SENTIMENT ANALYSIS OF SPORTS-RELATED CONTENT USING FINE-TUNED BERT MODEL: UNVEILING PUBLIC REACTIONS TO SPORTS EVENTS

Hongyi Duan
Data Science Master
Duke University
Durham, NC 27708, USA
h.duan@duke.edu

Mu Niu
Data Science Master
Duke University
Durham, NC 27708, USA
mu.niu@duke.edu

Zihan Xiao
Data Science Master
Duke University
Durham, NC 27708, USA
zihan.xiao@duke.edu

ABSTRACT

This project aims to conduct sentiment analysis on sports-related content to uncover patterns in public reactions to sports events. A fine-tuned BERT (Bidirectional Encoder Representations from Transformers) model is utilized for its superior ability to understand contextual nuances in text. To train the BERT model, the Google Emotions Dataset is employed as a labeled data source, given that sports-related tweets, reviews, and news headlines are unlabeled. The performance of the BERT model is benchmarked against Logistic Regression and Random Forest models using a confusion matrix, demonstrating its superiority over traditional sentiment analysis approaches. Additionally, ChatGPT-generated text is used to further evaluate the BERT model's accuracy. The model is then applied to analyze public reactions to sports events, particularly the UEFA Champions League, and their impact on different teams. This novel application of a BERT model to the sports domain captures unique sentiment markers in sports commentary, such as excitement, sadness, and pride, utilizing datasets sourced from Kaggle.

1 INTRODUCTION

Sports fandom is often characterized by intense emotions, with fans passionately supporting their teams and expressing their feelings on social media platforms. These emotions are not only deeply personal but also amplified in the public sphere, particularly during high-profile events like the UEFA Champions League. Between 2017 and 2023, Real Madrid emerged as the most successful team, securing championships in 2017, 2018, and 2022. Other prominent teams, such as Liverpool, Manchester City, and Bayern Munich, also delivered exceptional performances during this period. This dynamic provides a rich context for exploring the relationship between team success and public sentiment.

Our project delves into the sentiment landscape surrounding top-performing teams, focusing on how public opinions evolve as teams achieve major milestones, such as winning championships. Specifically, we aim to investigate whether successful teams enjoy a more positive and optimistic public sentiment or if the reality is more complex. By analyzing shifts in fan attitudes over time, we provide insights into the interplay between team performance and public opinion.

This research carries several significant implications. First, from a methodological perspective, our study employs seven emotion labels tailored specifically for sports contexts, enabling us to evaluate the performance of a fine-tuned BERT model in capturing nuanced sports-related sentiments compared to traditional sentiment analysis models. Second, from a marketing standpoint, elite sports teams today allocate substantial resources to managing their public image and guiding fan sentiment. Our analysis of seven years of data offers actionable insights into optimizing resource allocation for more effective marketing strategies. Finally, from a fan engagement perspective, our findings highlight how collective fan sentiment can influence team dynamics. By understanding the impact of their expressions, fans may be inspired to contribute more positively to their teams' successes.

In essence, this project bridges the gap between sports analytics, sentiment analysis, and fan engagement, shedding light on how emotions and opinions shape the narrative of sports teams and their journeys to greatness.

2 METHODOLOGY

2.1 DATA COLLECTION

For this project, we utilized three datasets to comprehensively address our research objectives. Each dataset served a distinct purpose, contributing to the training, validation, and application of our sentiment analysis model.

- **Go Emotions: Google Emotions Dataset:**
  The primary dataset for training our model was the Go Emotions dataset developed by Google AI. This dataset consists of Reddit user comments, meticulously labeled with emotional tones to facilitate deep analysis of textual sentiment. With a total of 31 columns, the dataset includes identifiers (‘id‘), textual data (‘text‘), a clarity flag (‘example_very_unclear‘), and 28 columns representing various emotional categories. Go Emotions is specifically designed to train neural networks for fine-grained sentiment analysis, making it an ideal choice for our task of understanding nuanced sports-related emotions.
- **ChatGPT-Generated Test Data:**
  To evaluate the robustness of our fine-tuned BERT model, we created a test dataset using ChatGPT. Although this dataset is relatively small, comprising only 50 data points, it plays a crucial role in assessing the model's ability to handle randomly generated and potentially ambiguous input. This step allowed us to validate the model's generalization capabilities beyond structured, pre-labeled data.
- **Sports Sentiment Dataset:**
  For the application phase, we employed the COMP5420-dataset, which contains sports-related tweets, reviews, and news headlines. This dataset includes seven sheets, each corresponding to the UEFA Champions League seasons from 2017 to 2023. For our analysis, we focused on two key columns: ‘text‘, containing the raw textual content, and ‘flair‘, identifying the teams associated with the content. By leveraging this dataset, we were able to apply our trained model to real-world data and analyze public sentiment towards specific teams over multiple seasons.

By combining these datasets, we ensured a comprehensive approach to training, validating, and applying our sentiment analysis model. This layered methodology not only strengthens the reliability of our results but also highlights the versatility of the BERT model in capturing the emotional dynamics of sports-related content.

2.2 PREPROCESSING

The preprocessing step focuses on preparing the Go Emotions Dataset and the COMP5420 Sports Sentiment Dataset for analysis, ensuring that the data is clean, relevant, and structured appropriately for model training and evaluation.

- **Go Emotions: Google Emotions Dataset:**
  We begin by filtering rows where the example_very_unclear column is set to False, ensuring only clear and reliable data is retained. From this filtered dataset, we extract entries associated with seven specific emotions: excitement, joy, pride, relief, anger, sadness, and nervousness. These emotions are chosen for their relevance in the context of sports sentiment analysis. Subsequently, each text entry is assigned a corresponding label based on its emotional tone. The dataset is then split into two subsets: train_df (95%) for training the model and val_df (5%) for validation, providing a balanced foundation for model evaluation.
- **Sports Sentiment Dataset:** We focuses on isolating team names mentioned within the column, enabling us to map textual data to specific teams for subsequent sentiment analysis.

By doing so, we ensure the dataset is tailored to analyze public sentiment towards individual teams, facilitating deeper insights into fan reactions.
Subsequently, We selected 12 prominent UEFA Champions League teams as the focus of the study: Real Madrid, FC Barcelona, Chelsea, Manchester United, Liverpool, Tottenham Hotspur, Arsenal, Manchester City, Juventus, Inter Milan, AC Milan, and Bayern Munich. This selection includes the traditional eight European football powerhouses alongside four emerging teams known for their outstanding performance and growing popularity. Geographically, these teams represent two from La Liga, six from the Premier League, three from Serie A, and one from the Bundesliga. For this analysis, we exclusively used data related to these teams and dropped all other records.

This preprocessing pipeline ensures that both datasets are optimized for model training and application, providing a robust foundation for achieving the project's objectives.

2.3 Model Development

The model development process in this project incorporates both traditional machine learning models and a state-of-the-art transformer-based model to perform sentiment analysis. For traditional models, we utilize Logistic Regression and Random Forest classifiers. Text data is preprocessed using TF-IDF vectorization to convert text into numerical features, capturing unigrams and bigrams with a vocabulary size of up to 10,000 terms. A grid search is performed for hyperparameter optimization for both models: C values for Logistic Regression and combinations of n_estimators and max_depth for Random Forest. This allows the models to learn optimal decision boundaries for classifying emotions effectively. Once trained, the models are evaluated on validation data, generating performance metrics such as F1 scores and classification reports for comparison.

After that, we leverage the fine-tuned BERT model (bert-base-uncased) from Hugging Face, which excels at capturing contextual relationships in text. The text data is tokenized using the BERT tokenizer with truncation and padding to a maximum length of 128 tokens. The labeled data is then converted into Hugging Face Dataset objects for efficient handling. The BERT model is fine-tuned for sequence classification, with training conducted over two epochs. Training parameters, such as batch size, evaluation strategy, and logging steps, are carefully set using the TrainingArguments class. Metrics like accuracy, precision, recall, and F1 score are computed to monitor model performance. The fine-tuning process ensures that the BERT model is tailored to the domain-specific sentiment classification tasks.

2.4 Evaluation Metrics

The performance of all models is assessed using a comprehensive set of evaluation metrics to ensure a robust comparison. For the Logistic Regression and Random Forest classifiers, key metrics such as F1 score (macro-averaged across all emotion labels) and classification reports are generated to measure their ability to handle imbalanced datasets. Similarly, for the BERT model, evaluation metrics include accuracy, precision, recall, and F1 score. These metrics are computed on the validation dataset during training and final evaluation to monitor the model's learning progress. Additionally, confusion matrices provide insights into specific classes where models may excel or struggle.

The model comparison is visualized using a bar plot of F1 scores for Logistic Regression, Random Forest, and BERT. This helps highlight the effectiveness of each model, with the transformer-based BERT model typically demonstrating superior performance due to its deep contextual understanding of text. This evaluation framework ensures a fair and detailed analysis of model capabilities, guiding the selection of the best-performing approach for sentiment analysis in the sports domain.

3 Experiments and Results

3.1 Data Visualization

After completing the Data Collection and Preprocessing steps, we conducted a Data Visualization Analysis on the Google Emotions Dataset and the Sports Sentiment Dataset. The results are shown in Figures 1 and 2.

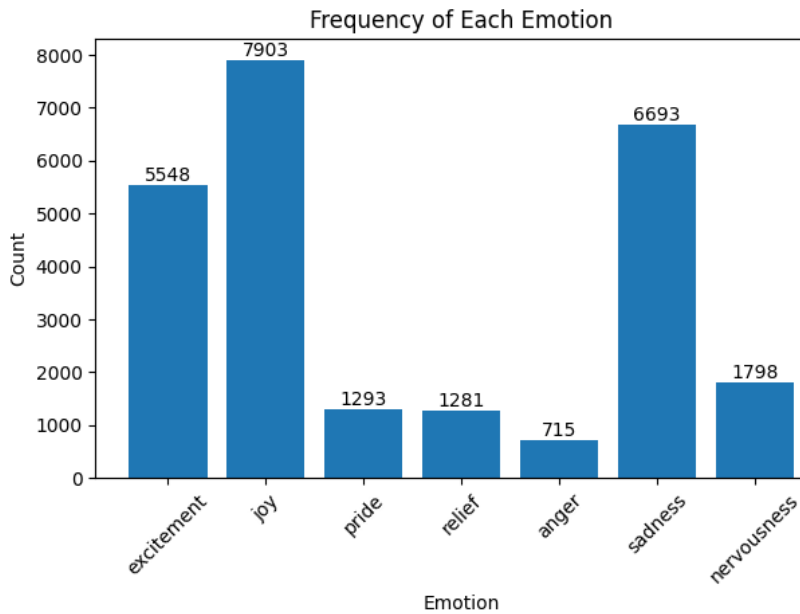

Figure 1: Frequency of Each Emotion in Google Emotions Dataset

Figure 1 illustrates the frequency of the seven emotions in the dataset. As shown, the most frequent emotion is joy, appearing 7,903 times, while the least frequent is anger, with only 715 occurrences—less than one-tenth of joy. This observation contradicts our intuition, as one might expect many individuals to express their anger on social media. On the other hand, the high frequencies of excitement and sadness align with common expectations. Although the distribution of emotions is uneven, the dataset is sufficiently large to train our model effectively.

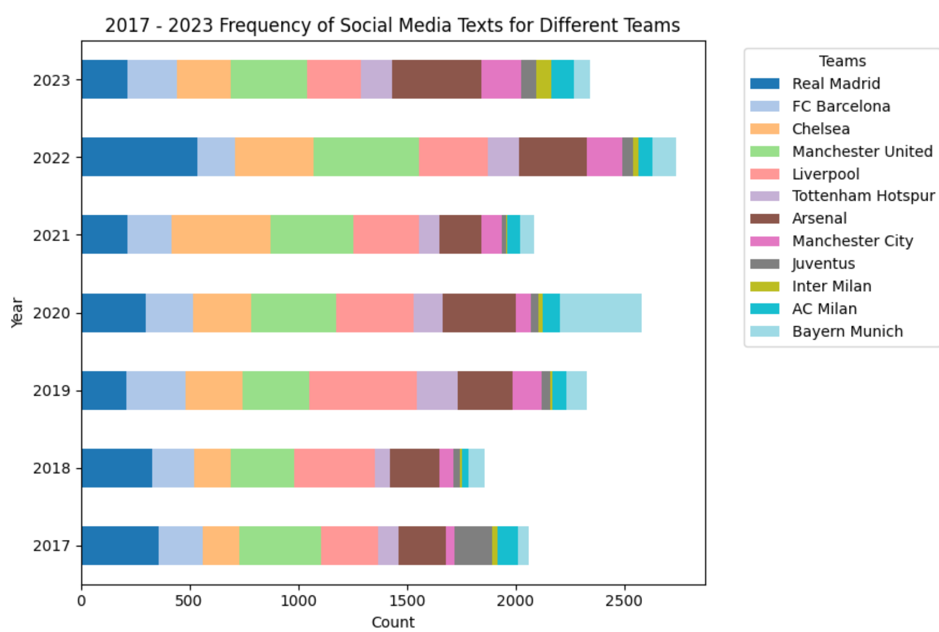

Figure 2: Frequency of Social Media Texts in Sports Sentiment Dataset

Figure 2 depicts the volume of social media data for different teams from 2017 to 2023. Overall, the data shows a fluctuating upward trend. Premier League teams generally dominate in popularity, with Manchester United and Liverpool being the two most-discussed teams. In contrast, Italy Serie A teams such as AC Milan, Inter Milan, and Juventus have relatively lower social media activity. Notably, the champion teams of each year tend to experience a significant spike in social media engagement during their championship year. For example, Real Madrid, as the UEFA Champions League champion in 2017, 2018, and 2022, had substantially higher engagement during those years compared to others. This trend is particularly pronounced for Bayern Munich, whose social media activity in 2020, when they won the championship, was at least twice as high as in any other year.

3.2 Performance of Logistic Regression and Random Forest

We evaluated the performance of the models using the classification_report function from sklearn.metrics. This function provides a detailed overview of the model's performance, including precision, recall, and F1-score for each category, as well as overall metrics, making it a powerful evaluation tool. For this task, we trained two traditional models commonly used in sentiment analysis: Logistic Regression and Random Forest. The performance of these models is presented in Tables 1 and 2.

Table 1: Classification Report for Logistic Regression

Emotion	Precision	Recall	F1-Score	Support
Anger	0.22	0.17	0.19	36
Excitement	0.62	0.58	0.60	277
Joy	0.60	0.66	0.63	351
Nervousness	0.41	0.34	0.37	87
Pride	0.42	0.31	0.36	58
Relief	0.58	0.38	0.46	58
Sadness	0.67	0.76	0.71	314
Accuracy			0.60	1181
Macro Avg	0.50	0.46	0.47	1181
Weighted Avg	0.59	0.60	0.59	1181

Table 2: Classification Report for Random Forest

Emotion	Precision	Recall	F1-Score	Support
Anger	0.11	0.06	0.07	36
Excitement	0.58	0.61	0.59	277
Joy	0.58	0.59	0.59	351
Nervousness	0.49	0.26	0.34	87
Pride	0.45	0.29	0.35	58
Relief	0.52	0.29	0.37	58
Sadness	0.62	0.78	0.69	314
Accuracy			0.58	1181
Macro Avg	0.48	0.41	0.43	1181
Weighted Avg	0.56	0.58	0.56	1181

As shown in Tables 1 and 2, the performance of both models on the Google Emotions Dataset is relatively similar. However, Logistic Regression demonstrates a slight advantage in metrics such as accuracy and precision. The accuracy scores of 0.60 for Logistic Regression and 0.58 for Random Forest are reasonable for the challenging task of classifying text into seven emotion categories. Notably, Random Forest performs particularly poorly in classifying the anger category, while both models achieve their best performance in classifying sadness, making it the most successfully predicted emotion among the seven.

3.3 Performance of BERT and Model Comparison

The classification report for the BERT model is presented in Table 3.

Table 3: Classification Report for BERT

Emotion	Precision	Recall	F1-Score	Support
Anger	0.32	0.31	0.31	36
Excitement	0.66	0.61	0.63	277
Joy	0.65	0.69	0.67	351
Nervousness	0.54	0.36	0.43	87
Pride	0.58	0.48	0.53	58
Relief	0.44	0.38	0.41	58
Sadness	0.72	0.85	0.78	314
Accuracy			0.65	1181
Macro Avg	0.56	0.52	0.54	1181
Weighted Avg	0.64	0.65	0.64	1181

As shown in Table 3, the BERT model demonstrates significant improvements across all metrics compared to Logistic Regression and Random Forest. The accuracy reaches 0.65, with other metrics improving by 0.05 to 0.11, establishing a clear performance advantage over the other two models. The BERT model also provides better predictions for each emotion, effectively addressing the poor performance of Logistic Regression and Random Forest in classifying the anger category.

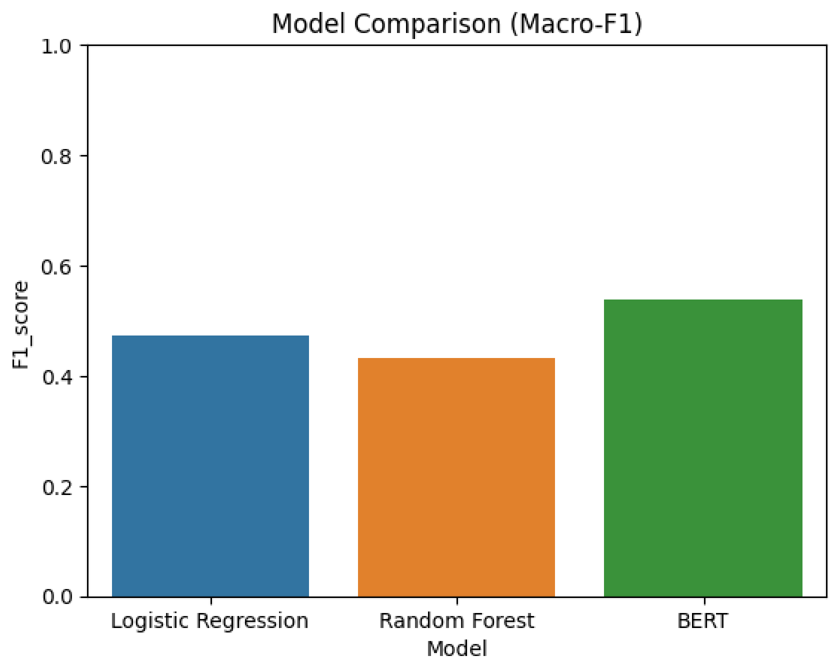

Figure 3: Model Comparison (Macro F1-Score)

Figure 3 further highlights the comparison of Macro F1-Scores, where the BERT model achieves the highest score, followed by Logistic Regression, which outperforms Random Forest. This reinforces the superiority of the BERT model for sentiment analysis tasks.

3.4 TESTING BERT WITH CHATGPT-GENERATED TEXTS FOR VALIDATION

After training the BERT model, we tested its performance using a randomly generated dataset of 50 texts created by ChatGPT. This evaluation aimed to determine whether the BERT model can effectively predict emotions in everyday textual data. The classification report is shown in Table

4, where the accuracy reached an impressive 76%, significantly higher than its performance on the Google Emotions Dataset.

Table 4: Classification Report for BERT with ChatGPT-generated texts

Emotion	Precision	Recall	F1-Score	Support
Anger	1.00	0.50	0.67	4
Excitement	0.86	0.75	0.80	8
Joy	0.67	0.86	0.75	7
Nervousness	0.67	0.80	0.73	10
Pride	0.80	0.67	0.73	6
Relief	1.00	0.33	0.50	3
Sadness	0.79	0.92	0.85	12
Accuracy			0.76	50
Macro Avg	0.83	0.69	0.72	50
Weighted Avg	0.79	0.76	0.75	50

The precision, recall, and F1-scores for all seven emotions were generally above 0.67, with relief and anger achieving perfect precision scores of 100%, and sadness attaining a recall of 92%. This discrepancy in performance may be attributed to the nature of the datasets: while the Google Emotions Dataset contains more diverse and irregular text, the ChatGPT-generated texts consist of well-structured sentences. Nevertheless, these results demonstrate the BERT model's strong and reliable performance across both structured and unstructured texts, highlighting its versatility in emotion prediction tasks.

4 Application and Analysis

4.1 Analysis of Emotion Proportions Over Time

After training the BERT model and testing it on the ChatGPT-generated dataset, we applied it to the Sports Sentiment Dataset to make emotion predictions. Our first objective was to analyze the proportion of different emotions in UEFA Champions League-related texts from 2017 to 2023. This analysis also served as a way to evaluate the model's stability on a new dataset. A highly unstable model would result in significant fluctuations in the proportions of different emotions.

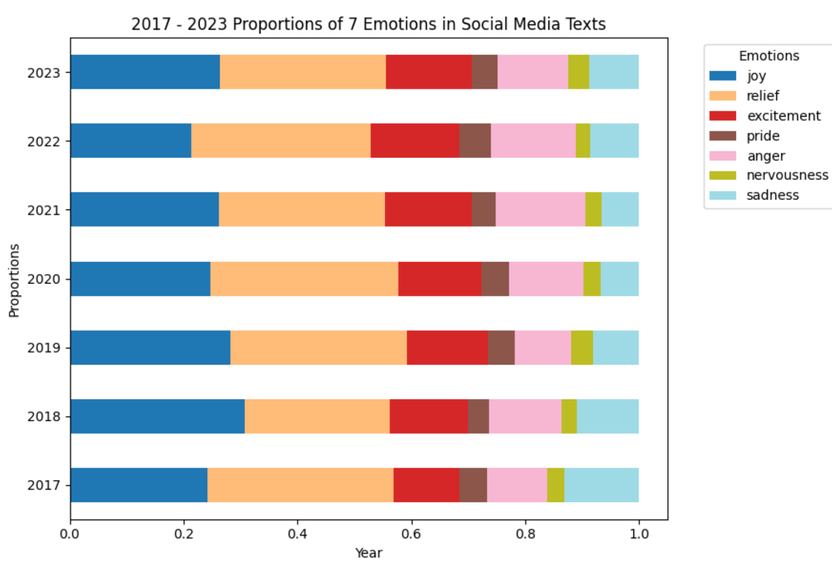

Figure 4: Frequency of Each Predicted Emotion in Sports Sentiment Dataset

The results, shown in Figure 4, indicate minimal variation in the proportions of the seven emotions over time, with no significant upward or downward trends. This demonstrates the BERT model's robustness when applied to distinctly different data across multiple years. Among the findings, the high proportion of relief is the most unexpected, while the low proportions of pride and nervousness align with expectations. This can be explained by the tendency of individuals to express more straightforward emotions, such as joy, excitement, or anger, on social media, whereas pride and nervousness are less direct. Overall, aside from the notably high proportion of relief, the distribution of other emotions aligns well with general emotional perception, and yearly variations appear random and unpredictable.

4.2 Comparison of Emotion Proportions Among Different Teams

The second application focuses on analyzing the UEFA Champions League champion teams for each year. The method involves calculating the average proportions of the seven emotions for all teams in a given year and then subtracting these averages from the emotion proportions of the champion team for that year. This measures the impact of winning the championship on the team's social media sentiment environment.

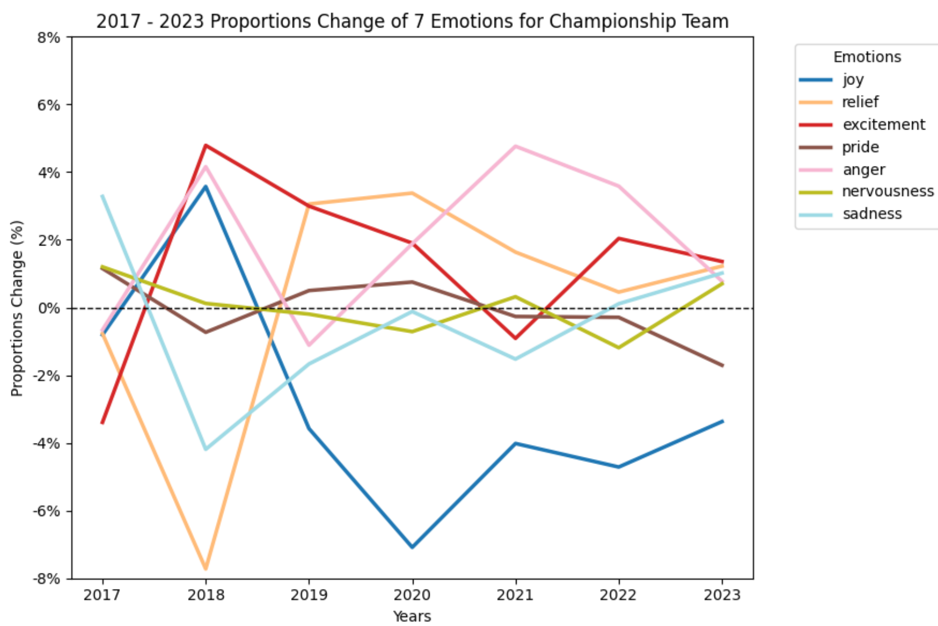

Figure 5: Proportions Change of Each Predicted Emotion for Championship Team

The results, shown in Figure 5, reveal that winning the championship does not necessarily lead to a more positive sentiment environment for the team. For instance, the proportions of joy, excitement, and pride, which are typically associated with positive outcomes, do not consistently increase. On the contrary, from 2020 to 2023, the proportion of joy was notably low. In 2020, this may be attributed to Bayern Munich's outstanding performance and Robert Lewandowski's stellar contributions being overshadowed by the cancellation of the Ballon d'Or due to the pandemic. Similarly, the high proportion of anger in 2021, the year Chelsea won the title, might be linked to the Ballon d'Or being awarded to someone other than Chelsea's key performers, Jorginho and N'Golo Kanté, despite their exceptional contributions.

These findings suggest that a team's social media sentiment environment is not solely determined by its performance or achievements, such as winning a championship. Other factors, including individual performances and awards, can significantly influence public sentiment. Social media reflects a vast and complex network of opinions. An interesting observation is the inverse trend of joy and anger from 2020 to 2023. The decline in joy alongside the rise in anger, as two opposing emotions, further validates the reliability of the BERT model in capturing nuanced sentiment dynamics.

4.3 Yearly Emotion Differences Between Championship Team and its Overall Average

The third application analyzes the emotion proportions of each team from 2017 to 2023. The calculation method involves dividing the total count of each emotion by the sum of all emotions for that team. Since the total data volume per year is relatively consistent, this method yields results similar to a weighted average. The results, presented in Figure 6, reveal an intriguing observation.

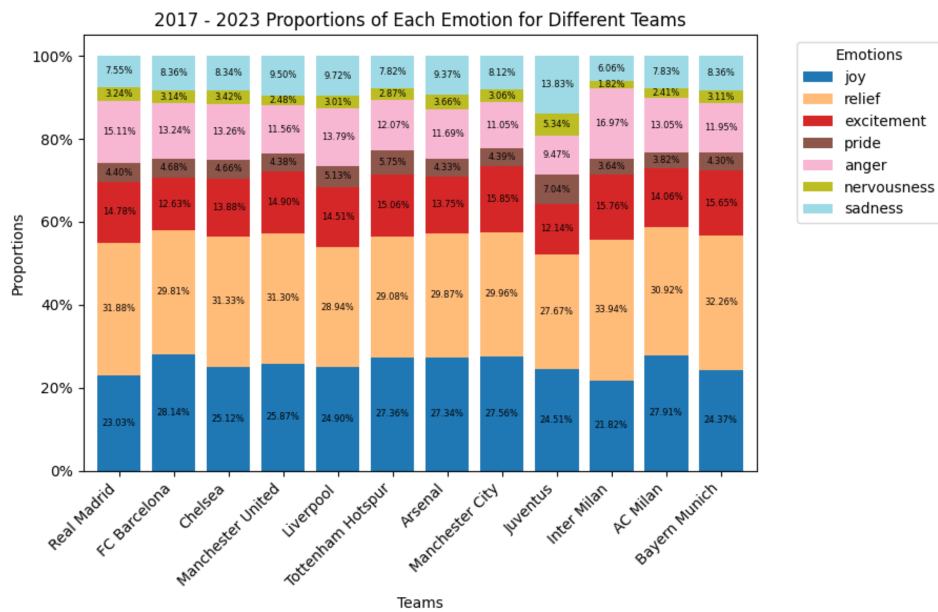

Figure 6: Proportions of Each Predicted Emotion for Different Teams

Real Madrid, the most successful team over these seven years with three UEFA Champions League titles (compared to one title each for four other teams), exhibits the second-lowest proportion of joy and the second-highest proportion of anger, trailing only Inter Milan in both cases. However, Inter Milan's data volume is the smallest among the 12 teams, and its emotion distribution differs significantly from the others, making it less representative. Thus, Real Madrid can effectively be considered the team with the least favorable sentiment environment among the 11 major teams analyzed.

This finding aligns closely with Real Madrid's well-known fan culture, characterized by high expectations, strict standards, and vocal criticism of their own players. Real Madrid fans, being accustomed to success, are not easily satisfied, and the players, in turn, have developed strong resilience and self-motivation in response to this demanding environment. This raises an intriguing hypothesis: **could a negative sentiment environment be more conducive to a team's progress than a positive one?** While this hypothesis requires further validation from multiple perspectives, the findings from the second application partially support this idea, adding depth to this line of inquiry.

5 Conclusion

This study demonstrates the power and versatility of sentiment analysis in understanding public reactions to sports events, particularly through the fine-tuned BERT model. By leveraging diverse datasets, including the Google Emotions Dataset, ChatGPT-generated texts, and the Sports Sentiment Dataset, we successfully trained, tested, and applied a robust sentiment analysis pipeline tailored to the sports domain.

The BERT model consistently outperformed traditional approaches like Logistic Regression and Random Forest, achieving higher accuracy and better handling nuanced emotions such as anger and sadness. The evaluation with ChatGPT-generated data further showcased its adaptability, achieving a remarkable 76% accuracy on well-structured texts, underscoring its reliability across different data types.

Our application phase provided significant insights into the sentiment environment surrounding UEFA Champions League teams from 2017 to 2023. The stability of emotion proportions over time highlighted the model's robustness, while the analysis of champion teams revealed that public sentiment is influenced by a complex interplay of factors beyond just team performance, such as individual achievements and external events. Notably, Real Madrid's unique sentiment dynamics reflected the high standards and demanding expectations of its fanbase, sparking an intriguing hypothesis about the potential benefits of negative sentiment environments on team development.

In conclusion, this study not only advances the application of BERT in sentiment analysis but also offers valuable perspectives on the relationship between sports performance and public sentiment. These findings pave the way for future research into the role of sentiment in sports marketing, fan engagement, and team dynamics, with potential applications extending beyond the sports domain into broader areas of public discourse analysis.

6 Limitations and Future Work

This study demonstrates the potential of sentiment analysis, powered by a fine-tuned BERT model, to uncover meaningful insights from sports-related content. By analyzing data from the UEFA Champions League and evaluating the model's performance on both structured and unstructured texts, the research provides valuable perspectives on public sentiment and its complex interplay with team performance. Despite its successes, the study is not without limitations, and opportunities for further exploration remain.

Several limitations emerged during the research. The datasets, while useful, were either general-purpose (Google Emotions Dataset) or narrowly focused (Sports Sentiment Dataset), limiting the generalizability of findings to other sports or contexts. The seven predefined emotions, though relevant, fail to capture more nuanced sentiments like frustration or hope, which are common in sports discourse. Additionally, the analysis did not consider regional and temporal variations in sentiment, which could influence the results. Lastly, while correlations between sentiment and events like championship wins were observed, causality could not be established without deeper qualitative or longitudinal analysis.

Future work could address these limitations by expanding the dataset to include other sports, leagues, and regions, as well as incorporating multilingual data to capture cultural differences. Enhancing emotion categories, either by adding more predefined labels or using unsupervised techniques to discover latent emotions, could provide richer insights. Real-time sentiment analysis for live sports events is another promising avenue, with applications in fan engagement and marketing. Exploring advanced transformer models or domain-specific embeddings could further improve performance. Finally, a deeper investigation into the relationship between sentiment, team performance, and external factors like awards or media coverage could provide actionable insights for teams and organizations. These directions build on this study's foundation, offering broader applications and deeper understanding of sentiment dynamics in sports.

7 Acknowledgments

We would like to express our sincere gratitude to Professor Patrick Wang and teaching assistants for their invaluable guidance, encouragement, and support throughout this research project. Their teaching and insights helped shape the direction of the work, and their patience and willingness to engage with questions and ideas provided a strong foundation for continuous learning and improvement. We are truly grateful for their expertise, mentorship, and the time they dedicated to helping us achieve our goals.

References

[1] Demszky, G., Ghosh, D., Guha, A., et al. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

Retrieved from https://aclanthology.org/2020.acl-main.372

[2] Gjurovic, T., et al. (2018). Reddit: A Gold Mine for Personality Prediction? Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Retrieved from https://aclanthology.org/W18-1112/

[3] Hada, K., et al. (2021). Ruddit: Norms of Offensiveness for English Reddit Comments. Proceedings of the 2021 Annual Meeting of the Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.acl-long.210/

[4] Liu, Y., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint. Retrieved from https://arxiv.org/abs/1907.11692

[5] Patel, S., & Passi, A. (2020). Sentiment Analysis on Twitter Data of World Cup Soccer Tournament Using Machine Learning. MDPI. Retrieved from https://www.mdpi.com/2624-831X/1/2/14

[6] Rothe, S., Narayan, S., & Severyn, A. (2019). Leveraging Pre-trained Checkpoints for Sequence Generation Tasks. arXiv preprint. Retrieved from https://arxiv.org/abs/1907.12461