Analyzing Public Sentiment on Controversial Sports Events in YouTube Comments

Yuvraj Singh
IIIT Bhubaneswar
Bhubaneswar, Orissa, India
yuvraj.mist@gmail.com

Devadripta Jadhav
Savitribai Phule Pune University
Pune, Maharashtra, India
devadripta@gmail.com

Samiksha Boduwar
Savitribai Phule Pune University
Pune, Maharashtra, India
samikshaboduwar@gmail.com

Kripabandhu Ghosh
IISER Kolkata
Mohanpur, West Bengal, India
kripaghosh@iiserkol.ac.in

Abstract

This paper presents an analysis of YouTube comments on famous and controversial Public Sports Events. We explore public stance (stance detection) on a total of 6 famous controversial sports incidents by extracting and processing YouTube comments. Stance detection is performed on multiple events, including *The Underarm Incident*, *Jonny Bairstow's Run-Out* Incident, *Ashwin's Mankadding* Event, *Luis Suarez Handball* Event etc. The complete event details, results and evaluation metrics will be discussed in detail in subsequent sections.

CCS Concepts

• **Information systems → Web mining**.

Keywords

Stance Detection, YouTube Comments, Social Media Analysis, Sports Controversies

ACM Reference Format:
Yuvraj Singh, Devadripta Jadhav, Samiksha Boduwar, and Kripabandhu Ghosh. 2018. Analyzing Public Sentiment on Controversial Sports Events in YouTube Comments. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, **??** pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Sports engages billions of followers worldwide[1] and impacts the economy Sports controversies often ignite passionate discussions among fans, analysts, and players. With the rise of social media, platforms like YouTube have become central to these discussions. This study aims to analyze the stances or perform opinion mining namely for, against, and neutral on comments from famous social media platforms like YouTube, focusing on events such as Jonny Bairstow's Run-Out Incident, Luis Suarez Handball Event etc. To our knowledge, the first-ever study of civic engagement in controversial sports events (cricket and football) spans around 40 years. LLMs (Llama3 family) were used for initial annotations (stance) of comments and later fine-tuned for comparative performance analysis (30% boost in accuracy).

Thereafter, a humanely verified dataset of 30,000 thousand comments is released. KRIPA: Cite these https://aclanthology.org/search/?q=sports and explain why our work is different

2 Methodology

2.1 Data Collection

First write about the six events in brief and state why did you choose these events (they had many comments, they span across a long duration of time, are from two very popular sports in the world and covering multiple continents. We identified YouTube videos related to each controversy based on quality and engagement (on average > 5k comments) metrics. Our methodology focussed on the creation of a curated playlist for each event, ensuring diverse opinions.

The data extraction process involved:

(1) Identifying a famous public sports controversy from relevant sources like YouTube, Wikipedia and news sources based on quality and engagement.
(2) Identifying relevant videos for the chosen public sports controversies.
(3) Extracting comments using the YouTube Data API for each of the identified videos.
(4) Sorting the videos by the number of comments. Selected the top 50 (average) with the highest amount of comments.
(5) Stored the comments for the top 50 videos in a structured CSV format.
(6) Repeated the process until we had such data for 10 such controversies.

A summarized version of the extraction code is available, and the full code repository link will be available at our Github repository.

2.2 Data Processing

The extracted comments were preprocessed as follows -

[1] https://www.statista.com/chart/14329/global-interest-in-football/

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

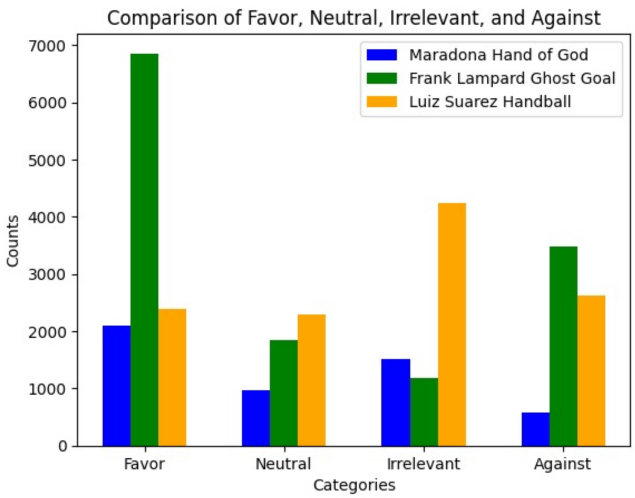

Figure 1: Number of *Favor*, *Neutral* and *Against* labels

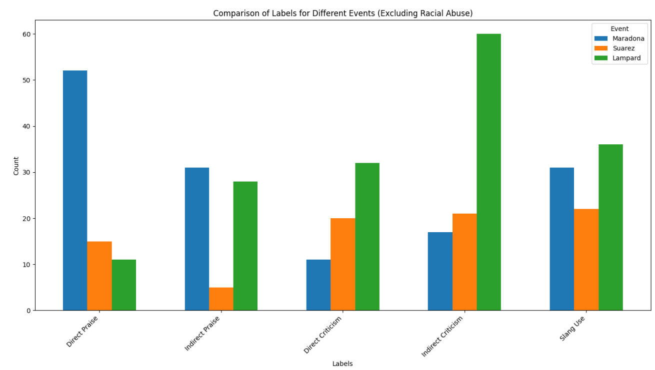

Figure 2: Comparison of Types of Praise (Favor) and Criticism (Against) for a sample of 200 comments.

(1) Removed special characters, stopwords, and non-English comments. We focussed on the evaluation of mostly English content.
(2) Irrelevant columns such as nested replies, time of comment and other metadata were removed.
(3) Additional cleaning steps included normalization and duplicate removal, which were essential to enhance the accuracy of the subsequent sentiment and stance analysis.

Controversy Name	Number of Comments
Ashwin Mankading	3785
Frank Lampard Ghost Goal	13520
Johny Bairstow Runout	6073
Luis Suarez Handball	11546
Maradona Hand of God	5159
The Underarm Incident	3676
Total	43759

Table 1: Name of the controversies used and their number of comments. KRIPA: Please put the values.

2.3 Stance Detection Pipeline

The process for stance detection on our curated dataset constitutes of two stages as follows -

(1) **Stage-I** (Preparation of "Gold Standard" labels:
  (a) **First pass** - In our first pass, we used an open-sourced LLM from HuggingFace, namely Llama 3.1-8b-Instruct. Custom prompts were constructed for each of the controversies from the dataset along with clear instructions to perform stance detection (for, against, neutral and irrelevant) on the comments for each of these controversies.
  (b) **Second pass** - Subsequently, the generated labels were humanely verified to account for misclassification and/or if further tuning of the prompt was needed to better the distribution of the data.
  (c) **Third Pass** - The tuned few shot prompts were used to perform the stance detection on our dataset and the final labels were generated thereafter.
(2) **Stage-II** (Fine Tuning of LLMs on the stance-labeled dataset) :
  (a) We used models from the Unsloth library which provides 70% reduction in memory usage and up to 2x inference speed.
  (b) We chose models like Llama 3.1-8b-Instruct model (same as used for labelling the dataset) to account for the difference made in the metrics when the same model is to be fined-tuned on our dataset.
  (c) Fine tuning on our dataset led to an average increment of 30% in terms of accuracy measured and drastic improvement in metrics such as F1, precision and recall as shown by confusion matrix and classification reports compared to its non-fine-tuned model(no data balancing techniques were employed).

The following details the above-mentioned pipeline for each of the controversies used to constitute our dataset KRIPA: there should NOT be separate strategies for different events. If it is done, it needs to be justified. -

2.3.1 The Underarm Incident. For the Underarm Incident, we employed a fine-tuned LLaMA-3 model from Unsloth. A structured prompt was used to classify comments into four categories: **For, Against, Neutral,** and **Irrelevant**. The responses were returned in a JSON format and then parsed to extract the stance label and the underlying rationale.

We followed the above-mentioned process for the *Maradona Hand of God* event, *the Luis Suarez Handball* Event and the *Frank Lampard Ghost Goal* Event.

2.3.2 Jonny Bairstow's Run-Out and Ashwin's Mankadding Events Using the OLLAMA Framework. For Jonny Bairstow's Run-Out Incident and Ashwin's Mankadding Event, we utilized the OLLAMA framework. Detailed API requests were sent with prompts explaining the context of each event, and JSON responses were parsed to extract the stance label and associated reason.

2.4 Algorithm Details and Rationale

The following outlines the overall pipeline and reasoning behind our stance detection approach:

- **Pipeline Design:** The pipeline starts with data extraction and preprocessing to ensure the quality of the input comments. Given the noisy nature of social media data, thorough cleaning was essential.
- **Model Selection:** For the Underarm Incident, the fine-tuned LLaMA-3.1 (Instruct) family of models were chosen for its ability to process long sequences (up to 2048 tokens) and follow the given instructions (as a few shot prompts) to generate coherent responses, making it suitable for detailed stance detection.
- **Structured Prompts:** We used structured prosmpts to guide the model in classifying comments. This method provided consistent JSON responses, ensuring ease of parsing and reliable extraction of stance labels and reasons.
- **OLLAMA Framework:** For Jonny Bairstow's and Ashwin's events, the OLLAMA framework allowed for scalable and concurrent processing of comments via API calls. This was critical in handling larger datasets and ensuring a rapid turnaround in analysis.
- **Evaluation Metrics:** In addition to the stance labels, we compute evaluation metrics such as accuracy, precision, recall, and F1-score to assess model performance.

Figure **??** presents a flowchart summarizing the stance detection pipeline.

For clarity, the pseudocode in Algorithm **??** summarizes the pipeline:

pipeline_flowchart.png

Figure 3: Stance Detection Pipeline Flowchart

Algorithm 1 Stance Detection Pipeline

Input: YouTube comments dataset
Preprocessing: Clean comments by removing noise and duplicated data
if Incident is Underarm **then**
    Use the Unsloth LLaMA-3(Instruct) family of models with a structured prompt
    Parse JSON response to extract stance label and reason
else
    Use the OLLAMA framework with API requests and detailed prompts
    Parse JSON response to extract stance label and reason
end if
Output: Stance labels and evaluation metrics (accuracy, precision, recall, F1-score)

3 Results and Discussion

Preliminary analysis indicates a significant division in public opinion across the six events within our dataset. Fine Tuning on our dataset improves the accuracy of the labels by a drastic margin along with other metrics such as F1 score, recall and precision as compared to the base instruct model.

Detailed results, including the distribution of stances (For, Against, Neutral, Irrelevant) and evaluation metrics (accuracy, precision, recall, F1-score).

3.1 Detailed Analysis

(1) *Number of samples* (**Favor**, **Against**, **Neutral** and **Irrelevant**)
   (a) The labels, *Favor* and *Against* is significantly higher for *Frank Lampard Ghost Goal* compared to other events with *Favor* being comparatively higher, followed by *Luiz Suarez Handball* event.
   (b) The *n*umber of samples for Neutral label is higher for *Luis Suarez Handball* event.
   (c) The label *Irrelevant* is significantly higher for *Luis Suarez Handball* event meaning the majority of the comments couldn't be classified into the other three labels.
(2) *Variations of Praise and Criticism*
   (a) Instances of *Direct Criticism* and *Racial Abuse* is highest for *Maradona Hand of God* event.
   (b) *Direct Praise* accounts highest for *Luis Suarez Handball* with equal instances for *Maradona Hand of God* and *Frank Lampard Ghost Goal.*

Event	Favor	Against	Neutral
Frank Lampard Ghost Goal	1. Germans can't say anything about unsporting behavior. 2. What a disgrace this Manuel is. 3. Could have been one of the greatest games.	1. The 1966 ghost goal had to be paid for. 2. Payback for the 1966 final 3. Even Geoff Hurst said his goal didn't count.	1. This was way more clear-cut than 1966. 2. Inconclusive! Just couldn't get a good angle. 3. I remember this game.
Luis Suárez Handball	1. Morality always loses, and nice guys finish last. 2. I just love Luis Suarez with all his behaviors I don't know why. 3. By apologizing, he would look stupid.	1. Suárez cost Ghana the World Cup semi-final. 2. This man was often mentally unstable. 3. Hand of Satan.	1. He will never step foot in Ghana. 2. That was back then, now there's goal-line technology. 3. It's part of the game.
Maradona Hand of God	1. Yup. Goalies s*cked these days. 2. The Real OG of all times. 3. Number 15 goal is something else...my favourite. Bravo.	1. Errrrr! Hand of god? Cheater. 2. He makes the opponent look so stupid and clumsy. 3. The most cheating player in football history.	1. When the football was football. 2. Back in the day nobody could play football, that's why he appeared to be that good. 3. You will never be able to pick one between Maradona and Messi.
Jonny Bairstow Run-Out	1. That's not cheating, that's the way of winning. 2. They gave Jonny ample time to stop walking out of his crease, so fair game! 3. Jonny was too quick to get out of the crease, so it is totally a fair decision.	1. Same old Aussies, always cheating. 2. Australia cheating again. Who could have seen that coming? 3. Not cheating, just bad sportsmanship.	1. The lesson for the players is "pay attention." 2. Growing up in Australia, we were taught to throw the ball at the batters' head if they left their crease like this. 3. This is just pure drama and I love it.
Ashwin Mankadding	1. If a bowler can keep his foot inside the crease, a batsman can wait with the bat inside the crease until the ball is bowled. What's wrong with that? 2. This is out according to the rule. 3. Great presence of mind... genius of the game.	1. If you Mankadding, you should be ashamed of yourself. 2. Not a fair runout considering sportsmanship and the spirit of the game. 3. This rule is absurd. How can you be allowed to run out someone without delivering the ball?	1. Ashwin merely expressed his disappointment but never wanted a wicket that way. 2. Whenever a batsman leaves the crease, he gets 1 to 2 ft benefits in running. 3. Batters don't do it on purpose; they walk out expecting the bowler to deliver the ball.
The Underarm Incident	1. That time, it was legal to bowl underarm according to the rules. 2. The bottom line is that there was nothing in the rules at the time saying you can't bowl underarm, so technically, nothing wrong was done. 3. Well, if they were allowed to do it, then fair play to them.	1. That was against the spirit of the game! Couldn't they just bowl a normal delivery? 2. The greatest cowards in the world... None other than the Aussies. 3. Poor sportsmanship.	1. What were the exact rules for underarm deliveries back then? 2. Carrying on about something decades later as if it's relevant today is the real injustice. 3. Greatest highlight in cricket history, great footage.

Table 2: Examples of Favor, Against, and Neutral Comments for Controversial Events

Event	Direct Praise	Indirect Praise	Direct Criticism	Indirect Criticism
Maradona Hand of God	1. Maradona is sooo good.	1. 13 was simply incredible.	1. 3 was handball. Not a goal. A cheat.	1. You forgot to add the hand of God goal.
	2. 13 was simply incredible.	2. The legend forever.	2. Goal 12 looks like a huge offside.	2. Back when players played for the crowd, not money.
	3. The legend forever.	3. Back when players played for the crowd, not money.	3. Cheating and poor goalkeeping.	3. Maradona's skill ended when he cheated.
Luis Suarez Handball	1. Suarez is a legend... but seeing Asamoah after the match broke my heart.	1. Suarez took matters into his own hands .	1. Suarez crushed so many African dreams.. absolutely criminal.	1. If you watch closely Suarez wasn't even the only one who tried to handball.
	2. Suarez is a genius.	2. He did what he had to do for his country.	2. Absolute scumbag play.	2. Ghana would have won if it weren't for the handball.
	3. Suarez hero or villain?	3. Anyone would have done that tho.	3. Suarez will forever be the biggest disgrace in modern football.	3. It's funny though because both Argentina and Uruguay cheated and both got what they had coming.
Frank Lampard Ghost Goal	1. They were voted the most entertaining team on FIFA.com.	1. It took 12 years but Germany got their karma at last.	1. I hate Germany for what they did, it is so sad.	1.wouldnt have made a difference, im english and we would have lost anyway. German were by far the better team.
	2. Wouldn't have made a difference, I'm English and we would have lost anyway. Germany were by far the better team.	2. OSCAR winning performance from football.	2. England defended like a bunch of girls.	2. With all the technology today, this would never happen now.
	3. My idol was Frank Lampard and I'm so happy to see, but that referee was an absolute fuck, blind and should go to the eye doctor.	3. This is the reason why FIFA needs VAR...	3. That's just poor by the officials..smh.	3.The revenge has been taken

Table 3: Examples of Direct Praise, Indirect Praise, Direct Criticism, and Indirect Criticism

Event	Direct Praise	Indirect Praise	Direct Criticism	Indirect Criticism
Maradona Hand of God	1. Maradona is sooo good.	1. 13 was simply incredible.	1. 3 was handball. Not a goal. A cheat.	1. You forgot to add the hand of God goal.
	2. 13 was simply incredible.	2. The legend forever.	2. Goal 12 looks like a huge offside.	2. Back when players played for the crowd, not money.
	3. The legend forever.	3. Back when players played for the crowd, not money.	3. Cheating and poor goalkeeping.	3. Maradona's skill ended when he cheated.

Table 4: Examples of Direct Praise, Indirect Praise, Direct Criticism, and Indirect Criticism for Maradona Hand of God

Event	Example Comments
	Favor 1. I wonder what Germans think to unsporting behaviour. They don't think. 2. What a disgrace this Manuel is. 3. Could had been one of the greatest games of all time but ref decided, no....
Frank Lampard Ghost Goal	**Against** 1. Even Geoff Hurst in 1966 said my goal didn't count even though it was a goal. This is just revenge, I believe. 2. Geoff Hurst's "ghost" winning goal in 1966 had to be paid. 1970 - Losing the QF to Germany after leading 2-0. 1990 - Losing the semi-final on penalties. 2010 - Lampard's goal not seen by the assistant referee. The curse started right after the final whistle of the 1966 World Cup Final. 3. Payback for the 1966 final
	Neutral 1. Inconclusive! Just couldn't get a good angle on that. 2. Didn't cross the line. 3. Not enough evidence to overturn the decision.

Table 5: Example sentiment-based comments for Frank Lampard's Ghost Goal event

(c) For *Indirect Criticism*, *Frank Lampard Ghost Goal* is highest followed by *Maradona Hand of God*.
(d) In terms of *Favor* label (Direct + Indirect Praise), Frank Lampard event is highest, followed by Maradona and then by Luis Suarez events.
(e) Similarly, for *Against* label (Direct + Indirect Criticism), Maradona event's count is highest followed by Frank Lampard and then Luis Suarez events.

(3) *Fine Tuning of LLMs*
(a) Without fine-tuning, Llama3.1-8b Instruct got an average of 35.6 % and with fine-tuning on our constructed dataset, there was a drastic improvement of over 40 % to 76.71 %.
(b) This improvement is majorly due to the quality of labels associated with the respective comments after a thorough human verification.

3.2 Challenges Encountered

During preprocessing, challenges such as handling noisy data, duplicate entries, and variations in comment formats were encountered. For stance detection, ensuring reliable automated classification and consistent JSON parsing proved difficult. These challenges motivated the use of structured prompts and robust frameworks like OLLAMA.

3.3 Future Directions

Future work will focus on:

- Refining sentiment classification models using advanced machine learning techniques.
- Expanding the dataset to include a broader range of sports controversies.
- Enhancing preprocessing methods and fine-tuning model parameters to improve overall performance.

4 Ethical Considerations

In this study, the ethical use of publicly available YouTube data was ensured. All data were anonymized and processed in accordance with ACM's policies on research involving human subjects. Informed consent was addressed by using publicly accessible data without any direct identification of individuals.

5 Conclusion

This study highlights the role of social media in shaping public perception of sports controversies. The integration of automated data extraction and stance detection provides a comprehensive view of audience sentiment. Future enhancements will aim to improve accuracy and broaden the scope of analysis.

References

Temporary page!

LaTeX was unable to guess the total number of pages correctly. As there was some unprocessed data that should have been added to the final page this extra page has been added to receive it.

If you rerun the document (without altering it) this surplus page will go away, because LaTeX now knows how many pages to expect for this document.