

ABSTRACT

Indian Premier League (IPL) is one of the more popular cricket world tournaments, and its financial is increasing each season, its viewership has increased markedly and the betting market for IPL is growing significantly every year. With cricket being a very dynamic game, bettors and bookies are incentivised to bet on the match results because it is a game that changes ball-by-ball. This project investigates machine learning technology to deal with the problem of predicting cricket match results based on historical match data of the IPL. Influential features of the dataset have been identified using filter-based methods including Correlation-based Feature Selection, Information Gain (IG), Relief and Wrapper. More importantly, machine learning techniques including Naïve Bayes, Random Forest, K-Nearest Neighbour (KNN) and Model Trees (classification via regression) have been adopted to generate predictive models from distinctive feature sets derived by the filter-based methods. Two featured subsets were formulated, one based on home team advantage and other based on Toss decision. Selected machine learning techniques were applied on both feature sets to determine a predictive model. Experimental tests show that tree-based models particularly Random Forest performed better in terms of accuracy, precision and recall metrics when compared to probabilistic and statistical models. And after creating the model, the model will be pickled and finally this will be served as an API, used for predicting score.