**Project Title: Sentiment Analysis of Cricket Tweets**

# Intent & Motivation

# Cricket is more than just a sport—it is a source of emotion, passion, and unity among millions of fans worldwide. During live matches, fans actively share their feelings on social media platforms like Twitter, expressing excitement, joy, frustration, and disappointment. These tweets reflect public sentiment and can be analysed to gain insights into fan engagement, team performance perception, and event popularity.

# The intent of this project is to develop a machine learning and deep learning-based sentiment analysis system that classifies cricket-related tweets into multiple emotional categories such as *Elated*, *Pleased*, *Unbiased*, *Displeased*, and *Frustrated*. This system will help broadcasters, sponsors, and analysts understand fan emotions in real-time, making it a valuable tool for sports analytics and audience engagement.

# Objectives

# To design and implement a multi-class sentiment classifier for cricket-related tweets.

# To build a preprocessing pipeline for cleaning, tokenizing, and vectorizing tweets.

# To address class imbalance using techniques such as SMOTE for better performance.

# To compare machine learning (Logistic Regression, RF) and deep learning models (LSTM, BERT).

# To visualize real-time sentiment trends during cricket matches.

# To evaluate models based on accuracy, precision, recall, and F1-score.

# To create an annotated domain-specific cricket tweet dataset for research and benchmarking.

# Expected Benefits

# Provides real-time insights into public sentiment during cricket matches and events.

# Helps broadcasters and sponsors understand audience engagement and brand perception.

# Assists teams and organizers in assessing fan reactions and improving strategies.

# Contributes to sports analytics research through the use of NLP and deep learning.

# Lays the foundation for multilingual and multimodal sentiment systems in the future.

# Proposed Data & Methods

Tweets related to cricket are collected using the **Tweepy API** with relevant hashtags and keywords such as *#Cricket*, *#IPL*, *#INDvAUS*, etc. The tweets undergo preprocessing steps like **lowercasing, stopword removal, punctuation removal, tokenization, and lemmatization**. The cleaned text is then converted into numerical features using **TF-IDF** and **word embeddings**.

# Outcome

The system successfully classifies cricket tweets into five sentiment categories with high accuracy.  
Deep learning models such as **BERT and LSTM** outperform traditional approaches by capturing context and emotion more effectively.  
The final system provides **real-time, data-driven insights** into fan sentiment, helping stakeholders understand audience behavior and event popularity.