***Sentiment Analysis (CricBuzz)***

*Introduction*

In this project, I conducted a sentiment analysis on IPL 2020 news articles sourced from Cricbuzz. The objective was to scrape news articles, process the text, and determine the sentiment (Positive, Negative, or Neutral) associated with each article. This analysis provides insights into how IPL 2020 was portrayed in media coverage.

*Data Collection*

Using Selenium, I scraped multiple articles from Cricbuzz, extracting headlines and full article text. Each article was saved into individual CSV files to avoid data loss and facilitate structured storage. The extracted data included:

- Title of the article

- Link to the original article

- Content of the article

*Data Preprocessing*

After scraping, all the individual CSV files were merged into a single dataset named `Cricbuzz\_All\_Articles.csv`. To ensure data quality, I checked for missing values, removed any empty or duplicate records, and formatted the text for consistency.

*Sentiment Analysis*

Using NLP techniques, I performed sentiment analysis on the collected articles. The sentiment of each article was categorized as:

- Positive: Indicates favorable or optimistic language.

- Neutral: Lacks strong emotion or opinion.

- Negative: Contains critical or unfavorable language.

*Results*

* Most articles had a Neutral sentiment, indicating balanced reporting or lack of emotional language.

A screenshot of a computer

AI-generated content may be incorrect.

* A significant number of articles were Positive, showing enthusiasm and optimism about IPL 2020.
* Very few articles were Negative, highlighting minimal critical coverage.

*Findings*

- The most positive article was: "The greatest T20 team of all time"

- The most negative article was: "Amidst the riches, the IPL retains its youthful..."

- Mumbai Indians had the highest number of positive articles, followed by Delhi Capitals and RCB.

*Visual Insights*

- A bar chart visualizing the sentiment distribution.

A bar graph with different colored bars

AI-generated content may be incorrect.

- A word cloud showing frequently occurring words in positive articles.

A close-up of words

AI-generated content may be incorrect.

*Conclusion*

The sentiment analysis of IPL 2020 articles reveals that media coverage was largely neutral, with a strong presence of positive articles, particularly about Mumbai Indians. Negative articles were minimal, indicating a generally favorable perception of the tournament.