Getting the Boot? Predicting the Dismissal of Managers in Football

Mounir Attié(✉), Diogo Pacheco, and Marcos Oliveira

Computer Science, University of Exeter, Exeter, UK
{maa250,d.pacheco,m.a.oliveira}@exeter.ac.uk

Abstract. Football club managers have a challenging and remarkably volatile job—the practice of sacking and replacing managers is widespread in the modern game. However, it is still unclear what exactly motivates managerial dismissal in clubs. More than ever, high-quality statistics are available to clubs, suggesting that dismissal decisions tend to be well informed. Likewise, supporters on social media might also influence clubs' decisions. Here we propose machine learning models to characterize the determinants of managerial dismissals. Is social media pressure associated with managerial sacking? Yes! We fit multiple ElasticNet regularised logistic regression models using features based on the social pressure of fans on Twitter and football statistics, showing that our best model obtains a balanced prediction accuracy of 0.75.

Keywords: Football dismissal · Impact performance and success prediction · Machine learning prediction · Social media

1 Introduction

Football is one of the most beloved sport worldwide. It unites people across borders, builds bridges, and provides an escape for so many from the hardships of life. The sport, however, can also be the creator of tribulations, notably for the manager, or first-team coach, due to the volatile nature of their job. Managers come and go way quicker than anybody else in football, but the reasons for this phenomenon are yet to be discovered [2,5,21]. With ever-increasing popularity, football clubs have become massive commercial entities with revenues in the millions [12,13]. The result of this is that top football clubs are run in a similar way to top businesses—with a certain ruthlessness towards underperforming at every level [17].

Yet most football managers have very little time to achieve their goals. The average tenure of a manager in England's top division (the Premier League) was less than 22 months in 2015, and in the second division (the Championship), just around 10 months [20]. Undeniably, for any business, constant failure to achieve company objectives is detrimental to growth. However, the frequency of sackings may lead us to believe that there may be a certain unfairness. Rowe et al. [18] call this the ritual scapegoating theory, in which succession at any given sports team

A. S. Teixeira et al. (Eds.): CompleNet 2023, SPCOM, pp. 132–140, 2023.
https://doi.org/10.1007/978-3-031-28276-8_12

is inevitable despite many times there being no practical reason for it. Nonetheless, decision-makers at football clubs are thought to make well-informed decisions based on state-of-the-art statistics to decide whether to sack managers or not. Conversely, every club is run differently and follows different processes to determine the fate of a manager. These factors make this area of research complex and active.

Managerial turnover has been an area of research for many academics in sports (not only in football), attempting to understand the causes and consequences of managerial turnover for decades [1–4,6–10,23]. Several researchers have attempted to understand why a manager gets sacked, and if it is the consequence of solely losing a few games. The general consensus is that it isn't the case, but the determinants of dismissal can vary a lot. They often conclude there is missing data in their models, such as fan or media pressure. Public relations are critical in football, and it is almost impossible to neglect the impact that the fans have on players, directors, managers and on the club as a whole. Football fans are known to be prominent users on social media to vent and express their views on different footballing topics and issues [16]. Twitter, for instance, has been used to characterise the world cup [14] and to unveil rivalries [15]. More specifically, the sentiment of tweets have been used to predict football scores [11,19]. Social media data has been used to justify general dismissals based on individual behaviour and privacy breaches [22], but to the best of our knowledge, it has never been used to predict sacking football managers or dismissals based on public opinion.

Previous works predicting dismissals mainly focused on probit and logit models, and survival models such as parametric proportional hazard. Unsurprisingly, they agree on the *team's win ratio* as an important determinant of a turnover. For instance, in NBA (basketball), the *team's win ratio* was the only significant determinant [9], but in NFL (American football) *player talent*, *coaching talent*, *on-field performance*, and *resource quality* were all important determinants [1]. Despite the lack of consensus, they offer valuable insights into other significant variables in their models. Forrest et al.'s [7] equally trained a probit model with football data from the Spanish first division and obtained quite different results. They concluded that features such as the *time of the season*, the result of the *last match*, a *previous dismissal* in the season, and a *relegation battle* are all determinants. These differences suggest singularities between different sports when sacking managers. Bachan et al. [3] found the *attendance at stadiums*, the *manager "internationalisation"*, and a *relegation battle* as relevant determinants of a dismissal. d'Addona et al. [8] concluded that *age* and *prior experience* of the manager, *recent* match *results*, and a *change* in league *position* as important factors to dismissal. Both works [3,8] exploit time-discrete logit models, enabling time-based analysis and modelling. Chase et al. [6] also use a logit model, but not very successful, only predicting 2 out of the 7 managerial dismissals correctly with NFL data. Frick et al. [10] garnered data from Germany's first division (the Bundesliga) and built a mixed logit model. Despite achieving good results, their model is very complex and relies on individual characteristics rather than statistics.

Another popular way to predict managerial turnover is with a parametric hazard model. A survival model analyses the duration of time until the event "a manager is sacked" occurs. Hence, utilising continuous time periods instead of discrete ones. This can be an issue when there is missing data. Many studies use variations of this model, such as Cox proportional model and the Weibull model, whose variation stem from different mathematical assumptions [2,4,8,23]. They draw similar conclusions to the researchers utilising probit and logit, but have also found other determinants. For instance, Audas et al. [2] conclude that results in recent games are more relevant than those from older games, but also consider the position of the team in the table before and after the manager was hired a crucial factor. Van Ours et al. [23] found the result of the last four games as the most important determinant for dismissal. The second most important factor was the "cumulative surprise indicator" – calculated by comparing actual results and expected results based on bookmaker's odds.

This work uses machine learning models to find the determinants of a managerial dismissal. We use traditional football statistics and analyse emotions in tweets. The results reveal that the sentiment of fans combined with match stats from previous ten games can accurately predict dismissals.

2 Results

To model and predict the relevant determinants of managerial turnover, we collected statistics from England's top two divisions: the Premier League and the Championship, including over 100 features about managers and teams. We also used sentiment analysis to gather the sentiment of the fans on Twitter about their club and about each manager individually for 4 different seasons ranging from 2014/15 to 2021/22. We use this data to understand the determinants of a managerial casualty.

2.1 Models

We construct models that categorizes the outcome of a managerial tenure (i.e., sacked or not sacked) based on different sets of features, allowing us to understand the determinants of dismissals. In all models, we used a regularized logistic regression model (i.e., ElasticNet), allowing us to select features for the models efficiently. We build 7 models of increasing complexity, as described in Table 1, which have the following feature categories:

Manager's career This category evaluates the manager's reputation as a manager, based on their previous and current tenures, including the record (i.e., wins, draws, losses, points, PPM) achieved for different periods.

Manager's playing career This category evaluates managers' reputation during their playing career, based principally on the leagues they played in and the games played in those leagues.

Table 1. Models with increased levels of complexity.

Model	Description	# Features
Naïve model #1	Points per match (PPM) of last 5 matches	1
Naïve model #2	Points per match (PPM) of last 10 matches	1
Group model #1	Tweets' sentiment scores only	8
Group Model #2	Manager's career and playing career	31
Complex Model #1	Match stats and results of last 10 matches	28
Complex Model #2	Match stats, results of last 10 matches, and manager's record and career	59
Complex Model #3	Match Stats, results of last 10 matches, manager's record and career, and sentiment scores	67

Match stats This category is the broadest, evaluating all the match statistics for a certain number of games. These include possession, passing accuracy, shot accuracy, touches and 22 other analytics from a game.

Sentiment scores The positive and negative sentiment about the manager and club as measured using Vader and TextBlob.

We note that ElasticNet combines both $L1$ and $L2$ regularizations with wheighing controlled by an α value. To find the best α, we use k-fold cross validation with $k = 10$. We also use stratified sampling to ensure that the rate of sacked and not sacked manager are the same in both the training set and test set.

2.2 Model Comparisons

First, we compare the naive models, finding that the model fitted with the PPM for 10 games outperforms the PPM for 5 games across different metrics (see Table 2). This finding supports our decision of using statistics of the last 10 games in all the other models. Next, we examine the group models; one fitted with uniquely sentiment score (Group Model #1) and the other using managers reputation as a player and manager (Group Model #2). These two models are based on the assumption that the sacking of a manager is based solely on the

Table 2. Performance metrics for all models.

Model	Precision	Recall	F1 score	Accuracy
Naïve Model #1	0.66	0.80	0.72	0.70
Naïve Model #2	0.79	0.84	0.82	0.82
Group Model #1	0.63	0.62	0.63	0.63
Group Model #2	0.77	0.83	0.80	0.79
Complex Model #1	0.82	0.92	0.87	0.84
Complex Model #2	0.86	0.91	0.89	0.89
Complex Model #3	0.82	0.96	0.88	0.85

sentiment of the fans and solely on the manager's past, respectively. We find that Group Model #2 outperforms Group Model #1 based on the AUC (0.88 and 0.67 respectively). When examining the complex models, we find that they perform better than these previous models, as shown in Table 2.

These models are all viable, but these analyses thus far have focused solely on the performance of the training set. We are interested in the generalizability of the model. For that purpose, we analyze the performance of these models in the test set as well. We use unweighted average recall (UAR) as the primary metric to evaluate the models' ability to predict outcomes, as it takes into account unbalance in the data. In this analysis, we focus on the Complex models and the Naive Model #2, as they had the best performance in the previous analysis.

To examine the UAR, we note that each model has an optimal cutoff for the predictions. We set the optimal cutoff at 0.45, slightly lower than 0.5, to quantify the slight unfairness of the field being studied. This means that a manager that has a 46% chance of being sacked will be considered sacked in the models.

Our results show that the most complex models fitted performs optimally in these circumstances with a UAR of 0.75, as shown by the UAR values for each model in Table 1. However, we have also observed that when we increase the optimal cutoff, the most naive model performs better than the more complex ones. This is expected, as if we are trying to base the sacking of a manager solely on results, the threshold is higher, as decision makers seem to not consider this the most important factor, as previously hypothesized, a sacking is more than losing a few games. Hence, the naïve models present missing features that we add progressively. Nonetheless, we have found that Complex Model #3, followed by Complex Model #2 are the most optimal models.

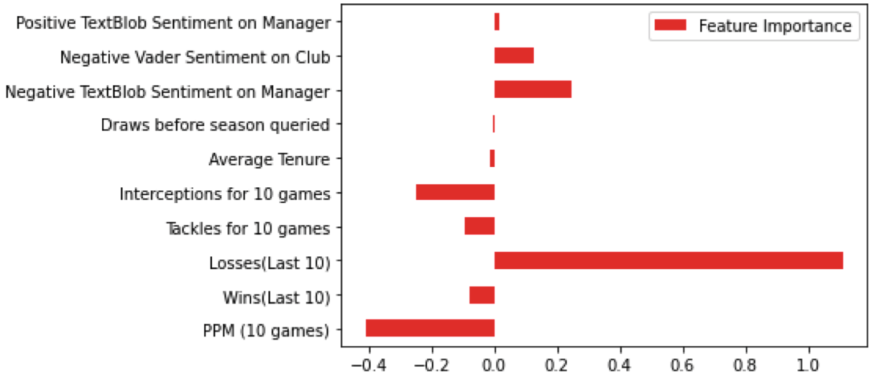
Table 3. UAR on unseen data.

Model	UAR of Test Set (0.45)	Number of features
Naïve model #2	0.72	1
Complex Model #1	0.70	28
Complex Model #2	0.73	59
Complex Model #3	0.75	67

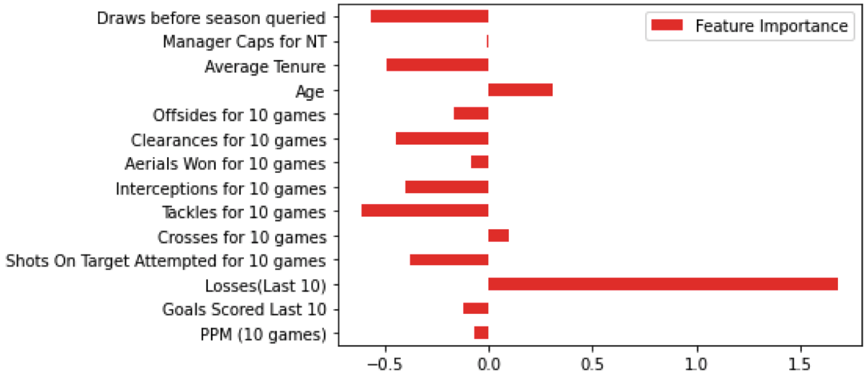
2.3 Determinants of a Dismissal

We now focus on understanding the relevant determinants of sacking in football. We focus on five aspects: sentiment of the fans, match statistics, manager as a manager, manager as a player, and human bias.

Sentiment Analysis. The addition of sentiment into the model has dramatically reduced the number of important features from the previous model. Figure 1(a) shows that Negative Vader Sentiment on the Club and Negative TextBlob Sentiment on the Manager are both relevant determinants of a dismissal. As the negative sentiment increases, the probability of being sacked



(a) Complex Model #3: Relative importance of features



(b) Complex Model #2: Relative Importance of Features

Fig. 1. Relative importance of features for best models

increases as well. This is in line with the discussion in the introduction which fans' anger towards the club and manager have an influence on a dismissal. Furthermore, the sentiment towards the manager has more of a weight on the outcome compared to the sentiment towards the club, which was equally assumed.

We cannot definitely conclude which of the two sentiment packages, TextBlob and Vader, is definitely better for our purpose. The results would suggest that they perform differently on tweets addressing the manager and the club, as the Negative Sentiment on the Club with TextBlob and the Negative Sentiment on the manager with Vader are both rendered irrelevant. It would be viable to explore the type of language used overwhelmingly in the respective categories. We also observe that Positive Sentiment is rendered almost irrelevant. We can hypothesize this is due to high displease (negative sentiment) drowning out the low praise (positive sentiment), which makes the sentiment closer to neutral.

Match Statistics. Many of the match statistics had been rendered irrelevant in the Complex Model #3. In Complex Model #2 however, the higher the average of Offsides, Clearances, Aerial Duals won, Interceptions, Tackles and Shots on Goal over the 10 games, the less chance of being sacked. The most significant determinants are the Clearances, Interceptions and Tackles. This is coherent with real world assumptions. The more offsides, crosses, and shots a team has a game, the more chances it is creating, which is linked to attacking performance and effort. In contrast, clearances, aerial duels won, interceptions and rackles represent a team's defensive ability and team organisation. All these statistics are associated with high player work rates, which is directly related to the manager's ability to motivate his players. It is noticeable, however, that crosses has a slight negative impact on the outcome. This may be because a cross can be a 'lazy' form of attack, instead of intricate passing.

In Complex Model #3 with the sentiment scores, the average tackles and interceptions over the span of 10 games stay determinants in the model, negatively impacting the outcome. They were previously the highest determinants in the match stats category from the previous model. This is because Tackles and interceptions are probably the best indicators of team effort and High work rate. However, interceptions have more weight on the outcome then tackles. We can say that the sentiment of the fans may reduce the prerogative of decision makers to look at different statistics to make a decision, instead focusing on public relations.

Manager Playing Career. Both models shows no correlation with the manager's previous career as a player. This means that managers are mostly held to the same standard as others even with a glamorous playing career.

Manager Managerial Career. According to our models, the manager's previous career is rendered irrelevant with sentiment but without, the average tenure and draws in his career have a negative weight on the outcome of being sacked. A longer average tenure is representative of a non-problematic and successful manager, whereas the number of draws the manager has culminated in his career is synonymous with not losing, and it is equally important in a relegation or title fight, as a draw is much better than a loss. In both models however, the most relevant metrics of a manager's career are the ones from the previous 10 games, and they are the most important determinants unsurprisingly. The number of losses suffered in the last 10 games is the most important determinant in our model, it has four times the weight of any other feature's importance in both models. The more games the manager has lost in the last 10, the higher the chance he has of being sacked. As opposed to the other determinants, the significant weight shows that this chance of being sacked is far superior to the others. The PPM for 10 games, is the driving determinant in our well performing naive model, Naive model #2, but loses its influence when we consider the losses, but regains some influence when including the sentiment in the model. This is in line with the naive presupposition that the more points a manager gains over

10 games, the less chance he has of being sacked. It proves to not be as naive as expected.

Human Bias. Interestingly, older managers seem to have an increased chance of being sacked. It was assumed that older age increased respect that decision makers have towards the manager. However, it seems that being older is less of a benefit than being younger. One possible hypothesis is that a younger manager brings more energy and can be instrumental towards a long term project, whereas an older manager can be perceived as a short term plan. It is rendered irrelevant however by the sentiment as the opinion of the fans renders this bias obsolete.

3 Conclusions

This paper investigated the determinants of managerial turnover in football. We collected statistics from England's top two divisions and tweets over four seasons. In addition to traditional performance indicators, we examined whether *social pressure* contributes to dismissals, or could be used to predict them. We used sentiment analysis on tweets as a proxy for the public opinion towards clubs and managers. We successfully created elasticnet regularised logistic regression models obtaining precision, recall, accuracy, and F1 scores over 0.80. The unweighted average recall (UAR) scores are over 0.70, with a maximum of 0.75, suggesting a good generalisation. As expected, managers' *points per match* and *losses* over the last ten games are the driving factors of a dismissal. Nonetheless, the model suggests that the fans anger towards the club and the manager could be the tipping point in a dismissal decision, and that team effort also weights into the decision (e.g., the number of *interceptions* and *tackles*). Hence, even if managers will mostly be sacked because of the number of losses and their lack of point in recent matches, the significance of social pressure indicators suggests a complex decision scenario. It is possible that a negative sentiment towards managers is an emergent collective effect driven by bad performance; It could also be driven by prejudice or previous encounters as an opponent. This study does not address the causes of this social pressure and further investigation is required.

References

1. Allen, W.D., Chadwick, C.: Performance, expectations, and managerial dismissal: evidence from the national football league. J. Sports Econ. **13**(4), 337–363 (2012)
2. Audas, R., Dobson, S., Goddard, J.: Organizational performance and managerial turnover. Manag. Decis. Econ. **20**(6), 305–318 (1999)
3. Bachan, R., Reilly, B., Witt, R.: The hazard of being an English football league manager: empirical estimates for three recent league seasons. J. Oper. Res. Soc. **59**(7), 884–891 (2008)
4. Barros, C.P., Frick, B., Passos, J.: Coaching for survival: the hazards of head coach careers in the German 'bundesliga'. Appl. Econ. **41**(25), 3303–3311 (2009)

5. Calvin, M.: Living on the Volcano: The Secrets of Surviving as a Football Manager. Random House (2015)
6. Chase, H.W., Glickman, M.E.: The analytics of getting sacked: coach firings in the national football league. Significance **13**(4), 34–37 (2016)
7. de Dios Tena, J., Forrest, D.: Within-season dismissal of football coaches: statistical analysis of causes and consequences. Eur. J. Oper. Res. **181**(1), 362–373 (2007)
8. d'Addona, S., Kind, A.: Forced manager turnovers in English soccer leagues: a long-term perspective. J. Sports Econ. **15**(2), 150–179 (2014)
9. Fizel, J.L., D'Itri, M.P.: Managerial efficiency, managerial succession and organizational performance. Manag. Decis. Econ. **18**(4), 295–308 (1997)
10. Frick, B., Barros, C.P., Prinz, J.: Analysing head coach dismissals in the German "bundesliga" with a mixed logit approach. Eur. J. Oper. Res. **200**(1), 151–159 (2010)
11. Godin, F., Zuallaert, J., Vandersmissen, B., De Neve, W., Van de Walle, R.: Beating the bookmakers: leveraging statistics and twitter microposts for predicting soccer results. In: KDD Workshop on Large-Scale Sports Analytics, New York, NY, USA, pp. 2–14. ACM (2014)
12. Kumar, A., Roshan, D.: Football market size, share & growth: Industry forecast - 2027, May 2021. https://www.alliedmarketresearch.com/football-market-A11328
13. Lange, D.: Football clubs revenue ranking 2019/20, October 2021. https://www.statista.com/statistics/271581/revenue-of-soccer-clubs-worldwide/
14. Pacheco, D., de Lima Neto, F.B., Moyano, L., Menezes, R.: Football conversations: what twitter reveals about the 2014 world cup. In: Anais do IV Brazilian Workshop on Social Network Analysis and Mining. SBC (2015)
15. Pacheco, D.F., Pinheiro, D., de Lima Neto, F.B., Ribeiro, E., Menezes, R.: Characterization of football supporters from twitter conversations. In: WI, pp. 169–176 (2016)
16. Price, J., Farrington, N., Hall, L.: Changing the game? the impact of twitter on relationships between football clubs, supporters and the sports media. Soccer Soc. **14**(4), 446–461 (2013)
17. Rieple, A., Vyakarnam, S.: The case for managerial ruthlessness 1. Br. J. Manag. **7**(1), 17–33 (1996)
18. Rowe, W.G., Cannella, A.A., Jr., Rankin, D., Gorman, D.: Leader succession and organizational performance: Integrating the common-sense, ritual scapegoating, and vicious-circle succession theories. Leadersh. Q. **16**(2), 197–219 (2005)
19. Schumaker, R.P., Jarmoszko, A.T., Labedz, C.S., Jr.: Predicting wins and spread in the premier league using a sentiment analysis of twitter. Decis. Support Syst. **88**, 76–84 (2016)
20. New statistics reveal average tenure of managers in England. https://www.skysports.com/football/news/11688/9875915/average-tenure-of-managers-in-england-just-123-years
21. Smith, A.: Football manager job security at all-time low, May 2020. https://www.skysports.com/football/news/11096/10811856/football-manager-job-security-at-all-time-low-sky-sports-study-finds
22. Thornthwaite, L.: Social media and dismissal: towards a reasonable expectation of privacy? J. Ind. Relat. **60**(1), 119–136 (2018)
23. Van Ours, J.C., Van Tuijl, M.A.: In-season head-coach dismissals and the performance of professional football teams. Econ. Inq. **54**(1), 591–604 (2016)