Can Social Media Predict Soccer Clubs' Stock Prices?
The Case of Turkish Teams

By:

Amirreza Safari Langroudi

Submitted to the Graduate School of Management

In partial fulfilment of the requirements for the degree of Master of Science in Business Analytics

Sabanci University

July, 2019

Can Social Media Predict Soccer Clubs' Stock Prices? The Case of Turkish Teams

Approved by:

Assoc. Prof. Raha Akhavan-Tabatabaei ....

(Thesis Supervisor)

Assoc. Prof. Abdullah Daşcı ....

Assoc. Prof. Mümtaz Karataş..........................

Date of Approval: 19/07/2019

ÖZET

Sosyal Medya Futbol Kulüplerinin Hisse Senedi Fiyatlarını Tahmin Edebilir mi?
Türk Futbol Takımları Vakası

Amirreza Safari Langroudi

İş Analitiği Yüksek Lisans Tezi, Temmuz 2019

Tez Danışmanı: Doç. Dr. Raha Akhavan-Tabatabaei

Anahtar Sözcükler: Futbol Kulüplerinin Hisse Senetleri, Hisse Senedi Getirisi, Maç Performansı, Maç Öncesi Beklentileri, Sosyal Medya, Duygu Analizi

Spor'da finans literatürü üç ana hisse senedi fiyat tahmin metodu üzerine odaklanmaktadır: maçın sonucuna, maç öncesi beklentilerine ya da maçın önemine göre. Maç öncesi beklentileri için bahis ihtimalleri yaygın olarak yatırımcıların duygularının göstergesi olarak kullanılmaktadır. Bu çalışma Twitter verisinin farklı bir gösterge olarak dahil edilmesini önermekte ve futbol maç sonuçları, duygular ve dört büyük Türk futbol takımının hisse fiyatlarının arasındaki bağlantıları analiz etmektedir. Sonuçlar hisse senedi fiyatlarının tahmininde sosyal medyanın güçlü bir maç öncesi beklentileri ve yatırımcı duygularının göstergesi olduğunu göstermektedir.

ABSTRACT

Can Social Media Predict Soccer Clubs’ Stock Prices?

The Case of Turkish Teams

Amirreza Safari Langroudi

Business Analytics M.Sc. Thesis, July 2019

Thesis Supervisor: Assoc. Prof. Raha Akhavan-Tabatabaei

Keywords: Soccer clubs' Stock, Stock Return, Stock Price Prediction, Match Performance, Pre-Match Expectation, Social Media, Sentiment Analysis

Finance literature in sports focuses on three main methods of stock price prediction in soccer: based on match results, pre-match expectations or match importance. For pre-match expectations, betting odds is commonly used as the indicator of investors' sentiments. We propose to include Twitter data as another indicator of this variable, and analyze the links between soccer match results, sentiments, and stock returns of the four major Turkish soccer teams. Our results show that social media can be a strong indicator of pre-match expectations and investors’ sentiments in stock price prediction.

Acknowledgments

This thesis was supported by the help and advice of many. I would first like to thank my thesis supervisor, Assoc. Prof. Raha Akhavan-Tabatabaei. Her advice and ideas were invaluable and provided guidance whenever I was struggling. This thesis would not have been possible without her direction and support.

I would also like to extend my sincerest gratitude to Altug Tanaltay for his support and provision of the Twitter dataset.

Finally, none of this would have been possible without the help and support of my family and friends.

Table of Content

List of Tables

List of Figures

List of Equations

Chapter 1: Introduction

A large number of professionals, businesses and organizations get involved in investing, producing, organizing and facilitating a variety of sport activities. The calculated size of the global sports industry is 1.3 trillion dollars (Plunkettresearch (2019)) and most of the sport-related businesses depend on professional leagues which have the major share of this global industry.

Soccer is one of the most popular sports with more than 4 billion followers, leading sports headlines in almost all the European countries. In 2018, the cumulative worth by the top 20 most valuable soccer teams was approximately $1.75 billion, with a 34% increase in comparison to the previous year (Rueters (2019)). Most of the soccer clubs around the world have their own private investors, but some of them have made an initial public offering and their stock can be publicly traded over the stock exchange market.

These soccer clubs with publicly tradable stocks, face many risks and challenges both in their team's match performance and the financial market. According to Szymanski (1998), the performance of a soccer club on the stock market is directly affected by its team's failure or success on the field. Winning a match can increase the club's stock price and make it a valuable asset, and on the other hand losing a match can cause depreciation of the stock leading to millions of dollars of loss. Since investing in soccer club markets is on the rise (Birkhauser et al. (2015)), researchers have been studying the impact of the team's match performance on the club's stock price. Arnold (1991) performed one of the earliest empirical studies on the relation between the sports team performance and their financial status, and found that there is a strong correlation between the revenues of the English soccer clubs and their team performance during 1905-1985.

Based on the finance literature in sports, there are three main methods of stock price prediction in soccer (Godinho and Cerqueira (2018)). The first method focuses on predicting the soccer clubs' stock prices based only on their match performance. The second type of approach focuses on the impacting factors of the match importance, including the match date, team rankings at the time of the match, and the level of rivalry between the two teams. The third method focuses on the pre-match expectations and investors' sentiments before the match, as compared to the match results. According to Edman et al. (2007), investors' pre-match expectation and their perception of the club status have a great impact on the clubs' stock prices.

Betting odds as an indicator of pre-match expectation and investors' sentiments, have been commonly used in the sports literature (Godinho and Cerqueira (2018)). Betting odds represent the probability of an event and show how much money one will win if his/ her bet wins. Each team has odds in favor and if a team is more likely to win, its odds will be lower and so is its gain. These odds for a match are usually determined by bookmakers who work as organizations or group of people that accept and payoff the bets in sports events. These bookmakers calculate the probability of each outcome and subtract their margin from the odds in order to increase their profits.

Although most researchers use betting odds as a representation of the pre-match expectation, due to the recent popularity of social media and advances in sentiment analysis through social media outputs, we propose to include Twitter data as another indicator of investors' sentiments, and analyze the links between soccer match results, sentiments, and stock returns of the soccer clubs in addition to betting odds.

For testing our argument, we use the financial data of four major Turkish soccer clubs with public stocks, and the vastly available Twitter data on them. Galatasaray, Fenerbahce, Besiktas and Trabzonspor are these four major Turkish clubs which have made an initial public offering. Our Twitter dataset also involves about 13 million real-time tweets for these four teams.

In this study, we aim to predict the amount and direction of the return in the stock price of these four clubs. To predict these variables, we run and compare three models: the first model is based on match performance and betting odds (Model 1), the second uses Twitter data as an indicator of the sentiments (Model 2) and the third combines Twitter sentiments and match performance data (Model 3). Our results display that social media can be a strong indicator of pre-match expectations and investors' sentiments in stock price prediction.

This study is structured as follows. Chapter 2 reviews the existing literature and related works on various approaches to predicting soccer clubs' stock prices. In chapter 3, we propose a brief introduction to Sentiment Analysis. In chapter 4, we describe our data collection, cleaning and structuring procedures. Chapter 5 presents the methodology used in this study. Chapter 6 discusses the predictive analysis models and their results, followed by the conclusion in Chapter 7.

Chapter 2: Literature Review

In this section, a review of the literature on various approaches to predicting soccer clubs' stock prices is presented. There are three main methods of stock price prediction in soccer: based on match results (subsection 2.1), based on match importance (subsection 2.2), and based on investors' sentiments and their pre-match expectations (subsection 2.3).

2.1 Match Performance

Among several studies focused on predicting the soccer clubs' stock prices, there is a concentration on the effects of off-field and on-field factors. Off-field factors include different aspects such as managerial decisions, coach changes, player transfers, and basically the features that is not related to the game itself. On the other hand, on-field factors focus on how the match performance can affect the clubs' stock price. In this study we focus on the influence of the team's on-field performance on changes in its stock price.

Szymanski (1998) focused on Manchester United becoming a financially successful club; later, following this article Szymanski and Kuypers (1999) discussed the relationship between the revenue and the team's league position among 69 clubs, and found that there is a positive correlation between the club revenue and its league performance.

Ronneboorg and Vanbrabant (2000) considered the effect of the weekly sporty performance on the stock price of soccer clubs. They focused on British clubs, and found that winning a match can result in positive abnormal returns of almost 1%. In contrast, defeats or draws can result in negative abnormal returns of 1.4% and 0.6%, respectively.

Devecioğlu (2004) studied the relationship between team performance and stock market price of Besiktas and Galatasaray as the first Turkish soccer clubs which went public. He investigated the relationship between match results and stock price performance in the 2002-2003 season.

Barajas et al. (2005) studied the relationship between team performance and expected income of the Spanish teams. They found that there is a non-linear relation between these two factors with about 55% explanatory degree.

Duque and Ferriera (2005) investigated the relationship between the stock price return and sport performance of the two major Portuguese teams (Sporting and Porto). They used data from 5 seasons (1998-2003) and the ARCH method to show that there is a positive relationship between winning and good share price performance. They also show that there is an association between draws and losses with negative stock price return.

Samagaio et al. (2009) studied the link between the financial performance and sporting performance of the English soccer clubs over 1995 to 2007. The study used cross-correlation analysis and regression analysis and concluded that there was a moderate correlation between stock market return and sporting performance.

Benkraiem, Louhichi, Marques, (2009) investigated the dates around 745 matches of different European soccer clubs. Their analysis demonstrated that around the dates of the matches, both the abnormal return and volume of the traded stock was affected by the sporting results.

Gollu (2012) investigated the impact of sportive performance of the four major Turkish teams in the domestic league on their financial performance. He used Beşiktaş, Fenerbahçe, Galatasaray and Trabzonspor data over the period of 2002-2009. The study indicates that there is no correlation between the sportive performances of the clubs and financial performances in the mentioned period. However, other papers contrast these results (e.g., Demir and Danis (2011) and Sarac and Zeren (2013)).

Floros (2014) considered the data from Porto, Benfica, Juventus, and Ajax to find the relationship between their European performance and their stock returns. They found that a draw has a positive effects on Benfica's and Ajax's stock returns, and draws and losses have a negative effect on Juventus’s stock returns. They also stated that the sport performance has no effect on stock returns for Porto club.

2.2 Match Importance

Some studies also took into account match importance measurements in addition to a mere consideration of the effect of team performance on the stock price.

Zuber et al. (2005) analyzed 10 English Premier League teams between 1997 and 2000. For the match importance measurement, they introduced a dummy variable for the current position of the teams in the national league to find out the importance of the matches between the top five or the bottom five teams. They found this variable statistically insignificant.

Palomino et al. (2009) studied English teams in the London Stock Exchange, and for the match importance measurements split the season into the matches played before April and between April and June**.** For matches between April and June, the effect of the match on the stock price was higher.

Bell et al. (2012) observed 19 clubs in the English league from 2000 to 2007. The study used two variables as match importance measurements: The first variable is a ‘‘degree of rivalry’’ between the two clubs playing a given match, which uses their final league positions in the last season and its difference with their current league positions. The second variable is their ‘‘final position”, which takes into account the number of remaining games and the extent to which the club’s league position differs from the mean. The results showed that each club acts differently, but in conclusion they stated that the importance of the game seems to have a moderate impact on the returns.

Godinho and Cerqueira (2018) took 13 teams from 6 European countries as their sample. They used a new measure of the match importance by giving weight to each match based on the expected and unexpected results obtained from the betting odds. Then they considered both the unweighted results and the results weighted by a new measure of match importance and found a significant relationship between the result and the stock performance of those teams.

2.3 Pre-match Expectations

The other type of the studies focuses on the pre-match expectations and investor's sentiments before the match and compare these sentiments with the match results.

Stadtmann (2004) investigated Borussia Dortmund between 2000 and 2002. He used models which apply different dummy variables like win, draw, and loss dummies and models that include the unexpected number of points variable, defined as the difference between the number of points a team gains in a match and the expected number of points in the same match. He concluded that all of the variables are statistically significant. He also stated that draw and loss dummies have a negative coefficients, win dummies have positive coefficient, and unexpected points have a positive coefficient.

Scholtens and Peenstra (2009) considered the effect of match results in the stock prices of 42 European clubs from 2000 till 2004. The study concluded that both expected and unexpected wins are followed by price increases and that both expected and unexpected losses are followed by price decreases. In the case of draw, if a win was expected the price will decline, if a loss was expected, coefficients are not significant.

Demir and Danis (2011) considered three major Turkish teams and used dummies for expected, weakly unexpected and strongly unexpected results. The coefficients are not significant when they did not use the expected results. When expectations are used, strongly unexpected wins are followed by significant price increases, and strongly unexpected defeats are followed by larger than expected price declines.

Bell et al. (2012) as we mentioned before, defined a variable named as ‘‘point-surprise’’ which measures the difference between the number of points obtained in the game and the expected number of points according to pre-match betting odds. They also used a variable defined as ‘‘goal-difference-surprise’’ which compares the goal difference in the match with the club's average goal difference in the last five matches. Point-surprise variable has a positive coefficient and a positive effect on the stock returns and goal-difference-surprise variable seems not to have a positive effect on the returns.

Sarac and Zeren (2013) investigated the effect of the team performance of three Turkish teams between 2005 and 2012. They used variables such as the match type, the betting odds prior to the match, the venue of the match, the lag between the match date and the market opening date and the market index return. They used a regression model to predict the stock return based on these variables.

Majewski (2014) considered different teams for Italy's A Series, from 2001 till 2014. He used betting odds to define the bookmarkers' expectations and find the relationship between the pre-match expectations and match results. The study showed a very clear relationship among financial variables (rates of return) and the variables representing match results and pre-match expectations.

Castellani et. al (2015) investigated the relationships between soccer match results, betting odds, and stock returns of 23 European soccer teams. The study concluded that wins usually lead to price increases and draws and defeats lead to price decreases with a higher effect on the case of defeats. They also concluded that unexpected results are followed by larger price changes compared to the expected ones.

Demir and Rigoni (2017) used the data of two major Italian teams, Roma and Lazio. They introduced the performance of the archrival and stated that the level of the archrival measure and the win of the archrival can have a negative influence on the mood of investors which can result in changes in the stock price.

In this study, we propose to include Twitter data as another indicator of these pre-match expectations, and analyze the links between soccer match results, sentiments, and stock returns of four major Turkish soccer clubs. In the next chapter, we give a brief introduction to Sentiment Analysis and review the literature on the role of social media in sentiment analysis.

Chapter 3: A brief introduction to Sentiment Analysis

Sentiment Analysis (SA) is a widely-studied research field, as the consequence of increased attention to social media platforms such as Twitter and Facebook in the last several decades. Sentiment Analysis is the process of recognizing and categorizing opinions expressed in a piece of text, especially in order to understand whether the writer's opinion is positive, negative, or neutral about a subject. Thus, the main objective of SA is to extract opinions about entities (products, services, etc.) in order to acquire useful information. Twitter can be regarded as a review platform where customers, manufacturers, service providers or any party are able to attain summarized information through sentiment analysis about their products and services. Twitter can also predict the stock market (Bollen et al., 2011). In the stock market prediction, sentiment polarity (positive and negative sentiments) can indicate stock price movements a few days in advance (Smailović et al., 2013).

Researchers studying SA need to deal with various types of subtasks and problems, some of which are aspect extraction, subjectivity detection, entity recognition or sarcasm detection by applying supervised or unsupervised machine learning, lexicon based, keyword based or concept based methodologies. By using these techniques, which are generally for solving problems of text mining, researchers try to find ways to process raw text, convert it to a structured form and attain information about a certain entity, like the public opinion about a certain product or a soccer club in our case. One of the objectives of this study is to extract the sentiments of soccer related tweets in Turkish language, on the four major teams in Turkey. Regarding the sentiment extraction phase, literature is reviewed for feature extraction strategies where unstructured text is transformed to a structured base, text annotation strategies where text is automatically labeled without human intervention data augmentation where unbalanced data is augmented to be balanced, and machine learning techniques for text classification of large amounts of data.

Naturally, the lifecycle of any data mining project is broken into six phases (Wirth, 2000): Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment. These phases form the industry standard named CRISP-DM. So, after defining the problem and before beginning any data related task, data must be collected from various sources. For instance, Pang et al. (2002) use Internet Movie Database (IMDB) archive for user reviews data, Pak and Paroubek (2010) use Twitter API to collect a text corpus and Agarwal et al. (2011) acquire labeled data from a commercial source. Ozturkcan et al. (2019) study the public usage of Twitter related with soccer by focusing on 2013 and 2019 leagues in Turkey. Prior to descriptive analysis, Ozturkcan et al. (2019) gets help from experts to define soccer-related keywords for search and collect purposefully selected tweets posted in Turkish for the 2018 and 2019 soccer leagues, which is the data collection methodology followed by this work.

After the data collection phase we need to prepare the data for the analysis. The data preparation phase includes all the activities for converting the raw data to the final dataset which is to be fed into the modeling tools. Regarding text mining, after removing all items that are not actual words (links, hashtags, URLs, numbers, stop-words, etc.), raw text data is converted into a tabular form. At this instance, each entry under examination (a tweet, a product review, etc.) becomes an observation, and each unique word (or a group of words) becomes a feature of that observation to be processed by a classification model, where values of each feature/word can be its frequency in the document, binary representation of its existence or its calculated weight in terms of frequency compared to the other documents. In short, each document is represented as a vector of words with their calculated frequencies or weights. While single words can be features, using a combination of adjacent words is also a common approach named as n-gram representation, where n is the number of adjacent words extracted. Part of speech (POS) labeling of n-grams, which displays the position of each n-gram in a sentence and their type as adjective, conjunctive, noun, etc. also represents the linguistic property of text, which can also be used as a feature. Assessment of these features helps to classify the observation as containing positive or negative sentiment.

Different values for *n* affect the precision of classification in different ways. Akaichi et al. (2013) tried different combinations of *n* and observed that Support Vector Machine (SVM) and Multinomial Naïve Bayes achieved the highest accuracy of classification when unigrams and bigrams are consolidated. On the other hand, in a similar study, Zhai et al. (2011) acquired less accurate results when using a mix of n-grams. They concluded that bigrams achieve better results than other n-gram features. Bermingham & Smeaton (2010) observed that representing text using n-grams with POS tags result in acquiring more information than using unigrams in classifying blogs, micro reviews or movie reviews when features are sent to SVM classifier. Moreover, they concluded that just using unigrams with Multinomial Naïve Bayes on the source of microblogs like Twitter perform better than the former case. Pak and Paroubek (2010) achieved the highest accuracy on classifying Twitter data by using bigrams with POS tags and their findings support that POS tags must be included as features in case of Twitter classification. They also examined that subjectivity (sentimentality) versus objectivity (neutrality) of a document can be detected getting use of the POS tags. Agarwal et al. (2011) found that combining prior polarity of words with their POS tags are important for classification tasks whereas Twitter specific features like emoticons or hashtags add a non-marginal value to the classifier. However, regarding Turkish language, conversion of raw text to POS tags is yet problematic because of the lacking of lexical libraries. Thus, for this study, raw text is converted to unigram vector representation before training the classifiers and after cleaning the text from non-words, hashtags, emoticons, and punctuation.

As mentioned above, documents can be represented by a vector of words with their frequencies, by their binary representation of existence or by a special weighing that implies the importance of each word in a certain document. As best results are achieved when the feature values are set as binary representation of a word's existence, followers of Pang et al. (2002) applied the same strategy when dealing with text sentiment classification. Some of the examples are Pak & Paroubek (2010), Barbosa & Feng (2010), Ye et al. (2009) and Habernal et al. (2014). However, it is also discussed in literature that when dealing with a corpus, in most of the cases it is not enough to represent documents as word frequency or binary vectors. Each word has a significance factor when its existence in other documents is compared. A very common word in a specific language will appear in most of the documents, thus its existence in a document will not make a significant difference than its existence in other documents. Thus, a weighing strategy for the word frequencies in each document might help to characterize them better. TF-IDF (Term Frequency – Inverse Document Frequency) is used to determine the significance of a word in a specific document by comparing its frequency in the whole corpus and weighing each word with a calculated index. Barnaghi et al. (2016), Martinez et al. (2011), Smailovic et al. (2013) are some classification examples applying TF-IDF conversion of word frequencies. In our work, prior to data training, unigram vector representation of raw text is converted to TF-IDF form and a significant gain in accuracy is achieved as a result.

Opinion and sentiment analysis usually start after the data preparation part. These analyses in literature generally apply supervised or unsupervised machine learning, lexicon based, keyword based and concept based approaches for classification of sentiments. Supervised methodologies mostly consist of Maximum Entropy, Naïve Bayes, Logistic Regression and SVM classifiers. These methodologies are applied by Pang et al. (2002), Pak & Paroubek (2019) and Barbosa & Feng (2010) previously. As mentioned before, after converting the unstructured raw text into a structured form (binary representation, frequency representation and TF-IDF representation), the tabular formed data is processed by a classifier and then a performance metric is calculated in order to evaluate the outcome. Unsupervised methodologies use clustering techniques for mining opinions. The most popular unsupervised methodology applied appears to be Lexicon Based classifiers where a word polarity source that provides polarity scores is used to calculate the cumulative polarity of a document. A threshold is determined for final classification of the document. If a document’s cumulative polarity score is over the threshold, then it is accepted as positive. If it is less than the threshold, then it is accepted as negative. Some studies worked on multi classes, adding neutral outcome to their results. The word polarity source can be an external source like Wordnet or the polarity scores can be calculated directly from the word frequencies of the corpus collected. There also appeared new approaches in the last 10 years applying semi-supervised techniques or neural networks & deep learning methods to sentiment classification.

During the process of sentiment analysis we deal with different problems. When training a classifier with the goal of maximizing overall accuracy, imbalanced training data cause the classifier to perform better on the class with more observations, and worse on the class with less observations (Seiffert et al., 2008). One of the proposed methods as a solution to this problem is applying sampling on the training data. By artificially balancing the class distributions, oversampling creates a more balanced dataset by increasing the number of observations in the minority class (BalakrishnanGokulakrishnan et al., 2012). By this way the skewness of the data is fixed to an extent by the duplication of the already existing minority class instances that helps the sizes of the classes becoming comparable (Pandey and Iyer, 2009). Pandey and Iyer (2009) have compared the performances of Alternative Naïve Bayes and SVM classifiers on an imbalanced dataset and observed that classifiers with no oversampling gave a lower recall with a relatively lower false positive rate. In our case, as neutral and negative number of tweets were nearly half of the positive tweets, oversampling on the neutral and negative classes was applied during the preprocessing phase.

Another problem of applying sentiment analysis using machine learning techniques is the need for human annotated data. Supervised algorithms are trained on text instances with labels that differ according to the problem studied. In the case of sentiment analysis, they are usually labeled as positive, negative or neutral. Moreover, supervised classifiers perform much better when run on a huge amount of labeled data. However, acquiring large amounts of labeled data is an expensive and time consuming task. When the actual text data is online reviews for a specific product or service, collected though a CRM system or a website, as the reviews are accompanied by a rating provided by the reviewers, one can easily generate classes through these rating “points” as negative or positive. For instance, as mentioned above, Pang and Lee (2002) used the movie review messages with ratings for the prior classification, and first applied subjectivity detection followed by sentiment classification. They tested Maximum Entropy, Naïve Bayes and SVM classifiers with support of POS tagged messages. Unfortunately, Twitter messages do not contain such a grading mechanism and in most of the cases researchers need to organize labeling teams prior to sentiment analysis. J. Read (2005) proposed an alternative approach for annotating microblogging messages. He analyzed

Usenet newsgroup messages and categorized messages according to the emoticons used in the message. Messages containing emoticons like "☺" or "☹" were used to create a training set for running classifiers. While happy emoticons made the message "positive", sad or angry emoticons made the message negative. J. Read achieved up to 70% accuracy by applying SVM and Naïve Bayes on the "emoji" labeled data. In Pak and Paroubek (2010), authors follow a similar strategy to construct corpora of emoji labeled positive and negative Twitter messages and run classifiers afterwards. They also apply objective text classification (classification of the third class: neutral messages) with the same technique on arbitrarily large data. They collected Twitter messages using the Twitter API for positive and negative messages, and also consumed messages of news agents as "New York Times" for classification of neutral tweets. As Twitter messages are limited containing around 250 words on average, they assumed that *"an emoticon within a message represents an emotion for the whole message and all the words of the message are related to this emotion"* (Pak and Paroubek, 2010). They apply a mixture of these techniques: pre-classification of Twitter messages according to their emoticon content, then applying machine learning classifiers on the automatically labeled corpora.

With this introduction and literature review, we will discuss our data collection and descriptive analysis in the next chapter.

Chapter 4: Collection and descriptive analysis of data

In this chapter, our aim is to present the data collection process and the descriptive analysis of this data for match performance, financial and Twitter data. First, we give a brief description of the four teams and their performances in the previous years in subsection 4.1. We also discuss the data collection process for match performance and financial data in this subsection. Then we describe the match performance data and financial data descriptive analysis. Subsection 4.2 gives a description of the Twitter data.

4.1 Team descriptions and performances

Founded in 1905, Galatasaray S.K. (GS) is the most successful Turkish team, consisting of the Galatasaray high school student members. They have won 22 Super Leagues and 18 Turkish Cups since their conception. They also won the UEFA (Union of European Football Associations) Cup in 2000 and became the only Turkish team to have won this title. This team is based in Istanbul and their stocks went public in 2002.

Fenerbahçe S.K. (FS) is also one of the most successful teams in Turkey, founded in 1907 and based in Istanbul. They also won 19 Super Leagues and 6 Turkish Cups. They won the most national championship titles among all the Turkish teams. Their stocks went public in 2004.

Beşiktaş J.K. (BJK) is also based in Istanbul and founded in 1903. It was first a gymnastics club but after 1910 with soccer becoming popular in the Ottoman Empire, the club focused more on soccer. Their stocks went public in 2004.

Trabzonspor (TS) is not an old club, founded in 1967 through the merger of some local teams. They have won 6 Super Leagues and 8 Turkish Cups and are the first club which is not based in Istanbul, winning the Super League. Their stocks went public in 2005.

We have accessed all the match results from 2004 till 2019 for these four Turkish teams retrieved in April 2019 from https://us.soccerway.com. The data contains the date of the match, type of the match and the game result. We consider different match types like Turkish Super League (SÜL), Turkish Super Cup (CUP), UEFA Championship League (UCL), UEFA Europa League (UEL) and Friendly matches. Figure 1 shows a snapshot of this data.

Day	Date	Match Type	Team.1	Result	Team.2
Thu	26/07/18	UEL	B36	0 - 2	Beşiktaş
Thu	02/08/18	UEL	Beşiktaş	6 - 0	B36
Thu	09/08/18	UEL	Beşiktaş	1 - 0	LASK
Sun	12/08/18	SÜL	Beşiktaş	2 - 1	Akhisarspor
Thu	16/08/18	UEL	LASK	2 - 1	Beşiktaş
Sun	19/08/18	SÜL	BB Erzurumspor	1 - 3	Beşiktaş
Thu	23/08/18	UEL	Partizan	1 - 1	Beşiktaş

Figure 1: Game Results

We also collect the betting odds for every match appearing in our teams' database retrieved in April, 2019 from https://www.oddsportal.com. This site calculates the average odds of different bookmakers for each match. Figure 2 shows a sketch of the data. This figure includes the match date and time, teams, match result, home team winning odd (H.odd), Draw odd (D.odd), Away team winning odd (A.odd).

Date & Time	Team.1 – Team.2	Result	H.odd	D.odd	A.odd
Soccer » Turkey » Turkish Cup 2004/2005			1	X	2
02/03, 17:00	**Fenerbahce** - Kayserispor	4:0	1.10	7.00	9.99
Soccer » Turkey » Super Lig 2004/2005			1	X	2
27/02, 16:00	**Fenerbahce** - Sebatspor	2:0	1.07	8.12	9.99
Soccer » Europe » UEFA Cup 2004/2005			1	X	2
24/02, 18:45	Zaragoza - **Fenerbahce**	2:1	1.94	3.20	3.51
17/02, 17:00	**Fenerbahce** - Zaragoza	0:1	1.62	3.41	4.96

Figure 2: Betting odds

We have merged the two above mentioned datasets and carried out descriptive statistics on it. Table 1 shows the descriptive statistics for each team's match performance:

Table 1: Descriptive analysis of match results

	Number of matches	Number of Wins	Number of Draws	Number of Losses	Number of Home Matches	Number of European Matches
Besiktaş	775	406	177	192	391	122
Fenerbahçe	806	472	170	164	408	126
Galatasaray	770	426	163	181	394	94
Trabzonspor	707	341	170	196	355	57
Total	3058	1645	680	733	1548	399

For the financial performance of the clubs, we have collected the daily stock market information for each team since the beginning of their stock's public initiation until March 2019, from Yahoo Finance. The table contains the date, stock's opening and closing prices, highest and lowest prices, and the volume of the stock sold on a given date. We have also collected the Istanbul Stock Exchange BIST 100 on the same dates, in order to consider the overall market changes. Figure 3 shows a snapshot of this team's daily stock market data.

Date	Open	High	Low	Close	Adj.Close	Volume
2002-02-21	0.83218	0.89162	0.83218	0.83218	-231.70781	1018158
2002-02-22	NA	NA	NA	NA	NA	NA
2002-02-25	NA	NA	NA	NA	NA	NA
2002-02-26	0.79255	0.83218	0.76283	0.79255	-220.67345	8734848
2002-02-27	0.75292	0.80246	0.75292	0.75292	-209.63907	1979636
2002-02-28	0.68357	0.73311	0.66376	0.68357	-190.32964	4153652
2002-03-01	0.69348	0.72320	0.66376	0.69348	-193.08891	3627229
2002-03-04	0.67367	0.77274	0.65385	0.67367	-187.57312	9283202
2002-03-05	0.65385	0.68357	0.65385	0.65385	-182.05455	1833121
2002-03-06	0.65385	0.66376	0.62413	0.65385	-182.05455	1682468

[1]

Figure 3: Teams' Daily Stock Market information

[1] The prices are in Turkish Liras

Table 2 presents the descriptive statistics on the stock's closing price in Turkish Liras for each team.

Table 2: Closing Price Descriptive Statistics

Team	Number of days	Mean	Standard Deviation	Coefficient of Variation	Median	Min	Max
Fenerbahçe	3867	17.375	7.770	0.44	17.141	4.949	53.299
Besiktaş	3879	2.167	1.327	0.61	1.900	0.386	6.500
Galatasaray	3858	2.954	1.855	0.62	2.331	1.180	10.393
Trabzonspor	3567	2.944	1.808	0.61	2.350	0.860	11.611

Fenerbahçe's stock has the highest standard deviation and range but it has the least coefficient of variation among all the teams. On the other hand, Besiktas's stock has the lowest standard deviation and range among all the teams. Galatasaray's stock has the highest coefficient of variation.

Table 3 presents the descriptive statistics for the stock price returns for each team:

Table 3: Return of stock descriptive statistics

Team	Number of Days	Mean	Standard Deviation	Median	Min	Max	Range	Skew	Kurtosis	Standard Error
FB	3866	0.0003847	0.0269824	0	-0.2321438	0.2000018	0.4321456	0.6017014	12.8373500	0.0004330
BJK	3878	0.0011464	0.0519760	0	-0.3232334	2.4397651	2.7629985	26.8857711	1249.1783500	0.0008300
GS	3858	0.0003386	0.0292967	0	-0.1750001	0.2035406	0.3785407	0.8882894	10.0870500	0.0004710
TS	3566	0.0002747	0.0297271	0	-0.2222207	0.2212392	0.4434599	0.7673386	10.3638600	0.0004978

4.2 Collection of the Twitter data and descriptive analysis

We also include the Twitter data for testing the effects of the fans' sentiments on our model. Regarding the collection of Twitter data, as in (Ozturkcan et Al., 2019), by getting use of Twitter's public API, we collected purposefully selected tweets posted in Turkish for the 2018 and 2019 soccer leagues using Logstash (for collecting) and Elasticsearch (for indexing). Regarding the 2018 and 2019 leagues, 172 keywords were separately chosen by 2 researchers, 2 soccer fans, and a sports consultant, which were then used to purposefully collect streaming data from Twitter. We acquired around 20,000,000 soccer related tweets between December 2017 and March 2019. Following the selection and clustering of the keywords specific to our four selected teams, and applied a second filter to distribute the twitter messages among these teams. As a result, a total of 12,814,581 tweets regarding these teams as displayed in Table 4, were collected. We then transferred the filtered data to a distributed computing environment backed up by Apache Hadoop for further processing.

Table 4: Twitter data description

Twitter data	
Data Start	12/1/2017
Data End	3/31/2019
Total Tweets	12,814,581.00
Total Tweets FB	4,987,408.00
Total Tweets GS	4,917,873.00
Total Tweets TS	1,011,830.00
Total Tweets BJK	3,190,178.00

The major proportion of the filtered data belongs to Fenerbahçe (FB) and Galatasaray (GS) teams followed by Beşiktaş (BJK) and Trabzonspor (TS), which also represents the fan-base for these four teams. As mentioned before, FB, GS and BJK are clubs from Istanbul, supported by the majority of the soccer fans in Turkey; whereas TS, although being among the top 4 teams, is local to the Black Sea region of Turkey and has a fan-base less than each of FB, GS and BJK.

Figure 4 shows the frequency of tweets in the time window of December 2017 and March 2019. Note that the data during the months of July, August, and September of 2018 is missing due to server shutdown.

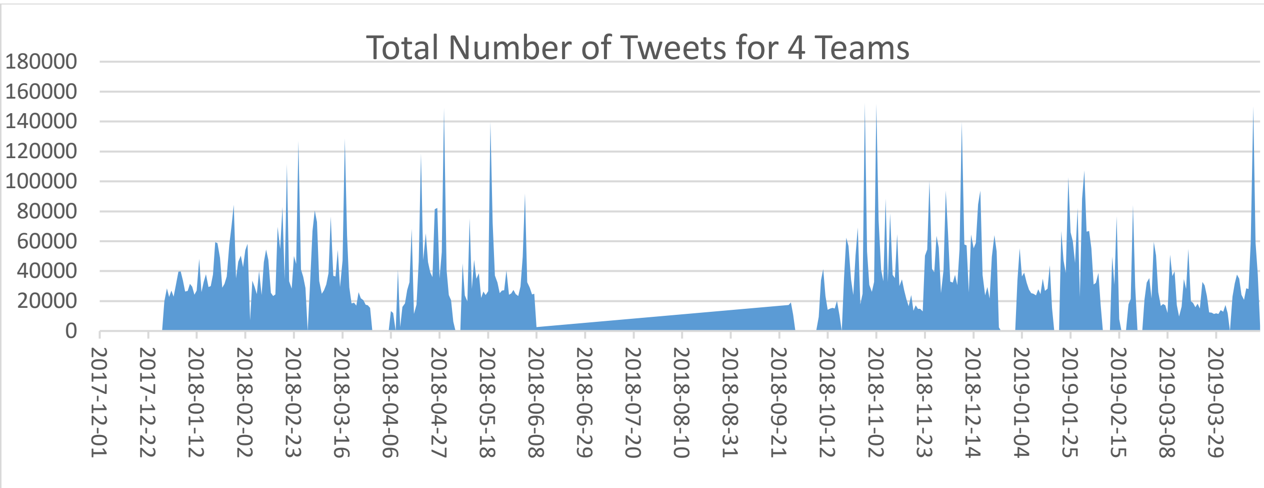


Figure 4: Tweet trends

Following the data collection phase of Twitter messages, we applied a similar approach to CRISP-DM, as detailed in the introduction section, for the sentiment extraction with the phases of emoticon extraction and tweet labeling, text cleaning, feature extraction from text and finally model building, validating and predicting.

From Emoticon Extraction and Message Labeling to predicting, we used Apache Spark distributed computing engine. Processing a total 20,000.000 soccer related tweets, 1,131 unique emoticons were extracted. Among these, some are not representing a sentiment or are not very frequent. Finally, we selected 50 emoticons with more than 80% frequency for each class (positive, negative and neutral). As an example, happy face emoticons are regarded as positive; angry or unhappy face emoticons are regarded as negative. Sports news accounts use flags, calendar signs or notification signs in their tweets. Thus, the most frequently used emoticons by these accounts are regarded as neutral. Some of the most frequently used emoticons are listed in Table 5.

Table 5: Top 15 Emoticons for Each Class

Positive	💚 👏 ❤ 💙 👍 😁 🙏 😉 💪 😎 ☺ □ 😍 😄 👊 😃
Negative	😭 😒 😫 😢 😱 😤 😦 👎 😳 😟 😑 😓 😐 😮 😞 😣
Neutral	🌟 ⚽ 👉 📌 ✅ ➡ ⚫ ❌ □ ✔ 📺 🔹 ⬅ 📢 🆚 📍

After the extraction and selection of significant emoticons, we applied a rule based approach for labeling the whole soccer related tweets. Tweets containing at least one negative emoticon were labelled as negative; tweets without any negative emoticon and having mostly positive emoticons were labelled as positive; and finally tweets having mostly neutral emoticons were labelled as neutral. After the labeling phase the data is distributed as displayed in Table 6.

Table 6: Distribution of Tweets after Labelling

Class	# of Tweets 2018	Percentage 2018	# of Tweets 2019	Percentage 2019
Positive	326,063	56%	420,681	61%
Negative	130,625	22%	109,036	16%
Neutral	130,207	22%	164,222	24%
TOTAL	586,895		693,939	

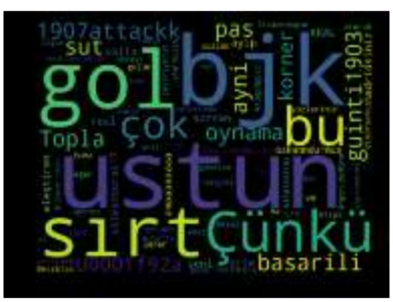
Positive	Negative	Neutral
		

Figure 5: Most Common Words for Positive, Negative and Neutral Datasets

In order to check the consistency of the content with their labels, word cloud plots of the most common phrases used in the three classes are shown in Figure 5. In the positive set, words with positive sentiment like "gol (goal)", "ustun (superior)", "çok (a lot), "basarili (successful)" can be observed. In the negative set, interestingly, "Galatasaray" and "GalatasaraySK" are the most common words which are directly related with the team Galatasaray. Apart from them, the negative set contains words like "saklabana (an insult in Turkish)", "kanser (cancer)" and "kiralik (for rent)". In the neutral set, we observe some player names (Erhan, Emre) and words like “lig (league)”, “maclardaki (at the matches)”, and “ortalama (average)”. The words are consistently distributed among the three sets and this distribution will directly affect the classifier algorithm’s tendency to classify a certain tweet. One can easily see that there is not an obvious intersection of words between these sets, which will increase the classifier’s performance. Another fact is that, in three of the datasets words like “1907attack”, “https”, ”UU001f92a”, “nKaynak” or “co” also appear. These words are related with user accounts, links in tweets and special characters like the emoticons, and do not directly represent the sentiment in the tweet text. This fact puts forward the necessity of cleaning the text, getting rid of such symbols or non-words. Thus, before training the classifier, all stop-words and non-words (punctuation, special characters, numbers, links, hashtags, emoticons) not representing a sentiment or a lexical meaning are removed from the text of all Twitter message instances, with the exception of exclamation marks which particularly indicate strong sentiments in Latin based languages. Moreover, words with two characters are intentionally not removed as they are frequently used in slang and swearwords by soccer fans.

As keywords and activities vary according to soccer seasons, 2018 and 2019 Twitter data has been treated separately in the sense of labeling and modeling. It is clearly seen on Table 6 for both seasons that positive tweets are several times more in number than negative or neutral tweets. If any model is trained on this distribution, it is certain that the model will predict the positive set much better than the others as it would have experienced the positive examples more. In order to solve the unbalanced dataset problem, as described in Seiffert et. al., (2008) and Pandey & Iyer (2009) oversampling on neutral and negative sets was applied separately for the data of two seasons: Negative and neutral number of tweets of 2018 season were oversampled by 190%; 2019 season negative tweets were oversampled by 380%; and 2019 season neutral tweets were oversampled by 285% randomly without replacement. As a result, all classes contain a similar number of tweets in the oversampled dataset. Our final model’s validation accuracy increased by 5% when we applied only random oversampling.

After the data collection and preparation phases, we propose our research methodology both on the sentiment analysis part and our prediction methods in the next chapter.

Chapter 5: Methodology

In this chapter we discuss two main parts. First, in subsection 5.1 we explain our methodology for sentiment analysis and the way we deal with the unstructured Twitter data and label it for our analysis. Then, in subsection 5.2 we present our predictive models for the stock price return of our selected four teams.

5.1 Sentiment Analysis

Following the text cleaning and oversampling operations, feature extraction is applied in order to transform the unstructured text to a structured form, firstly bag-of-words representation of the raw text is acquired prior to TF-IDF calculation. Similar to the work done in previous research, a dictionary is formed by the all words in the collected twitter training data, words appearing less than 20 times in the whole corpus are omitted. The words in the dictionary are the features for each tweet and a tweet is represented by a vector of the count of each word in this dictionary. As the importance of words is not reflected well in word counts, a further operation was applied for each tweet in order to calculate the TF-IDF values with the following formulas:

Term Frequency (TF) is calculated by $tf(t, d) = f_{t,d}$ which represents the number of times that term *t* occurs in document *d*, where each document is a tweet in our case. The Inverse Document Frequency (IDF) is a measure of how much information the word provides and it is basically the logarithmically scaled inverse fraction of the documents that contain the word. IDF calculation is as follows:

$$idf(t) = log \frac{total\ number\ of\ documents\ in\ corpus}{number\ of\ documents\ where\ term\ t\ appears}$$

Equation 1

Finally, the TF-IDF is calculated by $tfidf(t,d) = tf(t,d) \,.\, idf(t)$ which is able to give information about both the words' existence and its importance in each tweet. As a result of this transformation process, each tweet in our dataset was represented with 2043 unique words. This number is quite low and shows that sports related Twitter messages in Turkish do not contain a large vocabulary.

When the dataset is ready for training, it is split into train and validation sets by 70% and 30% proportions respectively. Naïve Bayes, SVM's and Logistic Regression classifiers provided by Apache Spark environment were trained with cross validation that helped to attain the best hyper-parameters for these classifier algorithms. Best accuracy on the validation set was achieved by Multinomial Logistic Regression classifier which is known for its good performance on large datasets. The performance of the algorithms tested are displayed in Table 7.

Table 7: Performance Summary

Classifier	Train Accuracy	Validation Accuracy	Processing Time (hours)
SVM	0.73	0.69	4.2
Naïve Bayes	0.72	0.70	1.2
Logistic Regression	0.81	0.75	1.1

While Naïve Bayes and Logistic Regression classifiers train approximately in 1 hour, SVM classifier completes training in 4 hours which is not surprising as SVM applies kernel transformation and increases the feature size. In our experiments, Logistic Regression model with regularization parameter of 0.01 and 100 maximum number of iterations was the best classifier acquired both in terms of performance and processing time. The Logistic Regression model was further trained on the whole data without splitting the validation set, on 2018 and 2019 data sets separately as the keywords differ between the two seasons.

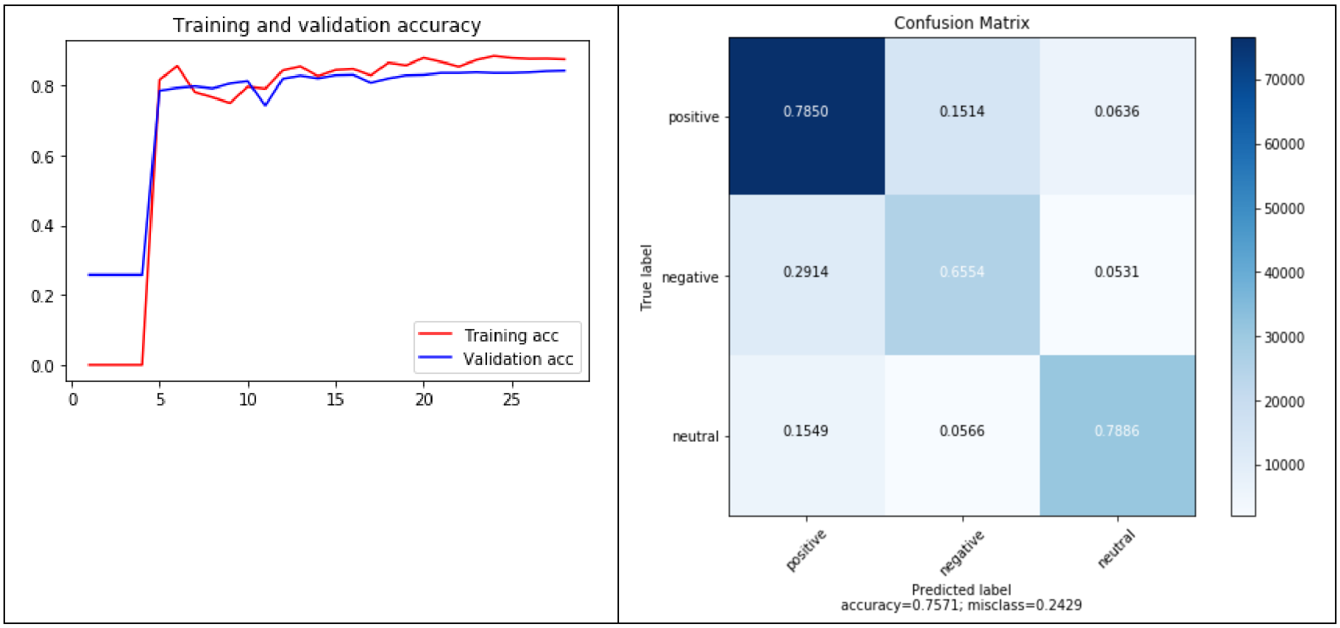


Figure 6: Learning Curves and Confusion Matrix

In Figure 6, learning curves of our Logistic Regression classifier for the first 25,000 training examples is presented on the left, and the confusion matrix provided by model's prediction on the validation set on the right. As it is clearly observed from the learning curve of the classifier's performance, the model stabilizes after being trained with 10,000 observations. Training accuracy is slightly higher than validation accuracy, without a large gap, which proves that the model does not overfit the training data. Moreover, when the model's performance on each class is separately examined, it is obvious that the model predicts the neutral class at best with 78.86% accuracy. It is followed by 78.5% accuracy for positive class and 65.54% accuracy for the negative class. Data augmentation applied with oversampling of the negative and neutral sets has worked well to increase the model's performance on the scarce classes. Interestingly, even though the oversampled number of observations for the negative and neutral sets are close to each other in the training dataset, the model predicts the negative class 10% worse than the neutral class.

In order to validate the performance of the final model, a ground truth dataset was prepared. The ground truth was sampled from 2018 and 2019 datasets and labeled by 20 graduate students. The students labeled the twitter texts in three categories: positive, negative and neutral. Same observations were given to several students in order to average out the personal bias. Our final model achieved the accuracy of 72% on the ground truth, which is not very different than the performance of the model on validation data of the automatically labeled tweets. As the last step of the work, after ensuring the performance of the model on ground truth data, the two models for the 2018 and the 2019 season were used to predict all of the 12,814,581 tweets for the four major teams.

5.2 Predictive modeling of stock price return

In this section, we describe our methodology to construct predictive models of stock price return. First, we tested the hypothesis to check if a match has an effect on the stock price. For this purpose, we divided the days based on the stock trade and labeled them as follows.

- First stock traded after the match: 0
- Last stock traded before the match: -1
- Stock traded 1 day after the match: 1
- Other days: 2

We ran Welch's two-sample t-test on the difference between the means of stock prices before the match and after the match.

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

Table 8 shows the *p*-values of this test, for the four teams and each day:

Table 8: Welch t-test p-values

	-1,0	-1,1	0,1	1,2	0,2	-1,2
Fenerbahce	5.49e-09	0.0986	1.403e-05	0.0407	1.03e-10	0.8208
Besiktas	0.0010	0.1118	0.6571	0.0987	0.0037	0.8580
Galatasaray	0.01395	0.0007	0.0048	0.0775	0.4132	0.0156
Trabzonspor	0.0074	0.3010	0.1077	0.2957	0.0023	0.8175

For all of the four teams, the last stock traded before the match is statistically different with the first stock traded after the match. Now we can proceed to present our models and predict the stock price return based on the match factors, betting odds and sentiment analysis.

In our predictive models, our dependent variable is the daily return in the stock's closing price for each team, defined as the percentage change between the first stock traded after the match and the last stock traded before the match divided by the first stock traded after the match, referred to as "*change*". The other dependent variable we predict in this study besides the amount of the stock return, is the direction of the stock price return, which is basically a classification problem. For this purpose, we define a binary variable named "*changedummy*" and if the return is positive we classify it as 1 and if the return is negative or zero, we classify it as 0. Table 9 presents the dependent variables that we are going to predict:

Table 9: Dependent Variables

Notation	Dependent Variables	Type
change	Stock return	Numeric
Changedummy	Direction of the return in the club's stock price	Binary

We aim to predict these two variables using three different models and compare the result of these models to find the effect of the match performance, betting odds and sentiment analysis, individually and together, on each club's stock price return. The first model is based on match performance and betting odds (Model 1), the second uses Twitter data as an indicator of the sentiments (Model 2) and the third combines Twitter sentiments and match performance (Model 3). We use different prediction methods like linear regression to predict the *change* and we used logistic regression, linear discriminant analysis (LDA), and Quadratic discriminant analysis (QDA) to predict the *changedummy*. We also remove the outliers which fall outside of ±1.5 times inter quartile range (IQR) of the stock data in our stock data for a better analysis. We run each model on each team and then we combine all the teams' data and run a model on the combined data.

Now we explain and compare each of our models, their independent variables and their other differences.

5.2.1 Model 1

The first model we propose for the soccer teams' stock price return and return direction prediction is to only use match performance data and betting odds in our model. This model analyzes the effect of match performance and betting odds on the *change* and *changedummy*. For this model we used different variables that we collect and Table 10 presents the description of these independent variables we used:

Table 10: Model 1 Independent Variables

Notation	Independent Variables	Type
Match Type	UCL, UEL, TL, CUP, Friendly	Categorical
Gdiff	Goal difference	Numeric
Extra	If the match went to extra time or penalty	Binary
Odds	Betting odds	Numeric
Price	Closing price of ISE	Numeric
Change	Change in ISE	Numeric
ISEchangelag1	Change in ISE with 1 day lag	Numeric
Vol	Volume of traded stock of ISE	Numeric
DDay1,2,3	If there is a lag between the match day and the next trade date	Binary
Dvenue	Home or away	Binary
Derby	If the opponent is from the same city	Binary
Drawwin	Unexpected draw when win is expected	Binary
DrawLoss	Unexpected draw when loss is expected	Binary
Winodd	Unexpected win when loss is expected	Binary
Lossodd	Unexpected loss when win is expected	Binary

5.2.2 Model 2

The second model we used for predicting the *change* and *changedummy* is to only use the sentiment analysis. At this stage, we used sentiments gathered from Twitter data to predict the stock's *change* and *changedummy* for each team. This model analyzes the effect of only Twitter sentiments on the *change* and *changedummy*. We used three different scores, the total number of the tweets, the number of positive, negative and neutral tweets in our models. We also define a one day lag for finding the effect of previous day tweets on the next day results. Table 11 presents the independent variables for our second model.

Table 11: Model 2 independent variable

Notation	Independent Variables	Type
Negative	Number of the negative tweets	Numeric
Positive	Number of the positive tweets	Numeric
Neutral	Number of neutral tweets	Numeric
Negativechange	Change in negative tweets between two days	Numeric
Positivechange	Change in positive tweets between two days	Numeric
Neutralchange	Change in neutral tweets between two days	Numeric
Sum	Total number of tweets	Numeric
Score1	(Positive – Negative)/ Sum	Numeric
Score2	(Positive – Negative)/(Sum – Neutral)	Numeric
Score3	Change in positive – Change in negative)/ Change in Sum	Numeric
Score1change	Change in score 1 between two days	Numeric
Sumchange	Change in sum between two days	Numeric
Score1lag1	Score 1 with one day lag	Numeric
Score2lag1	Score 2 with one day lag	Numeric
Score3lag1	Score 3 with one day lag	Numeric

5.2.3 Model 3

The third model we used in our study is a combination of the sentiment analysis and the match results with the financial data. At the last stage, we combined match data and the results of sentiment analysis on Twitter data to find the effect of this combination on *change* and *changedummy*. The independent variables for this model is the combination of the independent variables of Model 1 and Model 2.

In the next chapter we discuss the results of these models and compare the outputs.

Chapter 6: Results

In this chapter, we present the results of each of the three models, separately. We run the models in Rstudio and present the results for each team in the following subsections. We run each model for each team and also combine all of the teams' data in a model named *Total* to predict the amount of stock return (*change*) and the direction of the return (*changedummy*). We compare the result of these models with each other at the end.

6.1 Model 1 (Match Performance + Betting Odds)

As we discussed, this model is the combination of match performance and betting odds. In subsection 6.1.1 we show the results of Model 1 for *change* prediction and in Subsection 6.1.2 we show the results of Model 1 for *changedummy* prediction.

6.1.1 Predicting the value of return in Model 1 (*change*)

We used stepwise selection from both sides for variable selection and we select the variable based on exact AIC. After selecting the variables and running the model, we use 10-folds cross-validation with 3 repetitions to validate our results.

Table 12 presents the summary of the Model 1 results:

Table 12: Model 1 return prediction results

Teams	Multiple R-Squared	Adjusted R-Squared	CV R-Squared	Sarac and Zeren R-Squared	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Fenerbahçe	0.2255	0.2154	0.2270	0.0550	0.0258	0.0174
Besiktaş	0.1284	0.1214	0.1350	0.1250	0.0359	0.0229
Galatasaray	0.0882	0.08092	0.0861	0.0840	0.0307	0.0193
Trabzonspor	0.0795	0.06912	0.0812	-	0.0307	0.0198
Total	0.1136	0.1096	0.1114	-	0.0315	0.0199

In this model, all of the teams have a better accuracy than the previous study on this subject by Sarac and Zeren (2013). Fenerbahçe has the highest explanatory power. The model is highly significant and has a higher multiple R-Squared (22.5%) and adjusted R-Squared (21.5%). Compared to the previous studies, this result with only match performance and betting odds as an indicator of the pre-match expectation is noteworthy.Besiktas's model has explanatory power of 12.8% and adjusted R-Squared of 12.2%. This is also higher than the previous study. Galatasaray's model is also statistically significant and its explanatory power is about 9%. The RMSE and MAE is also low. In Trabzonspor's model the explanatory power is 8% and the model is statistically significant. When we combine all the teams' data together, the model is also significant and its explanatory power is about 11%. We can see the Rstudio outputs for Model 1 in Appendix 1.

6.1.2 Predicting the direction of return in Model 1 (*Changedummy*)

In this model we will predict the direction of each team's stock return and we also combine all of the teams' data to run the *Total* model. We ran LDA, QDA and logistic regression methods for this prediction and

Table 13 presents the results.

Table 13: The direction of return prediction results for Model 1

		LDA	QDA	Logistic Regression	Baseline
Fenerbahçe	**Accuracy**	0.6991	0.6931	0.7	0.6808
	Sensitivity	0.9176	0.6949	0.5221	-
	Specificity	0.2550	0.6892	0.7834	-
	CV Accuracy	0.6925	0.6849	0.7033	-
Beşiktaş	**Accuracy**	0.7118	0.6997	0.7185	0.6501
	Sensitivity	0.8784	0.7361	0.4176	-
	Specificity	0.4023	0.6322	0.8804	-
	CV Accuracy	0.7073	0.6853	0.7139	-
Galatasaray	**Accuracy**	0.7057	0.6751	0.7004	0.6644
	Sensitivity	0.9098	0.7455	0.3175	-
	Specificity	0.3016	0.5357	0.8938	-
	CV Accuracy	0.6982	0.6413	0.6928	-
Trabzonspor	**Accuracy**	0.6757	0.6741	0.6869	0.6438
	Sensitivity	0.8759	0.6700	0.5605	
	Specificity	0.3139	0.6816	0.7568	
	CV Accuracy	0.6699	0.6342	0.6693	
Total	**Accuracy**	0.6823	0.6905	0.6575	0.6602
	Sensitivity	0.9202	0.8691	0.0222	
	Specificity	0.2199	0.3435	0.9843	
	CV Accuracy	0.6803	0.6861	0.6945	

For Fenerbahce and Besiktas, all of the models work better than the baseline and they are statistically significant. For Galatasaray, LDA and Logistic Regression models work better than the baseline but QDA model has a lower cross-validation accuracy than the baseline. For Trabzonspor, LDA and Logistic

Regression models work better than the baseline but QDA model has a lower cross-validation accuracy than the baseline. For the *Total* model QDA performs better than the other predictive methods.

6.2 Model 2 (Twitter Sentiments)

As we mentioned before, in Model 2 we try to predict the stock return and also the direction of return using only the Twitter sentiments. We do not use any match performance data or betting odds in this model to find the effect of Twitter sentiments on the stock price return individually. We also run the *Total* model on the combination of all of the teams' data to compare the results.

6.2.1 Predicting the amount of the return in Model 2 (*change*)

Table 14 presents the summary of the Model 2 results for stock price return.

Table 14: Model 2 return prediction results

Teams	Multiple R-Squared	Adjusted R-Squared	CV R-Squared	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Fenerbahce	0.1413	0.1098	0.0768	0.0294	0.0196
Besiktas	0.0945	0.0612	0.0701	0.0201	0.0149
Galatasaray	0.0991	0.0772	0.0974	0.0240	0.0188
Trabzonspor	0.0668	0.0370	0.0802	0.0260	0.0176
Total	0.0452	0.03331	0.0316	0.0256	0.0175

Compared to Model 1, accuracies of Model 2 with only the use of sentiments for Fenerbahce, Besiktas and Trabzonspor is lower than Model 1. For Galatasaray this model works about 1% better than Model 1.

Fenerbahce's model is statistically significant and its explanatory power is 14% which is 8% lower than Model 1 results. Besiktas model is also statistically significant and its explanatory power is 9%. Galatasaray model is statistically significant and its explanatory power is 9.9% which is higher than Model 1 results.

Trabzonspor's model is statistically significant and its explanatory power is 6.6%. The *Total* model is also significant but its explanatory power is lower than the other models. We can see the Rstudio outputs for Model 2 in Appendix 1.

6.2.2 Predicting the direction of return in Model 2 (*Changedummy*)

In this model we will predict the direction of the stock price return. We run LDA, QDA and logistic regression models on each team separately and together. Table 15 presents the results:

Table 15: The direction of return prediction results for Model2

		LDA	QDA	Logistic Regression	Baseline
Fenerbahce	**Accuracy**	0.652	0.6476	0.6388	0.5683
	Sensitivity	0.8992	0.9225	0.3367	-
	Specificity	0.3265	0.2857	0.8682	-
	CV Accuracy	0.6074	0.6209	0.6039	-
Besiktas	**Accuracy**	0.6872	0.652	0.6828	0.6476
	Sensitivity	0.9932	0.9184	0.1375	-
	Specificity	0.1250	0.1625	0.9795	-
	CV Accuracy	0.6808	0.6519	0.6754	-
Galatasaray	**Accuracy**	0.6274	0.6274	0.6415	0.6274
	Sensitivity	1	1	0.1519	-
	Specificity	0	0	0.9323	-
	CV Accuracy	0.6242	0.6241	0.6226	-
Trabzonspor	**Accuracy**	0.7048	0.7313	0.7048	0.6784
	Sensitivity	0.9935	0.8896	0.12329	-
	Specificity	0.09589	0.3973	0.98052	-
	CV Accuracy	0.6783	0.6602	0.6760	-
Total	**Accuracy**	0.6405	0.6305	0.6473	0.6305
	Sensitivity	0.9627	0.9130	0.0606	
	Specificity	0.0909	0.1485	0.9911	
	CV Accuracy	0.6285	0.6166	0.6267	

For Fenerbahce all three prediction methods are statistically significant. The highest accuracy is for LDA and the highest CV accuracy is for QDA. For Besiktas, the models' p-values are not lower than 0.05 but all of the models' accuracies are better than the baseline. Cross validation accuracies are higher than the baseline and LDA is the best model. For Galatasaray, the accuracies are not good enough and based on the p-values, the models are not statistically significant. CV accuracies are lower than the baseline. For Trabzonspor, the QDA model works well in the training accuracy but not in the CV accuracy. The CV accuracies are lower than the baseline and their p-values are not lower than 0.05 and the models are not statistically significant. The *Total* model is also not significant and CV accuracies are lower than the baseline.

In general, for predicting the direction of return, Model 1 works better than Model 2 based on accuracies and CV accuracies.

6.3 Model 3 (Twitter Sentiments + Match performance + Betting odds)

In this model we combine match performance and betting odds data with the sentiments we acquired from Twitter and our aim is to predict the amount and the direction of the soccer clubs' stock return. Our initial hypothesis was to check if Twitter sentiment in addition to match performance and betting odds can improve the prediction accuracy of the amount and direction of each soccer clubs' stock return. We also run the *Total* model on the combination of all of the team data to compare the results.

6.3.1 Predicting the amount of the return in Model 3 (*change*)

In this subsection, we predict the amount of return for each team based on Model 3. Table 16 presents the summary of model results for change in stock price.

Table 16: Model 3 return prediction results

Teams	Multiple R-Squared	Adjusted R-Squared	CV R-Squared	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Fenerbahce	0.6491	0.4203	0.3192	0.0302	0.0239
Besiktas	0.8289	0.6986	0.6002	0.0229	0.0193
Galatasaray	0.6845	0.5688	0.5741	0.0208	0.0171
Trabzonspor	0.5035	0.3316	0.4260	0.0393	0.0324
Total	0.2326	0.1794	0.2331	0.0275	0.0198

As we can see in Table 16, the multiple R-squared and adjusted R-squared of Model 3 are better than both Model 1 and Model 2. These results can show that the Twitter sentiments in addition to match performance and betting odds data can improve the prediction of the amount of soccer clubs' stock return for our four Turkish teams. All of the models are statistically significant and the RMSE and MAE are low for every model. The *Total* model has the lowest explanatory power because each teams' stock acts differently so combining all the data would not increase the R-squared. Model 3 for predicting the *change* for Besiktas is the best achieved model in this study. The explanatory power is about 83% which is higher than Model 1 and Model 2 and the other studies.

In Appendix 1, we can see the R-studio output for each team.

6.3.2 Predicting the direction of return in Model 3 (*Changedummy*)

In this subsection, we predict the direction of stock return for each team using sentiments, match performance and betting odds data. We ran LDA, QDA and logistic regression methods for this prediction and Table 17 presents the results:

Table 17 :The direction of return prediction results for Model 3

		LDA	QDA	Logistic Regression	Baseline
Fenerbahce	**Accuracy**	0.9200	0.8974	1	0.6154
	Sensitivity	0.9167	1	1	-
	Specificity	0.9333	0.7300	1	-
	CV Accuracy	0.7630	0.7500	0.81	-
Besiktas	**Accuracy**	0.8684	0.8974	1	0.7632
	Sensitivity	0.9310	0.8966	1	-
	Specificity	0.6667	0.8889	1	-
	CV Accuracy	0.73	0.8031	0.78	-
Galatasaray	**Accuracy**	0.881	0.9048	1	0.6905
	Sensitivity	0.9655	0.8966	1	-
	Specificity	0.6923	0.9231	1	-
	CV Accuracy	0.7600	0.7100	0.7655	-
Trabzonspor	**Accuracy**	0.9444	0.9444	1	0.6944
	Sensitivity	0.9600	0.96	1	-
	Specificity	0.9091	0.9091	1	-
	CV Accuracy	0.7981	0.7012	0.76	-
Total	**Accuracy**	0.7484	0.7226	0.7871	0.6903
	Sensitivity	0.9346	0.9439	0.5417	
	Specificity	0.3333	0.2292	0.8972	
	CV Accuracy	0.7333	0.7200	0.7481	

In Fenerbahce's models, all of the predictive methods are statistically significant and the accuracies are better than the baseline with large difference. The best model we found is logistic regression with 0.81 CV accuracy. In Besiktas's models, QDA has the highest CV accuracy but logistic regression has the highest training accuracy. All of the models are statistically significant. In the Galatasary's models, as we can see the accuracies are more than the baseline and the models are statistically significant. The best model is logistic regression with 0.76 CV accuracy. In Trabzonspor's models, the logistic regression model works better than LDA and QDA models but CV accuracy of the LDA is better than the others. All of the models are statistically significant. The *Total* model is also significant and all the predictive methods' CV accuracies are better than the baseline.

Based on Model 3 accuracy and CV accuracy results for the prediction of the stock return direction, we can state that the combination of the match performance, betting odds and Twitter sentiments can predict the direction of the return better than our first two models. This means adding Twitter sentiments to the match performance and betting odds data can improve the model accuracy for predicting the direction of the stock price return.

Chapter 7: Conclusion and Future work

In this study we aimed to predict the amount and direction of change in the soccer clubs' stock return, using a database of four major Turkish teams. According to the finance literature in sports, there are three main methods of stock price prediction in soccer. First based on match results, second based on match importance and third based on pre-match expectation. We tested the hypothesis that whether a match has an effect on a soccer teams' stock price and we found that between the mean of the stock price before and after the match there is a statistically significant difference. After this hypothesis testing, we proposed the inclusion of fan sentiments expressed on Twitter in addition to betting odds as an indicator of the pre-match expectation and we hypothesized that it could improve the prediction models. We ran three main models to check this hypothesis. The first model contained match performance and betting odds data. This model could predict the amount of stock price return for the four chosen teams better than the previous studies. In the second model, we only used Twitter sentiments data to predict the amount and the direction of stock return for these four teams to check the effect of sentiments individually on the stock return. Although the results of Model 2 for Fenerbahce is significant, our results show that sentiments individually are not good predictors of the amount and direction of the stock price return for the other teams. In Model 3, we combined match performance and betting odds data with Twitter sentiments to check whether adding these sentiments to our first model can improve the prediction results. The results showed that sentiments in addition to match performance and betting odds data can improve our prediction models significantly. Although there is a difference between the cross-validation R-squared and the model R-squared due to the lack of match data in one year, still we can state that adding Twitter sentiments to the model can improve the accuracies both in the amount and the direction of our soccer clubs' stock return. Adding all of the teams' data together and run a model on the whole data would not give a high explanatory power to us because each team's stock act differently and combining the data together will mislead the prediction models.

As future work we propose several experiments to build upon our findings. A first proposal is to interpret weekly and monthly returns of the stock price besides only predicting the next day's return.

Another proposal is to consider match importance factors like: ranking of the playing teams, division of the season and giving more importance to the final matches, as well as the division of the on-season and off-season period. Finally we propose to take financial and other stock market factors into account. There are several influential factors like interest rate, number of investors, dividends, and economic situation which can affect the stock price and they can be included in the analysis.

List of References

Akaichi J., Dhouioui Z. and López-Huertas Pérez M. J. (2013), *"Text mining facebook status updates for sentiment classification"*, 2013 17th International Conference on System Theory, Control and Computing (ICSTCC), pp. 640-645.

Arnold, A. J. (1991). An industry in decline? The trend in football league gate receipts. *Service Industries Journal*, *11*(2), 179-188.

Barajas, A., Fernández-Jardón, C. M., & Crolley, L. (2005). Does sports performance influence revenues and economic results in Spanish football? Available at SSRN 986365.

Barbosa L. & Feng J. (2010). *"Robust sentiment detection on Twitter from biased and noisy data"*. In Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10). Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 36-44.

Barnaghi P., Breslin John G., Ghaffari P. (2016). *"Opinion Mining and Sentiment Polarity on Twitter and Correlation Between Events and Sentiment",* 2016 IEEE Second International Conference on Big Data Computing Service and Applications, pp. 52 – 57

Bell, A. R., Brooks, C., Matthews, D., & Sutcliffe, C. (2012). Over the moon or sick as a parrot? The effects of football results on a club's share price. Applied Economics, 44, 3435–3452. doi:10.1080/00036846.2011.577017

Benkraiem, R., Louhichi, W., & Marques, P. (2009). Market reaction to sporting results: The case of European listed football clubs. Management Decision, 47, 100–109. doi:10.1108/ 00251740910929722

Bermingham A. & Smeaton A. (2010). *"Classifying sentiment in microblogs: is brevity an advantage?".* In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, pp. 1833-1836.

Birkhäuser, S., Kaserer, C., & Urban, D. (2015), Investor Presence and Competition in Major European football Leagues, TUM Working Paper.

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

Castellani, M., Pattitoni, P., & Patuelli, R. (2015). Abnormal returns of soccer teams: Reassessing the informational value of betting odds. Journal of Sports Economics, 16, 735–759.

Demir, E., & Danis, H. (2011). The effect of performance of soccer clubs on their stock prices: Evidence from Turkey. Emerging Markets Finance and Trade, 47, 58–70. doi:10.2753/ REE1540-496X4705S404

Demir, E., & Rigoni, U. (2017). You lose, I feel better: Rivalry between soccer teams and the impact of schadenfreude on stock market. *Journal of Sports Economics*, *18*(1), 58-76.

Devecioğlu, S. (2004). Halka Arz Edilen Spor Kulüplerinin Sportif Başarıları ile Piyasa Değerleri Arasındaki İlişki. Spormetre Beden Eğitimi ve Spor Bilimleri Dergisi. 2(1), 11-18

Duque, J., & Ferreira, N. A. (2005). Explaining share price performance of football clubs listed on the Euronext Lisbon (Working Paper no 05-01). Lisbon, Portugal: Technical University of Lisbon.

Floros, C. (2014). Football and stock returns: New evidence. *Procedia Economics and Finance*, *14*, 201-209.

Godinho, P., & Cerqueira, P. (2018). The impact of expectations, match importance, and results in the stock prices of European football teams. *Journal of Sports Economics*, *19*(2), 230-278.

Gokulakrishnan B, Priyanthan P, Ragavan T, Prasath N, Perera A (2012) Opinion mining and sentiment analysis on a Twitter data stream. In: Proceedings of the 2012 international conference on advances in ICT for emerging regions (ICTer), Colombo, pp 182–188

Göllü, E. (2012). Impact of the financial performances of incorporations of football clubs in the domestic league on their sportive performances: A study covering four major football clubs in Turkey. *Pamukkale Spor Bilimleri Dergisi*, *3*(1), 20-29.

Habernal I., Ptáček T., Steinberger J. (2014), *Reprint of "Supervised sentiment analysis in Czech social media",* Information Processing & Management, Volume 51, Issue 4, pp 532-546.

Liu, B. (2010), "Sentiment analysis and subjectivity", in Indurkhya, N. and Damerau, F.J. (Eds), Handbook of Natural Language Processing, Vol. 2, CRC, Chapman and Hall, pp. 627-666.

Liu B., Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) Mining Text Data. Springer, Boston, MA

Majewski, S. (2014). Modelling of football companies' rates of return according to sport results and bookmakers' expectations on the example of serie A. *Business and Economic Horizons*, *10*(3), 214-222.

Martínez-Cámara E., Martín-Valdivia M.T. & Ureña-López, L.A. (2011*), "Opinion classification techniques applied to a Spanish corpus"*, Natural Language Processing and Information Systems, Springer, pp. 169-176.

Ozturkcan, S., Kasap, N., Tanaltay, A., & Ozdinc, M. (2019): Analysis of tweets about football: 2013 and 2018 leagues in Turkey, Behaviour & Information Technology, DOI: 10.1080/0144929X.2019.1583284

Pak A., & Paroubek P. (2010). *"Twitter as a Corpus for Sentiment Analysis and Opinion Mining".* LREC, Vol 10, Issue 4, pp. 1320-1326

Pandey, V., Iyer, C. (2009) Sentiment analysis of microblogs. <www.stanford.edu/class/cs229/proj2009/PandeyIyer.pdf>, unpublished

Palomino, F., Renneboog, L., & Zhang, C. (2009). Information salience, investor sentiment, and stock returns: The case of British soccer betting. Journal of Corporate Finance, 15, 368–387.

Pang B., Lee L. & Vaithyanathan S. (2002), *"Thumbs up? Sentiment classification using machine learning techniques"*, paper presented at the Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Vol. 10, Philadelphia, PA, July 6.

Pingue, F. (2019) Real Madrid unseat Man United as most valuable soccer team retrieved from https://uk.reuters.com/

Plunkett Research (2019), Sports Industry Statistic and Market Size Overview, Business and Industry Statistics retrieved from https://www.plunkettresearch.com

Qazi A., Tamjidyamcholo A., Raj R. G., Hardaker G., Standing C. (2017), *"Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach"*, Computers in Human Behavior, Volume 75, 2017, pp. 450-460

Read, J. (2005) Using emoticons to reduce depen- dency in machine learning techniques for sentiment clas- sification. In ACL. The Association for Computer Linguistics

Renneboog, L., & Vanbrabant, P. (2000). Share price reactions to sporting performances of soccer clubs listed on the London Stock Exchange and the AIM (CentER DP 2000-19). Tilburg, the Netherlands: University of Tilburg.

Samagaio, A., Couto, E., & Caiado, J. (2009). Sporting, financial and stock market performance in English football: an empirical analysis of structural relationships. *Centre for Applied Mathematics and Economics (CEMAPRE) working papers*, 1-41.

Saraç, M., & Zeren, F. (2013). The Effect of Soccer Performance on Stock Return: Empirical Evidence From" The Big Three Clubs" of Turkish Soccer League. *Journal of Applied Finance and Banking*, *3*(5), 299.

Scholtens, B., & Peenstra, W. (2009). Scoring on the stock exchange? The effect of football matches on stock market returns: An event study. Applied Economics, 41, 3231–3237.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J. ,and Napolitano, A. (2008), "Building useful models from imbalanced data with sampling and boosting," in Proc. 21st Int. FLAIRS Conf., , pp. 306–311.

Smailović J., Grčar M., Lavrač N. and Žnidaršič M. (2014), *"Stream-based active learning for sentiment analysis in the financial domain",* Information Sciences, Vol. 285 No. C, pp. 181-203.

Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2013). Predictive sentiment analysis of tweets: A stock market application. In *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 77-88). Springer, Berlin, Heidelberg.

Stadtmann, G. (2004). An empirical examination of the news model: The case of Borussia Dortmund GmbH & Co. KGaA. Zeitschrift fu¨r Betriebswirtschaft, 74, 165–185.

Turney P. D. (2002), *"Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews"*, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 417-424.

Szymanski, S. (1998). Why is Manchester United so successful? Business Strategy Review, 9(4), 47-54.

Szymanski, S., & Kuypers, T. (1999). Winners and losers. London: Viking

Wirth R. (2000), *"CRISP-DM: Towards a standard process model for data mining"*, Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, pp. 29-39

Ye Q., Zhang Z. and Law R. (2009), *"Sentiment classification of online reviews to travel destinations by supervised machine learning approaches"*, Expert Systems with Applications, Vol. 36 No. 3, pp. 6527-6535.

Zhai Z., Xu H., Kang B., Jia P. (2011), *"Exploiting effective features for chinese sentiment classification"*, Expert Systems with Applications, Volume 38, Issue 8, 2011, pp. 9139-9146

Zuber, R. A., Yiu, P., Lamb, R. P., & Gandar, J. M. (2005). Investor–fans? An examination of the performance of publicly traded English Premier League teams. Applied Financial Economics, 15, 305–313.m

Appendix 1

Model 1 Rstudio outputs:

1. **Fenerbahçe**

Predicting stock return

```
Coefficients:
            Estimate  Std. Error  t value Pr(>|t|)
(Intercept) -0.0537696  0.0057039  -9.427  < 2e-16 ***
Change       0.0034208  0.0005309   6.444  2.05e-10 ***
DDay1        0.0514950  0.0056157   9.170  < 2e-16 ***
DDay2        0.0550592  0.0058016   9.490  < 2e-16 ***
DDay3        0.0498576  0.0065508   7.611  7.96e-14 ***
GDiff        0.0020102  0.0007214   2.786  0.00546 **
Derby       -0.0047791  0.0028057  -1.703  0.08890 .
Dvenue      -0.0045082  0.0019247  -2.342  0.01942 *
Champ       -0.0061554  0.0039047  -1.576  0.11534
Loseodd     -0.0093071  0.0034830  -2.672  0.00770 **
Drawwin     -0.0129192  0.0027643  -4.674  3.49e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02596 on 771 degrees of freedom
Multiple R-squared:  0.2255,	Adjusted R-squared:  0.2154
F-statistic: 22.45 on 10 and 771 DF,  p-value: < 2.2e-16
```

Figure 7: R output for Model 1 Fenerbahce without outliers

2. **Besiktaş**
   Predicting stock return

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0009651  0.0024187   0.399 0.689987
Change       0.0045104  0.0007551   5.973 3.61e-09 ***
Dvenue      -0.0047093  0.0027384  -1.720 0.085903 .
Euro        -0.0091518  0.0042372  -2.160 0.031103 *
GDiff        0.0037926  0.0010237   3.705 0.000227 ***
Loseodd     -0.0113897  0.0046415  -2.454 0.014362 *
Drawwin     -0.0131013  0.0037442  -3.499 0.000495 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0363 on 739 degrees of freedom
Multiple R-squared:  0.1284,	Adjusted R-squared:  0.1214
F-statistic: 18.15 on 6 and 739 DF,  p-value: < 2.2e-16
```

Figure 8: R output for Model 1 Besiktas without outliers

3. **Galatasaray**

Predicting stock return

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.000524   0.001597    0.33    0.743
DDay2        0.004051   0.002661    1.52    0.128
Champ       -0.009013   0.004801   -1.88    0.061 .
Loseodd     -0.007490   0.003296   -2.27    0.023 *
Change       0.004112   0.000635    6.48  1.7e-10 ***
Drawwin     -0.014314   0.003040   -4.71  3.0e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0312 on 745 degrees of freedom
Multiple R-squared:  0.0869,	Adjusted R-squared:  0.0808
F-statistic: 14.2 on 5 and 745 DF,  p-value: 2.81e-13
```

Figure 9: R output for Model 1 Galatasaray without outliers

4. **Trabzonspor:**

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.019502   0.004615   -4.23  2.7e-05 ***
DDay1        0.017412   0.004820    3.61  0.00033 ***
DDay2        0.021324   0.005275    4.04  5.9e-05 ***
DDay3        0.025394   0.006116    4.15  3.8e-05 ***
Loseodd     -0.011287   0.003447   -3.27  0.00112 **
Change       0.003105   0.000739    4.20  3.1e-05 ***
Drawwin     -0.008079   0.003473   -2.33  0.02035 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0312 on 619 degrees of freedom
Multiple R-squared:  0.0779,	Adjusted R-squared:  0.069
F-statistic: 8.72 on 6 and 619 DF,  p-value: 3.95e-09
```

Figure 10: R output for Model 1 Trabzonspor without outliers

Model 2 Rstudio outputs:

1. **Fenerbahce**

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.419e-01  9.975e-02   2.426 0.016101 *
Score1lag1    -5.963e-02  2.775e-02  -2.149 0.032773 *
Score2        -8.602e-01  5.856e-01  -1.469 0.143301
Sum            5.226e-07  2.342e-07   2.231 0.026706 *
Score1         1.291e+00  6.706e-01   1.925 0.055510 .
Sumchange      4.781e-02  1.827e-02   2.616 0.009506 **
Score3         3.199e-03  8.094e-04   3.952 0.000105 ***
positivechange -4.807e-02  1.952e-02  -2.462 0.014590 *
S.positive    -5.454e-01  2.200e-01  -2.479 0.013950 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02769 on 218 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1413,	Adjusted R-squared:  0.1098
F-statistic: 4.486 on 8 and 218 DF,  p-value: 4.664e-05
```

Figure 11: R output for Model 2 Fenerbahce

2. **Besiktas**

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        2.063e-01  5.196e-02   3.971 9.73e-05 ***
Score2            -4.479e-01  2.882e-01  -1.555  0.12151
neutralchange     -5.555e-03  2.132e-03  -2.605  0.00981 **
Score1             7.292e-01  3.542e-01   2.059  0.04072 *
Sumchange          6.461e-03  2.611e-03   2.475  0.01410 *
Score1change       8.297e-05  5.858e-05   1.416  0.15811
Positivechangelag1  4.931e-03  2.837e-03   1.738  0.08358 .
negativechangelag1 -3.376e-03  2.049e-03  -1.648  0.10083
S.positive        -4.702e-01  1.153e-01  -4.078 6.37e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0201 on 218 degrees of freedom
Multiple R-squared:  0.0945,	Adjusted R-squared:  0.06127
F-statistic: 2.844 on 8 and 218 DF,  p-value: 0.005037
```

Figure 12: R output for Model 2 Besiktas

3. Galatasaray

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.021e-01  4.277e-02   2.387  0.01787 *
Score2             8.920e-02  4.408e-02   2.024  0.04429 *
Sum                4.012e-07  1.968e-07   2.039  0.04276 *
Positivechangelag1  8.230e-03  4.451e-03   1.849  0.06591 .
negativechangelag1 -9.942e-03  3.579e-03  -2.778  0.00597 **
S.positive        -2.433e-01  9.899e-02  -2.458  0.01480 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02481 on 206 degrees of freedom
Multiple R-squared:  0.0991,	Adjusted R-squared:  0.07724
F-statistic: 4.532 on 5 and 206 DF,  p-value: 0.0006092
```

Figure 13: R output for Model 2 Galatasaray

4. Trabzonspor

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        9.750e-02  6.335e-02   1.539  0.12524
Score2            -3.782e-01  2.323e-01  -1.628  0.10487
Sum                3.241e-06  1.105e-06   2.932  0.00373 **
Score1             6.220e-01  3.457e-01   1.799  0.07333 .
Sumchange          1.221e-02  5.922e-03   2.062  0.04043 *
positivechange    -1.749e-02  7.124e-03  -2.455  0.01486 *
Positivechangelag1 -3.150e-03  1.701e-03  -1.852  0.06534 .
S.positive        -2.701e-01  1.514e-01  -1.785  0.07570 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02753 on 219 degrees of freedom
Multiple R-squared:  0.06684,	Adjusted R-squared:  0.03701
F-statistic: 2.241 on 7 and 219 DF,  p-value: 0.03209
```

Figure 14: R output for Model 2 Trabzonspor

Model 3 Rstudio outputs:

1. **Fenerbahce:**

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.351291   0.290974   1.207   0.2396
ISEchangelag1     -0.006643   0.004260  -1.559   0.1325
Score1lag1         0.050297   0.043162   1.165   0.2558
neutralchange     -0.030333   0.012481  -2.430   0.0233 *
Score1             0.401151   0.333005   1.205   0.2406
Sumchange          0.099036   0.042907   2.308   0.0303 *
positivechange    -0.062285   0.038892  -1.602   0.1229
Positivechangelag1 -0.007122  0.003140  -2.268   0.0330 *
S.positive        -0.808287   0.639475  -1.264   0.2189
DDay1              0.017625   0.015036   1.172   0.2531
DDay2              0.035510   0.018106   1.961   0.0621 .
Besiktas          -0.026459   0.021995  -1.203   0.2412
Loseodd            0.038690   0.017291   2.238   0.0352 *
Change             0.007991   0.003145   2.541   0.0183 *
Drawwin           -0.017053   0.012646  -1.349   0.1906
Drawlose           0.054383   0.027014   2.013   0.0559 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02305 on 23 degrees of freedom
Multiple R-squared:  0.6491,	Adjusted R-squared:  0.4203
F-statistic: 2.837 on 15 and 23 DF,  p-value: 0.012
```

Figure 15: R output for Model 3 Fenerbahce

2. Besiktas:

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.075e-01  1.760e-01   2.884 0.008874 **
ISEchangelag1  -5.950e-03  2.933e-03  -2.028 0.055400 .
Score2        -4.645e+00  1.817e+00  -2.557 0.018379 *
Sumlag1        7.908e-07  3.849e-07   2.055 0.052573 .
neutralchange   6.846e-02  3.395e-02   2.017 0.056677 .
Score1         5.733e+00  2.080e+00   2.756 0.011834 *
Sumchange     -1.001e+00  4.293e-01  -2.332 0.029719 *
positivechange  4.358e-01  1.918e-01   2.272 0.033728 *
negativechange  4.847e-01  2.035e-01   2.381 0.026800 *
S.positive    -1.197e+00  3.943e-01  -3.036 0.006279 **
Champ          2.804e-02  1.877e-02   1.494 0.150140
Euro           3.357e-02  1.455e-02   2.307 0.031336 *
Derby         -4.130e-02  1.234e-02  -3.348 0.003049 **
Change         1.123e-02  2.555e-03   4.394 0.000253 ***
GDiff          3.240e-03  2.034e-03   1.593 0.126103
Drawwin        1.668e-02  9.593e-03   1.739 0.096682 .
Drawlose      -3.531e-02  2.107e-02  -1.676 0.108650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01543 on 21 degrees of freedom
Multiple R-squared:  0.8289,	Adjusted R-squared:  0.6986
F-statistic: 6.361 on 16 and 21 DF,  p-value: 6.788e-05
```

Figure 16: R output for Model 3 Besiktas

3. Galatasaray

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         3.012e-01  1.320e-01   2.282 0.029751 *
ISEchangelag1       5.870e-03  2.864e-03   2.050 0.049215 *
Score1lag1         -7.280e-02  3.699e-02  -1.968 0.058348 .
Sum                 2.403e-06  6.163e-07   3.899 0.000504 ***
Score1              3.703e-01  1.514e-01   2.446 0.020517 *
Sumchange           7.747e-03  3.505e-03   2.210 0.034840 *
negativechangelag1 -5.288e-03  2.961e-03  -1.786 0.084211 .
S.positive         -7.964e-01  3.074e-01  -2.590 0.014658 *
DDay1               2.419e-02  6.840e-03   3.537 0.001338 **
Champ              -2.770e-02  1.248e-02  -2.219 0.034214 *
Derby               2.432e-02  1.002e-02   2.427 0.021452 *
GDiff               4.671e-03  1.730e-03   2.701 0.011268 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01875 on 30 degrees of freedom
Multiple R-squared:  0.6845,	Adjusted R-squared:  0.5688
F-statistic: 5.917 on 11 and 30 DF,  p-value: 5.01e-05
```

Figure 17: R output for Model 3 Galatasaray

4. Trabzonspor

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.72142    0.23039   3.131  0.00427 **
neutralchange      -0.22652    0.08226  -2.754  0.01061 *
Score1              0.95337    0.27614   3.452  0.00191 **
Sumchange           1.34431    0.43080   3.121  0.00438 **
positivechange     -0.45495    0.18160  -2.505  0.01883 *
negativechange     -0.62753    0.18408  -3.409  0.00214 **
Positivechangelag1  0.12303    0.05456   2.255  0.03279 *
negativechangelag1 -0.10355    0.05115  -2.024  0.05332 .
S.positive         -1.69198    0.53609  -3.156  0.00402 **
Drawlose            0.07330    0.03235   2.266  0.03201 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03066 on 26 degrees of freedom
Multiple R-squared:  0.5035,	Adjusted R-squared:  0.3316
F-statistic:  2.93 on 9 and 26 DF,  p-value: 0.0154
```

Figure 18: R output for Model 3 Trabzonspor