Analyzing Public Sentiment on Social Media during FIFA World Cup 2022 using Deep Learning and Explainable AI

***Abstract*—Analysis of public sentiment is extremely useful for comprehending the responses of the general public during important events, and the FIFA World Cup 2022 was no exception. Within the scope of this study, we used deep learning models such as roBERTa, distilBERT, and XLNet to conduct an analysis of the views that were stated on Twitter during the first day of the tournament. These models were fine-tuned using a comprehensive dataset consisting of 30,000 tweets, which had been preprocessed. The performance of these models was assessed using measures such as accuracy, F1-score, precision, recall, etc. In addition, we used an Explainable AI known as Local Interpretable Model-Agnostic Explanations (LIME) so that we could better understand how model decisions were made in sentiment classification. Our research has shown that roBERTa is an excellent model for classifying sentiment, and it has also shown the significance of interpretability achieved using LIME. Our research enhances the understanding of sentiment analysis during major sports events and suggests future directions for research in this domain.**

***Index Terms*—Sentiment Analysis, Deep Learning, roBERTa, distilBERT, XLNet, LIME**

I. INTRODUCTION

Sporting events have long since evolved beyond the concept of straightforward competition to become worldwide gathering places for individuals of many backgrounds and ways of life. Among these noteworthy sports events, the FIFA World Cup stands out as a sight that can't be matched. This is the time when countries all over the world join together to enjoy the wonderful game of football. The 2022 FIFA World Cup, held in Qatar, marked a significant turning point in the long and illustrious history of this prestigious sport. Aside from the importance it has for sports, the World Cup also serves as a unique lens through which we are able to evaluate the general sentiment of mind of audiences all around the world in real-time.

In this age of digital technology, platforms for social media have quickly become the most important venues for public engagement and discussion. Particularly on social media platforms such as Twitter, people from all walks of life come together to share their views, opinions, and responses to big events. On the first day of the 2022 World Cup, an extraordinary increase in the amount of online conversation was seen. Moreover, thousands of individuals are communicating with one another about their views, feelings, and even criticisms on Twitter. These tweets not only convey the mood of the current situation but also provide a great supply of data for conducting an analysis of sentiment. They are symbolic of the public's emotional response to the tournament, which reveals how people felt about it as a whole.

The discipline of sentiment analysis, which is a subfield of Natural Language Processing, gives us the ability to methodically explore and understand the emotions conveyed in written material. We are able to go deeper into the intricate network of tweets pertaining to the World Cup by using the capabilities of deep learning models like roBERTa [1], [2], distilBERT [3], and XLNet [4]. By this, we are able to identify the sentiments that are driving the worldwide discourse around the tournament. These models have shown a remarkable capacity, after being pre-trained on massive text corpora, to accurately express nuanced feelings in a range of different environments. Our dataset of 30,000 tweets from the opening day of the 2022 World Cup was used to do fine-tuning on three different models: roBERTa, distilBERT, and XLNet. Their performance was meticulously evaluated using a broad variety of assessment indicators, which enabled us to compare how well each one captured the sensory nuances.

This research investigates the perspectives that were discussed on Twitter on the very first day of the 2022 FIFA World Cup using the power of Deep Learning. We carried out a sentiment analysis on 30,000 tweets in order to ascertain the most commonly held opinions that surfaced during this significant event. Through this study, we aimed to gather insight into how people responded to the opening of the World Cup in order to give a unique perspective on how people feel all around the world. Also, we take advantage of Explainable AI such as LIME Interpretability [5] to tackle the intuitive problem that lies behind the decisions that our deep learning models make. This interpretability technique sheds light on the categorization decisions made by our models as well as the traits that lie below those classifications.

This research expands our knowledge of the convergence of big sports events, social media, and natural language processing. Moreover, it highlights the relevance of sentiment analysis as a tool for assessing the collective sentiment of global audiences at unusual periods. Our study will help policy makers and stakeholders gain insights and extract necessary information from future big events like FIFA World Cup and other sport events.

II. RELATED WORKS

Numerous studies have been conducted in the field of sentiment analysis to investigate the significance of user-generated information, particularly on social media platforms such as Twitter, in relation to a variety of important events. The evaluation of how people felt during the first day of the FIFA World Cup 2022 has been a prominent subject of attention during the past few months. These studies have shown that data from social media platforms has the potential to be a useful source of insights into public sentiments and attitudes during times of crisis. These insights may be gleaned by analyzing enormous numbers of user-generated content in order to identify patterns and responses.

The author in [6] discusses a Twitter-based study on global sentiment towards Qatar's 2022 World Cup hosting. The research included three stages: before Qatar was chosen as host, after the selection, and during the event. This research found 84% positive Twitter sentiment and 16% negative. The model's 87% accuracy in predicting sentiment in unannotated data is noteworthy.

This study [7] collected data and developed a new dataset on which they used Logistic Regression, Random Forest, Naive Bayes, and SVM classifiers. The logistic regression technique achieved the highest average accuracy at 93%, followed by the random forest classifier at 92%. The Naive Bayes classifier averaged 88% accuracy and the SVM-Classifier 93%. These results show that machine learning can accurately measure Arabic-speaking Twitter sentiment on the Qatar World Cup 2022.

To determine global opinion, the researchers in [8] analyzed Twitter data from the 2014 World Cup in Brazil. They used WordNet's part of speech and context to create an algorithm to extract emotive words. Using naive Bayes, SVM, KNN, and random forest, sentiment polarities were further analyzed. Random forest had the best AUC of 0.97, while naive Bayes had the highest accuracy of 88.17%. This study shows that utilizing NLP and machine learning to analyze Twitter data from the 2014 World Cup can reveal users' sentiments.

The authors examined football fan Twitter sentiment in [9]. GloVe and other word vectors were used for sentiment analysis to collect semantic and syntactic information. They added context with a sentiment lexicon. The authors also used Random Forest, Support Vector Machine, Multinomial Naive Bayes, K-Nearest Neighbours, and XG Boost for sentiment analysis. Their 2018 FIFA World Cup Twitter experiment has great validation and testing accuracy, with Random Forest being the most consistent and robust classifier for football fans' sentiments.

The authors in [10] studied aspect-level sentiment analysis using transformer-based pre-trained models like BERT and RoBERTa. The aspect-based technique was applied to BERT and RoBERTa models after they had tested them without it. The aspect-based method outperformed standard models by approximately 1%. Among the models tested, the aspect-based BERT model had the highest accuracy and performance.

Sentiment analysis has been done not only on sports events but also on other important events like elections. The authors in [11] examined 2022 election-related Twitter discussions. They assessed voter preferences and identified major debate topics using sentiment analysis and topic modeling. They employed Naive Bayes and Support Vector Machines for sentiment analysis, with Naive Bayes doing better at 73% than 69%.

TABLE I
DATASET EXAMPLES BEFORE PREPROCESSING

Tweet	Sentiment
Ecuador players after 1-0 against Qatar...	neutral
Morgan Freeman too old damn can't believe...	negative
So, I guess we have a #WorldCup2022 opener...	neutral

The studies in [12] have attempted to demonstrate connections between shifts in sentiment and big events such as lockdown announcements, demonstrating the concrete impact that significant developments have on public sentiment, which frequently takes the form of mood deteriorations in response to certain occurrences.

In the field of word embeddings for sentiment analysis, the introduction of sentiment-specific word embeddings (SSWE) has addressed a problem in existing word embedding methods, specifically, their failure to properly capture sentiment polarity. This restriction was solved by the introduction of sentiment-specific word embeddings in [13]. The research conducted has shown that SSWE is competitive in sentiment analysis, highlighting the potential of these systems to improve the accuracy of sentiment classification.

The author introduces TWITA in [14], the first Italian tweet corpus produced automatically and transferable to any language. They test sentiment analysis on two TWITA datasets: generic and topic-specific. They only use a polarity lexicon, which they created by automatically matching three existing resources to create the first Italian polarity database.

These studies jointly highlight the value of data from social media platforms, approaches for analyzing sentiment, and emerging assessment tasks in the context of understanding public sentiments, reactions, and mood shifts, particularly during significant events and on platforms such as Twitter.

III. DATASET AND PREPROCESSING

Our dataset for doing sentiment analysis includes everything there is to talk about in relation to the first day of FIFA World Cup 2022 that can be found on social media such as Twitter [15]. It is a large collection that includes 30,000 tweets, all of which are centered around the first day of the FIFA World Cup 2022. The dataset contains only tweets written in English, all of which make consistent use of the #WorldCup2022 hashtag, and no tweets contain any other keywords. This dataset has Tweet, Sentiment, Number of Likes, Source of Tweet and Date Created.

In order to guarantee the accuracy and consistency of the data, we utilized systematic steps in our preprocessing pipeline. The content of the tweets was subjected to an extensive editing process. As part of this process, user mentions, URLs, stop words, emojis, etc. were removed from the text in order to preprocess it for further research. We changed every instance of text to lowercase in order to guarantee

TABLE II
DATASET EXAMPLES AFTER PREPROCESSING

preprocessed_tweet	labels
ecuador players losing qatar tonight fifaworld...	0
morgan freeman old damn cant believe...	1
guess world cup opener ahead...	0

text consistency and make text matching easier. Also, text cleaning removed unnecessary letters, symbols, and formatting to improve data quality. Tokenization, which segments text into discrete pieces, was done to prepare it for our deep-learning models. We added a new column to the database called "preprocessed_tweet" so that we could save the text of tweets that had been preprocessed and standardized. Each data point was carefully annotated with sentiment labels, and expressions were sorted into one of three basic groups: neutral (0), negative (1), and positive (2). In table I, we can see the Tweet and Sentiment before preprocessing, and in table II, we can see the preprocessed and labeled tweet data. In Fig. 1 we can visualize that the negatively labeled sentiment data portion is a bit lesser than the positive and neutral data in the dataset.

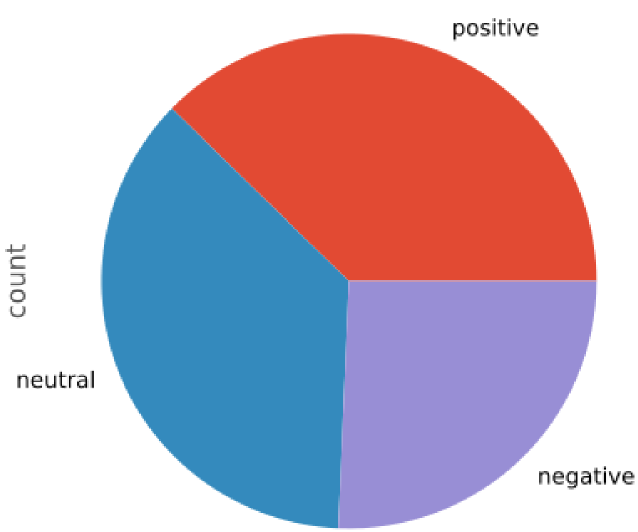

Fig. 1. Pie chart of Total Sentiment Classes in the Dataset.

We were able to make the dataset more understandable by producing a word cloud and highlighting the words that occurred most frequently in the corpus in Fig. 2. This robust dataset, which has been rigorously produced and preprocessed, serves as the cornerstone for our sentiment analysis, which enables us to delve deep into the sentiments expressed by users throughout the FIFA World Cup 2022.

Fig. 2. Word Cloud of the dataset.

IV. METHODOLOGY

In this part, we provide an overview of the approach that was used for doing sentiment analysis by using deep learning models, more especially roBERTa, distilBERT, and XLNet. Justifying our model selection, describing the model architectures, training method, and assessment metrics, and using Explainable AI such as LIME (Local Interpretable Model-agnostic Explanations) for model interpretation are the topics that we cover in this section.

A. Deep Learning Models Selection

The exceptional performance of deep learning models such as roBERTa [1], [2], distilBERT [3], and XLNet [4] in a variety of natural language processing (NLP) tasks, such as sentiment analysis, was a primary factor in their selection. Because these models have been pretrained on huge text corpora, they are able to pick up on delicate contextual information and nuances that are present in the text. Because of their accurate representations and their capacity for transfer learning, they are well-suited for doing sentiment analysis on data derived from social media.

B. Model Architectures and Fine-Tuning

Each model that was chosen was then fine-tuned in accordance with the particular objective of conducting sentiment analysis on tweets relating to the FIFA World Cup 2022. During the fine-tuning process, the pre-trained models were modified by adding a classification layer that was suited for the categorization of sentiment. The following is a condensed summary of each model's underlying architecture:

- **roBERTa:** roBERTa is a modification of the BERT architecture that includes improved training methods. It does this by using a bidirectional transformer architecture and by capturing information about the context from both directions. We began the process of fine-tuning the model by first training it on our preprocessed dataset, then we added a classification layer on top of the roBERTa model that had been pre-trained.
- **distilBERT:** distilBERT is a distilled version of BERT that was developed to maximize productivity without sacrificing performance. It employs a structure that is similar

to BERT's but has fewer parameters overall. DistillBERT was enhanced by the addition of a classification layer, which was then fine-tuned using our data.
- **XLNet:** XLNet presents a permutation-based training strategy to capture relationships between all input locations. XLNet is an extension of the convolutional neural network. It makes use of an architecture known as a transformer. We fine-tuned XLNet with a classification layer in order to do sentiment analysis, just as we did with the previous models.

C. Training Process and Evaluation Metrics

The preprocessed text data was tokenized, which meant that it was cut up into separate tokens or unit texts. Following the tokenization process, the texts were then turned into encoded tensor slices, which is a format that can be used as input for deep learning models. Each model was fine-tuned by making use of our dataset after it had been preprocessed. In the process of fine-tuning, we adjusted the parameters of the model in order to improve its performance on the particular sentiment analysis job we were working on. We made a great effort to select the appropriate hyperparameters, such as learning rates, batch sizes, and training epochs, in order to get the highest potential level of performance from our model.

We used a wide variety of assessment measures in order to evaluate the effectiveness of our models such as accuracy, F1 score, precision, recall, loss, etc. Accuracy refers to the fraction of properly categorized attitudes and measures how accurate the classifications are. The F1 score offers a compromise between accuracy and recall, which is especially helpful when working with datasets that are not evenly distributed. The model's ability to accurately categorize responses as either positive or negative is indicated by the term "Precision." Loss is a representation of the training loss for the model, which indicates how well the model matches the data. Because of the criteria that we've chosen to use, we are able to conduct an exhaustive analysis of how well the models perform when it comes to sentiment categorization.

D. LIME for Interpretability

In addition to using model performance measures, we interpreted the predictions that our deep learning models produced with the help of a tool called LIME (Local Interpretable Model-agnostic Explanations) [5] which is an Explainable AI. LIME sheds light on the underlying elements that are responsible for influencing predictions and offers insights into the categorization choices that these models make. Our comprehension of how the model arrives at its conclusions is improved thanks to LIME, which does this by drawing attention to the significance of certain words or phrases throughout the categorization process.

Our process includes the fine-tuning of cutting-edge deep learning models, painstaking training and assessment, and the deployment of LIME for model interpretation, which enables us to get deeper insights into the emotion represented in tweets relevant to the FIFA World Cup 2022.

TABLE III
COMPARISON OF EVALUATION METRICS OF DEEP LEARNING MODELS

Model	Accuracy	F1 Score	Precision	Recall
roBERTa	84.8%	0.83	0.83	0.82
distilBERT	83.5%	0.82	0.81	0.82
XLNet	79.6%	0.72	0.78	0.67

V. EXPERIMENTS AND RESULTS

In this part of the paper, we will discuss the outcomes of our experiments with the roBERTa, distilBERT, and XLNet models for doing sentiment analysis. We evaluate their accuracy, F1-score, precision, and recall in order to make a comparison of their performances. In addition, we examine the relevance of these measures within the framework of sentiment analysis, and we investigate the implications of our results for sentiment analysis during the FIFA World Cup 2022.

As outlined in the section under "Dataset and Preprocessing," the first step in the methodical experimental strategy that we used was to preprocess the dataset. This step included tokenization of texts and then making encoded tensor slices for the model input. We then fine-tuned three distinct models—roBERTa, distilBERT, and XLNet—utilizing specific pretrained models from Huggingface tailored to each, namely 'cardiffnlp/twitter-roberta-base-sentiment-latest' for roBERTa [1], [2], 'distilbert-base-uncased' for distilBERT [3], and 'sshleifer/tiny-xlnet-base-cased' for XLNet [4]. In order to get optimal performance from the model, hyperparameter tuning was carried out. This included making modifications such as warmup steps and weight decay. To determine how well the models were able to classify people's feelings, they were subjected to a stringent analysis that relied on many critical criteria, including accuracy, F1-score, precision, and recall.

Table III presents the evaluation metrics for roBERTa, distilBERT, and XLNet on the sentiment analysis task. These results showcase the performance of the models in terms of accuracy, F1-score, and precision.

A. Discussion of Metrics

- **Accuracy:** Accuracy can be defined as the proportion of a total number of predictions in which sentiments were accurately identified. During all of our evaluations, roBERTa demonstrated the highest level of precision, with an accuracy level of 84.8%. distilBERT finished in second place with 83.5% of the vote, while XLNet took third place with 79.6%.
- **F1-Score:** In order to determine the F1-score, one must first determine the harmonic mean of accuracy and recall scores. This creates a score that strikes a healthy balance between the two components of the examination. The fact that roBERTa was able to get the maximum possible score on the F1 test (0.83) indicates that it carried out its tasks quite capably in general. The value of distilBERT's F1 score was 0.82, which indicates that it performed very well. XLNet's F1 score was 0.72, which was lower than

the competition's score despite the fact that it had a lower accuracy.

- **Precision and Recall:** Precision is the capacity of the model to properly categorize positive or negative feelings, while recall is the ability of the model to detect actual positive situations. Both the accuracy (0.83) and recall (0.82) that roBERTa displayed were quite high, which is an indication of a well-balanced categorization. distilBERT also displayed balanced accuracy and recall, in contrast to XLNet, which had somewhat lower values for both parameters. This may be because of the use of the tiny version of XLNet pre-trained model.

B. Implications and Insights

In the context of sentiment analysis, the selection of assessment metrics is of utmost significance since it is one of the most important factors in determining how well a model is able to properly capture positive, negative, and neutral feelings about a topic. Our research has produced a number of significant conclusions that may be drawn with regard to the FIFA World Cup in 2022. RoBERTa comes out as the best, outperforming other models in terms of both accuracy and F1-score, demonstrating its superiority in terms of doing sentiment analysis for this specific event. A high level of precision ensures accurate sentiment classification and a high level of recall helps in identifying true positive cases, which ultimately contributes to the dependability of sentiment analysis. In addition, we stress the importance of precision and recall, particularly when dealing with imbalanced datasets, as a high level of precision ensures accurate sentiment classification and a high level of recall helps in identifying true positive cases. The results of our tests highlight the relevance of optimizing models by modifying hyperparameters and using other strategies, such as warmup steps and weight decay, which significantly improve model performance. In addition, the incorporation of LIME for model interpretation promises to give essential insights into the logic behind certain sentiment classifications, hence boosting our grasp of model decision-making processes. This is despite roBERTa's top performance. In conclusion, these results highlight the significant influence that model selection and optimization strategies can have on the classification of sentiments, which in turn enriches our knowledge of global sentiment during significant events.

VI. LIME Interpretability For Models

We used a technique known as Local Interpretable Model-Agnostic Explanations (LIME) [5] which is an Explainable AI because we wanted to obtain a more in-depth understanding of how the deep learning models we were working with behaved. The major purpose was to shed light on the aspects that are responsible for the predictions that were generated by our models, with the end goal of improving our comprehension of the decision-making procedures that they use.

To illustrate the application of LIME, we selected a sample instance for classification by our three models — roBERTa, distilBERT, and XLNet. The instance tweet was:"Wanna see MESSI x RONALDO together somehow! EXCITED! SIUUUU! #Qatar2022 #Messi #Ronaldo". From figures in 3, 5 and 7, the classifications by the models for this instance were as follows:

- **roBERTa:** Positive(0.98), Neutral(0.01), Negative(0.0)
- **distilBERT:**Positive(0.96), Neutral(0.02), Negative(0.02)
- **XLNet:** Positive(0.34), Neutral(0.34), Negative(0.32)

Importantly, in 4, 6 and 8, values are in the negative (Red) side mean "Not Negative" and values are in the positive (Green) side mean "Negative" as the plot is made for Class Negative.

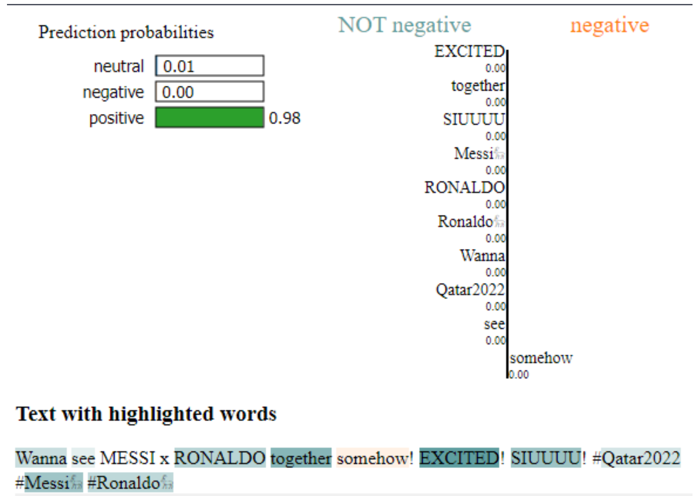

Fig. 3. LIME Explanations for roBERTa Model.

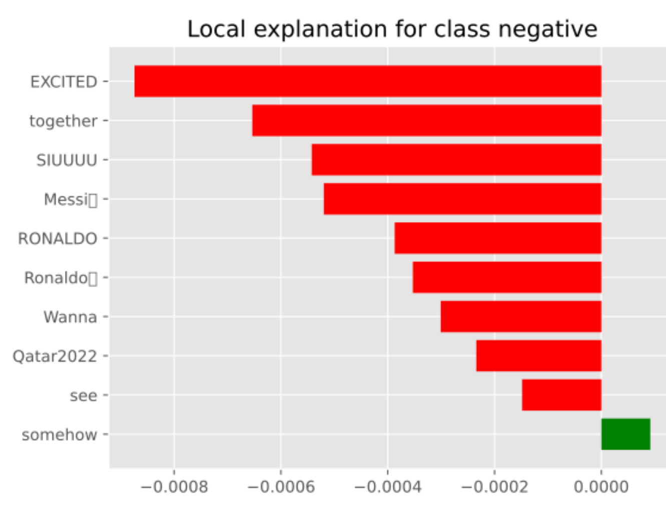

Fig. 4. roBERTa: Probability Plot of words in the Instance.

From fig. 3 and 4, we can see that it highlighted "somehow" as the only Negative word in that instance and highlighted all the other words as Not Negative, and "EXCITED" carries the highest probability value for being Not Negative and in second position comes to the word "together".

From fig. 5 and 6, it also highlighted "somehow" as the only Negative word in that instance and highlighted all the other

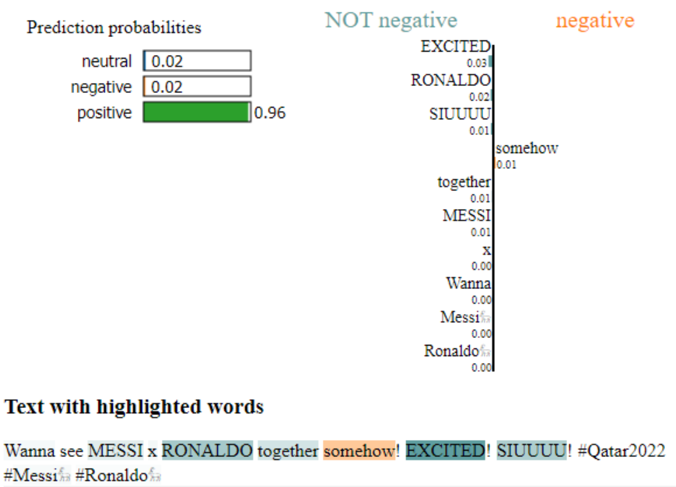

Fig. 5. LIME Explanations for distilBERT Model.

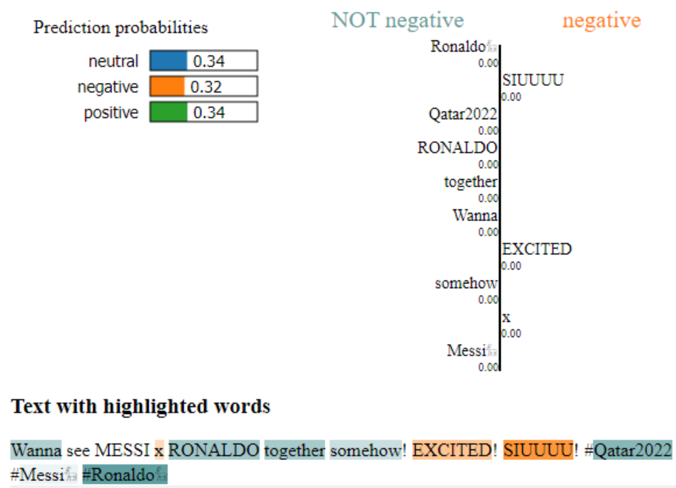

Fig. 7. LIME Explanations for XLNet Model.

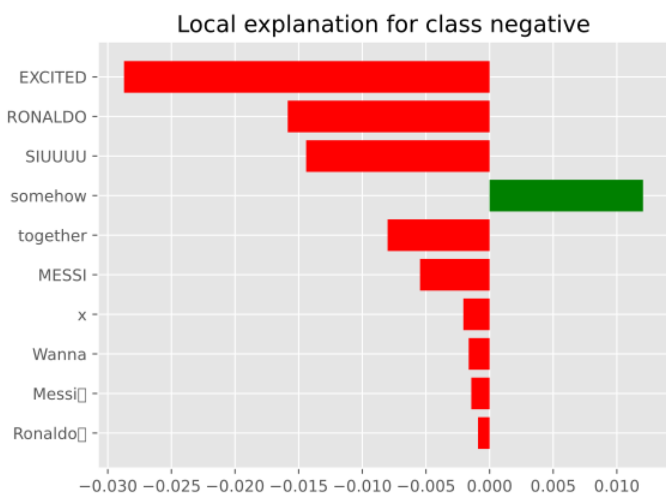

Fig. 6. distilBERT: Probability Plot of words in the Instance.

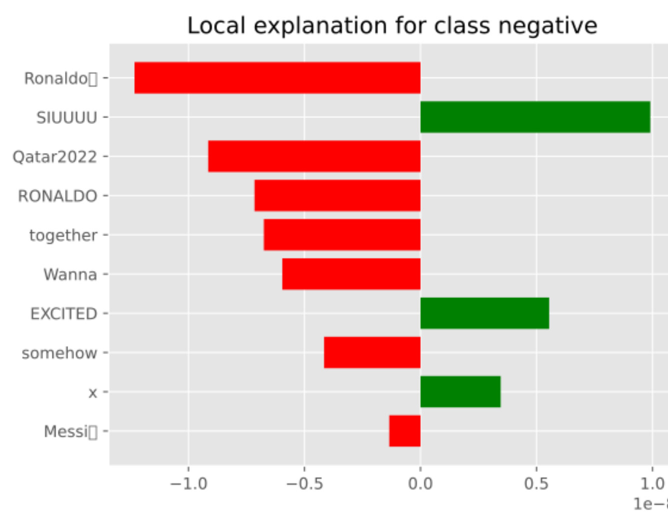

Fig. 8. XLNet: Probability Plot of words in the Instance.

words as Not Negative, and “EXCITED” carries the highest probability value for being Not Negative and second position comes the word “RONALDO”.

From fig. 7 and 8, we can see that it highlighted 3 words - “SIUUUU”, “EXCITED”, and “x” as the Negative words in that instance and highlighted all the other words as Not Negative, and “#Ronaldo” carries the highest probability value for being Not Negative and second position comes to the word “Qatar2022”.

LIME included explanations for the classifications it produced for each model, drawing attention to the significant role that individual words have in determining the outcome of the prediction. Notably, the word “somehow” was consistently recognized as the biggest driver of negative emotion across roBERTa and distilBERT models. This is an interesting finding. On the other hand, words such as “EXCITED” and “together” held the greatest probability of not being connected with a negative mood in either roBERTa or distilBERT. One more finding is that XLNet model identifies the “EXCITED” as Negative word while this word carries the highest probability values for being Not Negative in both roBERTa and distilBERT.

These LIME explanations offer light on the subtle choices made by our models, highlighting the relevance of certain words in molding predictions of mood. In addition, the identification of “somehow” as a negative factor on a regular basis highlights how sensitive the models are to context. The interpretability of LIME has shown to be quite helpful in illuminating the logic behind model predictions and boosting our grasp of how they behave.

VII. Discussion

In a nutshell, the results of our investigation into the use of deep learning models and LIME for the purpose of conducting sentiment analysis during the FIFA World Cup 2022 have produced a number of notable conclusions. For the purpose of analyzing the feelings communicated in tweets about the event, we made use of three cutting-edge deep-learning models. These models were named roBERTa, distilBERT, and XLNet. In addition, we used LIME for model interpretation so that we could better understand the thought process that went into producing their forecasts.

The results of our investigation demonstrate the usefulness of deep learning models for doing sentiment analysis, in particular when applied to the setting of a significant international event such as the FIFA World Cup. Notably, roBERTa emerged as the model with the best performance, sporting the greatest accuracy and F1 score out of all of the contenders. These findings provide further evidence that the algorithms are able to accurately capture the intricacies of attitudes that are conveyed by Twitter users in the midst of such occurrences.

In addition, the use of LIME in our study has shown itself to be a very helpful addition to the project. Because of LIME's interpretability characteristics, we were able to analyze the judgments that were produced by our deep learning models, which helped us to better understand the significance of certain words and context in the process of sentiment categorization. This increased interpretability of the model is vital not just for understanding why certain feelings were predicted but also for finding areas in which the models may be improved.

As a conclusion, the findings of our study have shown that deep learning models, when suitably fine-tuned and optimized, are capable of providing accurate sentiment analysis during big global events such as the FIFA World Cup. Additionally, the contributions that LIME has made toward improving the interpretability of models have the potential to boost the area of sentiment analysis. This would make it possible for these algorithms to make decisions that are more open to scrutiny and insightful. These results add to a better understanding of sentiment analysis as well as its usefulness in the context of significant athletic events and the dialogue that takes place on social media.

VIII. Conclusion and Future Work

In conclusion, the findings of our study have offered significant insights into the analysis of sentiment during important sporting events, with a particular emphasis on the FIFA World Cup 2022. In order to conduct an analysis of the feelings that were shared on Twitter in the course of the event, we made use of the capabilities offered by deep learning models, in particular roBERTa, distilBERT, and XLNet. In addition, we used LIME, which stands for Local Interpretable Model-Agnostic Explanations, in order to improve the interpretability of the judgments that these models produced.

Our most important discoveries illustrate the efficacy of deep learning models, in particular roBERTa, in effectively collecting attitudes in an environment characterized by constant change and activity on social media. These models achieved excellent levels of accuracy and F1 scores, demonstrating their capacity to comprehend and categorize the feelings stated by users in the course of a significant international athletic event.

In addition, the introduction of LIME has shed light on the decision-making process of these models, therefore offering invaluable insights into the primary characteristics and words that have an impact on the categorization of emotion. This interpretability is essential for both academics and practitioners in order to comprehend the behavior of models and come to judgments based on accurate information.

Nevertheless, our investigation had to contend with a few constraints. Although it contains a large amount of information, the dataset is restricted to tweets written in English and is focused on the opening day of the FIFA World Cup 2022. If the dataset was expanded to include tweets written in a variety of languages and collected over a longer length of time, then a more in-depth comprehension of how people across the world felt about the event may be achieved. In addition, LIME is helpful in gaining significant insights; nevertheless, further study is required to investigate various interpretability methodologies and their potential applications in sentiment analysis.

Researchers have the opportunity to investigate the generalization of deep learning models across a variety of sporting events as well as languages for their future work. During events, using real-time sentiment analysis may give up-to-the-minute insights about the reactions of the crowd. In addition, the creation of interpretability tools and methods that are even more cutting-edge may further contribute to our improved comprehension of model choices.

Overall, our study makes a contribution to the area of sentiment analysis by illuminating its use during important sporting events and the significance of ensuring that models can be interpreted in a variety of contexts. We are able to continue to expand our knowledge of global sentiment in the digital era by resolving the limits that have been recognized, as well as by pursuing new paths of inquiry.

References

[1] J. Camacho-collados, K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. Espinosa Anke, F. Liu, and E. Martínez Cámara, "TweetNLP: Cutting-edge natural language processing for social media."

[2] D. Loureiro, F. Barbieri, L. Neves, L. Espinosa Anke, and J. Camacho-collados, "TimeLMs: Diachronic language models from Twitter," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 251–260. [Online]. Available: https://aclanthology.org/2022.acl-demo.25

[3] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *ArXiv*, vol. abs/1910.01108, 2019.

[4] Z. Yang, Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *CoRR*, vol. abs/1906.08237, 2019. [Online]. Available: http://arxiv.org/abs/1906.08237

[5] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[6] S. Dewi and D. B. Arianto, "Twitter sentiment analysis towards qatar as host of the 2022 world cup using textblob," *Journal of Social Research*, vol. 2, no. 2, pp. 443–455, 2023.

[7] M. Faisal, Z. Abouelhassan, F. Alotaibi, R. Alsaeedi, F. Alazmi, and S. Alkanadari, "Sentiment analysis using machine learning model for qatar world cup 2022 among different arabic countries using twitter api," in *2023 IEEE World AI IoT Congress (AIIoT)*. IEEE, 2023, pp. 0222–0228.

[8] R. Patel and K. Passi, "Sentiment analysis on twitter data of world cup soccer tournament using machine learning," *IoT*, vol. 1, no. 2, p. 14, 2020.

[9] S. U. Hegde, S. B. Basapur *et al.*, "Distilbert-cnn-lstm model with glove for sentiment analysis on football specific tweets." *IAENG International Journal of Computer Science*, vol. 49, no. 2, 2022.

[10] G. R. Narayanaswamy, "Exploiting bert and roberta to improve performance for aspect based sentiment analysis," 2021.

[11] R. E. Demillo, G. Solano, and N. Oco, "Philippine national elections 2022: Voter preferences and topics of discussion on twitter," in *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 2023, pp. 724–729.

[12] A. Kruspe, M. Häberle, I. Kuhn, and X. X. Zhu, "Cross-language sentiment analysis of european twitter messages duringthe covid-19 pandemic," *arXiv preprint arXiv:2008.12172*, 2020.

[13] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin, "Learning sentiment-specific word embedding for twitter sentiment classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2014, pp. 1555–1565.

[14] V. Basile and M. Nissim, "Sentiment analysis on italian tweets," in *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2013, pp. 100–107.

[15] T. AI, "Fifa world cup 2022 tweets," Dec 2022. [Online]. Available: https://www.kaggle.com/datasets/tirendazacademy/fifa-world-cup-2022-tweets