# APPLICATION OF SENTIMENT ANALYSIS MODELS ON SOCCER COMMENTARY

*By*

MEHUL RAMASWAMI (Spire ID: 31678709)

*Produced as part of*

Independent Study (COMPSC 496)
Spring 2022

*Under the Guidance of*

Prof. Weiai Xu
Prof. Mohit Iyyer

May 12, 2022

*Manning College of Information and Computer Sciences*

# Table of Contents

# Abstract

This report performs a comprehensive comparison between popular sentiment analysis models using data from soccer commentary and analysis.  Specifically, the focus is on 3 models – Pointwise Mutual Information sentiment analysis model, Zero-shot text classification model, and the Bidirectional Encoder Representations from Transformers (BERT) model.

After running each model on the set of soccer commentary data (from 2 top clubs in the English Premier League), it appears that the Zero-shot text classification model works the best when compared to manual evaluation of content.  The primary reason for this model to stand out is that it has been built with a larger and more precise dataset.

Closer examination of the sentiments also indicated that commentors and pundits appear biased towards larger and well-known teams and continue to present them positively despite the result of the games. Furthermore, the margin of the win or loss does not seem to change the sentiments expressed.

# Background

This paper is inspired by past research conducted at Umass Amherst, specifically a paper on "Investigating Sports Commentator Bias within a Large Corpus of American Football Broadcasts" by esteemed NLP professors Mohit Iyyer and Brendan O'Connor (1). In this, a comprehensive method of investigating bias for sport commentators has been described in detail.

Initially the goal of this project was to analyze multitude of feedback, comments, and criticism that soccer players receive through social media, pundits, and commentators, and subsequently biases, and the impact said biases might have on players. But, access to certain social media APIs were restricted, which made it very difficult to gather such data.

Instead, the focus specifically turned to ESPN sports commentary of two soccer clubs in the English Premier League (Manchester United and Chelsea) from YouTube, and the data was used to do a comparison between popular sentiment analysis models. This was done by using YouTube transcripts, who's method of extraction has been explained in the paper mentioned above by professors Mohit Iyyer and Brendan O'Connor (1). Using these transcripts, extracted data to be fed into the sentiment analysis models, and the best-fit models were determined, and subsequently insights relating to bias towards certain team's performances was determined.

# Method

## Data Collection

For this study, YouTube transcript data was collected for all the English Premier League games involving Manchester United and Chelsea for the season of 2020-21. This was achieved by using the YoutubeTranscriptApi from python. Using this, the transcript was automatically loaded into the python script after passing in the ID of the YouTube video.

```python
import ytcc
import youtube_transcript_api

from youtube_transcript_api import YouTubeTranscriptApi

# assigning srt variable with the list
# of dictonaries obtained by the get_transcript() function

#In the get_transcript() function call below, add the ID of the youtube video
#For example, for the link, 'https://youtu.be/57uSMWTW414' the ID would be "57uSMWTW414"
srt = YouTubeTranscriptApi.get_transcript("MdJqSyEEqV4")
```

With this method, a total of 74 transcripts were collected, and stored as text files for easy access for analysis.

## Sentiment Analysis Method 1 – Pointwise Mutual Information

This method of sentiment analysis has been derived from a book written by Sudipta Mukherjee called "F# for Machine Learning Essentials" (2). Pointwise Mutual Information (PMI) essentially uses the information from a pair of words to compare the sentiment with respect to all the other words in the transcript of its origin, and all other transcripts in the dataset.

The first step in this process is define all the words based on the desired sentiment. An example of this is shown in the image below:

```python
#Defining the positive and negative words
pwords = ['nice', 'excellent', 'positive', 'fortunate', 'superior','fantastic', 'happy', 'great','brilliant'
nwords = ['disappointed','bad','poor', 'nasty', 'negative', 'wrong', 'inferior', 'critical', 'lack', 'lacks'
```

Next, we go through all pairs of adjacent words in the transcript and calculate their PMI.

Assuming the two words in a pair are w1 and w2, the PMI score is given by the following formula:

$$PMI(w1, w2) = P(w1, w2) / (P\_all(w1) + P\_all(w2))$$

Here, the numerator P(w1, w2) is the probability of both the words occurring in the same document (or transcript in this case). And, in the denominator, P_all(w1) represents the total probability of the word w1 occurring in all the documents put together; and the whole denominator is the sum of that probability for both the words.

After applying this formula on all the adjacent pair of words, and subsequently getting a PMI score for each, all the scores are added together, resulting in the final score for the document (transcript). If the resulting score in less than zero, then the overall sentiment of the document is Negative, and if above zero, the sentiment is Positive.

An example of a score is given below:

```
['1-3-2021.txt']
Document score = -42.112172881637086
```

## Sentiment Analysis Method 2 – Zero Shot Classification

The Zero Shot classification model is an extremely powerful model that was initially developed by Facebook. It now very easily accessible for use in the 'transformers' module in python (this is a hugging-face transformer).

```python
#Running the zero shot text classification model on the data
!pip install transformers
from transformers import pipeline

classifier = pipeline("zero-shot-classification")
```

This model consists of pre-trained data, which can be used in a way that enables the user to pass in any target labels, using which sentiment analysis can be performed. In other words, it is a model that allows classification of data which hasn't been used to build the model.

For the case of this project, two sets of labels were used:
                    ["Praise", "Criticism"] and ["negative", "positive"]

The resulting score would be the label, along with its probability:

```
['1-20-2021.txt']
positive, 0.6419761180877686
```

## Sentiment Analysis Model 3 – BERT

Bidirectional Encoder Representations from Transformers (BERT) is one of the most popularly used NLP machine learning models. This model is specifically used for detecting Positive or Negative sentiments. This model essentially uses word embeddings and tokenizers to classify the input text.

A pre-trained form of this model can be easily accessed using the same 'transformers' python module mentioned earlier.

```
!pip install transformers
from transformers import pipeline
classifier_2 = pipeline("sentiment-analysis")
```

This model is primarily built to classify a single, or a few sentences, and there is therefore a limit of around 1000 words that the model can take at a time. Due to this, for this project, each of the transcripts were broken into chunks of 500 words and passed into the BERT model.

Following this, each chunk would receive a sentiment, and after their combination, a final sentiment would be determined. The result from this model would just be the

```
['1-17-2021.txt']
NEGATIVE
```

# Findings

## Finding 1 – Zero Shot Classification is the Best Model

This results from all the models were compared with the manually tagged sentiments that were determined by looking at all the transcripts. Here is the result based on the models' comparison with manually tagged sentiments. In other words, higher the number, higher is the model's match to real world sentiments.

Pointwise Mutual Information Model = 52.7%

Zero Shot Classification Model (Positive, Negative) = 87.83%

Zero Shot Classification Model (Praise, Criticism) = 82.79%

BERT Model = 59.45%

From these, it was clear that the Zero Shot model performs best, even though the sentiment labels that are passed into the model are different. The main reason for this is the massive pre-trained dataset. Using this, this model is easily able to take in the desired sentiment labels and train itself to predicted based on them, essentially making use of all the occurrences of these sentiments in the pre-trained data.

As for the PMI approach, the main issue is that is only looks at pairs of words and not the whole sentence or paragraph. For example, if the pair was "not good", the resultant PMI score would not necessarily be negative since the pair has a positive and negative part to it. The other two models take this issue into account better, since they learn the other of phrases such as these through machine learning, and thereby correctly predict the sentiment.

In the case of the BERT model, the problem is that it is built only to take in only a few sentences at a time, and not a large paragraph, which is the case for all the transcripts in the extracted data.  Due to this, when the chunks of words are created, some information tends to be lost, and instead of calculating the overall sentiment, just that for the separate chunks is calculated. Now, if all the chunks show the same sentiment, then result is fine; but if not, it becomes an issue in determining which chunk is more important to the overall sentiment.

Based on these results, in can be said that the Zero Shot Classification model is the best for finding sentiments for sport commentary that was examined.

## Finding 2 – Commentators/Pundits Biased towards Manchester United

Using the results from the Zero-shot Model, it was found that the Positive sentiment is mostly observed when Manchester United wins, or in some cases if the match was a draw.
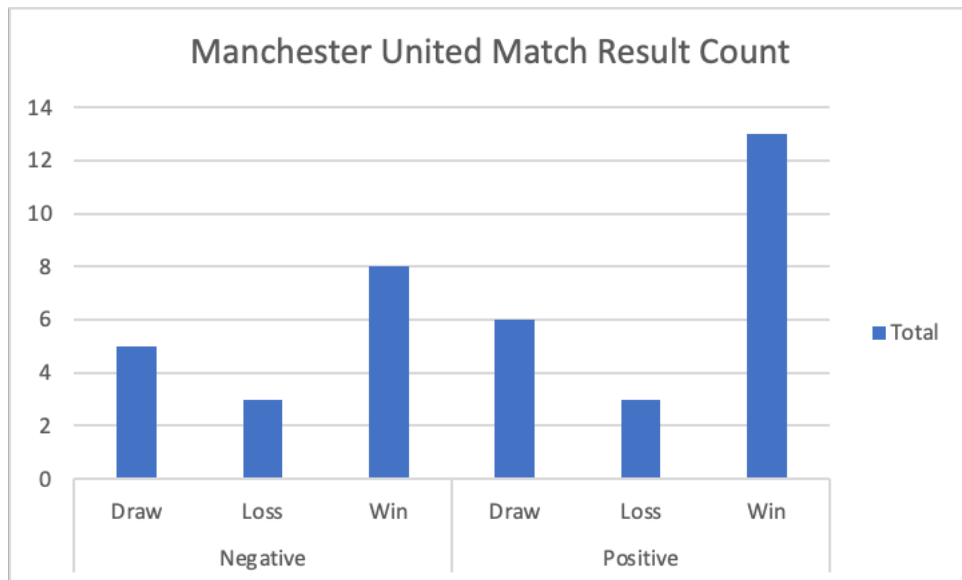


From the graph above, it can also be seen that the sentiment is negative only half the time when Manchester United wins, and for the rest of the cases, when they lose.

This essentially shows that the commentators are biased towards Manchester United, and their sentiments are very proportional to this club's result. This would not be a good sign for the overall audience of their show, since other teams would not necessarily be praised when they do well, and similarly not be criticized if they play poorly. Due to this, they might only get engagement from only those fans who support Manchester United.

## Finding 3 - Commentators/Pundits not reliant on Game Score for analysis

Using the results from the Zero-shot model again, it was found that the actual score of the game doesn't really have a say in what the pundits/commentators are saying.
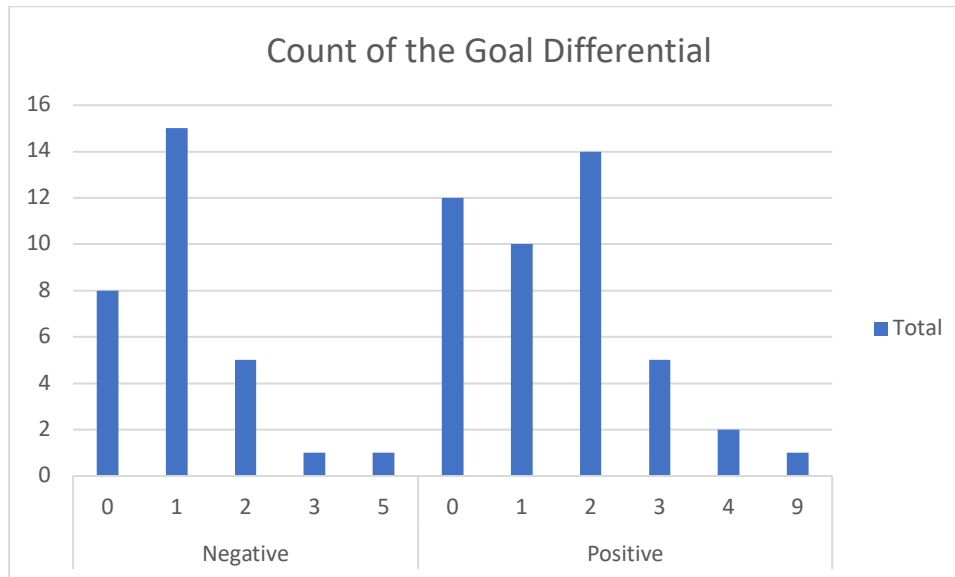


The graph above shows the difference between the goals scored by two teams in a game. For example, on the y-axis, the number "0" signifies that the game was a draw, or "1" signifies that a team won the game by a 1-goal margin.

From this graph the overall trend that can be seen is very similar for both the sentiments. In other words, the score differential is not impacting the sentiments expressed. This essentially means that the pundits are possibly not analyzing the game based on the score-line, but only on the result or outcome. This might also mean that they are not truly analyzing the events of the game, but just seeing it in the perspective of the so-called bigger team.

# Conclusion

The main finding of this paper was that the Zero-Shot Classification model is clearly the best for classification and sentiment analysis of soccer (or sport) commentary. The model can used by the analysts of sports broadcasting companies, such as ESPN, or even by those of certain soccer teams.

For the broadcasting companies, this would be very useful in determining how the sentiments from their commentators/pundits are affecting their views, or sales. For the analysts of the soccer clubs, it can be used to see if the sentiments from the commentators/pundits influence their gameplay. This could potentially be the case especially in current times where most of the players are very active on social media, and do listen to what critics say about them.

The flexibility of the Zero-shot model would also enable it to be useful in a wide range of genres, since its pre-trained data is vast.

All code and data gathered for this project can be found in the Google shared folder (3), whose link is in the References section.

# Acknowledgments

This independent study would not have been possible without the constant guidance and encouragement of Prof. Weiai Xu.  I am extremely grateful for his inputs from the several meetings we had.

A special call out to Prof. Mohit Iyyer. He facilitated the approvals such that this study could be performed as an independent study under the computer science departments.

Finally, my sincere gratitude to the Manning College of Computer Science for providing me the facilities and flexibility for me to solicit guidance from two professors, one from within CICS and another from the Department of Communication.

# References

(1) *Merullo, Jack, et al. "Investigating Sports Commentator Bias within a Large Corpus of American Football Broadcasts." ArXiv.org, 19 Oct. 2019, https://arxiv.org/abs/1909.03343.*

(2) *Mukherjee, Sudipta, "F# for Machine Learning Essentials", Packt Publishing Ltd., Feb 2016.*

(3) Link to shared folder: https://drive.google.com/drive/folders/1JCzP24T2UGGZvOvCSe-pfxsrAXRck3vF?usp=sharing