified in the text, the benefits of performing some pre-processing steps, such as removing stopwords, are lost. However, all pre-processing steps were necessary in the case of this analysis since a few methods of calculating sentiment score had to be tested before one was selected. As explained in chapter 3.2.1, using the LM dictionary has the drawback of having a large number of negative words and a small number of positive words. An optimal method would perhaps neutralise the overestimation of negative words when using this dictionary. However, no such method was found in previous literature. After comparing all three versions of the sentiment scores with the stock price movements using the statistical methods described in the following chapters, the method in sentiment score (3.3) was chosen since it provided the most accurate predictions of stock price movements without significant methodical drawbacks.

A final alteration of the data needs to be made at this point. Considering the findings of previous researchers that subjective news articles are better at predicting stock price movements than chance alone, it is only logical to remove neutral news articles from the data. This means that articles where the sentiment score is exactly zero are removed, which brings the final number of articles and matching stock prices from 4.127 to 3.516. This also causes the number of companies to drop from 126 to 124, meaning that two companies do not have any subjective articles written about them in the relevant time period. Although further data limitations can result in a worse performance, this was not the case when articles with a neutral sentiment were removed from the data since doing so improved predictions when tested.

At this point, it has been established which company each article covers, and each article has been assigned a sentiment score. This concludes the textual analysis part of the thesis. The following chapters will explain the statistical methods performed to answer the research question about whether a relationship exists between the sentiment of news articles and stock price movements in the Norwegian market. The theory behind each method will be explained as well as the implementation, followed by an introduction of the performance metrics used to evaluate the results of the analysis.

3.3 Classification

Supervised learning is a subcategory of machine learning, defined as predicting a response variable based on one or more predictor variables by training a statistical model. Two types of problems are included in supervised learning: regression and classification. Regression is used in many situations when the response variable is quantitative (James et al., 2021). Examples of quantitative variables are gross domestic product or years of education completed, which are both represented on a numeric scale. When the response variable is quantitative, the goal is often to predict the expected value of the dependent variable using the independent variable. However, the response variable is often qualitative, also referred to as categorical. Qualitative response variables can for example be which political party an individual will vote for or whether an individual is in debt. When working with qualitative variables, it is often the goal to predict the probability of the variable, for example the probability that an individual has tried a sport. Predicting qualitative response variables can be done with classification by assigning observations to factors, categories or classes (Gujarati, 2004).

Building a classifier, like in the regression setting, requires a set of training observations containing the dependent and independent variables used to train the model. This analysis uses the commonly used classifiers logistic regression and K-nearest neighbors, as well as the more computer-intensive classification methods gradient boosted trees and support vector classifier (James et al., 2021). Using only one method to analyse the data invites the possibility of wrongly rejecting the hypothesis that a relationship exists between media sentiment and stock price movements because of a bad choice of statistical method. Therefore, four methods are implemented to compare methods and find the best method for the data at hand.

To test the relationship between dependent and independent variables and evaluate the performance of the predictions made, the data is split into training and test sets. The training data is used to fit the model, essentially teaching a statistical learning method to predict an output based on one or more inputs. When the model has been trained, the performance of its predictions can be evaluated using the remaining data not used to fit the model, or the test data. The prediction performance can then be assessed by evaluating how well the model predicts a response on an entirely new observation in the test data (James et al., 2021). The performance metrics used in this analysis are discussed in chapter 3.4. In this analysis, the training set consists of 90% of the observations, or 3.164 observations, and the test set consists of the last 10% of observations, or 352 observations. This split was chosen since different splits, for example 80/20 or 70/30, provided worse results when tested with the statistical methods.

The split between training and testing data is also used to perform cross-validation, a type of resampling method that repeatedly draws samples from a training set and fits a model for each sample. Cross-validation holds out a subset of the training set and then performs a statistical method on the sample, obtaining information about each fitted model. There are several different cross-validation methods available, for example the leave-one-out and k-fold cross-validation. In this analysis, k-fold cross-validation is used. This method randomly divides a set of observations into *k* groups, also called folds, of roughly the same size. The first fold is the validation set and the statistical method is fit on the remaining folds. When classification is performed, the performance metric calculated on each fold is accuracy. This results in *k* accuracy estimates, which are then averaged to find the k-fold cross-validation estimate (James et al., 2021).

The k-fold cross-validation method can provide better estimates of the accuracy than, for example, leave-one-out cross-validation, or LOOCV. This can be explained with the bias-variance trade-off. The LOOCV method is often preferred from the perspective of bias reduction since each training set uses all observations except the one that is left out, giving it more or less an unbiased estimate. On the other hand, k-fold cross-validation has a lower variance than LOOCV since k-fold cross-validation finds the average estimate of *k* groups that are less correlated to each other than when LOOCV is used. K-fold cross-validation also requires less time for computation than LOOCV. Usually, k-fold cross-validation using 5 or 10 folds has been shown not to suffer from high bias or high variance. (James et al., 2021). This analysis therefore uses k-fold cross-validation with 10 folds.

3.3.1 Logistic Regression

Despite its name, logistic regression is a classification method. It is an efficient and simple method, making it easy to implement. When compared to other classification methods, logistic regression is generally known to have a good performance (Subasi, 2020). This method models the probability of a binary outcome given an independent variable. Most logistic models have two factors, which are given the classes 0 and 1 (Edgar and Manz, 2017). The functions used in classification give outcomes between 0 and 1 for all prediction values. In logistic regression, the logistic function (3.4) is used. The logistic function (3.4) will always provide a sensible prediction regardless of the value of the dependent variable.

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \tag{3.4}$$

The logistic function produces an S-shaped curve as shown in Figure 1 (James, et al., 2021).

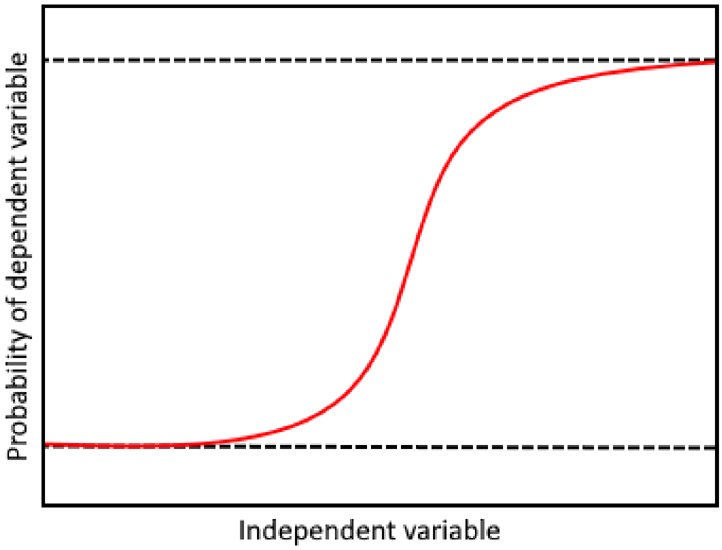

Figure 1. Predicted probability with logistic regression.

By using some manipulation of the logistic function (3.4) and taking the logarithm of both sides of the function, the logit function (3.5) can be found (James, et al., 2021):

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X \tag{3.5}$$

The left side of the logit function (3.5) is called log-odds or logit. The function is used to find the log-odds for each data point. In linear regression, β_1 represents the average change in the dependent variable associated with a one unit increase in the independent variable. However, in logistic regression, increasing the independent variable by one unit changes the logit by β_1. Finally, to estimate the coefficients β_0 and β_1 on the training data, the maximum likelihood function (3.6) is used. β_1 and β_2 are chosen based on the goal of maximizing the likelihood function (3.6). This approach is often used when fitting non-linear models (James, et al., 2021).

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} (1 - p(x_{i'}))$$

(3.6)

3.3.2 K-Nearest Neighbors

In an ideal world, the Bayes classifier would always be used to predict qualitative responses. However, this is not possible since the conditional distribution of the dependent variable given the independent variables are unknown for real data. There are several possible approaches to estimate the conditional distribution and classify each observation into the class which has the highest probability estimate. The K-nearest neighbors method is one of the simplest non-parametric methods that dominates methods such as linear discriminant analysis and logistic regression when the decision boundary is exceedingly non-linear. Using a test observation, x_0, K-nearest neighbors identifies the *K* points in the training data which are closest to the test observation. These *K* points are represented with N_0. Next, it estimates the probability that x_0 belongs to class *j* by calculating the fraction of points in N_0 that belong to class *j*, using the probability function (3.7). Then, the test observation x_0 is classified to the class with the highest estimated probability (James et al., 2021).

$$\Pr(Y = j|X = x_0|) = \frac{1}{K}\sum_{i \in N_0} I(y_i = j)$$

(3.7)

For example, assume that *K* has been chosen to be 5. The K-nearest neighbors will identify the five closest observations to the test observation. Assume that the classes available are *yes* and *no*, and the five observations consist of three *yes* observations and two *no* observations. The probability function will predict the test observation to be classed as *yes* with a likelihood of $\frac{3}{5}$, rather than classing the observation as *no* with a lower likelihood of $\frac{2}{5}$. The K-nearest neighbors classifier can be greatly affected by which value is chosen for *K* and the choice is dependent on the bias-variance trade-off. When *K* is small, the decision boundary is extremely flexible, which corresponds to low bias and high variance. With a larger value of *K*, the classifier gets less flexible and more linear which corresponds to higher bias and lower variance. Choosing the correct *K* level of flexibility is critical in order to achieve good results with K-nearest neighbors (James et al., 2021).

3.3.3 Gradient Boosted Trees

The classification trees method finds the most frequently occurring class of training observations in the section around each observation, predicting that the observation will belong to this class. Although classification trees are even easier to interpret than linear regression, they usually do not have the same accuracy as other statistical methods. However, using ensemble methods like boosting, random forests and bagging can substantially improve the accuracy. These ensemble methods, often referred to as weak learners, are used to improve base methods such as classification trees (James et al., 2021).

Like other classification methods, boosting includes the optimization of a loss function. The model is built by fitting, or “growing”, decision trees on a tweaked version of the original data. The trees are grown sequentially, meaning that each new tree uses information about previous trees to grow. Instead of fitting one large classification tree and risking overfitting, boosting learns slowly. The gradient boosted model (3.8) is computed where *B* represents the number of trees, λ is the shrinking parameter and $\hat{f}^b$ is a tree (James et al., 2021).

$$\hat{f}\,(x) = \sum_{b=1}^{B} \lambda \widehat{f^b}(x) \tag{3.8}$$

3.3.4 Support Vector Classifier

The support vector classifier is a statistical method based on the idea of a separating hyperplane. A separating hyperplane is a boundary that separates the dimensional space that the training observations exist in, with one class on each side of the hyperplane, in the most accurate way possible. A simple example of a hyperplane is shown in Figure 2 where the blue dots represent the observations that belong to class 1, and the orange dots represent the observations that belong to class 2. The blue observations fall on the left side of the separating hyperplane, pictured in red, and the orange observations fall on the right side.

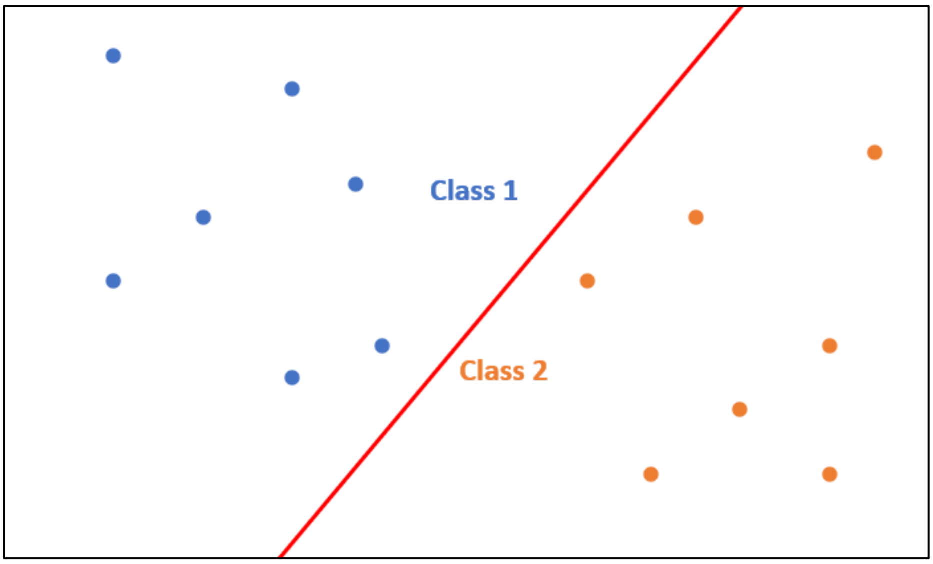

Figure 2. Example of a separating hyperplane.

A p-dimensional hyperplane (3.9) is a subspace of a dimension *p-1,* or in other words one dimension less than the data space. For instance, in two dimensions, the hyperplane is a one-dimensional subspace (James et al., 2021):

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p = 0$$

(3.9)

A separating hyperplane is constructed by separating the training observations to their classes. Maximal margin classifier is one method that implements a separating hyperplane where the hyperplane is the farthest distance from the training observations. The test observations are then classified based on which side of the maximal margin hyperplane they are located. The maximal margin classifier is however very sensitive to individual observations. Therefore, using an alternative method called support vector classifier can counteract this complication. The support vector classifier does not perfectly separate the classes but allows some observations to be on the incorrect side of the hyperplane. This results in better classification for most training observations. The support vector classifier is therefore a natural choice of method in a setting where the separating hyperplane is linear and has two classes (James et al., 2021).

3.3.5 Implementation of Classification Modelling

Implementing each of the four methods, logistic regression, K-nearest neighbors, gradient boosting machines and support vector classifier, is done by fitting statistical models using different tuning parameters. The parameters included in all models in this analysis are *formula, dataset, method* and *resampling method*. Some extra parameters are specific for certain methods only, which will be explained for each method.

Firstly, the *formul*a tells the function what the dependent and independent variables are. They are the same for all methods in this analysis: direction and sentiment. Secondly, the *dataset* provides a specific number of observations for all variables; this is the training set explained in chapter 3.3 for all methods. Thirdly, the type of *metho*d is defined, for example logistic regression. Fourthly, the type of resampling method is given, which is the 10-fold cross-validation for every method.

To implement logistic regression, the *method* parameter is set to fit generalized linear models. This parameter consists of different types of generalized linear models, which is why an extra parameter is included. The extra parameter chooses the *type* of generalized linear model used, which in this case is logistic regression. Logistic regression is chosen by setting *type* as “binomial”.

For K-nearest neighbors, the *method* parameter is set to K-nearest neighbors classification. The number of points closest to the test set, *K*, can be set manually. In this instance, cross-validation chooses the best value of *K* which is equal to 7.

When the *method* is set to fit generalized boosted methods, it requires several extra parameters to be defined. The first parameter chooses the type of generalized boosted method used. In this instance, a Bernoulli distribution is used which is a classification tree, also referred to as a decision tree. The four other extra parameters define how fast the algorithm adapts, sets the number of training set samples, decides on the number of iterations and sets the complexity of the tree. All these parameters are chosen by cross-validation. The values for the four extra parameters are set to 0.1, 10, 100 and 3, respectively.

With support vector machines, a type of *kernel* needs to be chosen. A *kernel* determines the decision boundary and the shape of the hyperplane by quantifying the similarity of two observations (James et al., 2021). The types of *kernels* are linear, radial, or polynomial. In this instance, a linear *kernel* is chosen based on the best fit and accuracy of the support vector classifier model. An extra parameter for the support vector classifier is *cost*, defined as the number of observations allowed to violate the constraints. This parameter is automatically chosen by using cross-validation and is in this instance set to 1.

When the necessary parameters have been defined for all models, predictions are made based on the input data. The input data consists of a model, for example logistic regression, and the test set. The predictions are then used to measure the relationship and performance on the test set. Before the results can be analysed, the performance metrics used are presented. The performance metric that measures the quality of the relationship between the variables is p-value. However, this value only exists in the summary output of the logistic regression method. In order to evaluate the prediction performance for all methods, the performance metrics accuracy, error rate, ROC curve and AUC are used.

3.4 Performance Metrics

A performance metric is, in a statistical representation, a value computed to analyse the performance of a statistical method. These measurements are used to evaluate the different methods and choose the best one for the problem at hand (Ferri et al., 2008). Testing the null hypothesis using a p-value is a common measure of relationship significance in both regression and classification, while commonly used performance metrics for classification are accuracy, error rate, ROC curve and AUC, calculated based on a confusion matrix (Sokolova and Lapalme, 2009).

3.4.1 P-Value

The null hypothesis suggests that there is no statistical relationship of significance between variables (Haldar, 2012). In order to accept or reject the null hypothesis, the p-value can be found to identify the statistical significance of a relationship between the dependent and independent variables. A p-value measures the probability that an observed statistical relationship could be random, on a scale between 0 and 1. A p-value of 0.01 means that there is 1% chance that the relationship between the variables is random. Traditionally, a p-value of below 0.05 is evidence that a null hypothesis can be rejected (Thiese et al., 2016).

This performance metric was chosen since it is a clear measurement of the significance of the relationship between the dependent and independent variables. It is however important to note that the p-value is a metric that cannot be found in all models. In the case of this analysis, logistic regression is the only method that provides a p-value as a measurement of significance in its summary output in R, displayed in appendix A1. P-value will therefore not be displayed when the results of K-nearest neighbors, gradient boosted trees or support vector classifier are discussed.

3.4.2 Accuracy and Error Rate

To define the performance of a classification problem, a confusion matrix is used. The confusion matrix consists of four values: True positive (TP) or correct positive predictions, true negative (TN) or correct negative predictions, false positive (FP) or incorrect positive predictions, and false negative (FN) or incorrect negative predictions (Singh, 2021). An example of a confusion matrix can be seen in Table 10:

		Actual	
		Down	Up
Prediction	Down	TN	FN
	Up	FP	TP

Table 10. Confusion matrix.

Based on TP, TN, FP and FN, several different performance metrics can be calculated. This includes accuracy and error rate. Accuracy (3.10) is a measure of performance represented as the ratio of correctly classified observations, $TP + TN$, to the total number of observations, $TP + FP + TN + FN$. Error rate (3.11) for binary classification calculates the sum of wrongly classified observations, $FN + FP$, and divides it by the total number of observations divided by the number of classes, $\frac{n}{2}$ (Singh, 2021).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \qquad (3.10)$$

$$Error\ rate = \frac{FN+FP}{\frac{n}{2}} \qquad (3.11)$$

Accuracy was selected since it is one of the most popular performance metrics for classification problems with two classes, calculated based on the results of a confusion matrix which evaluates the effectiveness of a classifier. Error rate is more commonly used as a performance metric in multi-class classification problems with more than two classes but it can however be a useful performance measure in binary classification problems as well (Sokolova and Lapalme, 2009). As discussed in chapter 2.3, the accuracy of these types of analyses often ranges between 55% and 65%, representing a weak but statistically significant relationship between sentiment and stock price movements. This is thought to be due to the extremely complicated nature of financial markets (Zhou et al., 2020). An accuracy above 55% is therefore thought to be satisfactory in this type of analysis when compared to relevant literature, supporting the hypothesis that a weak but significant relationship can be found between media sentiment and stock price movements. The error rate will be used as a comparison metric between the different models and no threshold will be placed for an acceptable or unacceptable error rate.

3.4.3 ROC Curve and AUC

Receiver Operating Characteristics curve, or ROC curve, visualises a model's ability to differentiate between binary classes. The ROC curve is formed by using the TP rate against the FP rate as the two parameters plot the curve. The TP rate is also known as sensitivity (3.12), which is the model's ability to identify observations that are true positives. The FP rate is equal to $1 - specificity$, where specificity (3.13) is the model's ability to identify observations that are actual negatives. The sensitivity (3.12) is plotted against *1 – specificity* to construct the ROC curve (Trucco et al., 2019).

$$Sensitivity = \frac{TP}{TP+FN} \tag{3.12}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3.13}$$

The AUC, or the Area Under the Curve, shows how well a model can determine between classes. The closer the ROC curve is to the top left corner and the higher the AUC value, the better the model is at predicting the different classes. The AUC value lies between 0 and 1 and a classifier with an AUC of 0.5 is equivalent to chance. As illustrated in Figure 3, the model with a higher AUC for all feasible thresholds is better than the other models (James et al., 2021), which in this example indicates that model 1 is better than models 2 and 3. In some instances it is harder to analyse which model performs better based on the ROC curve, for example between models 2 and 3, which is where the performance metric AUC comes in.

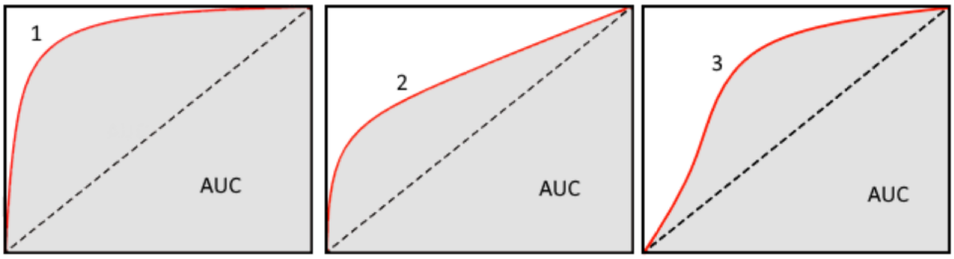

Figure 3. ROC curve and AUC examples.

ROC curve is often used to illustrate the connection between sensitivity and specificity while AUC gives an idea about the benefit of using ROC. Using ROC and AUC as performance metrics comes with both advantages and disadvantages. One disadvantage is that a suitable software is necessary to compute these metrics since they are difficult to calculate without involving some kind of programming. However, an advantage of the ROC curve and AUC is that they are easy to interpret, taking all feasible thresholds into consideration. This makes it an ideal performance metric to use in order to compare different classifiers (James et al., 2021).

Now that the theory behind all relevant performance measures has been explained, the results can be presented. Firstly, the performance of each model is discussed and subsequently, a comparison of the performance of all models is done to conclude which model has the best results. When the best model has been identified, the results will be briefly compared to the findings of previous researchers that have conducted similar analyses on the U.S. market.

4. Analysis

The subsequent chapters will introduce the results of the analysis based on the methodology presented in chapter 3. As discussed in previous chapters, the findings are used to identify a relationship between the sentiment of news articles and the movements of stock prices of Oslo Børs-listed companies. This is done by using different classification methods that can provide information about the relationship between the variables. The classifiers are then used to make predictions of the dependent variable using the independent variable. To find the quality of the predictions made and conclude whether a relationship is strong or weak, several performance metrics are calculated. The results will be presented by looking at the performance metrics for each method. Then, a comparison between the different methods is done, followed by a further exploration of the nature of the relationship between the variables.

Recall that 90% of the observations were used to train the four models, or 3.164 observations. The remaining 10% of observations were used as a test set to evaluate the performance of the models, or 352 observations.

4.1 Logstic Regression Results

As explained in chapter 3.4.1, logistic regression is the only method that provides the performance metric p-value in its summary output in R, displayed in appendix A1. This value can be used as evidence to accept or reject the null hypothesis which states that no statistical relationship or significance exists between variables. A p-value that falls below 0.05 is considered statistically significant and indicates evidence to reject the null hypothesis. Using the model created with the training data, the p-value for the relationship between sentiment and stock price direction is found to be 0.0087, which is low enough to reject the null hypothesis. The low p-value supports the alternative hypothesis that a relationship exists between the sentiment of news articles and stock price movements of Oslo Børs-listed companies. Recall that the p-value will not be provided for the other three classification methods.

The logistic regression model is then used to create predictions for the test data. The values in the confusion matrix for logistic regression, displayed in Table 11, represent the values of true positives, true negatives, false positives, and false negatives explained in chapter 3.4.2.

		Actual	
		Down	Up
Prediction	Down	93	76
	Up	74	109

Table 11. Confusion matrix for logistic regression.

The values in Table 11 are then used to calculate accuracy, error rate and AUC for the logistic regression model. These metrics can be seen in Table 12:

Logistic regression performance	
Accuracy	0.5739
Error rate	0.0853
AUC	0.5730

Table 12. Performance metrics for logistic regression.

Firstly, the performance accuracy of this method is 0.5739 or 57.39%. Recall that accuracy measures the ratio of correctly classified observations to the total number of observations. The accuracy for the logistic regression model exceeds 55% which is the threshold set for a satisfactory accuracy based on previous literature. Secondly, the error rate is 0.0853. Since it is difficult to conclude anything about a model's performance from a single error rate, the error rates of all four models will be compared in chapter 4.5. Thirdly, the AUC is 0.5730 which is also displayed in Figure 4. The red line, representing the ROC curve for logistic regression, lies above the dotted black line, which represents the performance of a model that would predict stock price movements at random:

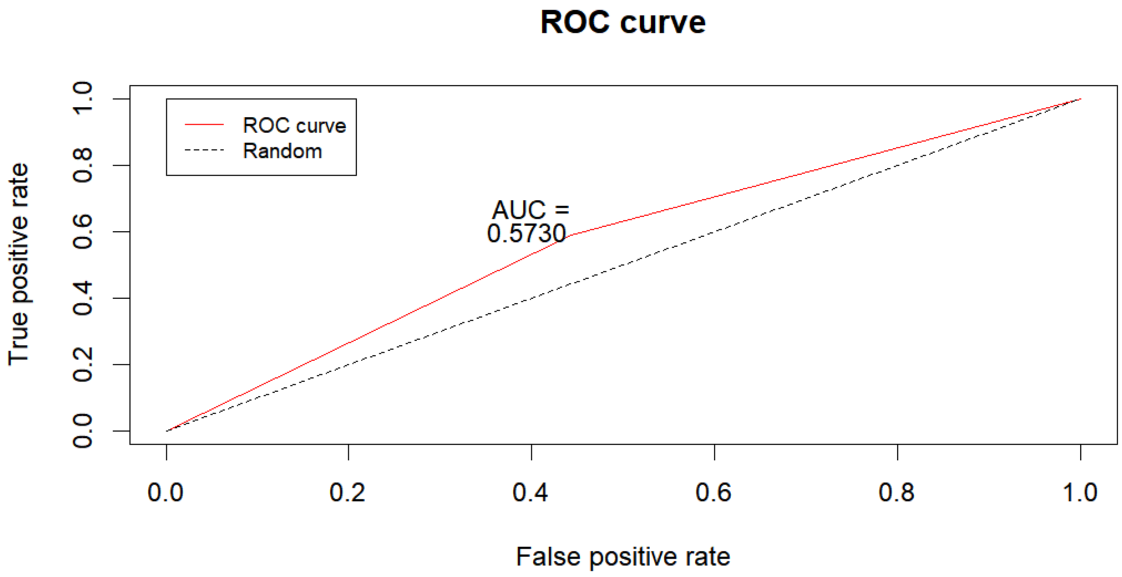

Figure 4. ROC curve for logistic regression.

4.2 K-Nearest Neighbors Results

The confusion matrix for the K-nearest neighbors method is displayed in Table 13:

		Actual	
		Down	Up
Prediction	Down	89	85
	Up	78	100

Table 13. Confusion matrix for K-nearest neighbors.

The performance metrics accuracy, error rate and AUC are then displayed in Table 14:

K-nearest neighbors performance	
Accuracy	0.5369
Error rate	0.0927
AUC	0.5367

Table 14. Performance metrics for K-nearest neighbors.

Firstly, an accuracy of 0.5369 or 53.69% is observed when K-nearest neighbors is used. The accuracy is lower than the accuracy of logistic regression and does not exceed the set threshold of an accuracy of 55%. Secondly, the error rate is 0.0927. This error rate is higher than the error rate of logistic regression. Thirdly, the AUC for K-nearest neighbors is 0.5367. This is lower than the AUC of logistic regression. The ROC curve for K-nearest neighbors is displayed in Figure 5 below. The ROC curve, displayed in yellow, lies slightly above the randomness curve. However, accuracy, error rate and AUC all indicate that the logistic regression model outperforms the K-nearest neighbors model. This will be explored further in chapter 4.5 when all models will be compared.

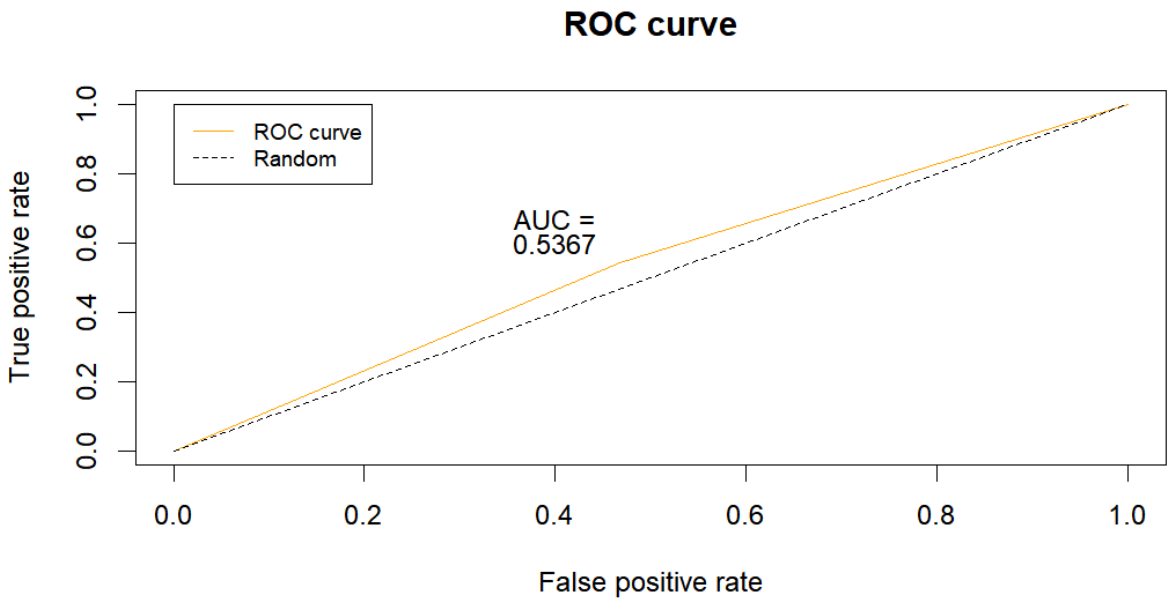

Figure 5. ROC curve for K-nearest neighbors.

4.3 Gradient Boosted Trees Results

The confusion matrix for the gradient boosted trees method is displayed in Table 15:

		Actual	
		Down	Up
Prediction	Down	84	70
	Up	83	115

Table 15. Confusion matrix for gradient boosted trees.

The performance metrics accuracy, error rate and AUC for gradient boosted trees are then displayed in Table 16:

Gradient boosted trees performance	
Accuracy	0.5653
Error rate	0.0870
AUC	0.5623

Table 16. Performance metrics for gradient boosted trees.

Firstly, the accuracy of gradient boosted trees is 0.5653 or 56.53%. This is higher than the accuracy of K-nearest neighbors but lower than the accuracy of logistic regression. Using the gradient boosted trees method results in an accuracy that exceeds the previously set threshold of a sufficient accuracy of 55%, like the accuracy of logistic regression does. Secondly, the error rate is 0.0870. This error rate is slightly higher than the error rate of logistic regression but lower than the error rate of K-nearest neighbors. Thirdly, the AUC is 0.5623 as can be seen on the ROC-curve for gradient boosted trees in Figure 6. The ROC curve lies slightly above the randomness curve as it does for the previous two models:

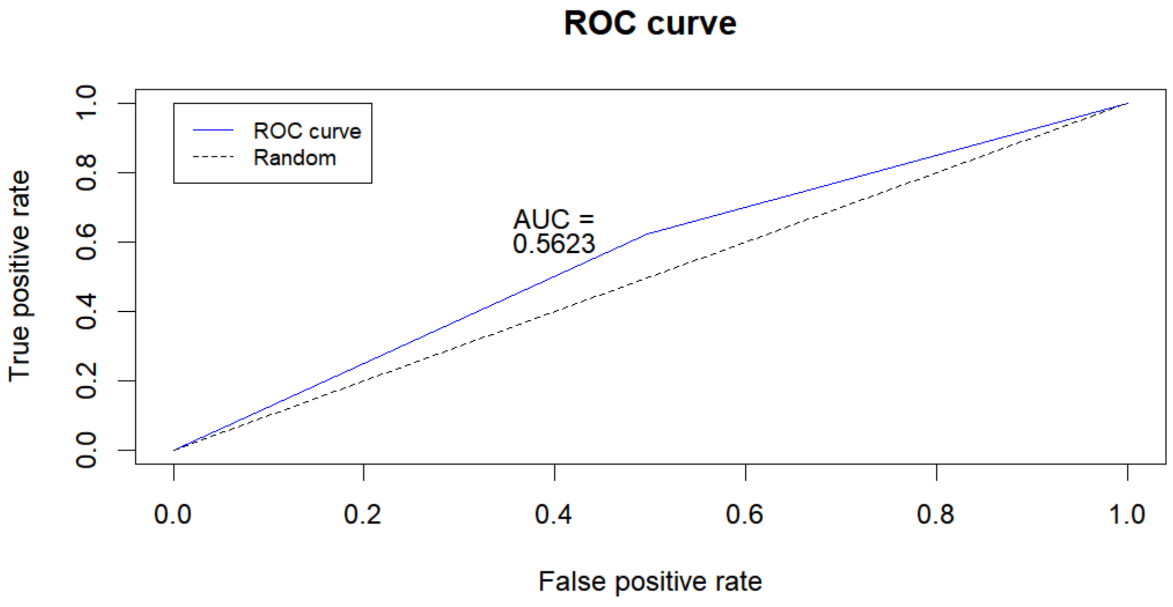

Figure 6. ROC curve for gradient boosted trees.

4.4 Support Vector Classifier Results

In the final support vector classifier model, the training observations are used to find the linear classification boundary. The model is then used to make predictions. The confusion matrix for the support vector classifier method is displayed in Table 17:

		Actual	
		Down	Up
Prediction	Down	64	52
	Up	103	133

Table 17. Confusion matrix for support vector classifier.

The performance metrics accuracy, error rate and AUC for support vector classifier are then displayed in Table 18:

Support vector classifier performance	
Accuracy	0.5597
Error rate	0.0882
AUC	0.5511

Table 18. Performance metrics for support vector classifier.

Firstly, an accuracy of 0.5597 or 55.97% is observed. This exceeds the threshold of 55% that previous literature has shown to be an acceptable accuracy rate for similar analyses. Secondly, the error rate is 0.0882. This is the second-highest error rate of all four methods. Thirdly, the AUC is 0.5511, the second-lowest AUC of all methods. Figure 7 displays the ROC curve of the support vector classifier method in green, which is slightly above the randomness curve.

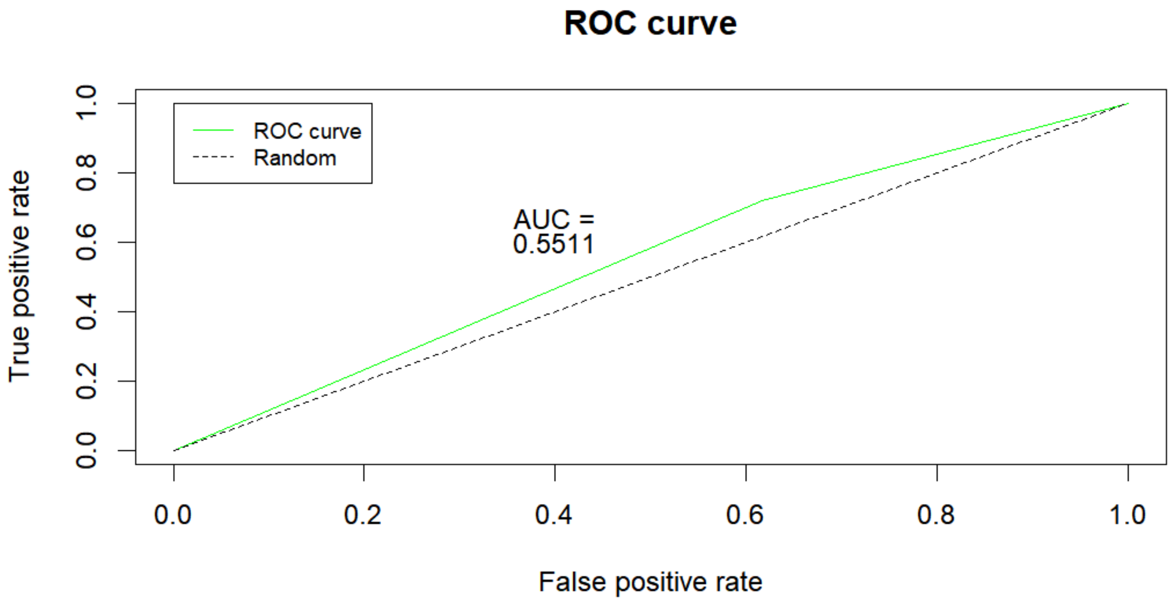

Figure 7. ROC curve for support vector classifier.

Now that the performance of all methods has been evaluated with the relevant metrics, a comparison will be made between the models to identify which model showed the best performance when predicting stock price changes using media sentiment.

4.5 Comparison

A comparison of the performance metrics accuracy, error rate and AUC can be seen in Table 19:

	Logistic regression	**K-nearest neighbors**	**Gradient boosted trees**	**Support vector classifier**
Accuracy	**0.5738**	0.5369	0.5653	0.5597
Error rate	**0.0853**	0.0927	0.0870	0.0882
AUC	**0.5730**	0.5367	0.5623	0.5511

Table 19. Comparison of performance metrics for all models.

As Table 19 shows, the highest accuracy, lowest error rate and highest AUC are observed when logistic regression is used. However, the difference between the AUC of logistic regression, gradient boosted tress and support vector classifier is very small**.** This is further supported by the comparison of the ROC curves of all models displayed in Figure 8. The curves are not drastically different and there does not seem to be one dominant method based on the ROC curves alone. It is however clear from Figure 8 that the K-nearest neighbors method, presented as the yellow line, has a smaller AUC than the other three methods. This is in line with other performance metrics as the K-nearest neighbors has the lowest accuracy and highest error rate of all four methods as Table 19 shows.

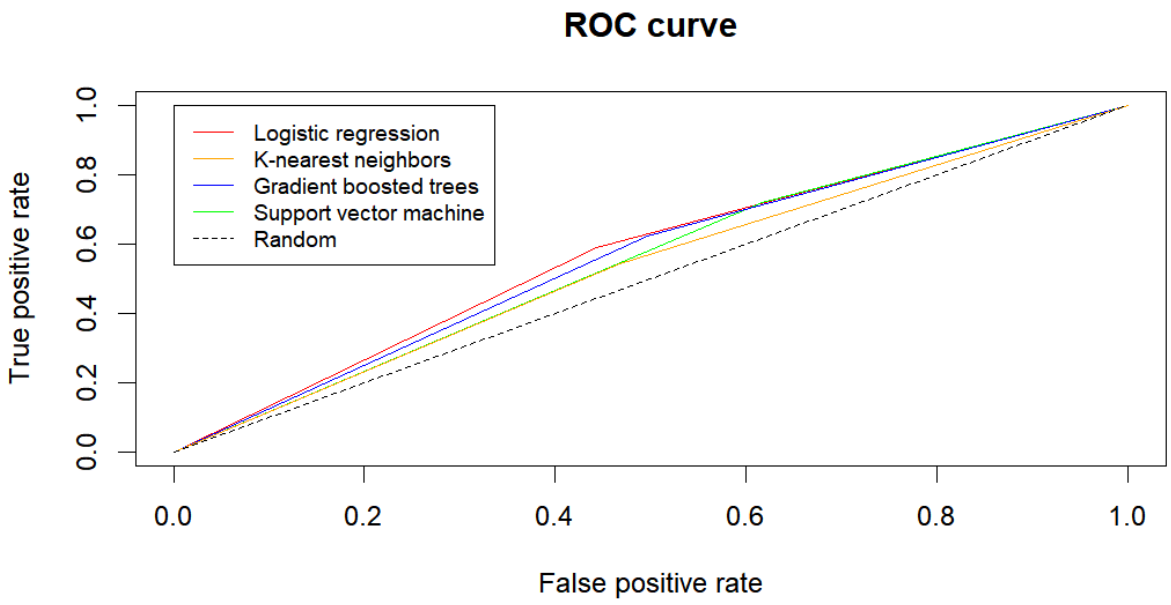

Figure 8. Comparison of ROC curve for all models.

Going back to the performance measures in Table 19, the dominating method seems to be logistic regression. As explained in chapter 3.4.2, a good accuracy for an analysis of this type is considered to be between 55% and 65%. Three of the four methods, logistic regression, gradient boosted trees and support vector classifier, have an accuracy above 55%. These results seem to be in line with what previous literature has found, supporting the hypothesis that a weak but significant relationship exists between news article sentiment and stock price movements. In addition, the p-value for logistic regression is below 0.05 as explained in chapter 4.1. This is further evidence that the null hypothesis can be rejected since the relationship identified between the two variables has significance.

4.6 Further Exploration of Relationship

As discussed in chapter 2.3, researchers have found conflicting evidence about the nature of the relationship between media sentiment and stock price movements. For example, Tetlock found that a negative media sentiment could forecast a downward swing of market prices (Tetlock, 2007) while Schumaker et al. found that a positive media sentiment was a better predictor of stock price downswings (Schumaker et al., 2012). Looking at the predictions made using logistic regression which is the method that provided the highest accuracy, lowest error rate and highest AUC, it is possible to explore the nature of the relationship to see if it is more in line with the findings of Tetlock or Schumaker et al. Table 20 shows the number of test observations based on whether their sentiment score is positive or negative, and which class was predicted with those sentiment scores using logistic regression.

		Sentiment	
		Negative (<0)	Positive (>0)
Prediction	Down	169	0
	Up	75	108

Table 20. Sentiment scores and predicted class using logistic regression.

The information in Table 20 seems to contradict the findings of Schumaker et al. that a positive media sentiment is a good predictor of stock price downswings. On the contrary, positive sentiment score never predicted that stock price would go down when logistic regression was used. Judging from this data, the results seem to be more in line with Tetlock's assumption that negative media sentiment forecasts a downward swing of stock prices. However, this is simply a rudimentary exploration of how the relationship between the two variables matches previous researchers' findings. The research question of this thesis is focused on the strength and significance of the relationship between the variables and a more detailed inspection of the nature of the relationship would therefore be outside the scope of this analysis.

Although the results are promising when it comes to significance and performance according to the performance metrics chosen, no analysis is without limitations. In the next chapter, there will be a discussion about the results including the potential drawbacks of the analysis and suggestions of further research that can be done on this topic.

5. Discussion

The findings are consistent with the hypothesis laid out at the beginning of the thesis that a weak but significant relationship exists between the sentiment of news articles and stock price movements. The conclusion that the relationship is significant is supported by the p-value of logistic regression, which is below the threshold of 0.05, indicating that there is evidence to reject the null hypothesis that no relationship exists between the variables. Furthermore, the accuracy when media sentiment is used to predict stock price changes varies from 53.69% to 57.38% using logistic regression, K-nearest neighbors, gradient boosted trees and support vector classifier. In three out of those four methods, the accuracy of the predictions exceeds the previously set threshold of 55%. The results therefore seem to be consistent with the results of previous research which has found an accuracy between 55% and 65% when using media sentiment of U.S. newspapers to predict stock price movements of New York Stock Exchange-listed companies. The results certainly support the idea that a significant relationship exists between the sentiment of Norwegian media and Oslo Børs-listed companies, although the relationship is weak like previous research has found for the U.S. market.

However, as the accuracy measurements show, the relationship between the two variables does not seem to be strong enough to be used as a good prediction method. It is therefore unlikely that sentiment analysis alone could be used as a profitable trading strategy. The analysis is also bound to certain limitations concerning both data and methodology that will be discussed in the following sections.

Firstly, it is likely that the choice of data greatly affects the results. This includes the choice of newspaper, the types of articles, the number of articles and the stock prices retrieved. The choice of scraping the online newspaper Dagens Næringsliv was based on it being one of the most popular online newspapers in Norway that has a vast selection of articles about finance. To cover a larger part of the Norwegian news scene, several different newspapers could have been included. As explained in chapter 3.1.1, certain categories of news articles were selected to find articles that were specifically related to finance, the stock market, business, and working life. If several different newspapers were included, it would be possible to include more categories, increasing the odds of retrieving more relevant articles to use in the analysis. As both the choice of newspapers and article categories affect the number of articles retrieved, a larger amount of news articles could give a better representation of the sample population, resulting in a more precise analysis. Additionally, if more frequent stock price information could be retrieved as well as the time stamps of each article, a more short-term analysis of news article sentiment's influence on stock prices could be conducted.

Several different types of classification can be chosen, including binary, multi-class, multi-topic and hierarchical classification (Sokolova and Lapalme, 2009). This analysis uses binary classification with two classes, up and down, meaning that the stock price on a given day either went up or down from the previous business day. Another option would have been performing multi-class classification by including more classes, like the instances where stock price does not change between days. However, when testing multi-class classification using those three classes, the results were much worse than when binary classification was used. Additionally, different types of performance metrics are used for different types of classification methods which could impact the analysis of the results. Therefore, choosing between different methods of classification could also limit or influence the results. This thesis has based this decision on research papers, cross-validation and performance metrics. There are other considerations which could influence this decision such as the size of training data, time used to train data and linearity. A lot of different types of classifiers exists and there are no right or wrong reasons for choosing a method. Other options for classifiers with a problem of this kind are for example linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), naïve Bayes, generalized additive models (GAM) as well as different types of classification trees.

In addition to the suggestions made above, further research could include looking at the sentiment of certain industries or certain companies to try to find a stronger relationship between sentiment and stock price movements. For example, it is possible that a negative article about *Equinor* has a much larger effect on the company's stock price movements than a negative article about *Aker BP*. Perhaps the relationship is stronger when only larger companies are included in the analysis, or if only companies within certain industries are considered. However, this could only be done with a larger amount of textual data that could provide company-specific or industry-specific information to draw conclusions from.

6. Conclusion

This thesis has analysed the relationship between Norwegian news articles that cover companies listed on Oslo Børs, and the stock price movements of said companies. To identify a relationship between the two variables, sentiment analysis was performed on articles written on Dagens Næringsliv's website. This included computing a sentiment score for each article based on a translated version of the Loughran-McDonald sentiment lexicon, as well as allocating each article to an Oslo Børs-listed company based on its content. The sentiment score of each article and the stock price movement of the relevant company was analysed using four statistical classification methods: logistic regression, K-nearest neighbors, gradient boosted trees and support vector classifier. Predictions of the stock price movements were made using the sentiment scores. The performance metrics accuracy, error rate, ROC and AUC were computed for each method based on the predictions. Additionally, p-value for logistic regression was found to evaluate the strength of the relationship between the variables.

The research question was framed as follows:

> *Can a relationship be found between media sentiment and stock price changes of Oslo Børs-listed companies?*

The results show that a relationship can certainly be found between the sentiment of news articles and stock price changes of Oslo Børs-listed companies. It was found that the p-value of logistic regression when applied to media sentiment and stock price changes was 0.0087. Typically, a p-value smaller than 0.05 is evidence to reject the assumption that no statistical relationship or significance exists between the variables. Therefore, sufficient evidence has been found to assume that a relationship does exist between the sentiment of news articles and stock price changes of Oslo Børs-listed companies.

In addition, the hypothesis laid out at the beginning of the thesis concerning the nature of the relationship was as follows:

> *A weak but significant relationship exists between media sentiment and stock price changes of Oslo Børs-listed companies, with a prediction accuracy consistent with what previous literature has found.*

As previously discussed, the threshold put in place for the prediction accuracy was 55%, meaning that an accuracy above 55% is strong enough to conclude a weak but significant relationship between media sentiment and stock price changes, as previous research has shown. The analysis found that the four statistical methods used to predict stock price changes using media sentiment had an accuracy ranging between 53.69% and 57.38%. Three out of the four methods implemented had an accuracy above the threshold of 55%, suggesting that the results match what previous research in this area has found. Researchers have previously suggested that subjective news articles are able to predict the direction of a stock price better than chance alone. Since all methods had an accuracy above 50%, the findings of previous research are supported with this analysis. It is safe to conclude that the relationship between the two variables is weak since the accuracy of the four methods lies only slightly above or below the threshold of 55%. However, the low p-value found for logistic regression suggests that the relationship is significant despite it being weak.

It was also found that the results of the most accurate method out of the four attempted, logistic regression, seemed to match the results of Tetlock who found that a negative media sentiment was a better predictor for downward swings in market prices than positive media sentiment (Tetlock, 2007). The results of the analysis contradict the results of Schumaker et al. who concluded that positive media sentiment could predict downward swings in market prices better than negative media sentiment (Schumaker et al., 2012). However, since the research question and hypothesis of this thesis are focused on the strength and significance of the relationship between the two variables, the specifics of the nature of this relationship are not explored closely. Considering that a significant relationship was identified between Norwegian media sentiment and stock price movements, further research on the specifics of this relationship could be an interesting topic for further exploration.

Lastly, it is important to note that even though the relationship between media sentiment and stock price movements is weak and this type of sentiment analysis might not be a profitable trading strategy, the relationship is significant and therefore not irrelevant. Including media sentiment could potentially increase the accuracy of predictions made using other relevant information about the stock market such as historical prices or trading volume. Sentiment analysis and its potential role in the stock market is therefore worthy of further exploration for those interested in the behavioural aspect of financial markets.

References

Angadi, M. and Kulkarni, A. (2015). Time Series Data Analysis for Stock Market Prediction using Data Mining Techniques with R. *International Journal of Advanced Research in Computer Science, 6*(6), 104-108. https://doi.org/10.13140/RG.2.1.1347.3360

Edgar, T.W. and Manz, D.O. (2017). *Research Methods for Cyber Security*. Elsevier.

Equinor. (2018, March 15). *Statoil to change name to Equinor*. https://www.equinor.com/en/news/15mar2018-statoil.html

Ferri, C., Hernández-Orallo, J. and Modroiu, R. (2008). An experimental comparison of performance measures for classification. *Pattern Recognition Letters, 30*(1), 27-38. https://doi.org/10.1016/j.patrec.2008.08.010

Finansleksikon. (2022). *Hovedlisten*. https://finansleksikon.no/finansleksikon/h/hovedlisten

Fung, G.P.C., Yu, J.X. and Lu, H. (2005). The Predicting Power of Textual Information on Financial Markets. *IEEE Intelligent Informatics Bulletin, 5*(1), 1-10.

Gujarati, D.N. (2004). *Basic Econometrics* (4th ed). McGraw Hill.

Gupta, E., Preetibedi, P., and Mlakra, P. (2014). Efficient Market Hypothesis V/S Behavioral Finance. *IOSR Journal of Business and Management, 16*(4), 56-60. https://doi.org/10.9790/487X-16445660

Haldar, S.K. (2012). *Mineral Exploration: Principles and Applications* (1st ed). Elsevier.

Hardeniya, T. and Borikar, D.A. (2016). Dictionary Based Approach to Sentiment Analysis-A Review. *International Journal of Advanced Engineering, Management and Science, 2*(5), 317-322.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). *An Introduction to Statistical Learning with Applications in R* (2nd ed). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis, 33*, 171-185. https://doi.org/10.1016/j.irfa.2014.02.006

Kudyba, S. (2014). *Big Data, Mining and Analytics: Components of Strategic Decision Making* (1st ed). Taylor & Francis Group. https://doi.org/10.1201/b16666

Loughran, T. and McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Malkiel, B.G. (2005). Reflections on the Efficient Market Hypothesis: 30 Years Later. *The Financial Review, 40*(1), 1-9. https://doi.org/10.1111/j.0732-8516.2005.00090.x

Medienorge. (2022). *Ti største nettaviser.* https://medienorge.uib.no/statistikk/medium/avis/395

Read, C. (2013). *The Efficient Market Hypothesists: Bachelier, Samuelson, Fama, Ross, Tobin and Shiller* (1st ed). Palgrave Macmillan. https://doi.org/10.1057/9781137292216

Rohleder, M., Scholz, H. and Wilkens, M. (2011). Survivorship Bias and Mutual Fund Performance: Relevance, Significance, and Methodical Differences. *Review of Finance, 15*(2), 441-474. https://doi.org/10.1093/rof/rfq023

Saif, H., Fernandez, M., He, Y. and Alani, H. (2014). On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter. *Proceedings of the 9th Language Resources and Evaluation Conference*, 810-817.

Schumaker, R.P., Zhang, Y., Huang, C. and Chen, H. (2012). Evaluating sentiment in financial news articles. *Decision Support Systems, 53*(3), 458-464. https://doi.org/10.1016/j.dss.2012.03.001

Singh, K., Elhoseny, M., Singh, A. and Elngar, A. (2021). *Machine Learning and the Internet of Medical Things in Healthcare* (1st ed). Academic Press. https://doi.org/10.1016/C2019-0-03077-4

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Statman, M. (2014). Behavioral finance: Finance with normal people. *Borsa Istanbul Review, 14*(2), 65-73. https://doi.org/10.1016/j.bir.2014.03.001

Subasi, A. (2020). *Practical Machine Learning for Data Analysis Using Python* (1st ed). Elsevier.

Tetlock, P.C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139-1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Thiese, M.S., Ronna, B. and Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease, 8*(9), 928-931. https://doi.org/10.21037/jtd.2016.08.16

Trucco, E., MacGilivray, T. and Xu, Y. (2019). Computational Retinal Image Analysis: Tools, Applications and Perspectives. Academic Press. https://doi.org/10.1016/C2018-0-00865-8

Zhou, Z., Gao, M., Liu, Q. and Xiao, H. (2020). Forecasting stock price movements with multiple data sources: Evidence from stock market in China. *Physica A: Statistical Mechanics and its Applications, 542*, 123389. https://doi.org/10.1016/j.physa.2019.123389

Appendix

A1. Summary Output

Summary output from the logistic regression model in R:

SUMMARY						
Deviance Residuals:						
Min	1Q	Median	3Q	Max		
-1.282	-1.179	1.008	1.168	1.410		
Coefficients:						
	Estimate	Std. Error	Z value	Pr(>|z|)		
(Intercept)	0.040159	0.038920	1.032	0.30215		
sentiment	0.018427	0.007022	2.624	0.00869 **		
Signif. Codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’	0.1 ‘ ‘	1
(Dispersion parameter for binomial family taken to be 1)						
Null deviance:	4386.2 on 3163 degrees of freedom					
Residual deviance:	4379.3 on 3163 degrees of freedom					
AIC:	4383.3					
Number of Fisher Scoring iterations: 3						