Computational Statistics
https://doi.org/10.1007/s00180-024-01584-0

ORIGINAL PAPER

The game beyond the field: on football players' performance through social media, sentiment and topic analysis

Marco Ortu[1] · **Francesco Mola**[1]

Received: 29 April 2024 / Accepted: 25 November 2024

Abstract

This study investigates the complex relationship between social media sentiment and football players' performance in the English Premier League (EPL). We adapt the TOpic modeling Based Index Assessment through Sentiment (TOBIAS) framework, originally developed for educational settings, to the domain of sports analytics. This novel application faces difficulties in handling the volume and variability of social media data, as well as in accurately linking pre-match sentiments to post-match performance metrics. Our methodology integrates advanced Natural Language Processing (NLP) techniques, including sentiment analysis and topic modeling, with Partial Least Squares Path Modeling (PLS-PM). We analyze a dataset of 167,841 tweets related to 512 English Premier League (EPL) players, collected from May 2022 to May 2023. The study is conducted in two phases: pre-match analysis to assess public expectations, and post-match analysis to evaluate reactions to player performances. Experimental analysis reveals significant correlations between pre-match sentiments and subsequent player performance, with negative sentiments showing a stronger predictive power than positive ones. Post-match, we observe a shift in the relationship between sentiments and performance metrics, indicating the public's responsiveness to match outcomes. Our findings contribute to the broader understanding of social media's role in sports performance and offer insights for potential applications in regulating online behaviors in sports contexts.

Keywords Multivariate analysis · Partial least squares · Complex analysis · Football players performance · TOBIAS

✉ Marco Ortu
marco.ortu@unica.it

Francesco Mola
mola@unica.it

[1] Dept. of Economics and Business Sciences, University of Cagliari, Viale Fra Ignazio 17, Cagliari 09123, Italy

Published online: 05 December 2024

1 Introduction

The application of novel methodologies in the field of performance assessment extends beyond the realm of academia, finding relevance in diverse contexts such as sports analytics. In this paper, we adapt the TOpic modeling Based Index Assessment through Sentiment (TOBIAS) methodology (Ortu et al. 2022), initially conceptualized for student quality assessment in educational settings, to analyze player performance in the English Premier League (EPL). The core objective of this study is to explore how public sentiment, as expressed in social media, correlates with both the topics discussed and the potential impact on player performance in the EPL. While previous studies have investigated sentiment's role in shaping public perception in various contexts (Du et al. 2023; Wunderlich and Memmert 2020), the link between sentiment and athletic performance remains an underexplored area. Some research suggests that social media discourse can influence player morale and public image(Smith and Sanderson 2015), but there is limited empirical evidence directly connecting sentiment to on-field performance. The adaptation of the TOBIAS model to analyze player performance in the English Premier League (EPL) based on social media sentiment not only contributes to the field of sports analytics but also holds significant implications for decision-makers concerned with regulating social media behaviors. By systematically quantifying the relationships between public sentiment expressed through social media and real-world outcomes like player performance, this model offers a structured approach to understanding the tangible impact of online discourse. This understanding is crucial for regulatory bodies and social media platforms as they navigate the complex landscape of online communication, where the line between benign speech and harmful rhetoric can often be blurred. The insights gained from this study can inform policy development and platform guidelines, aiming to foster a healthier digital environment. It underscores the potential of sophisticated analytical tools to guide decision-making processes in the digital age, highlighting the interplay between virtual interactions and their real-world consequences. Building upon the TOBIAS framework and adapting it to the dynamic world of sports and social media, this study aims to address the following key research questions:

RQ1: How does pre-game social media sentiment influence player performance in the English Premier League?

RQ2: To what extent does player performance impact post-game social media sentiment and topics?

RQ3: Can the adapted TOBIAS framework effectively model these bidirectional relationships in a sports context?

By answering these questions, we seek to demonstrate not only the versatility of the TOBIAS methodology but also to provide novel insights into the complex interplay between public sentiment and athletic performance in professional football. This study makes several contributions to the fields of sports analytics, social media analysis, and sentiment-based performance prediction. We present a novel application of the TOBIAS (TOpic modeling Based Index Assessment through Sentiment) framework to professional sports, demonstrating its versatility beyond its original educational context. Our temporal analysis, encompassing both pre-match and post-match sentiment, provides potential insights into the evolving relationship between public opinion and player performance. In light of our findings, which highlight stronger correlations between sentiment and topics rather than performance, this study emphasizes the role of sentiment in shaping the public discourse around players. Our results contribute to the literature by shedding light on how sentiment may reflect public expectations and reactions rather than directly predicting performance outcomes. This study consists of four sections besides the introduction. Section 2 gives the context and theoretical background and relevant literature. Section 3 presents the theoretical foundation of TOBIAS methodology. Section 4 shows the application of our framework to EPL. Finally, Sect. 5 ends the paper with some concluding remarks.

2 Background and related works

The application of Natural Language Processing (NLP) techniques in sports analytics has gained significant traction in recent years. These advanced computational methods offer innovative approaches to analyze vast amounts of textual data generated in the sports domain, particularly from social media platforms. Du et al. (2023) highlight the potential of machine learning techniques in understanding sport consumer behavior, emphasizing the rich insights that can be extracted from unstructured data sources. In a similar vein, Wunderlich and Memmert (2020) demonstrate the effectiveness of lexicon-based sentiment analysis in decoding sports-related Twitter communication, showcasing how NLP can provide valuable insights into fan engagement and public perception of sports events. These studies underscore the growing importance of NLP in bridging the gap between qualitative fan expressions and quantitative sports analytics, offering a more comprehensive understanding of the sports ecosystem.

In the following subsection we provide relevant literature and background about Sentiment Analysis and Topic Modeling

2.1 Sentiment analysis

Sentiment Analysis (SA) is characterized as an automated process for extracting and processing textual data to discern sentiment information embedded in opinions (Rahmadan et al. 2020). Its primary objective is to analyze, investigate, and distill subjective opinions and sentiments of humans, quantifying affective states and subjective content in text form.

Specifically, SA focuses on three key aspects: *Subjectivity/Objectivity*, which involves distinguishing subjective from objective text; *Polarity*, which assigns either a qualitative (positive/negative) or quantitative (numerical value within a range) sentiment rating to text; and *Discrete Emotions*, a finer-grained analysis aimed at extracting specific emotions like *joy* and *love* from textual content.

SA is applicable at various textual levels including document, sentence, and aspect, and employs diverse methodologies (Kaur et al. 2017; Liu 2012). One such method is the lexicon-based approach, subdivided into *Dictionary based*: using a linguist-constructed dictionary assigning sentiment scores to terms-and *Corpus based*, which depends on text corpora's co-occurrence statistics or syntactic patterns and predefined sentiment-laden seed words (Darwich et al. 2019). Another method is machine learning-based, categorized into three primary types (Madhoushi et al. 2015): supervised, unsupervised, and semi-supervised learning. *Supervised learning*, known for its efficacy in traditional document classification, has been successfully adapted for sentiment analysis (Sodanil 2016). *Unsupervised learning* approaches, such as *rule-based* classifiers, operate without pre-labeled training data to ascertain sentiment polarity (Vashishtha and Susan 2019; Hu et al. 2013). *Semi-supervised learning* leverages both labeled and unlabeled data. While unlabeled data does not provide class-specific information, it offers insights into the joint distribution over classification features, a fundamental concept in semi-supervised learning (Liu 2012).

SA's application spans numerous domains, becoming increasingly vital for extracting information critical to decision-making support, business applications, and predictions and trend analysis.

2.2 Topic modelling

Topic Modeling (TM) is a technique for extracting concise, multi-dimensional summaries from documents (Blei et al. 2003). It employs statistical methods to analyze text words, uncovering latent themes and their interrelations (Rahmadan et al. 2020). TM operates as an unsupervised learning technique, eliminating the need for document labeling (Blei 2012). Various TM approaches exist, with Latent Dirichlet Allocation (LDA) being particularly prominent due to its efficacy in handling large text document collections (Campbell et al. 2015). LDA is a generative probabilistic model, introduced by Blei et al. (2003), and has been applied in diverse contexts, including conference program organization (Frigau et al. 2021) and computational advertising (Soriano et al. 2013). It conceptualizes documents as random mixtures of latent topics, each characterized by a word distribution (Jelodar et al. 2019). LDA considers three elements: words, documents, and the corpus. However, while LDA provides useful insights into latent themes within large datasets, it often struggles with maintaining contextual coherence, particularly in dynamic datasets like social media posts. To overcome these limitations, this study employs BERTopic (Grootendorst 2022), a context-aware topic modeling approach. BERTopic leverages Bidirectional Encoder Representations from Transformers (BERT) to generate dense, context-aware embeddings of documents. This allows the model to better understand word meanings in their specific contexts, which is essential when analyzing social media data, where the same word can have different meanings based on its usage. Unlike LDA, which relies on word co-occurrence patterns, BERTopic clusters these BERT embeddings to group documents into coherent topics. It employs hierarchical clustering to discover topics that are both contextually and semantically meaningful. This approach is particularly effective in capturing evolving conversations and aspect-specific topics within social media discussions, where context plays a critical role in the interpretation of content. Additionally, BERTopic allows for better modeling of the temporal dynamics in pre-match and post-match social media discussions about football players. By incorporating BERTopic, we benefit from the use of deep learning techniques that significantly improve the coherence and relevance of the topics identified. This shift allows for more accurate analysis of the relationship between public sentiment and player performance, as well as deeper insights into how specific events influence online discussions. The adaptability of BERTopic makes it more suitable for large and unstructured datasets like social media, offering enhanced performance over traditional models like LDA.

2.3 Social media and sport performance

Performance psychology and the impact of public opinion on athlete behavior have been studied in many directions. Several studies pose that social media, can induce both positive and negative psychological effects on individuals exposed to its content. Studies such as Oberst et al. (2017) and Vannucci et al. (2017) demonstrate that heavy social media engagement can lead to stress, anxiety, and a heightened fear of judgment, which are particularly relevant for public figures, including athletes. Additionally, the concept of “choking under pressure”, as explored by Beilock and Carr (2005) and Baumeister (1984), underscores how heightened self-awareness, often exacerbated by external evaluation, can impair performance in high-stakes situations. This is especially pertinent in sports, where athletes are subject to constant scrutiny from fans and media alike (Cotterill 2017). Moreover, the relationship between social media feedback and athlete well-being has been specifically documented in the sports domain. Smith and Sanderson (2015) found that athletes’ interactions on platforms like Instagram can significantly influence their self-presentation and perceived performance. Furthermore, Kerr and Stirling (2008) noted the broader psychological impact of public feedback, which can shape not only athletes’ mental health but also their performance outcomes. Therefore, it is plausible that the sentiments expressed on social media before matches could have a measurable impact on athletes’ performance during games, either by increasing pressure or altering their emotional state in ways that affect their ability to perform optimally.

3 Methodology

In conceptualizing the research questions, we build upon established findings in performance psychology and the impact of public opinion on athlete behavior discussed in details in Sect. 2.3. By drawing on this body of literature, we justify the hypothesis that public social media sentiment can influence football players' performance and vice versa, providing a theoretical foundation for the research questions addressed in this study. To address these questions we use the TOBIAS methodology, which integrates Natural Language Processing techniques such as sentiment analysis and topic modeling with PLS Path Modeling (PLS-PM) to create a model that is both interpretable and informative. The objective is to harness the insights embedded in textual data to deduce and elucidate assessments of performances, ultimately aiding decision-makers in enhancing public service quality. The algorithm comprises five principal steps:

1. Extraction of emotional features
2. Assignment of topics
3. Definition of overall quality/performance indices
4. Feature selection using Adaptive LASSO
5. Estimation of the PLS-PM model

3.1 Notation

Let us denote $\mathcal{D} = \{d_1, \dots, d_N\}$, as a collection of N documents, where each one is referring to a single tweet c posted by a user referred to a player p indicated by the pair element $(c, p) \in \mathcal{O}$. In our model, every document d is viewed as a set of unordered sentences, where $d = \{s_1, s_2, \dots\}$, each characterized by emotional features and a common abstract topic. The aggregation of sentences from all N documents is denoted by the set $\mathcal{S} = \bigcup_{i=1}^{N} \{s \mid s \in d_i\}$, and its cardinality, or the total number of elements, is $|\mathcal{S}| = n_{\mathcal{S}}$. According to the tweet and player to which the sentence is referred, the set $\mathcal{S}$ can be partitioned into $|\mathcal{O}|$ subsets $\mathcal{S}_i = \{\forall s \in \mathcal{S} : \gamma(s) = (p, c)_i\}$ with $\gamma : \mathcal{S} \rightarrow \mathcal{O}$ is the function that maps the sentences to the corresponding tweet and player (p, c) considered in them.

The initial step of the algorithm involves the extraction of emotional features. Consider a set of observable emotional features $\mathcal{F} \in [0, n_{\mathcal{S}}]$ that are quantifiable from text and mirror a latent variable $\xi_{\mathcal{A}}$ representing *Affectiveness*. These manifest variables can be categorized into K subsets, such as *Positiveness* and *Negativeness*. Correspondingly, several functions $g(\cdot)$ are defined to measure the emotional features associated with each emotional state. Considering a subblock of emotional features A_k and an its i manifest feature $\mathcal{F}_{ki}$, the function $g_{ki}(\cdot)$ is defined as $g_{ki} : \mathcal{S} \rightarrow [0, n_{\mathcal{S}}]$.

According to the emotional states, these manifest variables $\mathcal{F}$ can be grouped into K subblocks A_k.

In the second phase topics are assigned. Let us define $\mathcal{W}$ as the collection of all words used in all sentences, that is the Word Embeddings of $\mathcal{S}$, and $\Theta = \{\theta_1, \dots, \theta_H\}$ the collection of the topics occurred in $\mathcal{S}$. Topic modeling method $\phi(\cdot)$ maps the words embeddings to topics such that $\phi : \mathcal{W} \rightarrow \Theta$. For each topic θ_h detected a latent variable ξ_{θ_h}, which is expressed by a single manifest variable $\mathcal{F}_{\theta_h}$, is defined.

The third step consists in determining the manifest indexes expressing the latent overall performance or quality $\xi^\star$ (called *Performance*) as perceived by users on social media.

Next, the values referred to the same pairs (p, c) are aggregated modifying the number of the observations from $n_{\mathcal{S}}$ to the cardinality of $\mathcal{O}$. Next, these manifest variables are filtered using Adaptive LASSO (Zou 2006) to retain only features useful to explain performance variables. Finally, the fifth step concerns the setting of the PLS-PM the framework of the formal model and its operative version, that is the model that fits the data.

3.2 Formal model definition

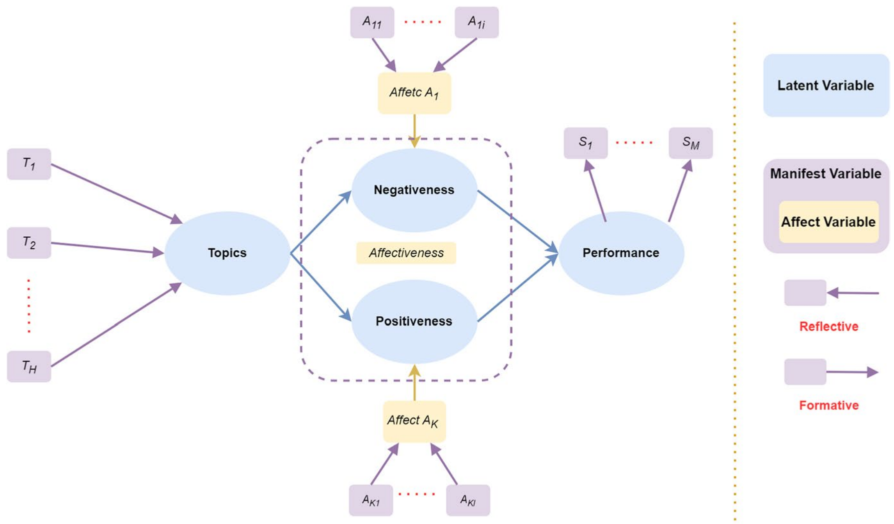
The TOBIAS' formal model is illustrated in Fig. 1. Figure 1a represents an adapted version of the framework to the pre-match tweets, namely the tweet collected the *week before* the match. Here, we hypothesize that public opinion discussed on social media influences players' performance, while Fig. 1b represents the post-match model, where we hypothesize that players' performance influences the public opinion expressed in tweets collected the *week after* the match. These two models are a modified version of the original TOBIAS model, where the *Affectiveness* block is split into two sub-blocks, namely *Positiveness* and *Negativeness*.

The inner model depicts the relation of *Performance* depending on the *Affectiveness*, both positive and negative, whilst the H latent variables ξ_{θ_h} representing *Topic*, are the main triggers of the *Affectiveness*. Regarding the outer model, the *Performance* block $\xi^{\star}$ is reflective of M manifest variables expressing the overall players' performance. The *Topic* blocks $\xi_{\theta_1}, \ldots, \xi_{\theta_H}$ are formative. Finally, *Affectiveness*, manifested through the sub-blocks *Positiveness* and *Negativeness*, comprises formative blocks. These blocks encompass all measurable indicators of emotional state from text, such as sentiment and emotion. The manifest variables within *Affectiveness* are organized into K homogeneous sub-blocks A, each corresponding to the emotional state they represent. Finally, the *Topics* block of latent variables are characterized by a single manifest variable. Considering $p(\theta_h|s)$ as the posterior probability of the topic θ_h in a document d provided by the topic modeling algorithm, the i-th values of the manifest variables of the topics are computed as follows:

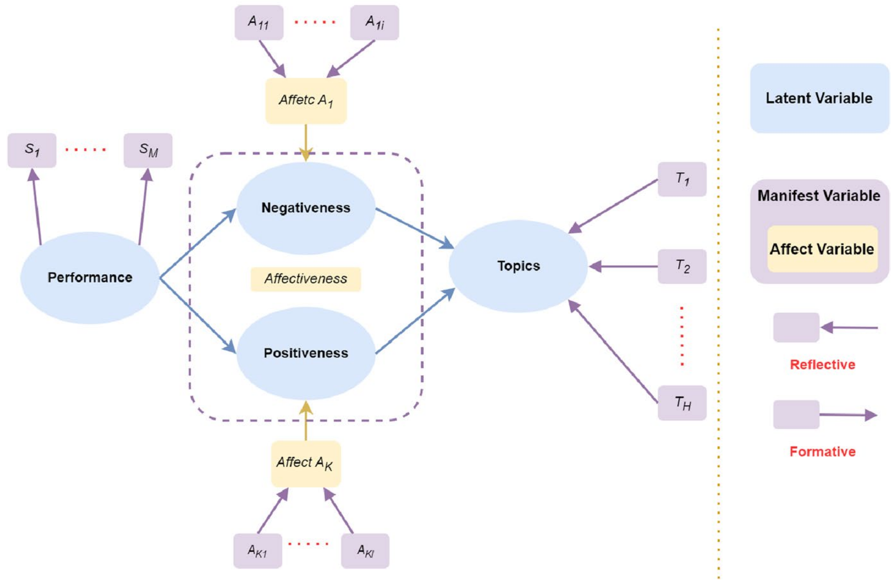
$$\mathcal{F}_{\theta_h,i} = \sum_{s\in\mathcal{S}_i} I\left(\operatorname*{argmax}_{j=1,\ldots,4} p(\theta_j|s) = h\right). \tag{1}$$

3.2.1 Pre-game and post-game influence modeling

To effectively capture the bidirectional relationship between social media sentiment and player performance in Fig. 1a, we have adapted the TOBIAS framework to accommodate both pre-game and post-game analyses. This approach allows us to model how pre-game sentiment influences performance and how performance, in turn, affects post-game sentiment and topics. For the pre-game model, we focus on tweets collected during the week leading up to each match. The process involves:



(a) Pre-Match model.



(b) Post-Match model

Fig. 1 Adapted TOBIAS framework. Blue nodes represent the latent variables, violet nodes represent manifest variables, *reflective* are drawn using an arrow from the latent variable to the manifest variable while *formative* manifest variables are drawn with an arrow from the manifest to the latent variable. Affective manifest variables are highlighted using yellow nodes, here the model is adapted to take into account positive and negative emotional state representation, namely *Positiveness* and *Negativeness* blocks. Each sub-block groups Affective reflective variables considering the dimension of the emotional state measured, such as sentiment, mood, emotions and so on. We consider H Topics, K affect measurements, and M performance measurements

- **Feature Selection**: We use Adaptive LASSO to select the most relevant pre-game sentiment and topic features that could potentially influence performance. This step helps in reducing noise and focusing on the most predictive variables.
- **Temporal Considerations**: We incorporate a time-decay factor to give more weight to sentiments expressed closer to the match day, reflecting the dynamic nature of public opinion.
- **Performance Prediction**: The selected features are then used in our PLS-PM model to predict player performance metrics (e.g., FotMob Rating, Goals, Assists).

The post-game model analyzes shown in Fig. 1b tweets collected in the week following each match. Key aspects include:

- **Performance Integration**: We incorporate actual performance metrics from the match as inputs to the model.
- **Sentiment and Topic Evolution**: We track how sentiments and topics evolve in response to the match outcome and individual player performances.
- **Feedback Loop Analysis**: By comparing pre-game predictions with post-game reactions, we can identify discrepancies between expectations and outcomes.

To provide deeper insights, we conduct a comparative analysis between the pre-game and post-game models: (i) Coefficient Comparison: We compare the coefficients of sentiment and topic variables between the two models to understand how their influence changes; (ii) Predictive Power Assessment: We evaluate and compare the predictive power of both models using metrics such as R-squared and Mean Absolute Error. (iii) Topic Shift Analysis: We analyze how the relevance and sentiment associated with different topics change from pre-game to post-game discussions. This bidirectional modeling approach allows us to capture the complex dynamics between public sentiment, player performance, and subsequent reactions, providing a more comprehensive understanding of the relationship between social media discourse and sports performance.

4 English premier league case study

In this study, we apply the adapted TOBIAS framework to the dynamic and high-stakes environment of the English Premier League (EPL), focusing on a case study that explores the relationship between player performance and social media opinion. Utilizing a data source previously derived from Twitter, now known as *X*, the model analyzes a dataset of public sentiments expressed about EPL players. This study serves as an investigation into how modern social media platforms, with their vast and rapid information exchange, influence and reflect the perceptions of sports performance.

4.1 Dataset description

Player performance metrics were sourced from the FotMob platform[1], a website frequented by over 14 million users. It provides real-time results and a comprehensive array of information on football teams and players, encompassing over 500 professional leagues across more than 100 nations. FotMob assigns a "FotMob Rating" to each player who participates in a match for a minimum of 10 min. This rating is derived from a specialized algorithm that incorporates over 300 individual statistics pertinent to the player's performance. The resulting score is confined within a numerical range from 0 to 10, precise to one decimal place. Data regarding public sentiment was extracted from *X* (formerly known as Twitter), utilizing player names as search keywords. The resultant dataset comprises 167,841 tweets, pertaining to 512 players from 20 Premier League football teams, collected over the period from May 11, 2022, to May 2, 2023.

4.2 Exploratory data analysis of features

We first performed an Exploratory Data Analysis on our features to construct a ground truth of the relationships among our features and players before and after the match. Table 1 summarises the manifest variables, namely the model's features, grouped by their block of latent variables in the PLS-PM.

Figs. 2 and 3 show the $\mathcal{F}_{\text{Sen}}$ (*sentiment*) and $\mathcal{F}_{\text{Emo}}$ (discrete *emotion*) evaluated by the RoBERTa model (Liu et al. 2019). We can see that the majority of tweets are neutral and similar numbers for positive and negative sentiment. For emotions, it can be noted that the most frequent emotion is Admiration followed by Amusement, in Fig. 3, Neutral emotion is removed to highlight the distribution that would be otherwise flattened as in this case too, Neutral emotion accounts for the majority with about 120K observations. Figures 4 and 5 show the Sentiment and Emotion distributions before and after the match. In Fig. 4 it can be noted that the proportion of positive tweets before the match is greater than negative while after the match the two proportions are similar. We tested the hypothesis that positive tweets decrease after the match using a proportion Z-test and we rejected the null hypothesis with a p-value of 2.62×10^{-8}, concluding that the difference is statistically significant. Figure 5 shows the proportion of emotions before and after the match. We can see that emotions such as Admiration, Disapproval, and Sadness tend to increase after the match while emotions such as Approval and Optimism tend to decrease after the match, other emotions related to the entertainment such as Amusement, Annoyance, and Excitement tend to remain at the same level before and after the match.

We then extracted the topics discussed before and after the match using BERTopic (Grootendorst 2022). After a refinement of the number of topics, we end up with 9 topics. Figure 6 shows the top five keywords per topic, for example, topic 0 is

[1] https://www.fotmob.com/.

Table 1 PLS-PM variables summary. Manifest variables are aggregated by teacher ID according to the description

Latent	Manifest	Description
Performance	*Rating*	Average player's score
	Minutes	Average number of minutes played per match
	Goal	Average number of goal scored per match
	Assist	Average number of assist per match
Positiveness	*Positive Sentiment*	Average number of tweet expressing positive sentiment
	Joy	Average number of tweet expressing Joy
	Love	Average number of tweet expressing Love
	Admiration	Average number of tweet expressing Admiration
	Amusement	Average number of tweet expressing Amusement
	Approval	Average number of tweet expressing Approval
	Caring	Average number of tweet expressing Caring
	Curiosity	Average number of tweet expressing Curiosity
	Desire	Average number of tweet expressing Desire
	Excitement	Average number of tweet expressing Excitement
	Gratitude	Average number of tweet expressing Gratitude
	Optimism	Average number of tweet expressing Optimism
	Pride	Average number of tweet expressing Pride
	Realization	Average number of tweet expressing Realization
	Relief	Average number of tweet expressing Relief
Negativeness	*Negative Sentiment*	Average number of tweet expressing negative sentiment
	Joy	Average number of tweet expressing Joy
	Anger	Average number of tweet expressing Anger
	Annoyance	Average number of tweet expressing Annoyance
	Confusion	Average number of tweet expressing Confusion
	Disapproval	Average number of tweet expressing Disapproval
	Disgust	Average number of tweet expressing Disgust
	Embarrassment	Average number of tweet expressing Embarrassment
	Fear	Average number of tweet expressing Fear
	Nervousness	Average number of tweet expressing Nervousness
	Sadness	Average number of tweet expressing Sadness
	Surprise	Average number of tweet expressing Surprise
	Disappointment	Average number of tweet expressing Disappointment
	Remorse	Average number of tweet expressing Remorse

Table 1 (continued)

Latent	Manifest	Description
Topics	T_0	Probability of tweet concerning T_0
	T_1	Probability of tweet concerning T_1
	T_2	Probability of tweet concerning T_2
	T_3	Probability of tweet concerning T_3
	T_4	Probability of tweet concerning T_4
	T_5	Probability of tweet concerning T_5
	T_6	Probability of tweet concerning T_6
	T_7	Probability of tweet concerning T_7
	T_8	Probability of tweet concerning T_8

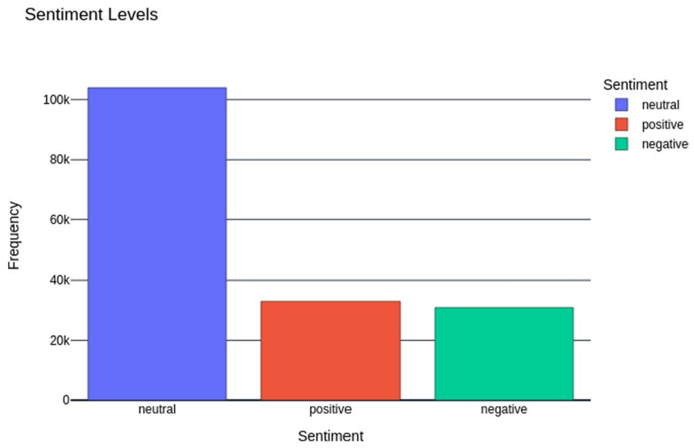


Fig. 2 Sentiment evaluated by RoBERTa model considering before and after the match

related to the performance of players during the season while other topics such as 4 and 6 are more related to particular players.

4.3 Results and discussion

In the first phase of our methodology, we applied a Multivariate Adaptive LASSO the reduce the number of manifest variable of the PLS model. Tables 2 and 3 show the results. We can se that **Topic 5** is not relevant in the pre-match model. This topic might indicate a metaphorical or literal application of gamification concepts (such as earning badges and leveling up, which are common in video games and apps like Untappd) to the experience of following or discussing the English Premier League on social media. Before a match, the discussion on social media is likely focused on

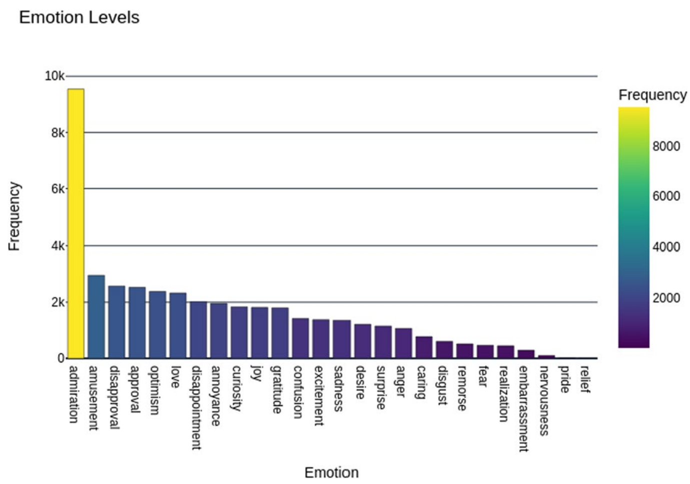


Fig. 3 Emotions spectrum

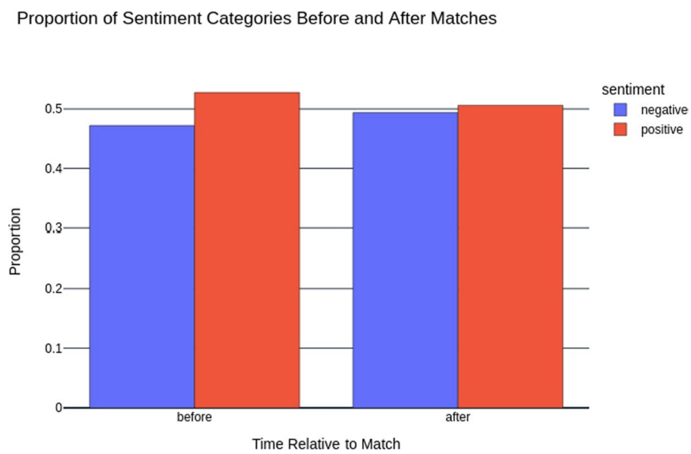


Fig. 4 Sentiment evaluated by RoBERTa model considering before and after the match

predictions, team strategies, player line-ups, and other aspects directly related to the upcoming game. These discussions are more tactical and anticipatory in nature. In contrast, post-match discussions shift towards reflection, celebration, and analysis of what occurred during the match. The emergence of this topic as significant in the post-match analysis indicates that social media users engage in celebrating achievements. Similar considerations can be made for **Topic 0**, **Topic 1** and **Topic 6** which are the only relevant topics in the post-match model.

By fitting TOBIAS model the results shown in Figs. 7a (the pre-match model) and b (the post-match model) are obtained. In the pre-match model, the latent construct 'Negativeness' is positively associated with *Topics* (0.362), suggesting that negative sentiments in tweets have a strong and direct relationship with the topics

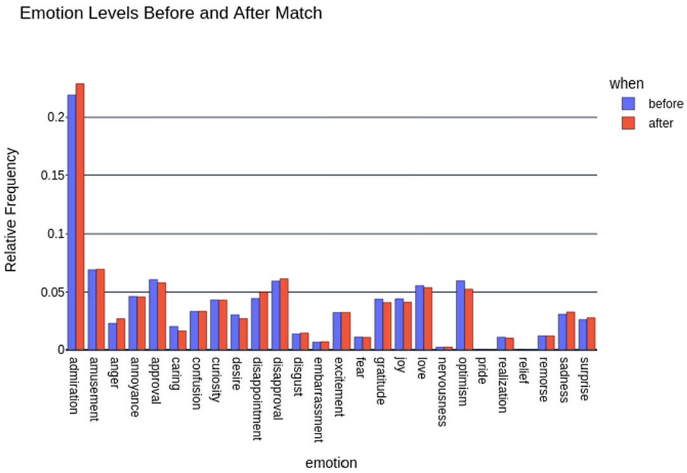


Fig. 5 Emotions evaluated by RoBERTa model considering before and after the match

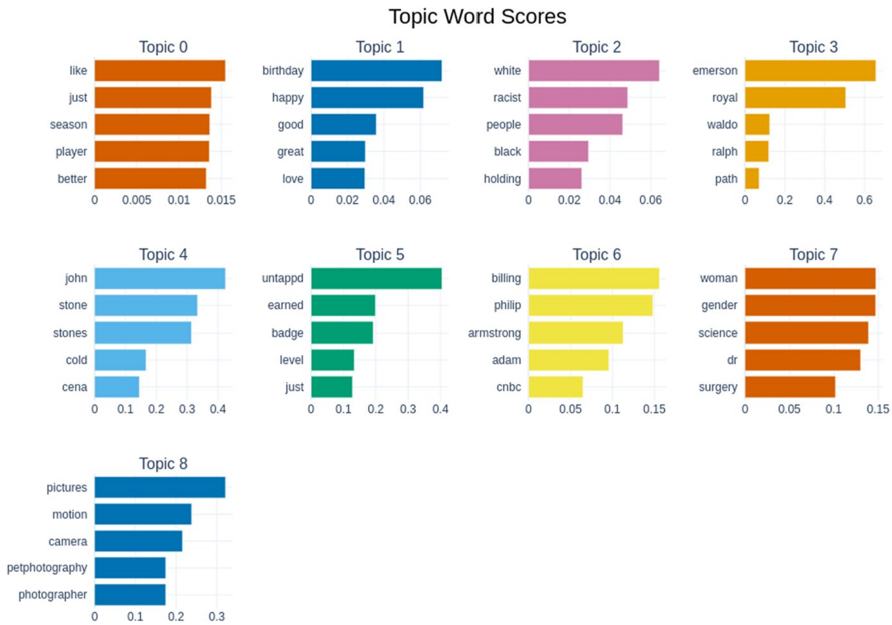


Fig. 6 Topics probabilities distributions

discussed before the match. This could indicate that negative sentiments are more influential or more likely to be expressed in relation to certain topics. Interestingly, *Negativeness* has a negative impact on *Performance* (−0.125), which might imply that a higher expression of negative sentiment could be associated with a lower subsequent performance, although the effect size is modest. *Positiveness* also has

Table 2 Adaptive LASSO results for Pre-Match model

Variable	Rating	Minutes	Goal	Assists
Topic 0	0.03109	1.33138	9.2000e−03	0.00889
Topic 1	−0.04836	1.01573	−1.3000e−02	0.00204
Topic 2	0.07871	1.76712	2.2900e−02	−0.01954
Topic 3	−0.03977	0.38011	−3.7600e−03	0.00696
Topic 4	0.06714	2.01475	7.7500e−02	0.03701
Topic 5	–	–	–	–
Topic 6	−0.15522	−3.87846	6.1400e−03	−0.00895
Topic 7	0.08157	−14.47262	1.5900e−01	−0.06795
Topic 8	−0.09813	3.92231	−7.1400e−02	0.00138
Negative Sentiment	−0.00759	−0.11204	−4.1400e−03	0.00045
Neutral Sentiment	–	–	–	–
Positive Sentiment	−0.00266	0.92537	1.3400e−02	0.00176
Admiration	0.00341	0.25138	−1.4500e−03	0.00113
Amusement	–	–	–	–
Anger	–	–	–	–
Annoyance	–	–	–	–
Approval	0.01115	0.56449	−8.9600e−04	−0.00229
Caring	−0.07755	−2.22999	−1.6400e−02	−0.00999
Confusion	0.02123	1.37544	−1.8500e−02	−0.00206
Curiosity	–	–	–	–
Desire	0.03046	2.96043	4.1800e−03	0.00994
Disappointment	–	–	–	–
Disapproval	0.00854	0.30763	−3.8500e−03	0.00370
Disgust	–	–	–	–
Embarrassment	−0.04162	−0.62953	2.6600e−04	−0.00844
Excitement	–	–	–	–
Fear	–	–	–	–
Gratitude	−0.02503	1.95572	−6.8600e−03	−0.00434
Joy	–	–	–	–
Love	0.00306	−0.14722	5.3800e−05	0.00325
Nervousness	0.01661	2.45933	1.0100e−02	0.01503
Neutral Emotion	–	–	–	–
Optimism	–	–	–	–
Pride	–	–	–	–
Realization	−0.00348	−0.08399	−2.7100e−04	0.00193
Relief	0.03951	0.89516	−6.8300e−03	−0.00408
Remorse	0.02882	−1.72515	5.8600e−03	0.02838
Sadness	−0.01163	−2.27207	3.5500e−03	−0.00286
Surprise	−0.03561	−1.00450	−5.6700e−03	0.00848

Table 3 Adaptive LASSO results for Post-Match model

Variable	Rating	Minutes	Goal	Assist
Topic 0	0.03273	1.23132	1.6195e−02	0.00303
Topic 1	−0.00835	−0.10682	−5.1510e−03	−0.00171
Topic 2	–	–	–	–
Topic 3	–	–	–	–
Topic 4	–	–	–	–
Topic 5	0.05430	−0.49502	1.5486e−02	0.00928
Topic 6	0.00056	−0.05931	−7.8620e−05	0.00036
Topic 7	–	–	–	–
Topic 8	–	–	–	–
Negative	−0.03123	−0.14189	−9.1626e−03	−0.00405
Neutral Sentiment	–	–	–	–
Positive	0.09569	1.09038	2.5412e−02	0.01500
Admiration	0.04865	0.82450	−2.5440e−03	0.00565
Amusement	–	–	–	–
Anger	–	–	–	–
Annoyance	–	–	–	–
Approval	0.01410	1.90460	−2.2654e−02	−0.01390
Caring	−0.08795	−3.72743	−2.3090e−02	−0.02222
Confusion	–	–	–	–
Curiosity	–	–	–	–
Desire	0.00190	0.10968	−2.2133e−03	0.00170
Disappointment	–	–	–	–
Disapproval	0.00559	0.46162	−3.8440e−03	0.00040
Disgust	–	–	–	–
Embarrassment	–	–	–	–
Excitement	−0.08328	−2.06340	−1.0483e−02	−0.03149
Fear	–	–	–	–
Gratitude	–	–	–	–
Joy	–	–	–	–
Love	–	–	–	–
Nervousness	–	–	–	–
Neutral Emotion	–	–	–	–
Optimism	−0.07802	−2.55529	−1.4562e−02	−0.01391
Pride	−0.04457	−1.10662	−1.3173e−02	−0.00820
Realization	–	–	–	–
Relief	–	–	–	–
Remorse	–	–	–	–
Sadness	–	–	–	–
Surprise	–	–	–	–

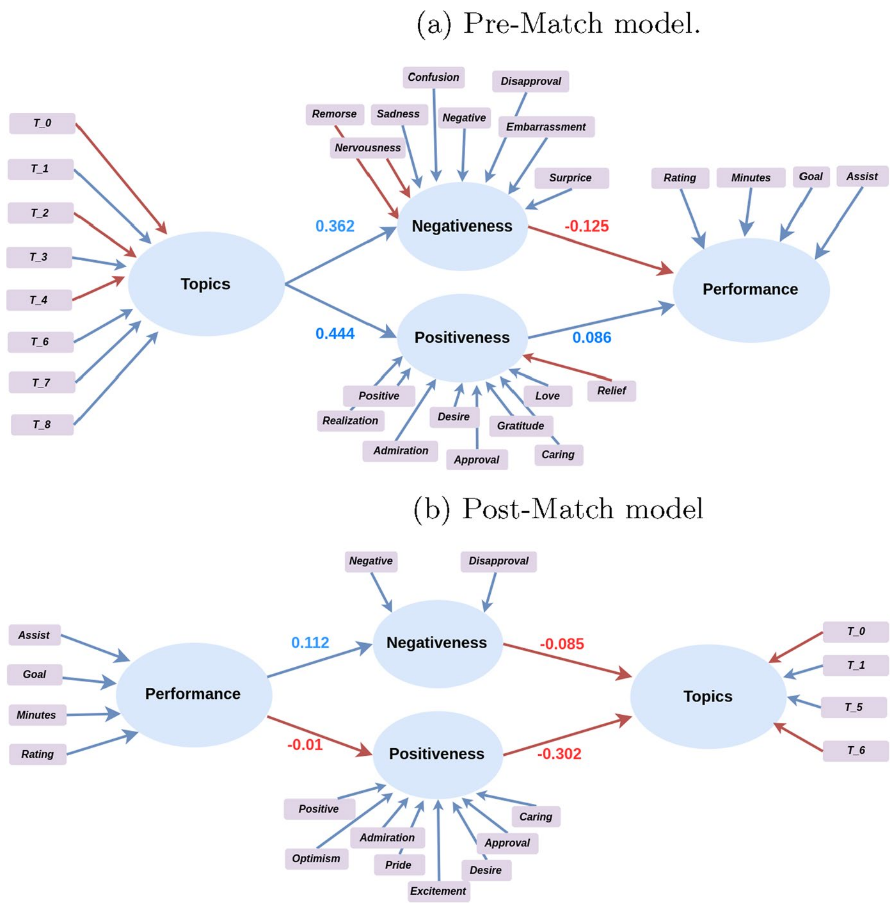


Fig. 7 Path Model definition. Blue nodes represent the latent variables, violet nodes represent manifest variables, *reflective* are drawn using an arrow from the latent variable to the manifest variable. Affective manifest variables are highlighted using yellow nodes. *Formative* manifest variable are drawn with an arrow from the manifest to latent variable

a strong positive relationship with *Topics* (0.444) in the pre-match model, which could suggest that positive sentiments are prevalent in discussions surrounding the match. However, the impact of *Positiveness* on *Performance* is positive but negligible (0.086), indicating that positive sentiments may not have a significant predictive power on performance outcomes.

We can now answer the first research question: *How does pre-game social media sentiment influence player performance in the English Premier League?*

While our initial hypothesis posited that social media sentiment would influence player performance, the results indicate only a modest relationship between these variables. Pre-game negative sentiments show a weak inverse relationship with performance (estimate −0.14, *p*-value 0.00), and positive sentiment demonstrates limited predictive power (estimate 0.08, *p*-value 0.08). More prominently, our analysis found stronger correlations between sentiment and the topics discussed in social media posts. This suggests that social media discourse is more reflective of public expectations and reactions to player actions than a direct influencer of performance. Future research may benefit from focusing on how social media amplifies or moderates public perceptions of sports performance through topics and narratives. Post-match, the relationships change significantly. *Negativeness* has a smaller positive association with 'Performance' (0.112), suggesting that negative sentiments post-match might reflect the performance observed, although the relationship is weak. Conversely, *Positiveness* shows a negative relationship with *Performance* (−0.01), which is counter-intuitive and may require further investigation to understand whether this reflects a reaction to unexpected performance outcomes or other factors. The most evident difference post-match is the negative association between *Positiveness* and *Topics* (−0.302). This substantial change might indicate that post-match, positive sentiments are less likely to be associated with the topics being discussed, possibly due to the outcome of the match not meeting the pre-match expectations. We can answer the second research question: *To what extent does player performance impact post-game social media sentiment and topics?*

Post-match, performance metrics show stronger correlations with ratings, indicating a closer alignment between actual outcomes and public perception. Positive sentiments become less associated with discussion topics (correlation −0.302), suggesting a potential disconnect between performance outcomes and positive fan reactions. Several general consideration can be made from these findings. Sentiment polarity has a stronger relationship with the topics of discussion pre-match than with the actual performance outcomes. Post-match, the positive sentiments seem to diverge from the topics discussed, which may suggest a dissonance between expectations and outcomes. The predictive power of positive sentiments on performance is minimal, indicating other factors may be more critical in influencing match outcomes. These findings have implications for understanding the dynamics of fan engagement and sentiment on social media platforms in relation to sports events. They also raise questions about the potential influence of public sentiment on the performance of athletes or teams, a topic that warrants further exploration. Additionally, the models suggest that post-match sentiments and topics are more closely aligned with the actual performance, which may provide valuable insights for teams and managers in managing public relations and expectations.

Table 4 reports the cross-loadings of the pre-match model. The cross-loadings indicate that while the *Positiveness* and *Negativeness* constructs have several indicators with moderate to strong correlations, the *Performance* construct is less influenced by the sentiment indicators, as seen by the generally lower correlation values. This could imply that sentiment expressed in pre-match tweets has limited direct impact on actual performance outcomes, although some aspects such as *Rating* and *Assist* are more closely related. These results underscore the complexity of the relationship between public sentiment and sports performance and highlight the multifaceted nature of sentiment expression in the context of sports events.

Table 5 shows the cross-loadings for the post-match model. There is a significance shift in the relationship between sentiment, topics, and performance when compared to the pre-match scenario. Post-match, performance metrics such as *Minutes*,

Table 4 Cross-loadings of latent variables and observed indicators for the pre-match model

Variable	Block	Topics	Positive	Negative	Performance
Topic 0	Topics	−0.58	−0.28	−0.19	0.03
Topic 1	Topics	0.91	0.43	0.32	0.04
Topic 2	Topics	−0.03	−0.02	−0.01	0.01
Topic 3	Topics	0.66	0.29	0.24	0.05
Topic 4	Topics	−0.01	−0.01	−0.00	0.11
Topic 5	Topics	−0.07	−0.01	−0.05	0.06
Topic 6	Topics	0.33	0.09	0.18	−0.00
Topic 7	Topics	0.07	0.03	0.02	−0.01
Topic 8	Topics	0.47	0.23	0.15	−0.03
Positive	Positive	0.25	0.67	−0.21	0.12
Admiration	Positive	0.10	0.41	−0.08	0.12
Amusement	Positive	0.11	0.13	0.15	−0.03
Approval	Positive	0.21	0.38	0.09	−0.00
Caring	Positive	0.36	0.65	0.06	−0.01
Curiosity	Positive	0.02	0.08	0.14	0.03
Desire	Positive	0.03	0.12	−0.01	0.04
Excitement	Positive	0.17	0.36	−0.05	0.02
Gratitude	Positive	0.26	0.52	−0.00	0.02
Joy	Positive	0.22	0.34	−0.06	−0.04
Love	Positive	0.11	0.19	−0.09	−0.01
Optimism	Positive	0.03	−0.02	−0.03	−0.04
Pride	Positive	0.01	0.18	−0.02	0.09
Negative	Negative	0.31	0.04	0.84	−0.10
Anger	Negative	0.15	−0.01	0.36	−0.03
Annoyance	Negative	0.09	0.01	0.24	−0.03
Confusion	Negative	0.05	−0.02	0.28	−0.09
Disapproval	Negative	0.11	0.06	0.27	−0.02
Disgust	Negative	0.08	0.00	0.30	−0.06
Embarrassment	Negative	0.08	0.08	0.20	−0.02
Fear	Negative	0.01	0.00	−0.08	0.05
Nervousness	Negative	−0.01	−0.02	−0.09	0.04
Sadness	Negative	0.13	0.05	0.34	−0.04
Surprise	Negative	−0.00	0.00	0.01	−0.00
Rating	Performance	−0.01	0.02	−0.07	0.68
Minutes	Performance	0.07	0.06	0.05	0.16
Goal	Performance	−0.07	0.04	−0.09	0.61
Assist	Performance	0.06	0.09	−0.12	0.83

Goal, and *Assist* show stronger correlations with *Rating*, indicating a closer alignment between actual outcomes and their evaluations. Positive sentiments post-match are more closely correlated with expressions of *Caring* and *Admiration*, possibly

Table 5 Cross-loadings of latent variables and observed indicators for the post-match model

Variable	Block	Rating	Positiveness	Negativeness	Topics
Rating	Rating	0.66	0.02	−0.05	−0.00
Minutes	Rating	0.26	0.05	0.01	−0.06
Goal	Rating	0.40	0.02	−0.03	0.06
Assist	Rating	0.91	0.12	−0.10	−0.07
Positive	Positive	0.13	0.74	−0.21	−0.25
Admiration	Positive	0.15	0.47	−0.04	−0.10
Approval	Positive	−0.01	0.33	0.12	−0.18
Caring	Positive	0.00	0.69	0.06	−0.35
Desire	Positive	0.04	0.12	−0.02	−0.02
Excitement	Positive	0.03	0.43	−0.02	−0.19
Optimism	Positive	−0.03	0.02	−0.05	−0.04
Pride	Positive	0.07	0.16	−0.00	−0.01
Negative	Negative	−0.10	−0.04	1.00	−0.28
Nisapproval	Negative	−0.02	0.05	0.26	−0.08
Topic 0	Topics	0.02	−0.24	−0.14	0.55
Topic 1	Topics	0.07	0.39	0.28	−0.97
Topic 5	Topics	0.05	−0.02	−0.05	0.11
Topic 6	Topics	−0.00	0.07	0.14	−0.30

Table 6 Structural assessment of TOBIAS model. Results of the regressions of *Topics*, *Positiveness*, *Negativeness* *Performance* Blocks. The statistical significance regression of blocks are indicated as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

	Estimate	Std. Error	t value	Pr(> $\|\|t\|\|$)
Positiveness				
Intercept	−0.00	0.04	−0.00	1.00
Topics	0.46	0.04	11.60	0.00
Negativeness				
Intercept	0.00	0.04	0.00	1.00
Topics	0.35	0.04	8.23	0.00
Performance				
Intercept	0.00	0.04	0.00	1.00
Positive	0.08	0.04	1.76	0.08
Negative	−0.14	0.04	−3.09	0.00

reflecting reactions to the match's results. *Topic 1* stands out with an even stronger negative correlation with *Topics* post-match, suggesting a concentrated focus on specific aspects of the match. The clear separation between Po*Positiveness* and *Negativeness* sentiments persists in the post-match analysis, with 'Negativeness' maintaining a high self-loading, underscoring its distinct nature. These dynamics suggest a transition from anticipatory discussions pre-match to more outcome-oriented conversations post-match, with performance metrics becoming more predictive of post-match ratings. Tables 6 and 7 report the PLS model evaluation of the TOBIAS.

Table 6 presents regression estimates, standard errors, t-values, and *p*-values for the constructs *Positiveness*, *Negativeness*, and *Performance* within a PLS-PM framework. For *Positiveness*, the estimate for *Topics* is 0.46 with a standard error of 0.04, yielding a t-value of 11.60. The associated *p*-value is 0.00, indicating a statistically significant positive relationship between the topics discussed and positive sentiment. *Negativeness* has a similar structure, with the estimate for *Topics* being 0.35 and a standard error of 0.04. The t-value here is 8.23, and the corresponding *p*-value is again 0.00, showing a significant positive relationship, though slightly weaker than that for *Positiveness*. In the *Performance* construct, *Positiveness* has an estimate of 0.08 with a t-value of 1.76, and a *p*-value of 0.08, which is marginally above the conventional threshold for significance, suggesting a positive but not statistically significant relationship with performance. In contrast, *Negative* has an estimate of −0.14 with a t-value of −3.09 and a *p*-value of 0.00, indicating a significant negative relationship with performance. The intercepts for all constructs are effectively zero, with corresponding *p*-values of 1.00, indicating that when the predictors are at their reference level, the constructs do not differ from their mean value. This is typical in PLS-PM where the primary interest lies in the relationships between constructs rather than the absolute values. In assessing the quality of the model, the significant *p*-values for the relationships between *Topics* and both *Positiveness* and *Negativeness* suggest that these constructs are well-explained by the topics discussed in the comments. However, for *Performance*, only the *Negative* sentiment shows a significant relationship, hinting at the potential greater impact of negative sentiments on performance outcomes. The *p*-value slightly higher than 0.05 for *Positiveness* suggests its relationship with *Performance* may be subject to other moderating factors not captured in this model. Overall, the model seems to have predictive relevance for *Negativeness* and *Positiveness* in relation to *Topics*, and *Negativeness* in relation to *Performance*.

Table 6 presents the structural assessment of the model, detailing the results of the regression analyses for each block of latent variables in the inner model. In addition to the regression outcomes, the quality of the structural model is assessed by examining three metrics: the determination coefficients R^2, block Communality, and Mean Redundancy, with Table 7 displaying these metrics for all latent variables. The R^2 value indicates the proportion of variance in an endogenous latent variable explained by its independent latent variables. For instance, the TOBIAS model

Table 7 Summary of the Inner Model. The coefficient R^2 quantifies the proportion of variance in an endogenous latent variable accounted for by its independent latent variables. *Communalities*, represented as squared loadings, quantify the shared variance between a latent variable and its indicators. *Redundancy* measures the amount of variance in an endogenous construct that is explained by its

	Type	R2	Communality	Redundancy
Topics	Exogenous	0.00	0.22	0.00
Positive	Endogenous	0.22	0.14	0.03
Negative	Endogenous	0.12	0.13	0.02
Performance	Endogenous	0.03	0.39	0.01

accounts for 34% of the total variance of *Affectiveness* using the *Topics*, while only 3% of the variance in *Satisfaction* is explained by *Affectiveness*. Communality measures the extent to which a block is represented by its indicators (manifest variables). *Performance* shows the highest communality at 39%, indicating that the latent variable is well represented by its indicators. The redundancy index quantifies the extent to which a set of independent latent variables accounts for variance in a dependent latent variable. In this case study, the redundancy index supports the interpretations derived from the R^2 values. With these finding we can answer the third research question: *Can the adapted TOBIAS framework effectively model these bidirectional relationships in a sports context?*

Yes, the adapted TOBIAS framework effectively models these relationships. It accounts for 34% of the total variance in Affectiveness using Topics, and while only explaining 3% of Performance variance, it captures significant relationships between sentiment, topics, and performance metrics.

5 Conclusions

This research presents an innovative approach to modeling the influence of expressed topics, moods, and sentiments on the overall performance rating of a subject, leveraging the TOpic Based Index Assessment through Sentiment (TOBIAS) framework. TOBIAS synthesizes various natural language processing techniques with statistical methodologies to delineate the effects of sentiment analysis and topic modeling on performance ratings via Partial Least Square with Path Modelling. This work focuses on three main research questions:

RQ1: How does pre-game social media sentiment influence player performance in the English Premier League?

RQ2: To what extent does player performance impact post-game social media sentiment and topics?

RQ3: Can the adapted TOBIAS framework effectively model these bidirectional relationships in a sports context?

Our study, applying the adapted TOBIAS framework to English Premier League data, yielded several key insights. Pre-game social media sentiment demonstrated only a modest predictive relationship with player performance, particularly in the case of negative sentiments. However, the more significant finding of this study lies in the relationship between sentiment and the topics discussed in social media posts, suggesting that public discourse around player performance is shaped more by broader social media narratives than direct sentiment-driven performance effects. Future work should focus on understanding how sentiment interacts with public discussions and shapes broader perceptions in the sports ecosystem. Post-match analysis revealed a stronger alignment between performance metrics and public perception, though positive sentiments became less associated with discussion topics. This suggests a complex relationship between fan expectations, player performance, and subsequent social media reactions. The TOBIAS framework proved effective in modeling these bidirectional relationships, accounting for a substantial portion of sentiment variance and capturing significant links between sentiment, topics, and performance metrics. These findings not only demonstrate the adaptability of TOBIAS to sports analytics but also provide valuable insights into the intricate interplay between social media discourse and athletic performance in professional football. The TOBIAS model underscored that, while pre-match discussions were dominated by anticipatory sentiments, post-match discussions were more reflective of the actual outcomes. For instance, the topic of 'Assist' was a strong predictor of performance, underscoring the applicability of TOBIAS in capturing the nuance of sentiment's impact on performance ratings. The implications of the TOBIAS framework extend beyond sports analytics, offering a robust tool for understanding a broad range of social phenomena where public sentiment is both influential and indicative. By capturing the nuanced interplay between sentiment and tangible outcomes, TOBIAS provides stakeholders with actionable insights into the collective psyche of communities engaged in discourse around events, policies, and social movements. For rule makers, such as policy developers, organizational leaders, or legislative bodies, the application of TOBIAS can be pivotal in gauging public reaction to decisions, anticipating the reception of new regulations, and understanding the emotional and topical undercurrents within public dialogue. By understanding the sentiment trends, rule makers can tailor their communication strategies, policy formulations, and interventions to align with or address the public's expectations and concerns. Moreover, the insights garnered from the TOBIAS analysis can lead to proactive measures in managing public relations, shaping organizational strategies, and crafting legislation that resonates with the emotional and rational tenor of the populace. The study acknowledges limitations, such as the correlations between the dimensions of sentiment, which may affect the model's results. Future research will aim to refine these aspects of the model to enhance its predictive capability.

Declarations

Conflict of interest/Conflict of interest The authors declare no Conflict of interest.

References

Baumeister RF (1984) Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. J Person Soc Psychol 46(3):610

Beilock SL, Carr TH (2005) When high-powered people fail: Working memory and "choking under pressure'' in math. Psycholo Sci 16(2):101–105

Blei DM (2012) Probabilistic topic models. Commun ACM 55(4):77–84

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Campbell JC, Hindle A, Stroulia E (2015) Latent dirichlet allocation: extracting topics from software engineering data. In: The art and science of analyzing software data. Elsevier, p 139–159

Cotterill S (2017) Team psychology in sports: Theory and practice. Routledge
Darwich M, Mohd SA, Omar N et al (2019) Corpus-based techniques for sentiment lexicon generation: A review. J Digit Inf Manag 17(5):296
Du James, Pifer Nathan David, Gulavani Susmit S. (2023) Harnessing the Power of Machine Learning in Sport Consumer Behavior Research. In: Leng Ho Keat, Zhang James J. (eds) Sports Sponsorship and Branding: Global Perspectives and Emerging Trends. Routledge, London, pp 165–183. https://doi.org/10.4324/9781003458890-12
Frigau L, Wu Q, Banks D (2021) Optimizing the jsm program. J Am Stat Association 117(538):617–626
Grootendorst M (2022) Bertopic: Neural topic modeling with a class-based tf-idf procedure. https://doi.org/10.48550/arXiv.2203.05794,
Hu X, Tang J, Gao H, et al (2013) Unsupervised sentiment analysis with emotional signals. In: Proceedings of the 22nd international conference on World Wide Web, pp 607–618
Jelodar H, Wang Y, Yuan C et al (2019) Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. Multimed Tools Appl 78(11):15,169-15,211
Kaur H, Mangat V, et al (2017) A survey of sentiment analysis techniques. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), IEEE, pp 921–925
Kerr G, Stirling A (2008) Child protection in sport: Implications of an athlete-centered philosophy. Quest 60(2):307–323
Liu B (2012) Sentiment analysis and opinion mining. Synthesis Lectures Human Lang Tech 5(1):1–167
Liu Y, Ott M, Goyal N, et al (2019) Roberta: A robustly optimized bert pretraining approach. arXiv:1907.11692 [cs.CL]
Madhoushi Z, Hamdan AR, Zainudin S (2015) Sentiment analysis techniques in recent works. In: 2015 science and information conference (SAI), IEEE, pp 288–291
Oberst U, Wegmann E, Stodt B et al (2017) Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. J Adolescence 55:51–60
Ortu M, Frigau L, Contu G (2022) Topic based quality indexes assessment through sentiment. Comput Stat 39(1):289–311
Rahmadan MC, Hidayanto AN, Ekasari DS, et al (2020) Sentiment analysis and topic modelling using the lda method related to the flood disaster in jakarta on twitter. In: 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), IEEE, pp 126–130
Smith LR, Sanderson J (2015) I'm going to instagram it! an analysis of athlete self-presentation on instagram. J Broadcast Electron Med 59(2):342–358
Sodanil M (2016) Multi-language sentiment analysis for hotel reviews. In: MATEC Web of Conferences, EDP Sciences, p 03002
Soriano J, Au T, Banks D (2013) Text mining in computational advertising. Stat Anal Data Mining: The ASA Data Sci J 6(4):273–285
Vannucci A, Flannery KM, Ohannessian CM (2017) Social media use and anxiety in emerging adults. J Affect Disord 207:163–166
Vashishtha S, Susan S (2019) Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Syst Appl 138(112):834
Wunderlich F, Memmert D (2020) Innovative approaches in sports science-lexicon-based sentiment analysis as a tool to analyze sports-related twitter communication. Appl Sci 10(2):431
Zou H (2006) The adaptive lasso and its oracle properties. J Am Stat Associat 101(476):1418–1429. https://doi.org/10.1198/016214506000000735