# World Cup Tweets Analysis

Under Mentorness Internship Program

**Abstract:**

This project focuses on leveraging Natural Language Processing (NLP) techniques to analyze and predict sentiments and trends in T20 World Cup 2021 tweets. Analyzing and predicting the sentiments and trends in tweets related to a major sporting event is a complex task that requires understanding the nuances of social media text and the dynamics of fan engagement. By harnessing NLP tools and machine learning algorithms, this project aims to develop a predictive model that can anticipate the sentiments, popular hashtags, and key insights from T20 World Cup 2021 tweets. The project is designed to assist cricket enthusiasts, sports analysts, and marketers in understanding the social media landscape surrounding this event.

**Problem Statement:**

The T20 World Cup 2021 was a highly anticipated sporting event that garnered significant attention on social media platforms. Analyzing and predicting sentiments, trends, and key insights from T20 World Cup 2021 tweets can provide valuable information for various stakeholders. The central challenges and questions addressed by this project include:

1. How can NLP techniques be employed to extract valuable information from T20 World Cup 2021 tweets, such as sentiment, trending topics, and user engagement?

2. Can we build predictive models that analyze tweet contents, user profiles, and tweet metadata to anticipate sentiments and trends during the T20 World Cup 2021?

3. What features and factors are most influential in determining the sentiments and trends in T20 World Cup 2021 tweets?

4. How accurate and reliable can an NLP-based prediction model be in the context of analyzing sentiments and trends in tweets related to a sporting event?

**Dataset Information:**

The dataset utilized for this project consists of a collection of T20 World Cup 2021 tweets, including information about user profiles, tweet timestamps, tweet text, hashtags, and more. This dataset serves as a comprehensive source of information for training and evaluating predictive models in the context of social media analysis.

# World Cup Tweets Analysis

Under Mentorness Internship Program

**Variable Description:**

The dataset comprises the following columns:

- user_name: The username of the tweet author.
- user_location: The location specified in the user's profile.
- user_description: The description provided in the user's profile.
- user_created: The date when the user's Twitter account was created.
- user_followers: The number of followers the user has.
- user_friends: The number of accounts the user is following.
- user_favourites: The number of tweets the user has marked as favorites.
- user_verified: A binary indicator of whether the user's account is verified.
- date: The timestamp of the tweet.
- text: The content of the tweet.
- hashtags: Any hashtags included in the tweet.
- source: The source from which the tweet was posted.
- is_retweet: A binary indicator of whether the tweet is a retweet.

**Scope**:

This project's scope encompasses several critical aspects:

1. Data Preprocessing: Cleaning and preparing the textual data from T20 World Cup 2021 tweets for NLP analysis.
2. Feature Engineering: Extracting relevant features from user profiles, tweet text, and tweet metadata.
3. Sentiment Analysis: Developing and training NLP models to determine the sentiments expressed in the tweets.
4. Trend Analysis: Identifying popular hashtags and trends in T20 World Cup 2021 tweets.
5. Model Evaluation: Assessing the accuracy, precision, recall, and F1-score of the predictive models.
6. Interpretability: Analyzing the factors and linguistic patterns contributing to sentiment and trend predictions.

7. Social Media Insights: Providing cricket enthusiasts, sports analysts, and marketers with insights into the social media landscape during the T20 World Cup 2021.

**Learning Outcome:**

By participating in this project, individuals will gain a variety of valuable skills and knowledge, including:

- NLP Techniques: Learning how to apply NLP techniques to social media text for sentiment and trend analysis.

- Feature Engineering: Extracting meaningful features from user profiles and tweet metadata.

- Machine Learning for Social Media Analysis: Building predictive models for sentiment and trend analysis in tweets.

- Social Media Domain Knowledge: Gaining insights into the dynamics of fan engagement and trending topics on Twitter.

- Interpretable AI: Understanding how AI models make predictions in the context of social media analysis.

This project offers a unique opportunity to combine NLP expertise with social media insights to address an important challenge in the context of a major sporting event: analyzing and predicting sentiments and trends in T20 World Cup 2021 tweets.