

• • •



Cricket Commentary Analysis

Team 10

“

Motivation:

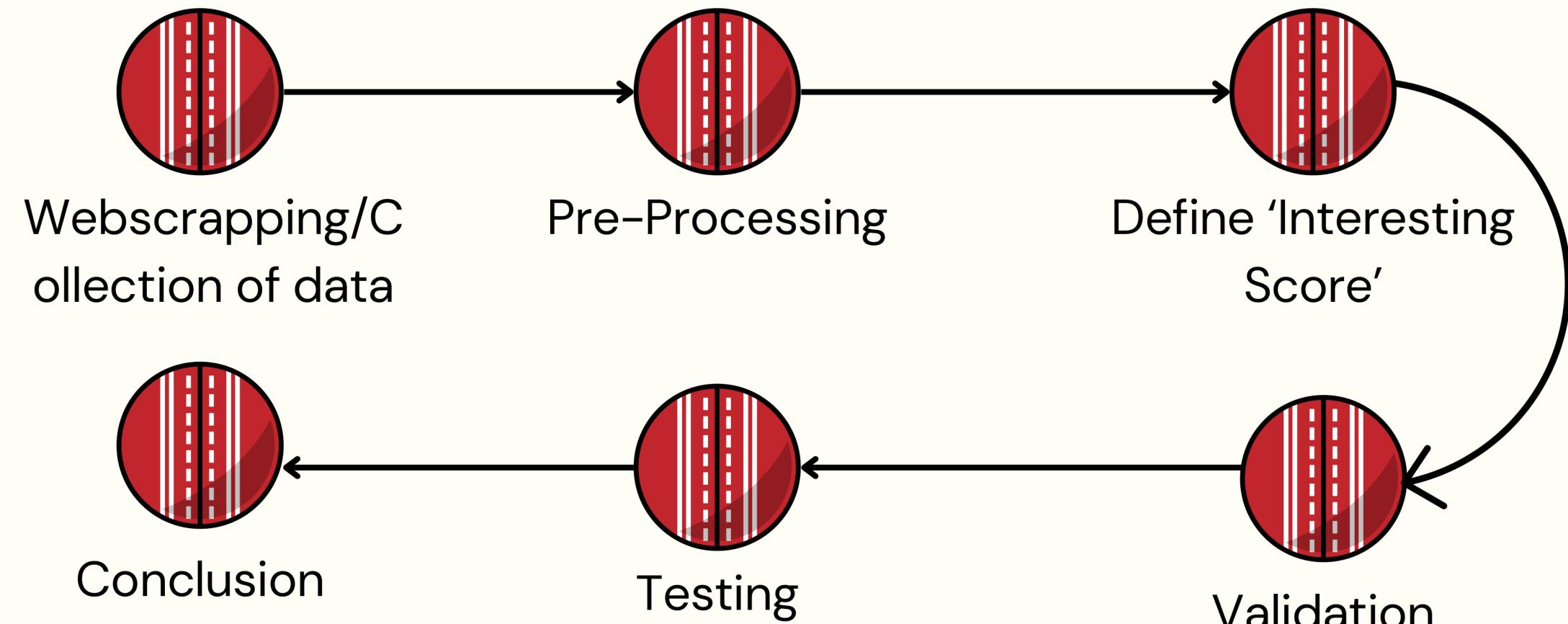
We chose this problem as this can help the user access a highlighted part of the match without having to watch the entire match footage/listen to the whole commentary.

Problem Statement:

To analyze transcripts of cricket commentary to recognize the most interesting parts of the match.



Pipeline



Pipeline In Detail



Collection of data:

Get the data by webscraping and use datasets from Kaggle



Data Pre-Processing

Clean the data by removing stop words and performing Tokenization

'Interesting' Score

We create an 'Interesting' score as a metric using strategies like NER, word2vec etc.

- 16.4 Jadeja to Bairstow, **FOUR, risky and four**, the arm-ball from RJ, kept coming into the right-hander - Bairstow backs away, plays late and then chops it off the stumps, beats backward point and races away to the ropes 4
- 16.3 Jadeja to Bairstow, **FOUR, nicely played**, dropped short and wide of off - Bairstow backs away and then cuts it square past cover-point, Jaiswal with the chase, but he comes up second best 4
- 15.1 Ashwin to Crawley, **out** Caught by Siraj!! Now, Siraj is claiming a catch at mid-off. Crawley's standing his ground, but India look convinced that it's a clean catch - a shimmy down the pitch 10

Other cricketers among the top-10, as per the study, were Rohit Sharma, MS Dhoni, George Mackay, Josh Richards, Hardik Pandya, Sachin Tendulkar, Chris Matthews and Shreyas Iyer. From January to June 2020, each cricketer, respectively, was searched an average of 9.7, 9.4, 9.1, 7.1, 6.7, 5.4, 4.1, and 3.4 lakh times.

With respect to the cricket teams, India led the chart followed by England, Australia, West Indies, Pakistan, South Africa, Bangladesh, New Zealand, Sri Lanka, Ireland, Afghanistan, and Zimbabwe. Each cricket team, respectively, was searched .66, .33, .29, .23, .16, .12, .12, .09, .05, .04, and .03 lakh times per month.

Source of the commentary : <https://m.cricbuzz.com/cricket-match-highlights/78628/ind-vs-eng-1st-test-england-tour-of-india-2024>

Popularity source: <https://www.indiatvnews.com/sports/cricket/virat-kohli-ms-dhoni-and-rohit-sharma-among-most-popular-cricketers-globally-study-641024>

...

Pipeline In Detail

Validation

Validate the model based on
the given outputs

Testing

Test the model on untrained
data

Conclusion

Optimized Interesting score
generator

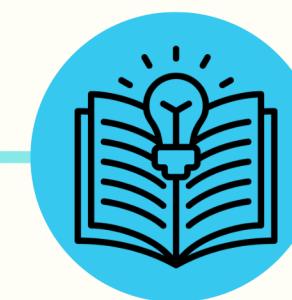
Timeline

March 31st



Obtaining corpus
and
Preprocessing

April 15th



Define
'Interesting Score'
Metric

April 30th



Validation/Modification of
"Interesting Score"
definition

May 15th



Testing on new corpus,
analyzing the results

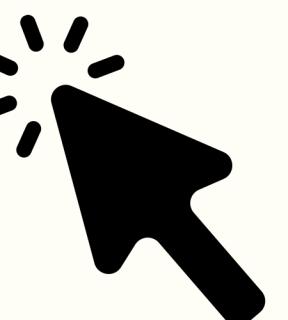
Outcomes

Display of the commentary that describes the most interesting parts of the match

Generation of highlights of a match, by over or by ball.

Thank you

kbyte :)





A black and white illustration of a smartphone. The phone has a black border and a white screen. On the screen, there is a large, bold, black question mark. At the top of the screen, there is a navigation bar with a black dot on the left, a back arrow on the right, and a forward arrow on the far right.

Questions?