INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Does Social Media Affect Football Clubs' Stock Prices?

João Pedro Dinis Mineiro

Master in Management

Supervisor:
PHD Professor Paulo Jorge Varela Lopes Dias, Assistant Professor at ISCTE Businiss School, Accounting Department

October 2022

BUSINESS
SCHOOL

Accounting Department

Does Social Media Affect Football Clubs' Stock Prices?

João Pedro Dinis Mineiro

Master in Management

Supervisor:
PHD Professor Paulo Jorge Varela Lopes Dias, Assistant Professor at ISCTE Businiss School, Accounting Department

October 2022

Resumo

Hoje em dia, as pessoas publicam milhões de atualizações nas redes sociais, onde expressam as suas opiniões, ideias e sentimentos. Este estudo visa avaliar o efeito do sentimento expresso no Twitter no retorno das ações dos clubes de futebol, através de uma análise empírica do caso português. O estudo baseia-se em dados do Sporting, Benfica, e Porto entre a época de 2016/2017 e a época de 2020/2021. Com base em estudos anteriores realizados por Saraç e Zeren são realizados modelos de Regressões Logísticas com o objetivo de analisar o efeito do sentimento expresso pelos adeptos de futebol na direção do retorno das ações dos clubes controlando variáveis como o índice de mercado PSI20, o sentimento geral neutro e negativo expresso no Twitter, o número total de Tweets por dia, o resultado geral do sentimento, a probabilidade de ganhar um determinado jogo, o resultado do jogo, o tipo de jogo, a diferença dos golos marcados, e a importância do jogo Nacional. A fim de obter o sentimento associado a cada tweet, foram analisados cerca de 79.000 tweets entre julho de 2016 e maio de 2021 através de uma Análise de Sentimento em Python. Os resultados indicam que o sentimento geral neutro e negativo expresso no Twitter é considerado estatisticamente significativo para o Sporting. Para o Sporting e Benfica, um modelo que combina variáveis de desempenho desportivo com o sentimento expresso no Twitter oferece uma melhor adequação dos dados quando comparado com um modelo que se baseia apenas em variáveis de sentimento.

Palavras-chave: *Indústria do Futebol, Análise desportiva, Redes sociais, Análise de sentimento, Twitter*

Classificação JEL: *G41 – Role and Effects of Psychological, Emotional, Social and Cognitive Factors on Decision Making in Financial Markets; L83 - Sports*

Abstract

Nowadays, people post millions of status updates on social media, where they express their opinions, ideas, and sentiments. This study aims to assess the effect of the sentiment expressed on Twitter on the football clubs' stock return through an empirical analysis of the Portuguese case. The study is based on data from Sporting, Benfica, and Porto, the biggest football clubs in Portugal, between the 2016/2017 season and the 2020/2021 season. Inspired by previous work done by Saraç and Zeren, Logistic Regressions models are performed to analyze the effect of the expressed sentiment by football fans on the direction of the clubs' stock returns controlling variables such as the market index, the overall neutral and negative sentiment expressed on Twitter, the total number of Tweets per day, the overall sentiment score, the probability of winning a given match, match result, match type, the difference of scored goals, and the importance of the National match. In order to obtain the associated sentiment to each tweet, around 79,000 tweets from July 2016 to May 2021 were analyzed through a Sentiment Analysis on Python. The findings indicate that the overall neutral and negative expressed sentiment on Twitter is found statistically significant for Sporting, and that for Sporting and Benfica, a model that combines sporting performance variables with the sentiment expressed on Twitter offers a better fit to the data when compared to a model that is based only on sentiment variables.

Key words: *Football industry, Sports analytics, Social media, Sentiment Analysis, Twitter*

JEL Classification: *G41 – Role and Effects of Psychological, Emotional, Social and Cognitive Factors on Decision Making in Financial Markets; L83 - Sports*

Index

List of Tables

List of Charts

List of Figueres

1. Introduction

Football leads the top list of the world's most popular sports with an estimated 3.5 billion supporting fans. As a result, nowadays, besides a sport, it is also a massive industry with significant economic importance[1], involving many professionals, businesses and organizations involved in investing, producing, organizing, and facilitating a variety of sports activities (Tanaltay et al., 2021).

Football clubs' revenues do not all come from transferring players among the different teams but also from another different sources, such as tickets, associate's fees, sponsoring, broadcasting rights, marketing, merchandising and others. Thus, since football clubs need money to finance their activities, they have two main options: increasing financial debt (for example, by getting credit from banks) or attracting inventors in the bond or stock markets. Therefore, to be financed and be able to maintain a high level in national competitions and fight for European trophies, some clubs made an initial public offering, and their stock can be publicly traded over the stock exchange market.

As any other publicly traded company, football clubs desire their shares to be an attractive investment and a secure form of financing. The debate arises on whether these publicly listed football clubs' stock prices depend not only on teams' performance but also on investors' sentiments. In fact, some literature concluded that the investors are no longer concerned only about financial performance but also with the club's success.

Regarding team performance, multiple independent variables might cause an impact on the stock prices, such as match type (for example National match or European match), goal difference between the two teams in a game, if the match went to extra time or penalties, home or away contests, derbies, unexpected results, among others (Tanaltay et al., 2021). Moreover, some studies showed that investors actions are likely also to be driven by their emotions (Nofer & Hinz, 2015) and passion for seeing their club being successful. Therefore, social media appears to be a promising technology instrument to investigate since it is there that most people write, share their ideas and thoughts, comment, discuss, interact, and communicate.

During the last decades, internet and social media have evolved and have been changing the society. The massive use of the internet and social media as a source of information triggered an increasing online activity[2]. This increment of the use of technological systems generates massive datasets that document collective behavior in a previously unimaginable fashion. Thus, in this repertory of internet activity, it is possible to find the global population's interests, concerns or intentions concerning various phenomena (Ranco et al., 2015) and analyze several impacts on different

[1] According to the Annual Review of Football Finance by Deloitte (2021), the combined European football market in the 2019/2020 season generated around €25.2bn in revenues.

[2] A recent study by DatePortal (2022) reveals that the average amount of time that internet users aged 16 to 64 spend using the internet each day on any device is 7 hours and 56 minutes in Portugal, and 6 hours and 58 minutes worldwide (DatePortal website: www.dateportal.com)

markets. Many researchers believe that the public mood or sentiment expressed in social media is related to financial markets performance (Ruan et al., 2018).

This study aims to understand if the sentiments expressed by the clubs' fans on social media impact the stock prices of the three most successful Portuguese football clubs publicly listed: Sport Lisboa e Benfica, Futebol Clube do Porto and Sporting Clube de Portugal. Furthermore, it will be possible to understand how it can be helpful in terms of predicting events based on people's opinions and sentiments.

The most common social media platforms in Portugal are Facebook, Instagram, and Twitter[3]. However, Instagram is mainly used for posting pictures. The purpose of Twitter is precisely for people to use its space to interact, share and discuss their ideas and thoughts with other users and express their feelings, making it the best social media platform for this case study. Additionally, most of the previous literature used Twitter data to analyze the sentiment of investors[4]. Twitter is a popular real time microblogging, with millions of daily active users[5], that allows them to share short information known as tweets, where they express their opinions about various topics. Twitter's popularity became a valuable asset to analyze the sentiment of the tweets by millions of users (Sarlan et al., 2014).

Regarding these three football clubs, this study sets out to find answers to the following questions:

1) Do sentiments expressed on Twitter provoke a change in stock prices?
2) Can the "wisdom of the crowd" (Surowiecki, 2005)[6] tell us something about stock returns?
3) How valuable and accurate is social media as a forecasting tool?

In order to find answers to the central questions of this dissertation, two models based on Logistic Regressions were performed for each of the three football clubs. Both models were created with the objective of investigating whether the direction of stock returns of these clubs is impacted by the overall sentiment expressed on Twitter or not. The difference between the first and the second model are the independent variables used in each model to test such hypothesis. The first model combines sentiment, sporting performance and match result probabilities independent variables, whereas the second model only includes independent variables related with sentiments. This is to also test whether the variables combination of model 1 offers a better goodness of fit when compared to the second

[3] Portugal in early 2022: Facebook – 5.95 million users; Instagram – 5.40 million users; Twitter – 1.40 million users. Source: DatePortal (website: www.dateportal.com)

[4] For example: Jai-Andaloussi et al. (2016); Broadstock and Zhang (2019)

[5] According to Duz Tan and Tas (2021), by the end of 2017, Twitter had around 330 million monthly active users.

[6] Based on Surowiecki's book, not all crowds (groups) are wise (e.g., mobs or crazed investors in a stock market bubble). He defined the key criteria that separate wise crowds from irrational ones: diversity of opinion, independence, decentralization, aggregation, and trust.

model. In order to classify the users' posted tweets into sentiments, a set of data was extracted from Twitter API, and a Sentiment Analysis[7] of the collected data was performed.

Regarding the dissertation structure, in this first chapter an introduction to the football industry and social media was made, as well as a brief overview of the dissertation. In the last subsection of this chapter, a brief introduction to each football club under analysis is also done. In the second chapter, a chronological literature review is done. This chapter is divided into two subsections. The first one consists of literature related to the football industry and its connection with stock markets, whereas the second subsection consists of literature related to social media outcome data assessment and its correlation with different stock markets. The third chapter presents the methodology used, the studied data, and its sources. In the fourth chapter, all results achieved by using the two Logistic Regression models and analysis are presented, and finally, the fifth chapter concludes the dissertation.

1.1. The "Big Three" Football Clubs in Portugal

Football has been the most popular sport in Portugal for many years since it has a long history in the country. The first organized game took place in 1875, and soon after the start of the 20th century, most precisely in 1934, was founded the National domestic league called Primeira Liga where the "Big three" – Sporting Clube de Portugal, Sport Lisboa e Benfica, and Futebol Clube do Porto – emerged and early started to dominate the league.

Futebol Clube do Porto, also known as FCP or Porto, is the oldest football club among "the biggest three" in Portugal. Since its foundation in 1893, FCP has already conquered 82 trophies – 75 in National competitions and 7 played internationally. They won 30 Primeira Liga trophies (five consecutively conquered between the 1994/1995 to 1998-1999 seasons), 18 Portuguese Cups, 23 Cândido de Oliveira Super Cups, and 4 Capeonatos de Portugal. Regarding international competitions, FCP stands out from SCP and SLB. They won UEFA Champions League in 1987 and 2004, the International Cup in 1987 and 2004, the UEFA Super Cup in 1897, and the Europe League in 2003 and 2011. Moreover, Porto is ranked 19th by UEFA club ranking. Today, according to TransferMarket[8] FCP has about 127,000 associates.

Sport Lisboa e Benfica, commonly known as SLB or simply as Benfica, was founded in 1904. According to Finance Football, Benfica is listed among the top 10 clubs worldwide with more associates (approximately 244,000, according to TransferMarket). Based in Lisbon, Benfica is the most decorated club in Portugal, with 83 conquered trophies since its foundation. Better than Futebol Clube do Porto

[7] Sentiment Analysis is a process that automates the mining of attitudes, opinions, views, and emotions from text, speech, tweets, and database sources through Natural Language Processing (NPL), classifying them into three categories: positive, neutral, or negative (Kharde & Sonawane, 2016).

[8] Data available at https://bityli.com/UX4jY.

in National competitions since they achieved 80 trophies in Portugal – 37 Primeira Liga trophies, 26 Portuguese Cups, 7 League Cups, 8 Cândido de Oliveira Super Cups, and 3 Campeonatos de Portugal. However, when it comes out to International competitions, Benfica is not as well succeeded as FCP – Benfica conquered twice the UEFA Champions League, first in the 1960/1961 and 1961/1962 seasons. However, they are ranked 22nd by the UEFA club ranking, which is an excellent place. Moreover, in 2008, Benfica was the first Portuguese football club to launch a channel with its own sports-oriented television network called Benfica TV.

The last of the "biggest three" Portuguese football clubs to be founded was Sporting Clube de Portugal. Founded in 1906, Sporting Clube de Portugal, commonly known as Sporting, is the third most decorated Portuguese football club. Since its foundation, they have conquered 54 trophies. They won 19 Primeira Liga trophies, 17 Portuguese Cups, 4 Campeonatos de Portugal, 4 League Cups, and 9 Super Cups. Internationally, Sporting only won one European Cup title in the 1963/1964 season. Nevertheless, Sporting Clube de Portugal is ranked 33rd by the UEFA club ranking. In 2014, Sporting also founded its tv channel, SportingTV, which broadcasts live Sporting matches at Academia Sporting, where the younger players play, and matches played at Pavilhão João Rocha, where are played other sports such as indoor football, basketball, and volleyball. Five years ago, more precisely, on the 15th of May 2018, a group of supporters invaded the training center of Sporting at Alcochete and attacked the football players because they were dissatisfied with the performance of the team. This event had an enormous negative impact on Sporting since the most important and valuable football players left the club by unilateral termination alleging a lack of security to maintain the contractual link with the club. However, the club recovered from that disastrous situation and won the National competition in the 2020/2021 season. Sporting is based in Lisbon, just as Benfica is, making them big city rivals. Today, according to TransferMarket, Sporting has approximately 117,000 associates.

Football is firmly incorporated into Portuguese culture, which makes it an exciting topic to study. Many researchers did much work related to the topic of football, trying to find some correlation between the stock price returns of a football club and its sporting performance, match importance, variables such as the odds, or even sentiments captured on social media. Also, social media is extremely interesting nowadays since many people's opinions are formed and consolidated by the immense information available online. The following chapter thus presents some of the work already done related to these topics in order to contextualize the investigation done in this dissertation.

2. Literature Review

In order to contextualize the indeed investigation, a review of the literature on various approaches to predicting stock prices is presented. The investment strategy in stocks in companies operating in other markets, most of the times purely involves an economic rational. When investing in football clubs, part of the literature state that while investors are also likely to carefully analyze the economic and financial information available, their decisions are also based on an emotional component that is not predominant or a driven in other stock market investments.

In summary, when it comes to sports, investors are more likely to be affected by emotions while investing in stocks. There are two main methods of stock price prediction in sports (not only football, but also other sports such as basketball): based on on-field and off-field factors and investors' sentiments and expectations.

Thus, the following paragraphs will be divided as follows:

i) On-field and Off-field factors' impact on the Football Clubs' Stock Market Returns.

ii) Social Media Sentiments' impact on the Stock Market Returns.

2.1. On-field and Off-field factors impact on the Football Clubs' Stock Market Returns

Several studies focused on predicting the football clubs' stock prices, considering the effects of off-field and on-field factors. Off-field factors are associated with managerial decisions, coach changes, player transfers, betting odds, generally, features unrelated to the match itself. In contrast, on-field factors are related to the match performance itself and how it can affect the clubs' stock price (Tanaltay et al., 2021).

Renneboog and Vanbrabant were two of the first authors studying the effect of the performance of a football team on stock prices. They considered the football teams' weekly sporty performance to investigate whether it influences the share prices of all the football clubs listed on the London Stock Exchange and the Alternative Investment Market from 1995 to 1998 using the event study methodology. They concluded that the first day of trading after a match win presents abnormal returns of almost 1% on the clubs' stock exchange. Contrarily, after a defeat or a draw, they are penalized with negative abnormal returns of 1.4% and 0.6%, respectively.

Later, Ashton et al. (2003), tried to understand the relationship between the performance of the England football national team and daily changes in the FTSE100 index, instead of focusing on a specific football club. By using an event study methodology, they found a statistically significant relationship between the performance of the English national football team and the change in the price of shares on the London stock exchange. In short, the national team's good (bad) performances are followed by good (bad) market returns.

Using a regression model, Zuber et al. (2005) analyzed a group of ten professional football clubs in the English Premier League between 1997 and 2000. The financial data was extracted from Data Stream. Beyond the FTSE index variable and other game-related variables such as the difference in goals scored incorporated in their model, they introduced a dummy variable for the teams' current position in the national league to determine the importance of the matches between the bottom five teams. However, they found this variable statistically insignificant. Further, they argue that football club investors are less active compared to traditional investors, trading only half as much as measured by average daily volume as a percentage of outstanding shares.

Shortly after, Ferreira (2005) investigated whether share prices of two football clubs listed on the Portuguese Stock Exchange – Sporting and Porto – are influenced by the football team's weekly sporting performances, extending the analysis to stock price volatility. Using Autoregressive Conditional Heteroskedasticity (ARCH) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) methodologies, they found that share prices always presented a positive mean return whenever the team wins the national championship and vice-versa. They concluded that success on the pitch is positively associated with good share price performances in the Stock Exchange. In contrast, negative performances, such as defeats and draws, are associated with negative stock price returns. Taking volatility into account, they concluded that it is related with trading volume, which is in line with some findings of previous studies.

Berument and Ceylan (2012) assessed the relationship between football clubs' performances against foreign rivals and stock market returns and return volatility using an ARCH methodology. They argued that investors become more risk-averse after a defeat and less risk-averse after a win. They found evidence that football match scores affect stock market returns and stock market return-volatility relationship. In countries with high football success (England and Spain), agents become more risk-averse after a loss. On the other hand, agents become more risk-loving after a win for the countries where football success is lower (Chile and Turkey). However, they could not find statistically significant shreds of evidence for the opposite situations for both cases.

According to Bell et al. (2012), the share price reaction depends on the importance of matches. Thus, they included two different dimensions in their regression model in order to measure that. 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 one is their "final position", which considers any match towards the end of the season and the extent to which the club's league position differs from the mean. Although all the clubs act differently, they concluded that the importance of the game seems to have a moderate impact on the returns.

Saraç and Zeren (2013) investigated the effect of soccer performance on the clubs' stock returns of the considered big three football clubs in Turkey – Besiktas, Fenerbahçe, and Galatasaray. Using multiple regression models, they controlled variables such as the market index, the type of match, betting odds, the venue of the match, the lag between the match date and market opening date and the market index return for the period between 2005 to 2012. They concluded that for all the three clubs, the sporting performance is statistically significant and positively related with the stock returns. Moreover, their findings indicate that this relationship is stronger for Besiktas compared to the other two football clubs. Also, the authors concluded that due to the unstable sporting performance of the team, the volatility in stock prices is higher, making the relationship of some variables (e.g., goal difference) higher for Besiktas. Another common finding for the three clubs is regarding the international matches (i.e., Champions League and Europe League matches). The international matches have had a negative effect on the stock price returns of the three clubs because of their weak performances in these competitions during recent years.

Floros (2014) used a threshold GARCH (TGARCH) method to capture good and bad news in terms of football results in order to test the link between some European football clubs' performances – Ajax, Benfica, Juventus and Porto - and their stock returns. The results explain the economic importance of football clubs' performance in finance: positive effects of draws on Benfica and Ajax stock returns, negative effects of draws and losses on Juventus stock returns, and no effects for Porto. Based on these results, the author concluded that these differences are based on the way investors behave and react differently after a draw and loss.

Sun and Wu (2015) tested the news model in order to study the stock prices of an Italian publicly traded football club, Juventus Football Club. They found out that unexpected match results affect the club's stock price. Additionally, when comparing Champions League with National League matches, results indicated that Champions League matches are more relevant for investors, showing a stronger effect on the share price. Finally, while testing the reversed news model, they concluded that corporate governance news can also be important in driving stock prices.

Sakinc et al. (2017) used the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) analysis method and Spearman's rank correlation methods to analyze the link between financial variables and sporting performance of 22 football clubs listed on various European stock markets from 2005 to 2014. The value of Spearman's rank coefficient is positive; however, it is weak. So, the authors concluded that their results are insufficient to conclude a positive and statistically significant correlation between sporting and financial performances.

Using a probability model, Godinho and Cerqueira (2018) analyzed the relation between stock returns and results in national league matches for 13 clubs of 6 European countries. They argue that they did not feel comfortable with the approaches used before. Thus, instead of associating the importance of a match with matches at the end of a season as previous authors did, they measured 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 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.

Shortly after, Berkowitz and Depken (2018) used regression models to analyze a large dataset of matches between two publicly traded English football clubs. They found out that match outcomes influence the football clubs' financial performance since results indicate that the market asymmetrically responds to winning and losing a match. Further, they argue that when it comes to elite clubs, the financial performance is more impacted by season losses than by winnings because investors might turn negative on a club's stock after a loss which impacts its future financial performance.

Dimic et al. (2018) studied the stock price reactions of 13 publicly traded football clubs (including Futebol Clube do Porto, Sporting Clube de Portugal, and Sport Lisboa e Benfica) following league matches from the 2000/2001 season to the 2012/2013 season. They used a regression model to calculate the abnormal stock returns and converted the betting odds into probabilities of outcomes to capture the ex-ante expectations of the football matches' results. Even though the results suggest significant first-day abnormal returns for both bad and good news, bad news generates a more significant price response than good news. Further, their findings reveal that the response to positive information increases in surprise to the resolution of uncertainty. However, bad information connotes negative regardless of the embedded surprise component, which is explained by the post-event irrational investor behavior, according to the authors.

Boțoc et al. (2019) used ARCH and GARCH methodologies to study the case of the three main Italian football clubs – Juventus, Lazio and AS Roma. They used two different approaches. The first studied the impact of the football match results over a market index that tracks the share prices of the listed football clubs in Europe. The second one studied a similar impact but over the clubs' individual stocks. They concluded that the clubs' share prices are sensitive to positive football match results rather for individual stocks. They considered their results twofold since, on the one hand, they make sense for emotional investments. However, in parallel, several investors might consider speculative investments when any predictions about the expected results are available.

In the same year, Škrinjarić and Barišić (2019) used event study methodology to analyze the effects of football match results of the Croatian national team on stock returns on the Zagreb Stock Exchange. As the first authors studying such effects on the Croatian stock market, they used a sample of 60 stocks on the Zagreb Stock Exchange from 2014 until 2018 and concluded that, for that period, neither win, draw nor lose have an impact on stock price reaction on the Zagreb Stock Exchange.

Busalini (2020) studied the relationship between football match results and stock price performance for a set of 4 different football clubs with high market capitalization: Manchester United, Juventus, Borussia Dortmund, and AS Roma. By running several regression models for each team while controlling a large set of variables, the results proved that football match outcomes influence stock price returns. Further, the author found that unexpected points extrapolated from betting odds have an impact on stock returns. Additionally, regarding European competitions, the author argues that UEFA Champions League matches have a more considerable impact on stock prices which is reasonable according to the importance of the game.

Similarly, Galloppo and Boido (2020) assessed 25 publicly listed football companies by the relationship between the outcome of their football events and their stock returns. Their database included results in three different types of competitions (National Championship, National Cup, and Continental Cups) from different countries between 2003 and 2015. Using an event study method, they concluded that football clubs, on average, provoke negative share price reactions since the sharp deterioration in equity returns resulting from a defeat is not offset by the rise recorded in the face of a win. Regarding the type of competition, the authors concluded that investors put greater emphasis on continental competitions such as Copa Libertadores or UEFA Champions League, meaning that a defeat in one of these competitions causes a fall in the football club's shares.

Dwi et al. (2021) investigated the stock market reaction to football match results for Manchester United, Juventus FC, and Borussia Dortmund. Considering a parametric paired sample t-test, they demonstrated that the market reacts differently to international and domestic matches and home and away matches. The market reactions to wins at international matches are more significant than at domestic matches, and market reactions to winning at home matches are higher than winning at away contests.

More recently, using an event study methodology, Gao et al. (2022) analyzed a sample of publicly listed firms in Chinese stock exchanges that are owners or sponsors of football teams in the national league of China between 2014 and 2017. They found out that these firms experience positive returns if the clubs they own/sponsor win a match and vice versa. They argue that firms should carefully choose which football clubs to invest in, even if their purpose is only to increase their visibility. In short, firms should invest in or sponsor a football club with solid sporting performances and avoid clubs with weak sporting performances since their future financial performance will depend on it.

2.2. Social Media Sentiments impact on the Stock Market Returns

Over the last years, given the importance social media has gained, many studies have been done based on its outcome data. Regarding financial markets, some of these studies were explicitly applied to football clubs' stock returns, and others were applied to other stock markets. In this study, Twitter data will be an indicator of investors' expectations to analyze the links between football match results, sentiments, and stock returns. Thus, literature regarding Twitter and other microblog platforms' sentiment analysis and its correlation with the stock market behavior is now presented.

Rao and Srivastava (2012) studied the relationship between twitter sentiments and stock prices. Using a Granger's Causality Analysis and Model Mining System they analyzed more than 4 million tweets from June 2010 to July 2011 for DJIA, NASDAQ-100 and 13 other big cap technological stocks. They argue that results show high correlation between stock prices and Twitter sentiments, which means that negative and positive dimensions of public mood carry strong cause-effect relationship with price movements of individual stocks.

Using an Profile of Mood States (POMS) model, Nofer and Hinz (2015) analyzed a sample of 100 million tweets published in Germany between 2011 and 2013 to study the relationship between Twitter mood states and the stock returns. They concluded that there is evidence that follower-weighted social mood levels can predict share returns. However, it is necessary to consider the community structure, in other words, the number of followers, explaining that the emotional contagion among internet users might be related with the number of followers of the club.

Applied to the Retail Industry, Souza et al. (2015) investigated whether there is a statistically significant relationship between Twitter sentiment and volume of tweets, with stock returns and volatility. They analyzed five publicly listed retail brands in the US: ABERCROMBIE & FITCH CO., NIKE INC., HOME DEPOT INC., MATTEL INC. and GAMESTOP CORP between November 2013 to September 2014. Using Granger Causality and Auto-Regressive models, they concluded that social media analytics have an important role in the retail sector financial market. Additionally, they found that compared to the traditional news, Twitter presents a stronger Granger Causality with stock returns.

In the same year, Ranco et al. (2015) investigated the relations between Twitter sentiments and financial markets of 30 stock companies that form the Dow Jones Industrial Average index using a regression model. For 15 months, they identified different events marked by increased activity on Twitter users and then observed market behavior in the days following the events. They argue that their main result is that the aggregate Twitter sentiment during the events implies the market evolution trend. In short, there is significant evidence of dependence between stock price returns and Twitter sentiment in tweets about the considered companies.

Nguyen et al. (2015) studied the relationship between stock price movements and the sentiment expressed on the Yahoo Finance Message Board for 18 companies. To accomplish this purpose, they used six different machine learining models based on Support Vector Machine (SVM), Latent Dirichlet Allocation (LDA), and Joint Sentiment Topic (JST) – four baseline models and introduced two new ones. Regarding the baseline models, they studied the correlation between stock price movements and: historical prices (model 1), sentiments (models 2 and 3), and main topics of comments based on probabilities (model 4). The two new models presented aim to understand whether a relationship exists between stock price movements and the topic-sentiment of users' comments. Authors claim that when users express their opinion regarding a stock, they also tend to express their opinions about the topic, such as profit and dividends and their belief on the stock's future price (i.e., if it goes up or down). The topic-sentiment is introduced to the predicting model by using JST based method and Aspect-based sentiment (models 5 and 6, respectively). Results showed that the average accuracy was only 43,31% due to stock price prediction being a challenging task because of the factors that affect it. However, their method predicted the stock price movement with more than 60% for a few stocks and achieved 2,07% better performance than the model based on historical prices.

Shortly after, Jai-Andaloussi et al. (2016) assessed the sentiments expressed on Twitter to detect and predict the club supported by each fan and the details associated with each event. They performed a Sentiment Analysis and text mining. Through the Knowledge Discovery in Databases (KDD) approach, machine learning algorithms, and the moving threshold burst detection algorithm, they concluded that the proposed framework is efficient with a mean precision of 90,04%. So, it plays an important role in improving the quality of summarization of soccer events. Also, they suggest that this framework can be applied to other sports.

Shutes et al. (2016) focused their study on the tweets of stocks on the US markets – NYSE and NASDAQ – by several financial micro-bloggers to understand whether their posts are reflected in stock price movements. They used Twitter API to extract 31780 tweets posted by 14 micro-bloggers between September 2011 through June 2013. By applying an event study, they tested whether these shared data carry any significant information or merely misinformation. They found that a considerable number of tweets are associated with price movements. The information is widespread and not limited to classic financial discussions of concepts but is less concrete when compared to traditional sources. However, it was found that Twitter has some limitations, such as the maximum tweet length of 140 characters, which suggests that posts are unlikely to carry significant specific or trading information. Even though the study focused purely on micro-bloggers with a large following within the online financial community, the conclusion was that Twitter is not a replacement for the traditional sources.

By using Pearson product-moment correlation coefficient, Li et al. (2016) investigated the relationship between Twitter users' moods and stock market behavior. From Twitter API, they extracted tweets from November 9 to November 20, 2015 (two weeks) and categorized them in different kinds of moods (happy, sad, anger, fear, disgust, and surprise). In parallel, they retrieved the NASDAQ market closing price over the same period. Their findings suggest that emotional-related words do not meet expectations due to their degrees of correlation with the overall stock market trend. However, happy mood words showed a relatively strong correlation, and unfortunate kind of vocabulary showed a significantly higher impact on the stock market. They argue that, regarding the worst moods, the reason may be due to the ease of making investments propensity adjustment action.

Pagolu et al. (2017) used sentiment analysis, machine learning and Logistic Regressions in order to understand the correlation between the changes – rises and falls – in stock prices of a company and the public opinions expressed on Twitter about that company. In this case study, the chosen company was Microsoft, from which a total of 250,000 tweets were extracted from Twitter API between August 2015 to August 2016. After classifying the tweets into positive, negative, or neutral, they concluded that a strong relationship exists between a company's rise or fall in stock prices and public opinions, or emotions expressed on Twitter about it.

Bartov et al. (2017) tested whether individuals' shared opinions on Twitter prior to a firm's earnings announcement predict its earnings and announcement returns. To study the case of Russel, the authors collected historical Twitter data from GNIP and applied a regression model. After controlling other determinants of earnings, such as the aggregate opinion in traditional media sources, they found out that the aggregate Twitter opinion helps predict quarterly earnings instead of misleading investors. Further, they argue that their results trump any concerns about the lack of credibility of information on Twitter.

Ruan et al. (2018) studied the correlation between Twitter sentiment valence and abnormal stock returns for eight firms in the S&P500 using ARCH and GARCH methodologies. In addition to previous studies, they took into consideration the source of tweets. They used a user-to-user network to calculate users' power or reputation. They analyzed the Pearson correlation test for eight months. They concluded that compared with treating all the tweets' authors equally important or weighting them by their number of followers, their user-to-user network amplifies the correlation between a specific firm's Twitter sentiment valence and its abnormal stock returns. To consolidate their findings, they constructed a linear regression model, which included historical abnormal returns to test the relation between Twitter sentiment valence and abnormal stock returns again, and the results showed that while using their trust network, Twitter sentiment valence can better reflect abnormal stock returns than the other methods.

Broadstock and Zhang (2019) tested the pricing power of sentiments expressed on Twitter towards the stock market for August 2018. Their analysis was developed around a Capital Asset Pricing Model (CAPM). They used S&P500 index data to obtain market returns of a sample of US companies and Twitter API to extract tweets related to these companies – including hashtags and cashtags (e.g., $SPY) – and classified them into emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, and trust) based on the number of terms related to such sentiments. They argue that, although the results seem a bit mixed due to the diversity among the stocks under investigation, results demonstrate that stocks considered show significant reactions to sentiment extracted from social media.

Teti et al. (2019) explored Twitter as a tool for investing, verifying the relationship with stock prices of the technology industry in the U.S. using Ordinary Least Squared (OLS) models and the Fixed Effects with binary treatment model. According to their results, they concluded that social media is part of today’s business environment and can hardly be ignored. This article proves that Twitter has a strong statistical association with stock prices, tacking the technology industry as a benchmark. However, they argue that social media are not the ultimate solution for the long-lasting aspiration of looking into the future. Instead, they claim that sentiment can be combined with traditional theories to make and support more informed investment decisions.

Derakhshan and Beigy (2019) replicated the study of Nguyen et al. (2015) for the U.S. and Iran’s stock markets, using similar models. Two different datasets were extracted, one in English and another in Persian. The English one was collected from Yahoo Finance Message Board with a total of 787,547 comments between July 2012 and March 2013, and the U.S. stock market information was extracted from Yahoo Finance. The 21,205 Persian comments were extracted between April 2016 and November 2016 from SAHAMYAB, which is a website that allows users to post comments on different stocks. The stock market dataset was extracted from the Tehran Securities Exchange Technology Management Co. website. They introduced a new model, called the LDA-POS Method, which is a new version of the LDA-based Method. The average accuracy of their model in the English dataset and Persian dataset was 56.24% and 55.33%, respectively. Further, they argue that usually, the sentiment extraction methods, which perform flawlessly in English, do not perform well in the Persian language; however, their LDA-POS method results were similar for both languages.

The novelty of the research of Sóti et al. (2020) was to show which daily ratio and their mood can predict the stock market of Juventus, Manchester United, and Ajax. Based on their findings, they argue that sport managers could use social media strategically to build better relationships with consumers and stakeholders. Thus, using an OLS model, they analyzed tweets exchanged immediately after a sporting event of each club and correlated with the changes in their stock prices to find out how investors respond to different events. They concluded that public information on Twitter influences fans’ behaviours, and since investor-fans invest based on temporary emotions, variables such as the number of tweets posted, fan groups locations, different languages used, and motivations must be controlled by managers to counterbalance the effects of an adverse event.

Since Twitter was blocked in China in 2009, Y. Sun et al. (2020) used Sina Weibo, the largest microblog platform in China, to study the impact of both investors’ sentiments and the news media on stock returns via regression models and sentiment analysis. To accomplish this purpose, they used Octopus Collector software to capture around 22,504 tweets from October 2015 to December 2015 and CSMAR database to obtain financial data for the same period. The conclusion was quite clear, both investors’ sentiments and the news media have a great impact on the stock market returns. Also, they argue that companies are more focused on shareholders, ignoring the interests of other stakeholders when making business decisions. From the authors’ point of view, managers should identify which groups are more likely to be influenced by a proposed action to prioritize groups according to their impacts on the action. Further, they argue that social media platforms are now the center for collecting public opinions and, in the long term, will impact a company's performance. Therefore, enterprises should pay more attention to the information released since it is mainly there that their reputation is created.

Duz Tan and Tas (2021) studied the social media effect on international stock returns and trading activity using regression models. The authors used the Twitter sentiment on stock returns for S&P500, S&P350 Europe, and S&P Emerging Markets Core Index stocks from January 2015 to December 2017. They found that daily Twitter activity and sentiment affect trading volume and predict subsequent-day trading volume. Additionally, they found that Twitter sentiment contains information for predicting future stock returns for a specific firm, but no such relationship exists in the number of tweets. Specifically, regarding the sentiment tone, their results showed that the positive tone of Twitter sentiment has more predictive power in small and emerging market firms. Finally, they argue that Twitter sentiment information is increasing and may provide valuable information and proxies for investor behaviours for financial market participants. Thus, investors could incorporate Twitter sentiment in their trading strategies, and firms could monitor it to manage firm-specific investor sentiment that may affect their performance.

Tanaltay et al. (2021) included Twitter data as an indicator of investors' sentiments beyond match performance and betting odds to analyze the links among football match results, sentiments, and stock returns of the four major Turkish football clubs. Using Sentiment Analysis and predictive statistical models, they concluded that adding Twitter data can improve the accuracy in predicting the amount and direction of change in the stock prices of these clubs. However, when running a model solely on Twitter sentiments to predict the amount and the direction of stock return, the results showed that sentiments are not good predictors for such dependent variables. Thus, the best option is to use a model combining match performance, betting odds data and Twitter sentiments.

Mehta et al. (2021) analyzed how movements in a company's stock prices correlate with the expressed opinions and sentiments of the public about that company. By using machine-learning and deep-learning methods, including Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB) classifier, linear regression, Naïve Bayes, and Long Short-Term Memory, they assessed the news and stock data of the Indian stock market. They collected the news data from Moneycontrol, IIFL, Economic Times, and Twitter. After, they performed a sentiment analysis to categorize them into negative, neutral, or positive sentiment and tried to predict the stock price based on it together with the historical and past stock data. Their findings suggested that the positive news effect is likely to reflect that the share market values are high, and if the news is negative, then the trend's impact is low. Also, they argue that using sentiment with the historical stock prices makes it possible to get the stock's accuracy so consumers can sell and buy their stock with stock movement.

More recently, Valle-Cruz et al. (2022) studied the impact of social media reactions and emotions on the stock market and vice versa during H1N1 and COVID-19 pandemics based on a lexicon approach. To carry out their analysis, for the H1N1 pandemic, they collected data from June to July 2009, and for the COVID-19 pandemic, the period for data collection was from January to May 2020. The stock market indices were IPC, S&P500, NASDAQ100, Dow Jones, FTSE100, BOVESPA, CAC40, DAX, Hang Seng, Nikkei 225, and SSE Composite, and the financial data used for this study were the adjusted closing prices. The results showed sufficient evidence to support the notion that Twitter posts have influenced financial indices during both pandemics – more significant during the COVID-19 pandemic than the H1N1 pandemic. They claim that the drop in stock prices during the COVID-19 era was more dramatic because there was more speculation, rumours, and negative news. Further, they identified a high correlation between Twitter sentiments and stock market behaviour for The New York Times, Bloomberg, CNN News, and Investing.com. Finally, they claim that sentiments on Twitter have an important effect on financial indices, which is observed a few days after the information is posted on Twitter.

Yilmaz et al. (2022) investigated the effects of social media activities on stock prices for the energy sector. They created three models, using S&P500 index, stock market volatility index (VIX), trade-weighted USD index (USD), and Brent oil prices (OIL) as the control variables to examine 20 energy companies traded in the S&P500 between June 2015 and May 2020. After, they tested these models with Augmented Mean Group (AMG) analysis and concluded that Twitter sensitivity does not affect firms' returns and volatility. However, the trade volume is affected. They claim that, according to the results, positive tweets do not affect the trade volume. In contrast, negative tweets have an impact on the investment decisions of individuals. Nevertheless, this effect is not effective on volatility and returns. Thus, they concluded that Twitter sentiments of the companies under study have a limited effect on their investment decisions.

3. Methodology

This study follows the methodology of Saraç & Zeren (2013) applied in Portugal. They used an econometric model based on a Multiple Linear Regression (MLR) that was applied for each of the three clubs' data separately (Besiktas, Galatasaray, and Fenerbahce) in order to examine the effect of football performance on stock returns. The novelty of this study is that instead of an MLR model, two Logistic Regression (LR) models are performed since the dependent variable of the data is a dummy variable and not a continuous one[9]. Based on these two LR models, three different Portuguese football clubs are investigated and further compared. Both these models include the sentiment expressed by fans on Twitter; however, there are some differences between them.

The first LR model combines sporting performance and winning probability variables with the overall stock market returns, financial information variables, and the expressed sentiments by people on Twitter posts in order to investigate whether these variables, and particularly whether the variables related to sentiments have any impact on the club's stock price returns. The following equation gives it:

$$\begin{aligned} Club = \beta_0 + \beta_1 ClosePSI20 + \beta_2 PSI20 + \beta_3 MatchResult + \beta_4 WinProb \\ + \beta_5 GoalDiff + \beta_6 MatchType + \beta_7 DerClass + \beta_8 Neutral \\ + \beta_9 Negative + \beta_{10} TotTweets + \beta_{11} SentScore + \varepsilon \end{aligned} \quad (1)$$

where $Club$ is a binary or dummy variable that classifies the direction of the club's stock return, being classified as 1 if the direction of the stock return of the club under analysis is positive, and if it is negative or equal to zero it is classified as 0, $ClosePSI20$ is a numeric variable that indicates the last price at which a PSI20 stock is traded and $PSI20$[10] is a dummy variable that classifies the direction of the overall stock price return of PSI20, being 1 it if is positive, and classified as 0 otherwise, $MatchResult$ is a dummy variable that quantifies the club's match result of a given match, if the result is a victory it is classified as 1, if the result is a draw or a defeat, it is classified as 0, $WinProb$ is a dummy variable that is classified as 1 if the winning match probability of the club is greater than the loss probability, and 0 otherwise, $GoalDiff$ is a numeric variable that measures the difference between the scored goals by the winning club and the goals scored by the losing club, $MatchType$ is a dummy variable that quantifies the type of match, if the match is National it is classified as 1 and if it is International, it is classified as 0, $DerClass$ is a dummy variable that quantifies the importance of the match, it is classified as 1 if it is a derby or a classic match, and 0 otherwise, $Neutral$ is a dummy

[9] The choice of considering a dummy variable instead of a continuous one is due to the low liquidity and high volatility of the Portuguese clubs' stock returns.

[10] $ClosePSI20$ and $PSI20$ enables to see the degree to which a given stock's return is affected by the overall changes in the market.

variable that is classified as 1 if the number of neutral tweets of the day is greater than the sum of the number of positive and negative tweets, otherwise it is classified as 0, $Negative$ is a dummy variable that is classified as 1 if the number of negative tweets of the day is greater than the sum of the number of positive and neutral tweets, otherwise it is classified as 0, $TotTweets$ is a numeric variable that measures the number of tweets, $SentScore$ is a numeric variable that measures the general mood of the crowd, which is computed by the number of positive tweets minus the number of negative tweets divided by the total number of tweets times 100.

The decision to build an LR model combining all these independent variables non-relatable to the sentiment expressed on Twitter is due to other previous study conclusions by Tanaltay et al.. They argue that a model has a stronger explanatory power when sentiment variables are incorporated with other independent variables related to sporting performance, probabilities, and overall stock market returns. At the same time, besides contributing to test this assumption, this model also allows to investigate whether any or even some of these variables is statistically significant in the model.

However, to test this hypothesis, it is also necessary to build a second model to compare its results with the results of the first model. The second model is built based only on sentiment variables, and the following equation gives it:

$$\begin{aligned} Club = \beta_0 + \beta_1 ClosePSI20 + \beta_2 PSI20 + \beta_3 Neutral + \beta_4 Negative + \\ \beta_5 TotTweets + \beta_6 SentScore + \varepsilon, \end{aligned} \tag{2}$$

Logistic Regressions (LRs) are similar to Multiple Linear Regressions (MLRs) because both may include only one or multiple independent variables and also because LR retains many features of MLR in its analysis of binary outcomes (Stoltzfus, 2011). However, MLR models analyze continuous outcomes (independent variables can potentially take any number), whereas LR models analyze binary ones, making LR the adequate choice for this study. Before running the models, there are basic assumptions for conducting Logistic Regression that must always be met. The first assumption is independence of errors, which means that data is independent (i.e., there is no relationship between the observations). The second assumption is a linear relationship between any continuous independent variable and its logit transformation outcome. The third assumption is the absence of multicollinearity among independent variables, in other words, independent variables are not highly correlated, and finally, the fourth assumption is the lack of strongly influential outliers (Stoltzfus, 2011).

Recalling that the main objective of this dissertation is to conclude whether the sentiment of football fans expressed on Twitter affects the stock price returns in the Portuguese football industry, after ensuring the assumptions stated above, the Logistic Regressions are run in STATA in order to test the following hypothesis:

H_1: The sentiment expressed on Twitter impacts the club's stock price returns.

H_2: When the sentiment expressed on Twitter is combined with sporting performance variables in the same model, this offers a better fit to the data when compared to a model based only on expressed sentiment.

The database needed in this study is composed of football sporting performance indicators, betting odds, and financial and Twitter data of the three clubs under analysis – Futebol Clube do Porto, Sporting Clube de Portugal, and Sport Lisboa e Benfica. All the data were collected between the 2016/2017 season and the 2020/2021 season.

For the financial data, the daily stock market information for each team were collected form the Finance Yahoo dataset that can be downloaded from its website[11]. Also, the Portuguese Stock Exchange PSI20 prices were collected from the same database, in order to consider the overall market in each individual club share price.

Regarding the sporting performance, the data were collected from various databases – BBC Football, ESPN Soccer, Sportinglife, Bundesliga, Gazzeta, Xscores – provided by Football-Data on their website[12]. Also, some statistics regarding match results were collected from Flashscore database, which can be downloaded on their website[13].

Betting odds for every match were collected from the Oddsportal database, which can be downloaded from its website[14]. The free version of Oddsportal database is based on 17 different bookmarkers such as Bet365, 1XBet, and Bwin. After, the betting odds were converted into probabilities of winning and losing, just as Dimic et al. did:

$$ProbWin_i = \frac{x_{iw}^{-1}}{x_{iw}^{-1} + x_{id}^{-1} + x_{il}^{-1}} \tag{3}$$

$$ProbLoss_i = \frac{x_{iw}^{-1}}{x_{iw}^{-1} + x_{id}^{-1} + x_{il}^{-1}} \tag{4}$$

where x_{iw} is the match-winning probability on the day i, x_{id} is the match-draw probability on the day i, and x_{il} is the match-losing probability on the day i. Giving a specific example, the average odds for

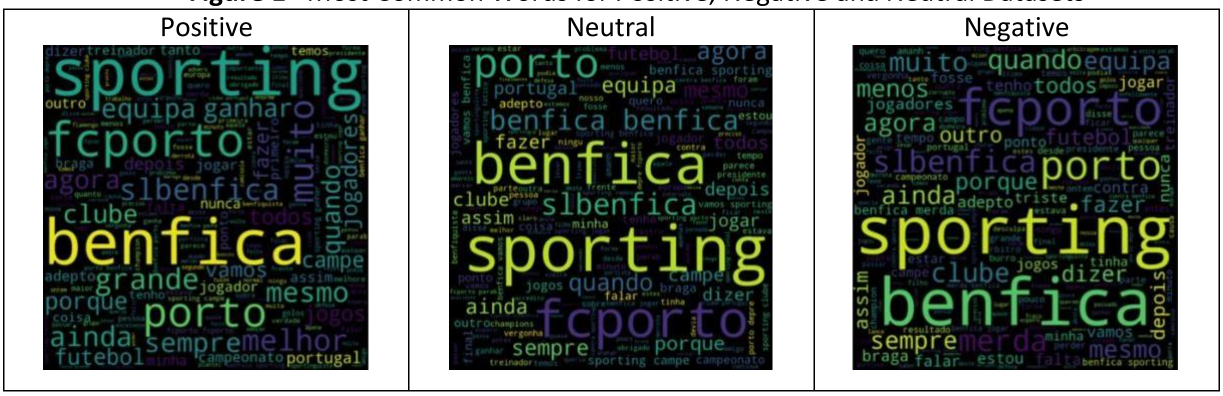
[11] Yahoo Finance website: www.finance.yahoo.com
[12] Football-Data website: www.football-data.co.uk
[13] Flashscore website: www.flashscore.co.uk
[14] OddsPortal website: www.oddsportal.com

Sporting against Porto at Sporting's stadium on the 28th of August 2016 were 2.1, 3.2 and 3.6 for the home win, draw and away win, respectively. Thus, Sporting winning probability is given by (1/2.1)/((1/2.1)+(1/3.2)+(1/3.6)) = 0.45, and losing probability by (1/3.6)/((1/2.1)+(1/3.2)+(1/3.6)) = 0.26.

Finally, Twitter text data were extracted from Twitter API database. This data is included for testing the effects of the fans' sentiments on the stock price changes. Around 79,000 tweets from July 2016 to May 2021 were analyzed on Python. In order to obtain the associated sentiment to each tweet, three Python libraries were used – Pandas, TextBlob and DeepTranslator. Python libraries provide a set of functions that were used as follows: firstly, by using the Pandas library, Twitter text data was imported; after, by using the DeepTransalator library, tweets were translated from Portuguese to English in order to be able to be measured (Tweets are translated because TextBlob only works with the English language); finally, TextBlob was used to measure the sentiment present in each tweet (values were categorized into "Negative" for values lower than -0.1, "Neutral" for results between -0.1 and 0.1, and "Positive" for values higher than 0.1).

In order to check the consistency of the text data, word cloud plots of the most common words used in the three types of sentiments are shown in Figure 1.

Figure 1 - Most Common Words for Positive, Negative and Neutral Datasets



In the positive set, it is possible to observe words with positiveness such as 'ganhar' (win), 'grande' (huge) or 'muito' (a lot). The negative set contains words with negative sentiment like 'triste' (sad), 'perder' (lose), or even some bad words and insults. Regarding the neutral dataset, it is possible to observe some random verbs such as 'dizer' (to say), 'estou' (I am) or 'quero' (I want) and words like 'adepto' (supporter), and 'clube' (club).

4. Results

Before the proper analysis of the results, for each of the three football clubs, the Logistic Regression assumptions are tested. In order to test the first assumption of no relationship between the observations, in other words, that all data come from different responses so there are no repetitions, the isid[15] command was applied. For the second assumption, a Box Tidwell test was done in order to check whether there is a linear relationship between any continuous independent variable and its logit transformation outcome. The third assumption of the absence of multicollinearity[16] among independent variables was tested based on the correlation matrix, which measures the correlation between the independent variables. Finally, right before running the LR models, the last assumption of lack of strongly influential outliers is tested by the least like probabilities.

After testing all these assumptions, the results are revealed, and the analysis is done per club, starting with Sporting, then Benfica, and finally Porto. After the individual analysis, the models are all compared between clubs, and to finish the chapter, a summary of the results is made.

4.1. Sporting Clube de Portugal (SCP)

Before revealing the results of the LR models, some tests were made to ensure its assumptions. For the first assumption, the isid command was applied in order to detect if duplicate observations exist. According to STATA, if isid reports no error, then the variables uniquely identify the observations. It is possible to check this in Figure A1 in the appendix. In order to check the second assumption, a Box Tidwell test was done. By looking at Tables A1 and A4 in the appendix, it is possible to check that, according to the p-values, there is no linear relationship between any continuous independent variable and its logit transformation outcome. The assumption of multicollinearity was tested by a correlation matrix. It is possible to check that there are no correlations above 0.70 between the independent variables for both models. Thus, there is no multicollinearity (see tables A7 and A10). Finally, regarding the last assumption, the outliers were detected through the least likely probabilities (see tables A13 and A16 in the Appendix), and the observations with probabilities below 0.10 were removed.

4.1.1. SCP – Findings of Model 1 and Model 2

Table 1 - Model 1: LR Results for Sporting Clube de Portugal

	Coefficient	Std. Error	z-Statistic	Prob.
$Constant$	-1.349751	0.6332989	-2.13	0.033
$ClosePSI20$	0.0001106	0.0000474	2.33	0.020

[15] Isid is a STATA command that can detect duplicate observations. If there are duplicate observations, isid returns an error. Otherwise it reports no error.

[16] Multicollinearity exists when there is a correlation between multiple independent variables.

$PSI20$	0.1920521	0.1286514	1.49	0.135
$MatchResult$	0.9594187	0.4606916	2.08	0.037
$WinProb$	-1.305232	0.5069708	-2.57	0.010
$GoalDiff$	0.2386459	0.1216015	1.96	0.049
$MatchType$	-0.6918208	0.5780598	-1.20	0.231
$DerClass$	-2.522115	1.129991	-2.23	0.026
$Neutral$	-1.10786	0.4442468	-2.49	0.013
$Negative$	-0.5647198	0.2930372	-1.93	0.045
$TotTweets$	-0.0036171	0.002466	-1.47	0.142
$SentScore$	-0.006156	0.0036545	-1.68	0.092
$Obs.$	1,675			
$LR\ chi2$	40.02			
$Prob > chi2$	0.0000			
$Pseudo\ R^2$	0.0249			
$Log\ likelihood$	-782.32			

Starting by looking at the $Prob > chi2$, is it possible to assume that the overall model is statistically significant, since this value is less than 0.05. Also, as for the individual effects of the variables, and according to their p-values (smaller than 0.05), the last price at which a PSI20 stock is traded, the match result, the club's winning probability, the goal difference, the importance of the match, the overall neutral sentiment and the overall negative sentiment have a significant effect on the direction of the stock return of Sporting Clube de Portugal.

Regarding the sentiment variables, findings imply that if the overall expressed sentiment is neutral, the direction of the stock price returns of Sporting has minus 67.0%[17] odds of being positive. Still regarding sentiments, when the overall expressed sentiment is negative, there is minus 43.1% odds of Sporting's stock return being positive.

When it comes to the financial variables, findings imply that if the last price at which a PSI20 stock is traded increases by 1 unit, there is a 0.01% increase in the odds of the direction of Sporting's stock return being positive.

Finally, regarding the sporting performance and winning probabilities, findings indicate that if the match result is a win, the direction of the stock returns of Sporting has 2.61 times the odds of a draw or a loss result being positive. The winning match probability has a curious effect on the direction of Sporting's stock returns. When this probability increases 1%, there are minus 72,9% odds of Sporting's stock return being positive. This might well be explained by the poor performances of the club during the last years. Sporting did not win any Primeira Liga title for almost two decades[18]. Regarding the difference of goals, when the difference of goals between the winning and losing teams increases by

[17] For all the interpretations, it is assumed that the other variables remain fixed. All the calculations for the coefficient's interpretations are in table A19 in the Appendix.

[18] Last Sporting's Primeira Liga titles: 2020/2021 and 2001/2002 seasons.

1 unit, there are 27,0% greater odds of Sporting's stock return being positive. Finally, derbies and classic matches are associated with a 92.0% reduction in the direction of the stock returns of Sporting being positive.

Table 2 - Model 2: LR Results for Sporting Clube de Portugal

	Coefficient	Std. Error	t-Statistic	Prob.
$Constant$	-1.409953	0.3302998	-4.27	0.000
$ClosePSI20$	0.0001036	0.0000466	2.22	0.026
$PSI20$	0.1825139	0.1274178	1.43	0.152
$Neutral$	-1.101293	0.4423329	-2.49	0.013
$Negative$	-0.6064826	0.2888547	-2.10	0.036
$TotTweets$	-0.0032139	0.0023659	-1.36	0.174
$SentScore$	-0.0064102	0.0036193	-1.77	0.077
$Obs.$	1,675			
$LR\ chi2$	18.84			
$Prob > chi2$	0.0044			
$Pseudo\ R^2$	0.0177			
$Log\ likelihood$	-792.91			

The table above presents the results of the Linear Regression for model 2. Looking at the $Prob > chi2$ of 0.0044 it is possible to assume that the overall model is statistically significant – p-value is smaller than 0.05. When considering individually, the last price at which a PSI20 stock is traded, and the overall neutral and negative sentiments are statistically significant.

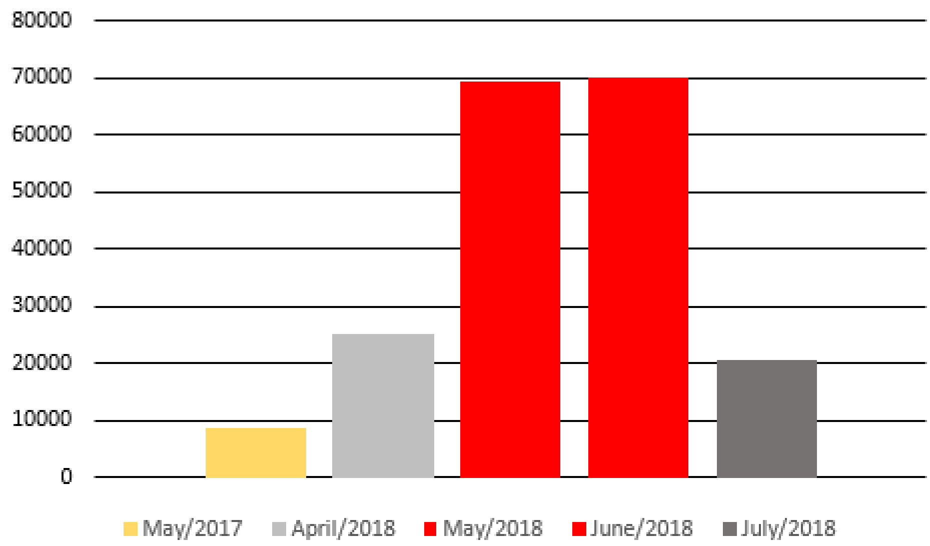
As for the model 1, although the last price at which a PSI20 stock is traded is statistically significant, it has a small impact on the dependent variable of the model. According to its coefficient, if the last price at which a PSI20 stock is traded increases by 1 unit, there is a 0.01% increase in the odds of the direction of Sporting's stock return being positive.

Regarding sentiment in this second model, both the neutral and negative sentiment expressed have an effect on the dependent variable. According to its coefficients, if the overall expressed sentiment is neutral, there is a reduction of 66.8% of the odds of the direction of Sporting's stock return being positive, and if the overall expressed sentiment is negative, then the reduction will be 45.5%.

Some conclusions can be made by comparing these two-run models. By comparing the $Log\ likelihood$ of both of them, it is possible to argue that the first model offers a better fit to the data when compared to model 2, which goes along with the second hypothesis of the dissertation. This is explained by the fact that model 1 includes sporting performance and winning match probabilities, whereas model 2 does not.

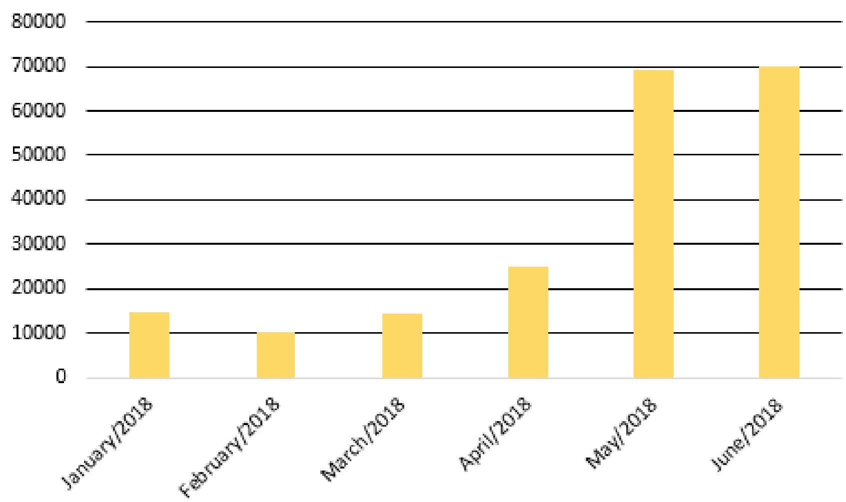
An extra analysis that can be made is regarding the supporters' invasion in the Sporting's training center at Alcochete. The chart below shows the stock trading volume during May 2018 (the month of the invasion) versus the homologous month – May 2017 – and versus the previous and next months – April and June 2018, respectively.

Chart 1 - Stock trading Volume SCP: Comparison



Looking at Chart 1, an unusual event is noticed due to the abnormal amount of volume traded during the month of the attack and the following one. Also, the trading volume in May 2017 was much smaller (8,855 shares vs 69,289 shares). Chart 2 shows that the normal trading volume from the beginning of that year until the following month of this event, from January 2018 to June 2018, was from 10,000 to 25,000 approximately. This might be explained by Sporting's stock investors' behaviour of trying to rapidly sell their shares during these months (May and June). This is an interesting topic that can be explored in future studies.

Chart 2 - Stock trading Volume SCP: Historical data



4.2. Sport Lisboa e Benfica (SLB)

Similarly to what has been done for Sporting was also done for Benfica. Before revealing the results of the LR models, tests are made to ensure its assumptions. For the first assumption, the isid command was applied in order to detect if duplicate observations exist (see figure A2 in the Appendix). In order to check the second assumption, a Box Tidwell test was done. By looking at Tables A2 and A5 in the appendix, it is possible to check that, according to the p-values, there is no linear relationship between any continuous independent variable and its logit transformation outcome. The assumption of multicollinearity was tested by a correlation matrix. It is possible to check that there are no correlations above 0.70 between the independent variables for both models. Thus, there is no multicollinearity (see tables A8 and A11). Finally, regarding the last assumption, the outliers were detected through the least likely probabilities (see tables A14 and A17 in the Appendix), and the observations with probabilities below 0.10 were removed.

4.2.1. SLB - Findings of Model 1 and Model 2

Table 3 - Model 1: LR Results for Sport Lisboa e Benfica

	Coefficient	Std. Error	z-Statistic	Prob.
$Constant$	-5.166247	0.8872619	-5.82	0.000
$ClosePSI20$	0.0003055	0.0000494	6.19	0.000
$PSI20$	0.1689513	0.110257	1.53	0.125
$MatchResult$	0.9061261	0.3745291	2.42	0.016
$WinProb$	1.737022	0.7598507	2.29	0.022
$GoalDiff$	-0.1777146	0.1013871	-1.75	0.080
$MatchType$	1.885194	0.644201	2.93	0.003
$DerClass$	1.929734	0.9162126	2.11	0.035
$Neutral$	0.1854246	0.183163	1.01	0.311
$Negative$	0.2919437	0.4626296	0.63	0.528
$TotTweets$	0.0026992	0.0012358	2.18	0.029
$SentScore$	0.0014469	0.0041005	0.35	0.724
$Obs.$	1,709			
$LR\ chi2$	72.82			
$Prob > chi2$	0.0000			
$Pseudo\ R^2$	0.0357			
$Log\ likelihood$	-984.85			

The p-value of the first LR model performed for Benfica is 0.000, meaning that the overall model is statistically significant – since this value is smaller than 0.05. When considered individually, the last price at which a PSI20 stock is traded, the match result of a given match, the Benfica’s winning probability, the match type, the importance of the match, and the total number of tweets have a significant effect on the direction of the stock return of Sport Lisboa e Benfica.

Regarding sentiments, according to the p-values presented in the table above, only the total number of tweets affects the direction of the stock return of Benfica. When the total of tweets increases by 1 unit, there are 0,27% greater odds of Benfica's stock return being positive. The other independent variables related to sentiments (i.e., $SentScore$, $Neutral$ and $Negative$ variables) are insignificant since their p-values are way above 0.05.

When it comes to the sporting performance variables, if the result of a given match of Benfica is a win, then the odds of the direction of Benfica's stock returns being positive are 5.68 times greater than the odds when the result is a draw or a loss. Also, International matches multiply by 6.59 the odds of the direction of Benfica's stock returns being positive compared to National matches. If the match is a derby/classic, then the odds of the direction of Benfica's stock return being positive are multiplied by 6.89 in comparison to the matches against the other non-rivals clubs. Regarding the winning match probability, when it increases by 1% the odds of the direction of Benfica's stock return being positive are multiplied by 5.68 compared to a losing match probability.

In the matter of the financial variables, if the last price at which a PSI20 stock is traded increases by 1 unit, there is a 0.03% increase in the odds of the direction of Benfica's stock return being positive.

Table 4 - Model 2: LR Results for Sport Lisboa e Benfica

	Coefficient	Std. Error	t-Statistic	Prob.
$Constant$	-2.694008	0.309539	-8.70	0.000
$ClosePSI20$	0.0002976	0.0000489	6.09	0.000
$PSI20$	0.1724329	0.109617	1.57	0.116
$Neutral$	0.2078799	0.1813036	1.15	0.252
$Negative$	0.3785008	0.460104	0.82	0.411
$TotTweets$	0.0019539	0.001085	1.80	0.042
$SentScore$	0.0018094	0.0040808	0.44	0.657
$Obs.$	1,709			
$LR\ chi2\ (6)$	56.40			
$Prob > chi2$	0.0000			
$Pseudo\ R^2$	0.0276			
$Log\ likelihood$	-993.06			

When it comes to the second LR model performed for Benfica, the scenario changes a bit. Although the overall model is statistically significant, since its p-value is smaller than 0.05, only the last price at which a PSI20 stock is traded and the total number of tweets are statistically significant when considered individually. The coefficient interpretation for $ClosePSI20$ is the same one made for model 1 – if it increases by 1 unit, there is a 0.03% increase in the odds of the positive direction of Benfica's stock return. The p-values presented for most sentiment variables indicate that they are not statistically significant. Precisely, $Neutral$, $Negative$ and $SentScore$ due to their high p-values. The total number of tweets is again statistically significant, meaning that when the total of tweets increases by 1 unit, there are 0,2% greater odds of Benfica's stock return being positive.

Comparing the two models for Benfica, the conclusions are straight. According to its greater $Log\ likelihood$ – -984.85 in Model 1 vs -993.06 in Model 2 – it is possible to conclude that the first model offers a better fit to the data, which also goes along with the second hypothesis of the dissertation.

4.3. Futebol Clube do Porto (FCP)

Before revealing the results of the models for Porto, the LR assumptions were tested, just like was made for Sporting and Benfica. For assumption one, the isid command was applied. Since that does not return any error, as it is possible to see in Figure A3 in the Appendix, the conclusion is that the variables uniquely identify the observations. For the second assumption, a Box Tidwell test was done. Tables A3 and A6 in the Appendix show that, according to the p-values presented, there is no linear relationship between any continuous independent variable and its logit transformation outcome. Assumption three, regarding the absence of multicollinearity, was tested by correlation matrixes. Since there are no correlations above 0.70 between the independent variables, it is possible to argue that there is no multicollinearity (see tables A9 and A12). To finally check the last assumption, the outliers were detected by the least likely probabilities (see tables A15 and A18 in the Appendix), and the observations with probabilities below 0.10 were removed.

4.3.1. FCP - Findings of Model 1 and Model 2

Table 5 - Model 1: LR Results for Futebol Clube do Porto

	Coefficient	Std. Error	z-Statistic	Prob.
$Constant$	-1.212667	0.651926	-1.86	0.063
$ClosePSI20$	0.0000677	0.0000443	1.53	0.126
$PSI20$	0.3121291	0.1309631	2.38	0.017
$MatchResult$	0.7730429	0.4244574	1.82	0.069
$WinProb$	-1.204509	0.580114	-2.08	0.038
$GoalDiff$	-0.3295511	0.1511435	-2.18	0.029
$MatchType$	1.147744	0.5272416	2.18	0.029
$DerClass$	-1.688136	0.8862214	-1.90	0.057
$Neutral$	0.1020259	0.2003529	0.51	0.611
$Negative$	-0.1271009	0.4260559	-0.30	0.765
$TotTweets$	-0.0079118	0.0053702	-1.47	0.141
$SentScore$	0.0012246	0.0054065	0.23	0.821
$Obs.$	1,711			
$LR\ chi2$	38.72			
$Prob > chi2$	0.0001			
$Pseudo\ R^2$	0.0245			

$Log\ likelihood$	-771.86			

Although the overall model is statistically significant, since the p-value is 0.0001, the results are less friendly than the ones for Sporting and Benfica for this first model. This is because none of the independent variables related to sentiments (i.e.,$Neutral$, $Negative$, $TotTweets$, and $SentScore$), when considered individually, are statistically significant[19]. However, there are still some variables that are statistically significant when considered individually. These are the direction of the overall stock price return of PSI20, the Porto's winning probability of a given match, the goal difference, the match type. Also, although its p-value is pretty close to 0.05, the importance of the match variable is not considered statistically significant.

Regarding these statistically significant variables, findings imply that if the direction of the overall stock price return of PSI20 is positive, the odds of Porto's stock return being positive are 1.37 times greater when compared to negative or zero returns of PSI20. As for Sporting, the winning match probability has a curious effect on the direction of Porto's stock returns. When this probability increases by 1%, there are minus 70,0% odds of Porto's stock return being positive. When it comes to the type of match, findings indicate that an international match multiplies by 3.15 the odds of the direction of Porto's stock return being positive. Regarding the difference of goals, when the difference of goals between the winning and losing teams increases by 1 unit, there are minus 28,1% odds of Porto's stock return being positive.

Table 6 - Model 2: LR Results for Futebol Clube do Porto

	Coefficient	Std. Error	t-Statistic	Prob.
$Constant$	-1.911239	0.317716	-6.02	0.000
$ClosePSI20$	0.0000682	0.0000438	1.56	0.119
$PSI20$	0.2776129	0.1292828	2.15	0.032
$Neutral$	0.03927	0.1965209	0.20	0.842
$Negative$	-0.2355756	0.4208827	-0.56	0.576
$TotTweets$	-0.0061661	0.0052667	-1.17	0.242
$SentScore$	-0.0003694	0.0053257	-0.07	0.945
$Obs.$	1,711			
$LR\ chi2$	9.14			
$Prob > chi2$	0.1660			
$Pseudo\ R^2$	0.0058			
$Log\ likelihood$	-786.65			

It was expected that model 2 for Porto would not be statistically significant due to the results of model 1 – because none of the variables related to sentiments were statistically significant in model 1. It is possible to verify that from the p-value of the overall model shown in table 6, which is way

[19] Table 5 shows that the p-values of these variables are way above 0.05.

above 0.05 (the p-value of model 2 is 0.1660). As for model 1, in model 2 none of the expressed sentiment variables are individually statistically significant. Thus, the conclusions are evident – the explanation towards response is very low or negligible (i.e., the insignificance of the explanatory variables).

4.4. Summary of Results

Table 7 - Models' Log Likelihood

Model 1		
SCP: $Log\ likelihood$	SLB: $Log\ likelihood$	FCP: $Log\ likelihood$
-782.32	-984.85	-771.86
Model 2		
SCP: $Log\ likelihood$	SLB: $Log\ likelihood$	FCP: $Log\ likelihood$
-792.91	-993.06	-*

*The $Log\ likelihood$ of model 2 for Porto is hide because the overall model is statistically insignificant.

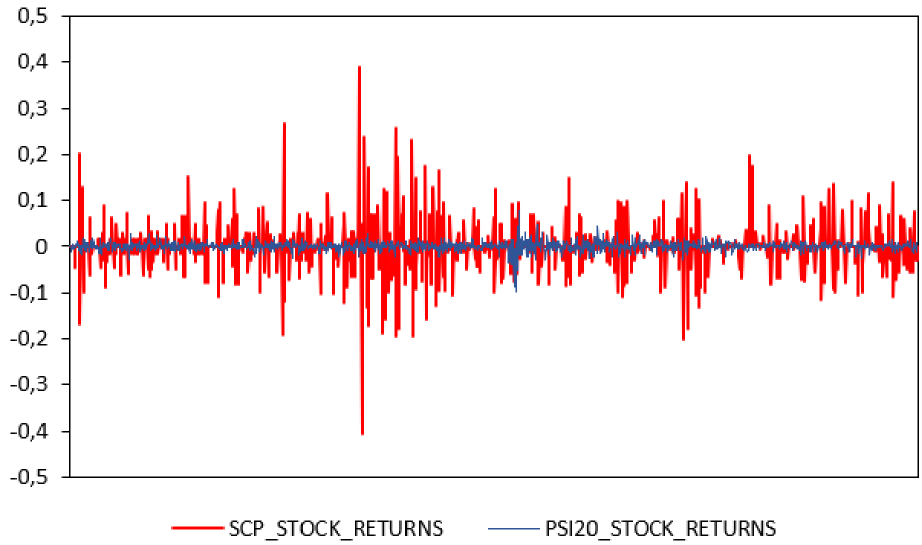
From the comparison of the $Log\ likelihood$ of each of the models, it is possible to make some conclusions:

1) For model 1, the one performed for Porto is the one that offers a better fit to the data since it has the highest $Log\ likelihood$. However, it is necessary to highlight that none of the variables related to the expressed sentiment on Twitter is statistically significant in Porto's first model. On the contrary, for Sporting, two of these variables are statistically significant, giving way to better responses to the research questions. Thus, for model 1, the model with the best findings is the one performed for Sporting.
2) Overall, when model 1 is compared to model 2 for each of the clubs, the conclusions are that model 1 offers a better fit to the data, since its $Log\ likelihood$s are higher than the ones for model 2. This gives responses to the second hypothesis of this dissertation – when the sentiment expressed on Twitter is combined with sporting performance variables in the same model, this offers a better fit to the data when compared to a model based only on expressed sentiment.

Also, in each model, it is possible to observe that many variables related to sporting performance – such as the match result, the winning probability of a given match, the match type, the importance of the match, and the goal difference – have an important effect on the direction of the clubs' stock returns.

However, the low liquidity is one of the biggest problems when investing in Portuguese football clubs' shares. During a normal trading day, there is a small number of shares transacted and there are even days with no transactions – according to Yahoo Finance, for Sporting Clube de Portugal, the average volume of stocks (i.e., number of shares transacted) during a regular trading day is 773, for Futebol Clube do Porto 966, and for Sport Lisboa e Benfica is 1,664. Besides increasing volatility, it also makes it difficult to find other investors to unwind the initial position. From the charts above, it is possible to observe the high volatility[20] of each of the three football clubs' stock returns versus the low volatility[21] of the PSI20 stock returns[22].

Chart 3 - SCP vs PSI20 Stock Returns



[20] Price fluctuates rapidly in a short period of time, hitting new highs and lows.
[21] Price does not fluctuate dramatically and tends to be steadier.
[22] Period of time: from 2016 to 2021.

Chart 4 - SLB vs PSI20 Stock Returns

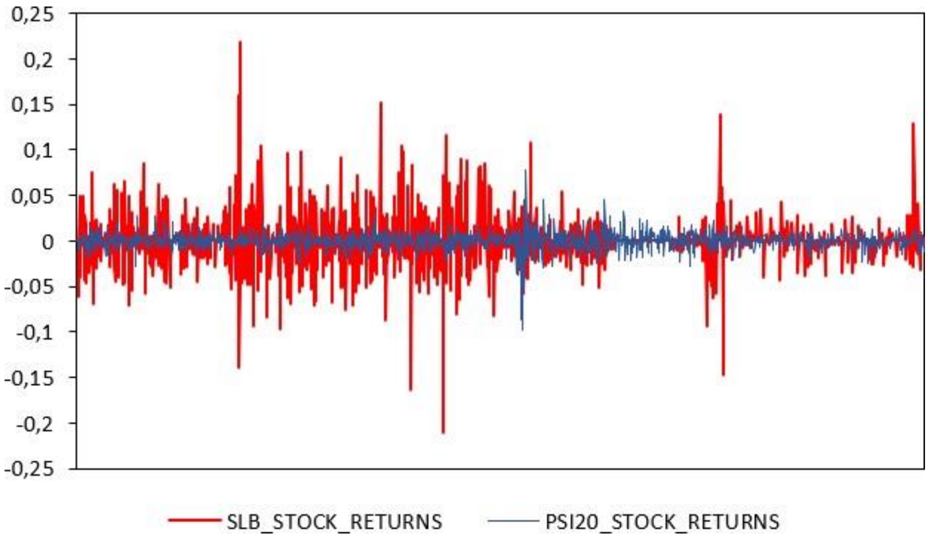
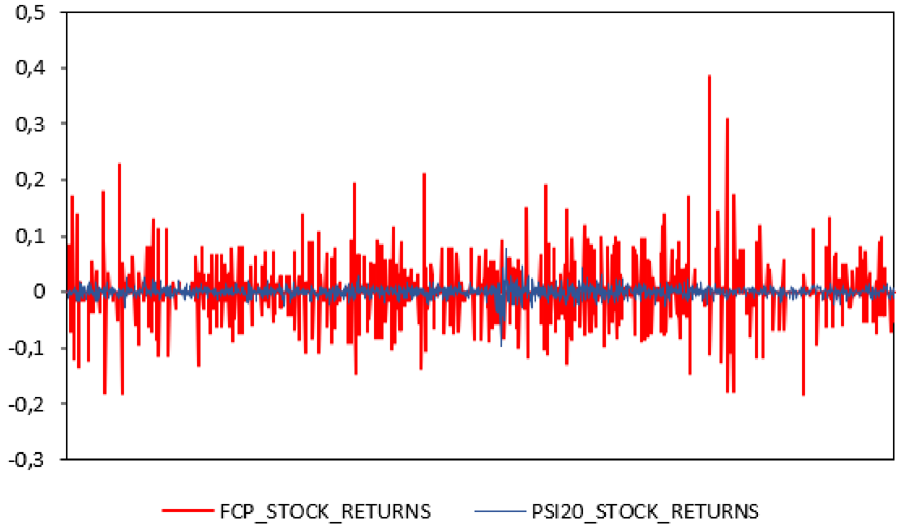


Chart 5 - FCP vs PSI20 Stock Returns



5. Conclusion

In this dissertation, two Logistic Regression models were applied for each football club in order to understand how the sentiment of football fans expressed on Twitter affects the direction of the stock prices in the Portuguese football industry – namely Sporting Clube de Portugal, Sport Lisboa e Benfica, and Futebol Clube do Porto. Both models were created with the objective of investigating whether the stock price returns of these clubs are impacted by the overall sentiment expressed on Twitter or not. However, there are some differences between them. The first model combines sentiment, financial information, sporting performance and match result probabilities independent variables, while the second one was built based only on financial and sentiment variables.

All models are found globally statistically significant for Sporting and Benfica. However, for Porto, only the first model is found statistically significant. Regarding the first model, explicitly concerning the expressed sentiment variables on social media, for Sporting, the overall neutral and negative expressed sentiments are statistically significant. For Benfica, only the total number of tweets is statistically significant. When it comes to Porto, none of the variables related to sentiment are statistically significant. Concerning the second model, results are only analyzed for Sporting and Benfica since this model is not globally statistically significant for Porto. For Sporting, as for model 1, in model 2, the overall neutral and negative expressed sentiments are found statistically significant. However, since the log likelihood of model 1 is greater than the log likelihood for model 2, it is assumed that the data fits better in model 1. For Benfica, only the total number of tweets is found statistically significant. From these results, the first conclusion that can be made is that the first hypothesis of this dissertation (i.e., the sentiment expressed on Twitter impacts the club's stock price returns) is not rejected for Sporting. However, for Benfica and Porto it is rejected. For Benfica, because the number of posted tweets is not directly related with the expressed sentiment on Twitter, and for Porto because any expressed sentiment variable is found statistically significant.

A common result for all three clubs in this study is that the effect of the probability of winning a given match is found to be statistically significant. Regarding the sporting performance variables, findings indicate that the difference in goals scored between the winning and losing teams is found to be statistically significant for Porto and Sporting, the match type is statistically significant for Benfica and Porto, and the match result and the importance of the match are statistically significant for Sporting and Benfica. Regarding the financial variables, the direction of the overall stock price return of PSI20 is found statistically significant for Sporting and Benfica, and the last price at which a PSI20 stock is traded is found statistically significant for Porto. Moreover, when the log likelihoods of the performed models are compared, the conclusion is that model 1 offers a better fit of the data. This analysis leads to the second finding of this study – hypothesis two (i.e., when the sentiment expressed on Twitter is combined with sporting performance variables in the same model, this offers a better fit to the data when compared to a model based only on expressed sentiment) is not rejected for Sporting nor Benfica. Since there is no evidence of statistical significance for the expressed sentiment variables for Porto, it turns out to be inconclusive.

Social media is increasingly taking a huge part of people's days. Most people wake up, and the first thing they do is to go on their social media accounts to check their feeds. This is because there it is possible to find information regarding any topic. Nowadays, since there is much available data, it is possible to make data-driven decisions, in other words, using facts, metrics, and data to guide strategic decisions according to the respective goals, objectives, and initiatives. Thus, this study contributes to the literature since it offers an analysis of the expressed sentiments in users' posts on Twitter regarding Portuguese football clubs that are publicly listed in order to understand whether their impact on the clubs' stock return is statistically significant and consequently, a metric to take in consideration when investing in the Portuguese football industry stocks.

In order to improve the results in further studies, it would be interesting to assess a more extensive set of tweets to get more reliable results. Also, in order to improve the results, another proposal is to include other financial and stock market factors into account, such as interest rate, number of investors, and the overall economic situation. However, the assessment of the Portuguese football industry is complex due to its low liquidity, because of its small size of the stock market, and high volatility, because of its inconstancy – price fluctuates rapidly in a short period of time, hitting new highs and lows. By way of example, when compared to other football industries, such as the Turkish one, according to Yahoo Finance, for Fenarbahçe, the average volume of traded stocks during a regular trading day is 3,992,853, for Besiktas 22,381,236, and for Galatasaray 37,114,638. When it comes to the Portuguese clubs, for Sporting Clube de Portugal, the average volume of traded stocks during a regular trading day is 773, for Futebol Clube do Porto 966, and for Sport Lisboa e Benfica is 1,664, representing a limitation for the present study. Another limitation of this study is the number of clubs under analysis. Thus, an additional recommendation for future studies would be to increase the number of countries under analysis (e.g., a study covering the football industry in European countries).

6. References

Ashton, J. K., Gerrard, B., & Hudson, R. (2003). Economic impact of national sporting success: Evidence from the London stock exchange. *Applied Economics Letters*, *10*(12), 783–785. https://doi.org/10.1080/1350485032000126712

Bartov, E., Stern, L. N., Faurel, L., Carey, W. P., & Mohanram, P. (2017). *Can Twitter Help Predict Firm-Level Earnings and Stock Returns?* http://fortune.com/2015/04/02/hedge-fund-twitter/

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*(26), 3435–3452. https://doi.org/10.1080/00036846.2011.577017

Berkowitz, J. P., & Depken, C. A. (2018). A rational asymmetric reaction to news: evidence from English football clubs. *Review of Quantitative Finance and Accounting*, *51*(2), 347–374. https://doi.org/10.1007/s11156-017-0673-6

Berument, M. H., & Ceylan, N. B. (2012). Effects of soccer on stock markets: The return-volatility relationship. *Social Science Journal*, *49*(3), 368–374. https://doi.org/10.1016/j.soscij.2012.03.003

Boțoc, C., Mihancea, E., & Molcuț, A. (2019). Football and Stock Market Performance Correlation: Evidence from Italy. *Scientific Annals of Economics and Business*, *66*(4), 525–539. https://doi.org/10.2478/saeb-2019-0044

Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters*, *30*, 116–123. https://doi.org/10.1016/j.frl.2019.03.030

Busalini, G. (2020). *The impact of soccer results on publicly traded soccer clubs Title: The impact of soccer results on publicly traded soccer clubs*.

Derakhshan, A., & Beigy, H. (2019). Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, *85*, 569–578. https://doi.org/10.1016/j.engappai.2019.07.002

Dimic, N., Neudl, M., Orlov, V., & Äijö, J. (2018). Investor sentiment, soccer games and stock returns. *Research in International Business and Finance*, *43*, 90–98. https://doi.org/10.1016/j.ribaf.2017.07.134

Ferreira, N. (2005). *Explaining share price performance of football clubs listed on the Euronext Lisbon*. Master's Dissertation, Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão.

Duz Tan, S., & Tas, O. (2021). Social Media Sentiment in International Stock Returns and Trading Activity. *Journal of Behavioral Finance*, *22*(2), 221–234. https://doi.org/10.1080/15427560.2020.1772261

Dwi, A., Dorkas, R., & Dolfriandra, H. (2021). Market reactions to football match results: The effect of venues and competition types. *The European Journal of Applied Economics*, *18*(1), 55–72. https://doi.org/10.5937/ejae18-28284

Floros, C. (2014). Football and Stock Returns: New Evidence. *Procedia Economics and Finance*, *14*, 201–209. https://doi.org/10.1016/s2212-5671(14)00703-5

Galloppo, G., & Boido, C. (2020). How much is a goal in the football championship worth? Match results and stock price reaction. *International Journal of Sport Finance*, *15*(2), 83–92. https://doi.org/10.32731/IJSF/152-052020.03

Gao, Y., Wang, Y., & Hafsi, T. (2022). Stock market reaction to affiliated sports teams' performance: evidence from China. *Chinese Management Studies*. https://doi.org/10.1108/CMS-06-2021-0262

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. https://doi.org/10.1177/1527002515626222

Jai-Andaloussi, S., el Mourabit, I., Madrane, N., Chaouni, S. B., & Sekkaki, A. (2016). Soccer events summarization by using sentiment analysis. *Proceedings - 2015 International Conference on Computational Science and Computational Intelligence, CSCI 2015*, 398–403. https://doi.org/10.1109/CSCI.2015.59

Kharde, V. A., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. In *International Journal of Computer Applications* (Vol. 139, Issue 11). http://ai.stanford.

Li, Q., Zhou, B., & Liu, Q. (2016). Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion. *Proceedings of 2016 IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2016*, 359–364. https://doi.org/10.1109/ICCCBDA.2016.7529584

Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, *7*, 1–21. https://doi.org/10.7717/peerj-cs.476

Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications*, *42*(24), 9603–9611. https://doi.org/10.1016/j.eswa.2015.07.052

Nofer, M., & Hinz, O. (2015). Using Twitter to Predict the Stock Market: Where is the Mood Effect? *Business and Information Systems Engineering*, *57*(4), 229–242. https://doi.org/10.1007/s12599-015-0390-4

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2017). Sentiment analysis of Twitter data for predicting stock market movements. *International Conference on Signal Processing, Communication, Power and Embedded System, SCOPES 2016 - Proceedings*, 1345–1350. https://doi.org/10.1109/SCOPES.2016.7955659

Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of twitter sentiment on stock price returns. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138441

Rao, T., & Srivastava, S. (2012). *Analyzing Stock Market Movements Using Twitter Sentiment Analysis*. https://dev.twitter.com/docs.

Renneboog, L., & Vanbrabant, P. (2000). Share Price Reactions to Sporty Performances of Soccer Clubs Listed on The London Stock Exchange and the AIM. *Discussion Paper No. 19, Tilburg University*.

Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). *Using Twitter Trust Network for Stock Market Analysis*. http://www.elsevier.com/open-access/userlicense/1.0/

Sakinc, I., Acikalin, S., & Soyguden, A. (2017). Evaluation of the relationship between financial performance and sport success in European football. *Journal of Physical Education and Sport*, *17*, 16–22. https://doi.org/10.7752/jpes.2017.s1003

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

Sarlan, A., Nadam, C., & Basri, S. (2014). *Information Technology and Multimedia (ICIMU), 2014 International Conference on.* [publisher not identified].

Shutes, K., McGrath, K., Lis, P., & Riegler, R. (2016). Twitter and the US stock market: The influence of micro-bloggers on share prices. *Economics and Business Review*, *2 (16)*(3), 57–77. https://doi.org/10.18559/ebr.2016.3.5

Škrinjarić, T., & Barišić, P. (2019). Effects of Football Match Results of Croatian National Team on Stock Returns: Evidence from Zagreb Stock Exchange. *Zagreb International Review of Economics and Business*, *22*(1), 13–45. https://doi.org/10.2478/zireb-2019-0010

Sóti, A., Ancarani, A., Stéger, J., & Vattay, G. (2020). Influence of Twitter activity on the stock price of soccer clubs. *Social Network Analysis and Mining*, *10*(1). https://doi.org/10.1007/s13278-020-00691-2

Souza, T. T. P., Kolchyna, O., Treleaven, P. C., & Aste, T. (2015). Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry. *Handbook of Sentiment Analysis in Finance*. http://arxiv.org/abs/1507.00784

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine*, *18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Sun, T., & Wu, M. (2015). Stock Market Reaction to News: Evidence from Juventus Revisited. *European Scientific Journal*, *11*(10).

Sun, Y., Liu, X., Chen, G., Hao, Y., & Zhang, Z. (Justin). (2020). How mood affects the stock market: Empirical evidence from microblogs. *Information and Management*, *57*(5). https://doi.org/10.1016/j.im.2019.103181

Surowiecki, J. (2005). *The Wisdom of Crowds.* Knopf Doubleday Publishing Group.

Tanaltay, A., Langroudi, A. S., Akhavan-Tabatabaei, R., & Kasap, N. (2021). Can Social Media Predict Soccer Clubs' Stock Prices? The Case of Turkish Teams and Twitter. *SAGE Open*, *11*(2). https://doi.org/10.1177/21582440211004153

Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between twitter and stock prices. Evidence from the US technology industry. *Technological Forecasting and Social Change*, *149*. https://doi.org/10.1016/j.techfore.2019.119747

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2022). Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods. *Cognitive Computation*, *14*(1), 372–387. https://doi.org/10.1007/s12559-021-09819-8

Yilmaz, E. S., Ozpolat, A., & Destek, M. A. (2022). Do Twitter sentiments really effective on energy stocks? Evidence from the intercompany dependency. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-022-21269-9

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*(5), 305–313. https://doi.org/10.1080/0960310042000338713

7. Appendix

Figure A1 - Isid Command Output SCP

```
. isid PSI20 ClosePSI20 MatchResult WinProb GoalDiff MatchType DerClass Neutral Negative TotTweets SentScore

.
```

Figure A2 - Isid Command Output SLB

```
. isid PSI20 ClosePSI20 MatchResult WinProb GoalDiff MatchType DerClass Neutral Negative TotTweets SentScore

.
```

Figure A3 - Isid Command Output FCP

```
. isid PSI20 ClosePSI20 MatchResult WinProb GoalDiff MatchType DerClass Neutral Negative TotTweets SentScore

.
```

Table A1 – Box Tidwell test results for SCP (Model 1)

Variables	P-value
$ClosePSI20$	0.087
$GoalDiff$	0.425
$TotTweets$	0.966
$SentScore$	0.750

Table A2 – Box Tidwell test results for SLB (Model 1)

Variables	P-value
$ClosePSI20$	0.902
$GoalDiff$	0.761
$TotTweets$	0.217
$SentScore$	0.196

Table A3 – Box Tidwell test results for FCP (Model 1)

Variables	P-value
$ClosePSI20$	0.573
$GoalDiff$	0.346
$TotTweets$	0.155
$SentScore$	1.000

Table A4 – Box Tidwell test results for SCP (Model 2)

Variables	P-value
$ClosePSI20$	0.087
$SentScore$	0.774
$TotTweets$	0.961

Table A5 – Box Tidwell test results for SLB (Model 2)

Variables	P-value
$ClosePSI20$	0.977
$SentScore$	0.183
$TotTweets$	0.194

Table A6 – Box Tidwell test results for FCP (Model 2)

Variables	P-value
$ClosePSI20$	0.510
$SentScore$	0.137
$TotTweets$	0.855

Table A7 - Correlation Matrix SCP: Model 1

	PSI20	ClosePSI20	MatchResult	WinProb	GoalDiff	MatchType	DerClass	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000										
ClosePSI20	-0.0215	1.0000									
MatchResult	-0.0450	0.1037	1.0000								
WinProb	-0.0597	0.1753	0.5434	1.0000							
GoalDiff	-0.0152	-0.0263	0.1598	-0.0813	1.0000						
MatchType	0.0774	-0.0033	-0.3325	-0.4733	0.1477	1.0000					
DerClass	0.0276	-0.1234	-0.3863	-0.6137	-0.0617	-0.0228	1.0000				
Neutral	0.0130	0.0133	0.0405	0.0340	-0.0077	-0.0128	-0.0304	1.0000			
Negative	-0.0255	0.0350	-0.0221	-0.0345	-0.0232	0.0183	0.0277	-0.4542	1.0000		
TotTweets	0.0171	-0.0223	-0.0050	-0.0038	0.0167	-0.0135	0.0099	-0.1267	0.1845	1.0000	
SentScore	0.0198	-0.0140	0.0037	0.0076	0.0135	-0.0366	0.0034	0.0566	-0.6877	-0.0770	1.0000

Table A8 - Correlation Matrix SLB: Model 1

	PSI20	ClosePSI20	MatchResult	WinProb	GoalDiff	MatchType	DerClass	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000										
ClosePSI20	-0.0241	1.0000									
MatchResult	-0.0163	0.0557	1.0000								
WinProb	-0.0197	0.0740	0.5570	1.0000							
GoalDiff	-0.0001	-0.0054	0.1060	-0.0317	1.0000						
MatchType	0.0193	-0.0345	-0.4740	-0.5361	0.0870	1.0000					
DerClass	0.0500	-0.1088	-0.3427	-0.6661	-0.0657	-0.0224	1.0000				
Neutral	-0.0258	0.0915	-0.0265	-0.0359	0.0164	-0.0235	0.0577	1.0000			
Negative	0.0508	-0.0597	0.0322	0.0391	-0.0229	0.0330	-0.0356	-0.5049	1.0000		
TotTweets	-0.0459	0.0008	-0.2168	-0.2869	0.0656	0.2031	0.2359	0.2720	-0.1776	1.0000	
SentScore	-0.0140	0.0113	-0.0227	-0.0225	0.0434	-0.0303	0.0309	0.1826	-0.6854	0.1421	1.0000

Table A9 - Correlation Matrix FCP: Model 1

	PSI20	ClosePSI20	MatchResult	WinProb	GoalDiff	MatchType	DerClass	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000										
ClosePSI20	-0.0000	1.0000									
MatchResult	0.0385	0.0579	1.0000								
WinProb	-0.0159	-0.0662	0.0002	1.0000							
GoalDiff	-0.0335	0.0903	0.0157	-0.0081	1.0000						
MatchType	-0.0428	0.1217	-0.0548	-0.0237	0.5714	1.0000					
DerClass	-0.0150	-0.0164	-0.0199	-0.0266	0.1630	-0.0026	1.0000				
Neutral	-0.0075	-0.0505	0.0104	0.6853	-0.0396	-0.0297	-0.0386	1.0000			
Negative	0.0348	-0.0185	-0.0139	-0.0960	-0.0078	-0.0026	0.0041	-0.1115	1.0000		
TotTweets	-0.0187	-0.0214	0.0244	0.2154	-0.0082	-0.0223	0.0243	0.1300	-0.6903	1.0000	
SentScore	-0.0087	0.0658	-0.0182	-0.1745	0.0353	0.0392	-0.0256	0.0011	-0.1016	-0.4358	1.0000

Table A10 - Correlation Matrix SCP: Model 2

	PSI20	ClosePSI20	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000					
ClosePSI20	-0.0215	1.0000				
Neutral	0.0130	0.0133	1.0000			
Negative	-0.0255	0.0350	-0.4542	1.0000		
TotTweeets	0.0171	-0.0223	-0.1267	0.1845	1.0000	
SentScore	0.0198	-0.0140	0.0566	-0.6877	-0.0770	1.0000

Table A11 - Correlation Matrix SLB: Model 2

	PSI20	ClosePSI20	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000					
ClosePSI20	-0.0241	1.0000				
Neutral	-0.0258	0.0915	1.0000			
Negative	0.0508	-0.0597	-0.5049	1.0000		
TotTweeets	-0.0459	0.0008	0.2720	-0.1776	1.0000	
SentScore	-0.0140	0.0113	0.1826	-0.6854	0.1421	1.0000

Table A12 - Correlation Matrix FCP: Model 2

	PSI20	ClosePSI20	Neutral	Negative	TotTweets	SentScore
PSI20	1.0000					
ClosePSI20	-0.0000	1.0000				
Neutral	-0.0075	-0.0505	1.0000			
Negative	0.0348	-0.0185	-0.1115	1.0000		
TotTweeets	-0.0187	-0.0214	0.1300	-0.6903	1.0000	
SentScore	-0.0087	0.0658	0.0011	-0.1016	-0.4358	1.0000

Table A13 – Least likely SCP (Model 1)

Outcome: 0	
Observation	Prob.
438	0.5109
921	0.4479
998	0.4972
1103	0.4205
1177	0.3884
Outcome: 1	
648	0.0942
1180	0.0832
1240	0.1054
1258	0.1054
1260	0.1054
1268	0.1054
1282	0.1054
1334	0.1054
1636	0.0560
1736	0.0914

Table A14 – Leastlikely SLB (Model 1)

Outcome: 0	
Observation	Prob.
83	0.4795
1169	0.4289
1204	0.4073
1217	0.4349
1596	0.4240
Outcome: 1	
1240	0.1215
1244	0.1215
1259	0.1215
1268	0.1215
1274	0.1215
1282	0.1215
1300	0.0749

Table A15 – Least likely FCP (Model 1)

Outcome: 0	
Observation	Prob.
69	0.4154
104	0.5173
573	0.4885
1205	0.4546
1574	0.5066
Outcome: 1	
88	0.0968
616	0.0817
1229	0.1048
1300	0.1086
1393	0.1115

Table A16 – Least likely SCP (Model 2)

Outcome: 0	
Observation	Prob.
438	0.6792
466	0.6762
644	0.6834
745	0.6720
827	0.6840
Outcome: 1	
378	0.1035
1240	0.1082
1258	0.1082
1260	0.1082
1268	0.1082
1282	0.1082
1334	0.1082
1544	0.1057
1736	0.0974

Table A17 – Leastlikely SLB (Model 2)

Outcome: 0	
Observation	Prob.
469	0.5551419
865	0.5296471
1204	0.5810519
1596	0.5018555
1735	0.5733825
Outcome: 1	
1240	0.1191075
1244	0.1191075
1259	0.1191075
1268	0.1191075
1274	0.1191075
1282	0.1191075
1300	0.1191075

Table A18 – Least likely FCP (Model 2)

Outcome: 0	
Observation	Prob.
449	0.7822
464	0.7796
466	0.7820
650	0.7803
651	0.7814
Outcome: 1	
1229	0.1250
1244	0.1256
1260	0.1256
1271	0.1256
1282	0.1256
1289	0.1256
1300	0.1256
1652	0.1080
1672	0.0887

Table A19 - Calculation for coefficients' intrepertations

Sporting			
Model 1		Model 2	
Coeffecient	Calculation for Intrepretation	Coeffecient	Calculation for Intrepretation
ClosePSI20: 0.0001106	$e^{0.0001106} - 1 = 0.0001 = 0.01\%$	ClosePSI20: 0.0001036	$e^{0.0001036} - 1 = 0.0001 = 0.01\%$
MatchResult: 0.9594187	$e^{0.9594187} = 2.61$	Neutral: -1.101293	$e^{-1.101293} - 1 = -0.668 = -66.8\%$
WinProb: -1.305232	$e^{-1.305232} - 1 = -0.729 = -72.9\%$	Negative: -0.6064826	$e^{-0.6064826} - 1 = -0.455 = -45.5\%$
GoalDiff: 0.2386459	$e^{0.2386459} - 1 = 0.27 = 27.0\%$	-	-
DerClass: -2.522115	$e^{-2.522115} - 1 = -0.92 = -92.0\%$	-	-
Neutral: -1.10786	$e^{-1.1078659} - 1 = -0.67 = -67.0\%$	-	-
Negative: -0.5647198	$e^{-0.5647198} - 1 = -0.431 = -43.1\%$	-	-
Benfica			
Model 1		Model 2	
Coeffecient	Calculation for Intrepretation	Coeffecient	Calculation for Intrepretation
ClosePSI20: 0.0003055	$e^{0.0003055} - 1 = 0.0003 = 0.03\%$	ClosePSI20: 0.0002976	$e^{0.0002976} - 1 = 0.0003 = 0.03\%$
MatchResult: 0.9061261	$e^{0.9061261} = 2.47$	TotTweets: 0.0019539	$e^{0.0019539} - 1 = 0.00196 = 0.02\%$
WinProb: 1.737022	$e^{1.737022} = 5.68$	-	-
MatchType: 1.885194	$e^{1.885194} = 6.59$	-	-
DerClass: 1.929734	$e^{1.929734} = 6.89$	-	-
TotTweets: 0.0026992	$e^{0.0026992} - 1 = 0.0027 = 0.27\%$	-	-

Table A19 (Continuation) - Calculation for coefficients' intrepertations

Porto			
Model 1		Model 2	
Coeffecient	Calculation for Intrepretation	Coeffecient	Calculation for Intrepretation
PSI20: 0.3121291	$e^{0.3121291} = 1.36$	-	-
WinProb: -1.204509	$e^{-1.204509} - 1 = -0.700 = 70.0\%$	-	-
GoalDiff: -0.3295511	$e^{-0.3295511} - 1 = -0.281 = -28.1\%$	-	-
MatchType: 1.147744	$e^{1.147744} = 3.15$	-	-