Cricket

- Played in more than 100 countries
- Second largest viewership by population after soccer
- Multi-million dollar industry
- Tough to model the sport – Numerous parameters like moisture, dew, soil condition influence the outcome
- Three formats – Tests, ODI, Twenty20
- ODI - 6 balls in an over, 50 overs and 10 wickets in an innings, two innings in a game.
- Runs scored through non-home runs (running) and home runs (fours & sixes)
- Team with maximum runs in the end wins

Problem Formulation

- 50 over window split into 10 segments of 5 overs each
- Given instantaneous match data upto a segment, predict game progression and outcome
- Formally, given \( R_{\text{known}} = \sum_{i=1}^{n} R_i, W_{\text{known}} = \sum_{i=1}^{n} W_i \) till \( S_n \), predict \( \hat{R}_i \) for \( S_{n+1} \leq i \leq 10 \)
- Predicted end-of-innings score \( \hat{R}_{eoi} = R_{\text{known}} + \sum_{i=n+1}^{10} \hat{R}_i \)
- Winner = \( \max(\hat{R}_{eoi}) \)

Features

- Historical Features
  - Batting strength
  - Bowling strength
  - Avg runs scored in an innings
  - Avg runs conceded in innings
  - Avg # wickets lost
  - Avg # opponent wickets taken
  - Frequency of being all-out
  - Frequency of getting opposition all-out

- Instantaneous Features
  - Home/Away
  - Powerplay
  - Target
  - Game snapshot
  - Bowler Class

Algorithm 1: Prediction of Future Segments runs \( \hat{R}_i \) & End of Innings Score \( \hat{R}_{eoi} \)

Experiment & Results

- Dataset consists of 125 complete matches played by top 9 full-time ICC teams between Jan 2011 - July 2012
- Winner prediction accuracy – 70%

Future Work

- Wicket Predictions
- Prediction of Twenty20 and Test matches