

Twitter Sentiment Analysis: A Survey in Cricket and Bollywood



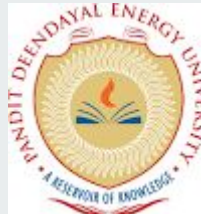
Authors:

Nayantara Kotoky, Smiti Singhal, Anushka Sharma, Dhara Ajudia

Presenter Details

Smiti Singhal - M.Tech. in CSE, IIIT Delhi, India

Anushka Sharma - B.Tech. in CSE, PDEU, India



Presenter Details



Smiti Singhal -

- B.Tech. in CSE from PDEU, Gandhinagar
- M.Tech in CSE from IIIT Delhi

Other published papers-

- Explainable AI (XAI): Core Ideas, Techniques and Solutions
- Intelligent Recruitment System using NLP

Anushka Sharma -

- B.Tech. in CSE from PDEU, Gandhinagar
- Work Experience in a Client-based Company in fields of Automation and Development

Other published papers-

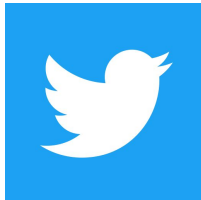
- Database Management Systems—An Efficient, Effective, and Augmented Approach for Organizations
- Intelligent Recruitment System using NLP

AIMS and CONTRIBUTIONS

Twitter has been the voice of the public for a long time now. With the rise in the usage of Twitter, the active participation of its users in expressing their views across all domains has significantly increased.

The aim of the project is to compare the influence of **Bollywood Stars** and **Indian Cricketers** on people.

- Classification of sentiments
- Twitter as a reflection of their performance
- Relative frequency of joint mentions inter and intra category



MOTIVATION



Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics.

Expression of opinions in tweets provides an opportunity for new explorations to find collective likes and dislikes of the community.

This will give us an idea of how powerful social media is when it comes to these widely discussed topics and whether or not these discussions result in something useful or valuable.

TOOLS USED



For Sentiment Analysis -

The tools we used for sentiment analysis were TextBlob , Empath, Pattern, Sentiwordnet and VADER. The most precise results were shown by VADER which accurately categorized the sentiments into positive and negative.

For Twitter as a reflection of performance -

Orange tool was used for the sentiment analysis of tweets based on the seven basic emotions proposed by Ekman : fear, anger, joy, sadness, disgust, appreciation, and surprise. Along with this tool, VADER and KHCoder are also used.

For Relative frequency of joint mentions inter-category and intra-category -

KH Coder is a text mining tool which is typically used for finding the potential relationships between entities represented within a document. Here, we first load the required file as the data and then pre-process it as a necessity for analysis. The tool is implemented to obtain co-occurrence matrices and networks to get insights into major themes from the text and analyze the associations between the text that appear together.

IMPLEMENTATION DETAILS - Data Collection

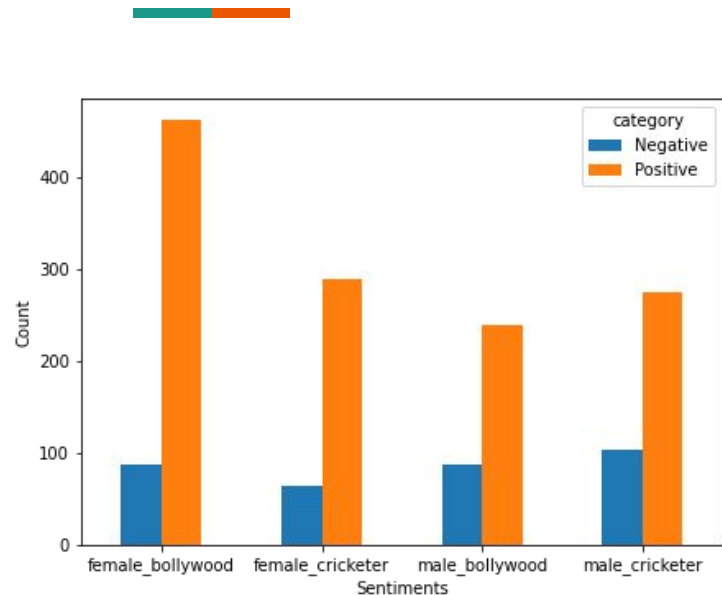


- ❖ Number of tweets collected: Around 6000 (All categories included)
- ❖ **Methodology** of Data Collection:

With the use of Twitter API, *Recent tweets* for **most popular** persons of all 4 categories were collected and merged together.

- ❖ **Steps** for Data Collection:
 - Importing **Tweepy** - an easy-to-use Python library for accessing the Twitter API
 - Authentication for Twitter Developer account
 - Defining List of hashtags for every category
 - Defining the date_since date as variables
 - Filtering retweets
 - Output data as csv file

IMPLEMENTATION DETAILS - Sentiment Analysis



Vader Results for input data

The tools we analysed for Sentiment Analysis were- Textblob, **Empath**, **Pattern**, **Sentiwordnet** and **Vader**.

The most precise results were shown by **Vader** which accurately categorized the sentiments into Positive and Negative.

We can conclude:

- The **positive** sentiments for the almost same amount of tweets is the *highest* for **Female Bollywood Stars**
- Similarly, the **negative** sentiments reflected through the tweets are the *highest* for **Male Cricketers**

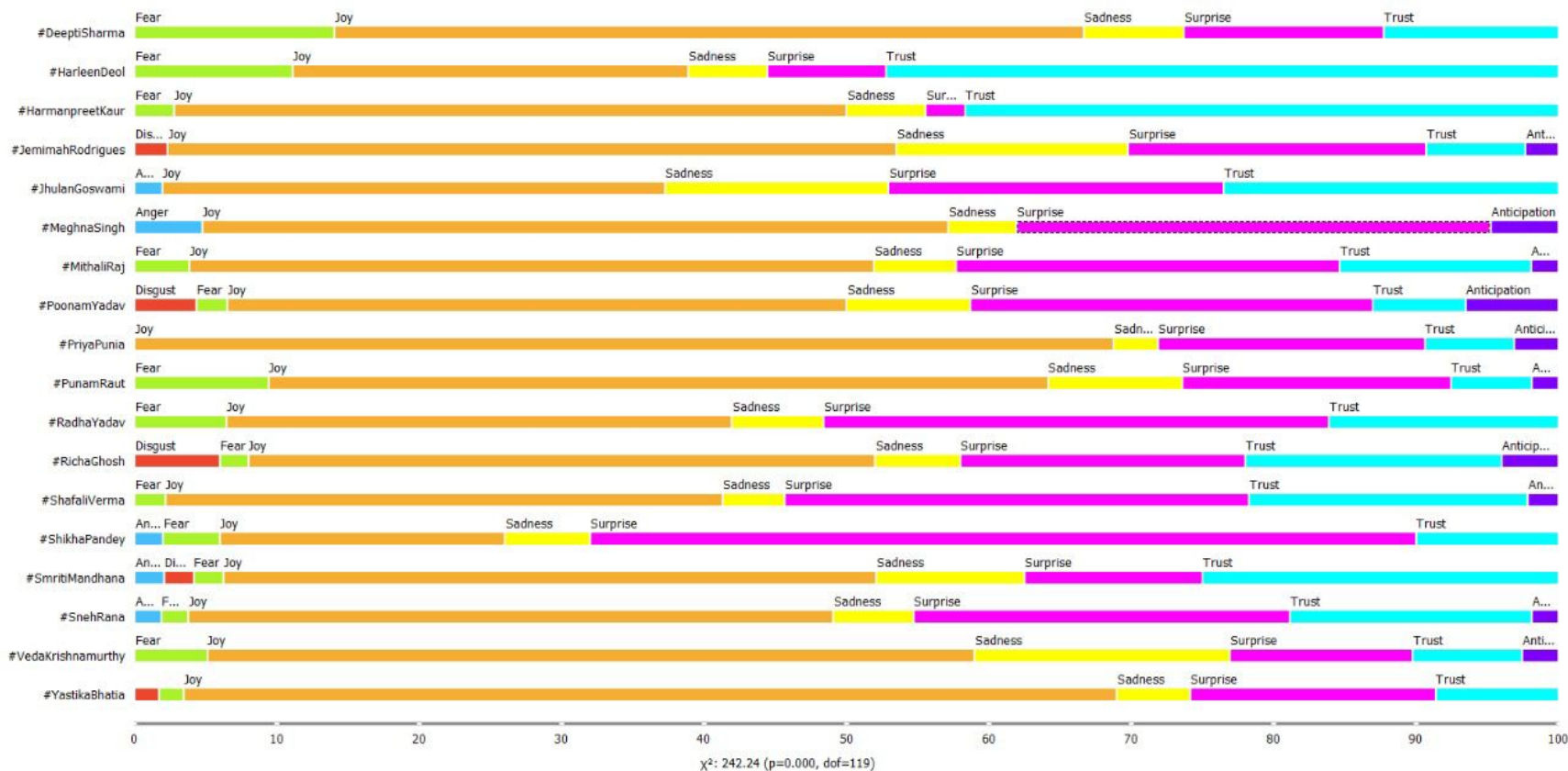
Twitter as a reflection of performance

Sentiment Analysis using TweetProfiler



- Tweet Profiler, a widget provided by orange, retrieves information about the emotions attached to the sentiment by sending data to the server where a model calculates the emotion scores/probabilities according to the text. This is then plotted with the help of a Box Plot. This analysis provides seven different sentiments in comparison to only positive and negative sentiments.
- Sentiment Analysis of Female Cricketers using Tweet Profiler is shown in figure (next slide). The figure clearly explains that Yastika Bhatia and Priya Punia have the highest percentage of joyful tweets while the tweets mentioning Shikha Pandey depicted highest percentage of surprise emotion.
- Yastika Bhatia and Priya Punia are indicated as having the highest positive sentiments using VADER as well as the most joyful emotions using Orange.

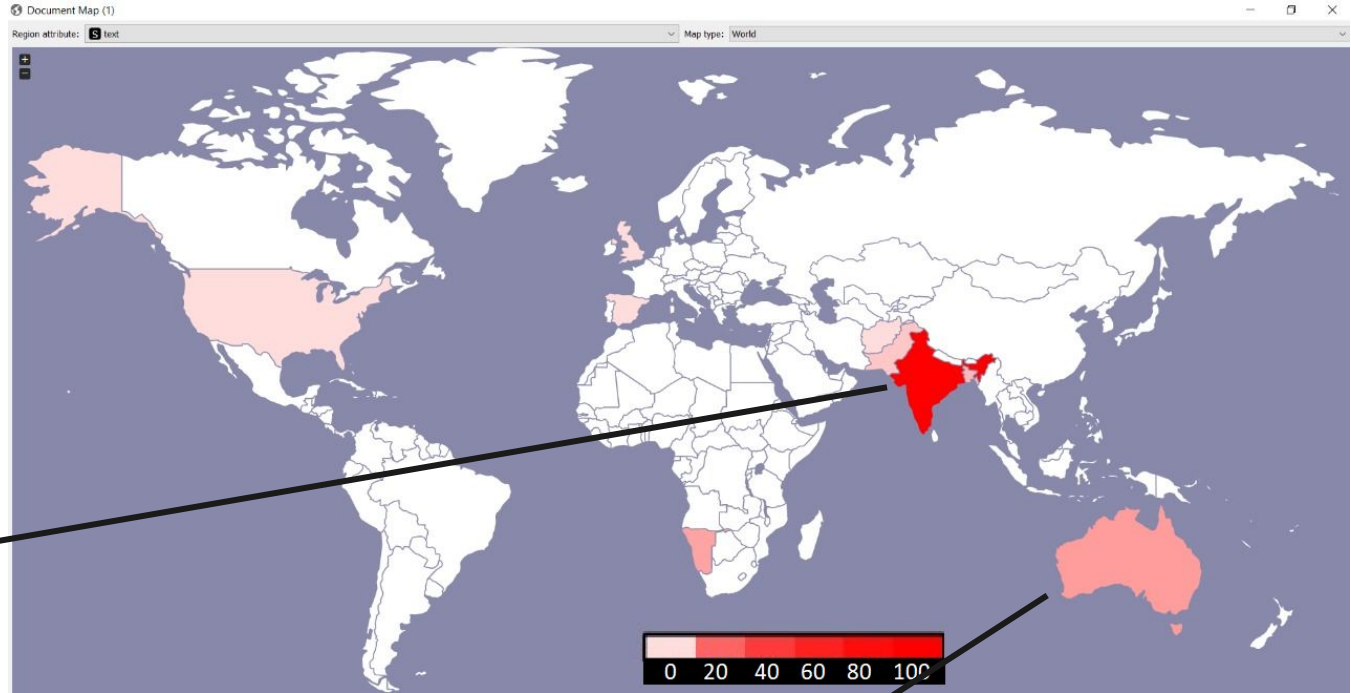
Sentiment Analysis using TweetProfiler



Note - Similar analysis has been done for other categories.

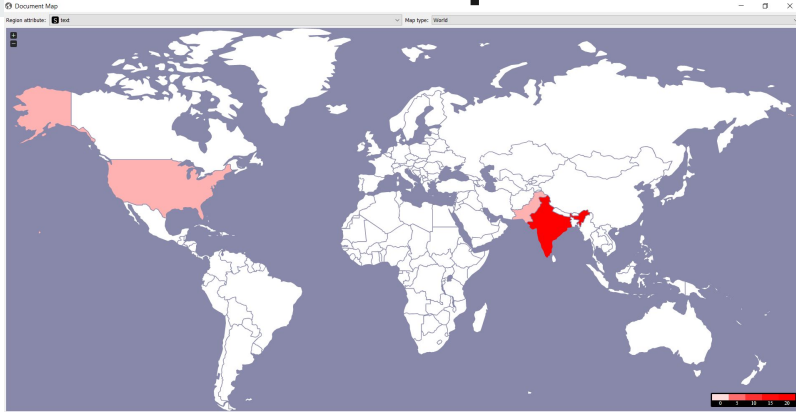
Document Map for Male Cricketers

Here, we can see that the number of mentions of India (Red) is considerably higher than other countries and these countries have been accurately displayed because of the recent T20 World Cup. The participants in the T20 World cup were Namibia, Pakistan, Afghanistan, England, Bangladesh, Australia, etc. which are highlighted on the world map too.

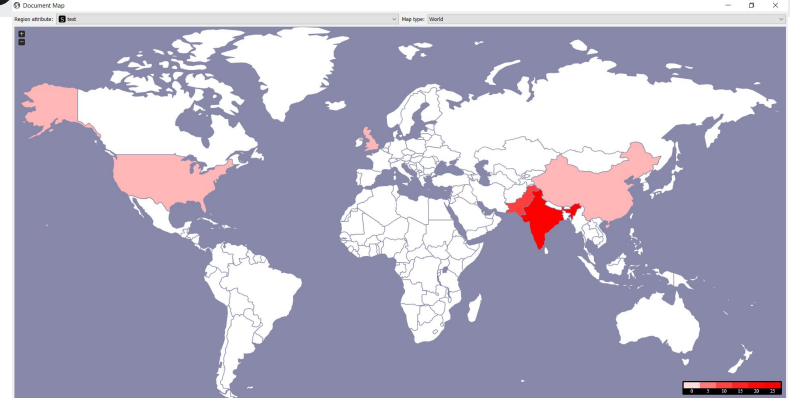


This shows that there were some tweets in which Australia was mentioned in the text.

Document Map for Other Categories



Male Bollywood

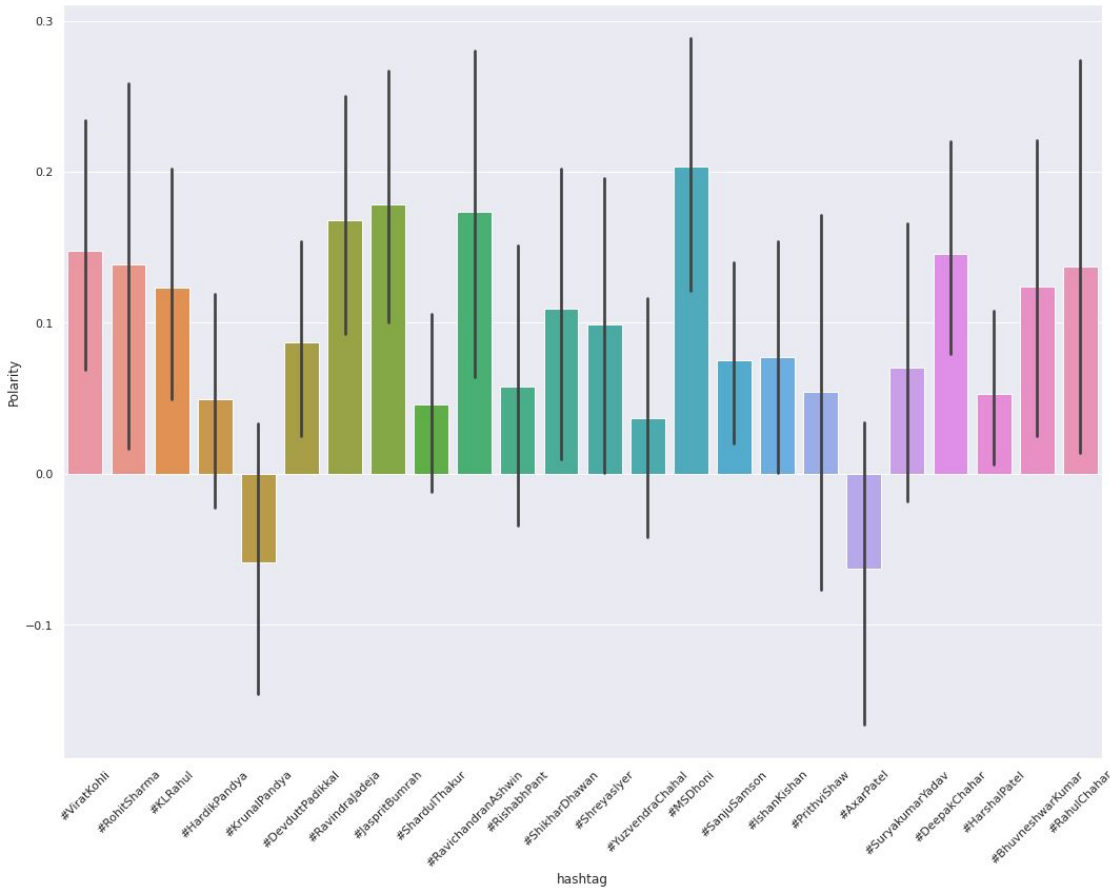


Female Bollywood



Female Cricketers

Male Cricketers



The results obtained were as follows -

- **Highly Appreciated : MS Dhoni**

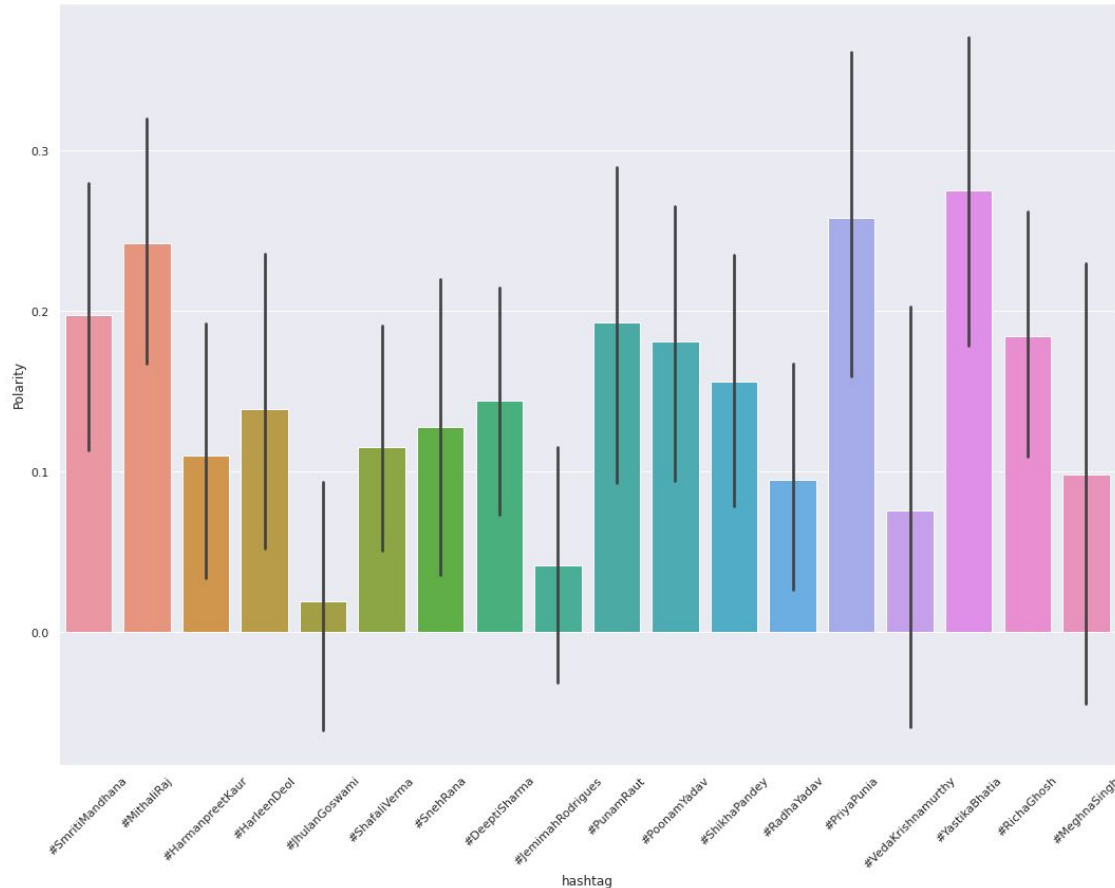
Why? In the light of the 2021 T20 World Cup, people were excited and happy because of Dhoni's presence as the mentor for Team India.

- **Involved in Negative Discussion : Axar Patel and Krunal Pandya**

BCCI had announced via their social media handles that Axar Patel would be replaced by Shardul Thakur in the 2021 T20 matches which clearly did not go well with a lot of people and hence people shared their disappointment through Twitter.

Krunal Pandya's poor performance in the IPL matches made the Twitterati furious.

Female Cricketers



The results obtained were as follows -

- **Highly Appreciated : Yastika Bhatia**

Yastika's brilliant performance in the 2021 India Vs Australia ODIs and the WBL matches was highly praised

Male Bollywood Stars



The results obtained were as follows -

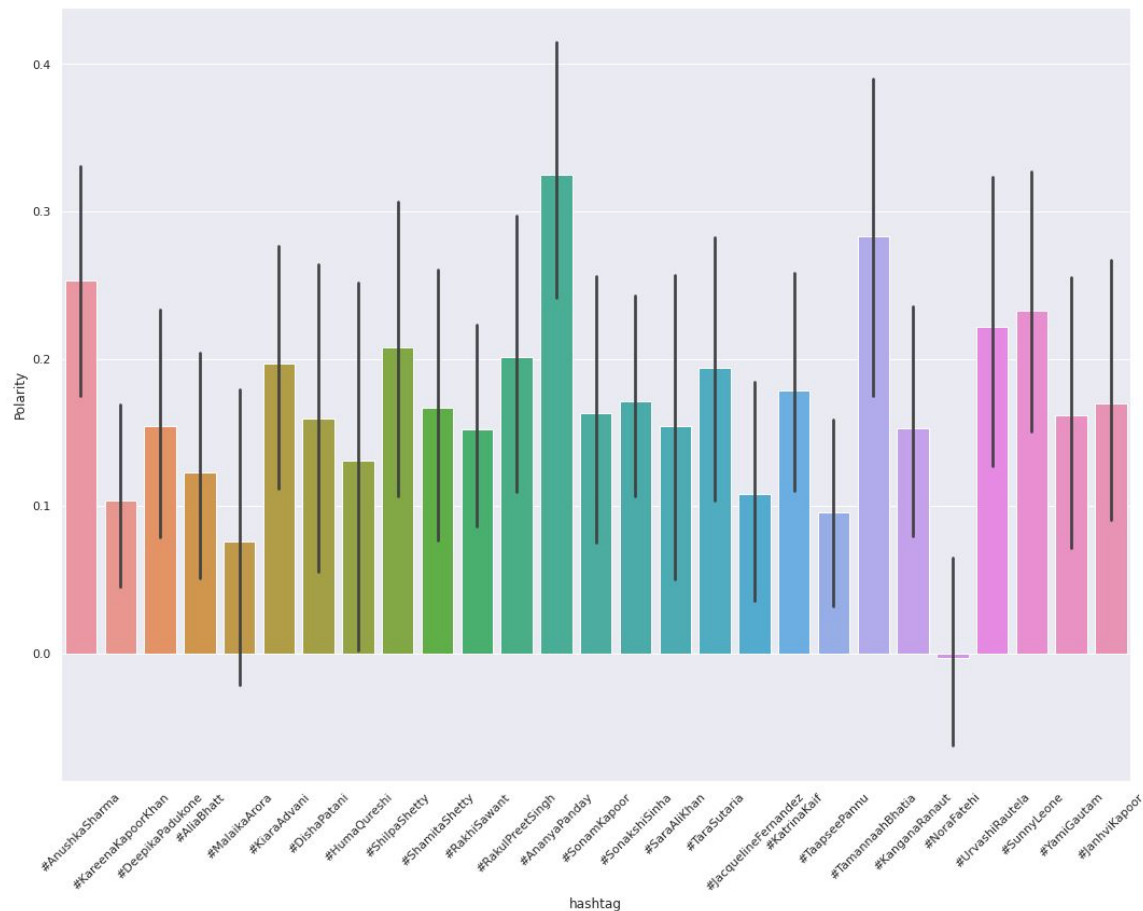
- **Highly Appreciated : Ranbir Kapoor**

Despite any recent project announcements or any other controversies, Ranbir still remains the most highly appreciated actor because of his huge fan following.

- **Involved in Negative Discussion : Nawazuddin Siddiqui**

Nawazuddin Siddiqui's statement garnered attention due to its critical comment on Racism being a bigger issue in the Bollywood Industry as compared to nepotism. People supported him and highly *criticized* the Bollywood industry.

Female Bollywood Stars



The results obtained were as follows -

- **Highly Appreciated : Ananya Panday**

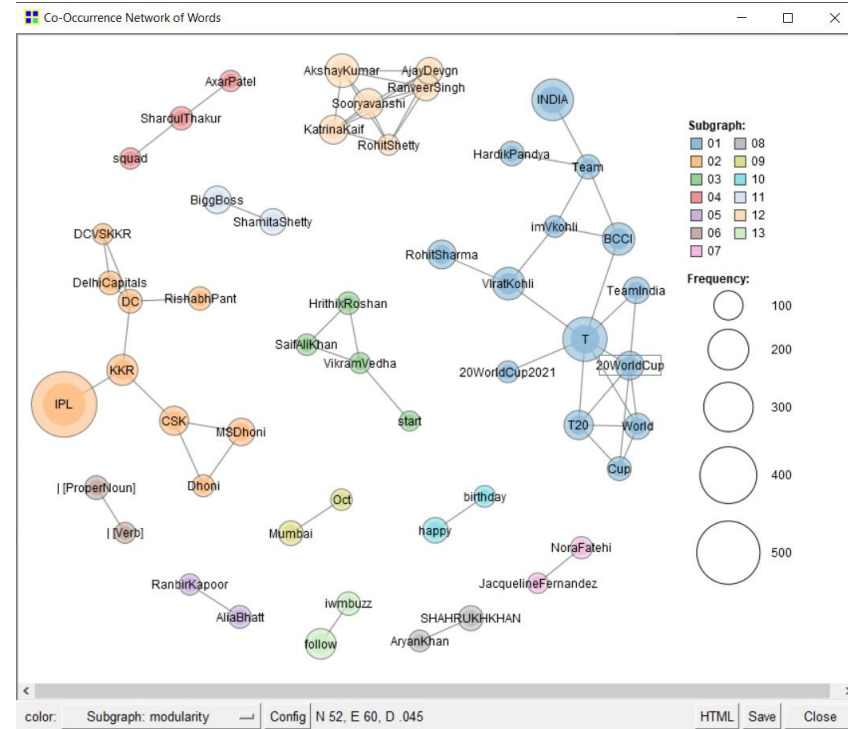
Ananya is highly active on social media and her fans regularly share positive comments on her posts which makes her highly appreciated

- **Involved in Negative Discussion : Nora Fatehi**

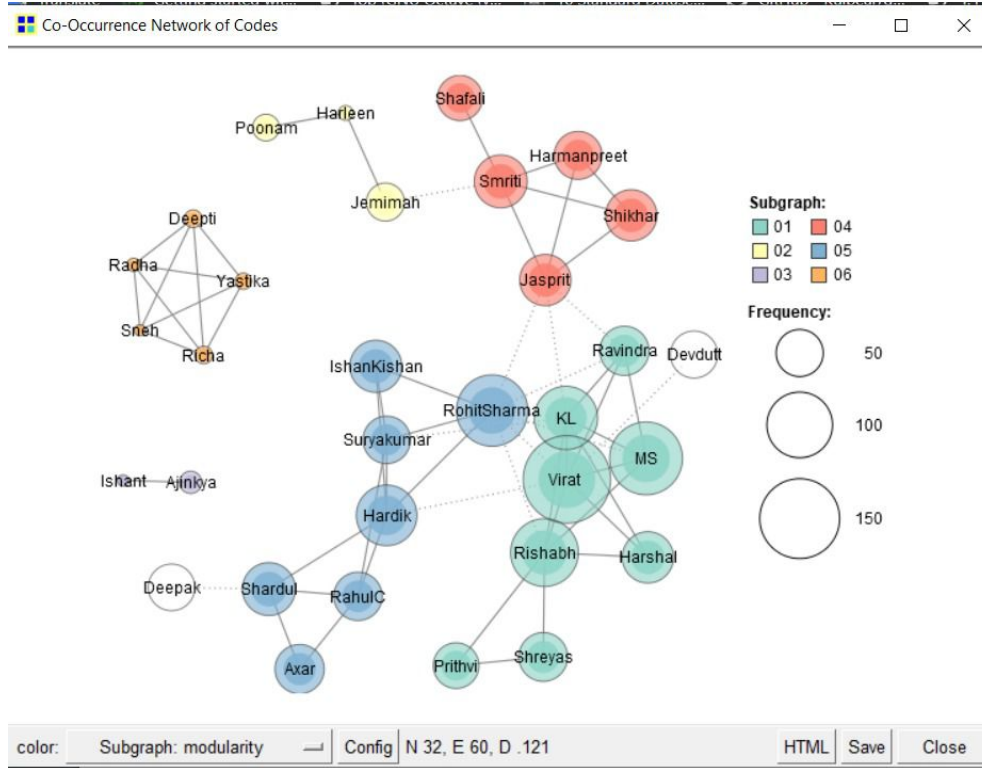
News of Nora being involved in money laundering and being summoned by the Enforcement Directorate.

Implementation Details - Relative frequency of joint mentions inter-category and intra-category

- This analysis clusters the nodes that have similar occurrence.
- In this work, it means that nodes which are often mentioned together by people's tweets are clustered, where nodes represent individuals from the four categories.
- The representation shows the words with similar appearance patterns, that is, with high degrees of co-occurrence, connected by lines.
- To determine edge strength, Jaccard coefficients are calculated for all possible combinations of target words.
- It was carried out on a combined document for all categories.



Network of Codes for Male and Female Cricketers.



- The size of the circles also shows that the frequency with which the Male Cricketers are discussed with each other within the category is much greater than when they are mentioned with the Female Cricketers.

- Also, the frequency with which the Female Cricketers are mentioned together within their category is much lower in frequency (which can be deduced by the size of the circle), indicating that

Female Cricketers do not receive as much popularity among the Twitter people as Male Cricketers.

Note - Similar analysis has been done for other categories.

Relative frequency of joint mentions inter and intra category

Co-occurrence Matrix using the tool - KHCoder

Coding: Jaccard Coefficients

Entry

Coding Rule File: Coding Unit:

Result

	*Smriti	*Mithali	*Harmanpreet	*Harleen	*JHulan	*Shafali
*Smriti	1.000	0.012	0.172	0.046	0.000	0.122
*Mithali	0.012	1.000	0.014	0.000	0.000	0.000
*Harmanpreet	0.172	0.014	1.000	0.000	0.000	0.032
*Harleen	0.046	0.000	0.000	1.000	0.000	0.000
*JHulan	0.000	0.000	0.000	0.000	1.000	0.000
*Shafali	0.122	0.000	0.032	0.000	0.000	1.000
*Sneh	0.015	0.000	0.019	0.000	0.000	0.021
*Deepthi	0.014	0.000	0.036	0.000	0.000	0.040
*Jemimah	0.079	0.000	0.024	0.091	0.000	0.026
*Virat	0.004	0.005	0.004	0.000	0.000	0.000
*RohitSharma	0.011	0.007	0.006	0.000	0.000	0.000
*KL	0.000	0.000	0.000	0.000	0.000	0.000
*Hardik	0.000	0.000	0.000	0.000	0.000	0.000
*Krunal	0.000	0.000	0.000	0.000	0.000	0.000
*Ravindra	0.000	0.000	0.000	0.000	0.000	0.000
*Jasprit	0.097	0.000	0.109	0.000	0.000	0.000
*Shardul	0.000	0.000	0.000	0.000	0.000	0.000
*Ravichandran	0.000	0.000	0.000	0.000	0.000	0.000
*Rishabh	0.000	0.000	0.000	0.000	0.000	0.000

The figure shows the Jaccard coefficient that shows the association of two individuals with respect to how much they are discussed together in the tweets.

- The red circle on the previous slide depicts the Jaccard coefficient for Virat Kohli and Smriti Mandhana. It means that:

$$|\text{Virat} \cap \text{Smriti}| / |\text{Virat} \cup \text{Smriti}| = 0.004$$

(No. of times both were mentioned together divided by No. of times either of them were mentioned)

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

INTERPRETATIONS -



1. While the maximum joint mentions are of male cricketers with fellow male cricketers (233), the joint mentions of female cricketers with female actors (1) and male actors (2) are negligible.
2. It can be concluded that women cricketers, in general are tweeted very less with other men as well as women compared to men cricketers.
3. On the other hand, there are quite a high number of tweets mentioning male and female actors together. This analysis shows resonance to the real-life phenomenon where movies include both genders, and hence their count of joint mentions is much higher than the other category.

	Actor Female	Actor Male	Cricketer Female	Cricketer Male
Actor Female	163	176	1	17
Actor Male	176	136	2	18
Cricketer Female	1	2	54	15
Cricketer Male	17	18	15	233

Conclusion and Future Work




Three different analyses have been presented on how discussions are held regarding Indian actors and cricketers using Twitter as a platform of expressing opinions. We have identified several interpretations of how the Twitter people view the four categories, that is, Male Cricketers, Female Cricketers, male actors and female actors.


Analyzing the most appreciated and criticized celebrities helps the brands publicize their products to customers by connecting with those celebrities for marketing purposes, which eventually helps the brands in making their product famous. Furthermore, we also observed that a well-balanced overview of current affairs can be acquired by looking at the significant amount of tweets in light of the latest happenings.

Although the analysis is performed at the Indian context, the methodology is generic and can be extended to other countries or world-wide topics. In addition, the analysis techniques can be used for identifying people's emotions on various other topics like war, usage of specific technology, natural phenomenon like climate change, etc.

REFERENCES

- 
- [1] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welp, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," The International AAAI Conference on Web and Social Media 16 vol.4, No.1, May 2010, pp. 178-185, doi:10.1609/icwsm.v4i1.14009.
- [2] M. Kukasvadiya and N. Divecha, "Analysis of data using data mining tool orange," International Journal of Engineering Development and Research 5.2, June 2017, pp. 1836-1840, ISSN - 2321-9939.
- [3] S. Hori, "An exploratory analysis of the text mining of news articles about water and society," WIT Transactions on The Built Environment, 168, 2015, pp. 501-508, doi:10.2495/SD150441.
- [4] Automate Getting Twitter Data in Python Using Tweepy and API Access [Online], Available from: <https://www.earthdatascience.org/courses/use-data-open-source-python/intro-to-apis/twitter-data-in-python/>
- [5] A. Beri, "Sentimental Analysis Using Vader, interpretation and classification of emotions". [Online] Available from: Aditya Beri, <https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664> [retrieved: November, 2022].
- [6] Orange widget catalog, <https://orangedatamining.com/widget-catalog/text-mining/twitter-widget/> [retrieved: November, 2022].
- [7] KH Coder 3 Reference Manual, Available from: [https://kncoder.net/en/manual en v3.pdf](https://kncoder.net/en/manual%20en%20v3.pdf)[retrieved: November, 2022].
- [8] Sentiment Analysis using TextBlob. Parthvi Shah.Available From: <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>.

REFERENCES

- 
- [9] P. Nattuthurai, and A. Aryal, "Content Analysis of Dark Net: Academic Journals from 2010-2017 Using KH Coder," ACET Journal of Computer Education and Research 11, 2018, pp. 25-35.
- [10] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and VADER sentiment." Proceedings of the international multiconference of engineers and computer scientists. Vol. 122, 2019, pp. 16.
- [11] Y. Yu and X. Wang, "World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets, Computers in Human Behavior", Volume 48, 2015, pp. 392-400, ISSN 0747-5632, doi:<https://doi.org/10.1016/j.chb.2015.01.075>.
- [12] D. Marcu and M. Danubianu, "Sentiment Analysis from Students' Feedback : A Romanian High School Case Study," 2020 International Conference on Development and Application Systems (DAS), 2020, pp. 204-209, doi:10.1109/DAS49615.2020.9108927.
- [13] U. Thange, V. Shukla, R. Punhani and W. Grobbelaar, "Analyzing COVID-19 Dataset through Data Mining Tool Orange," 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2021, pp. 198-203, doi:10.1109/ICCAKM50778.2021.9357754.
- [14] P. Ekman, Basic emotions Handbook of cognition and emotion, 1999, pp. 16.
- [15] BCCI Official Website. [Online]. Available from: <https://www.bcci.tv/>
- [16] KH Coder. [Online]. Available from: <http://khcoder.net/en/>