**Report**

Our model aims to forecast the outcome of a One-Day International cricket match in while the game is still on by building a dynamic logistic regression model. The various parameters used to predict this outcome are the venue (home match or not home), runs, wickets, toss winning team, runs per over, winning team, outcome etc.

The data files were taken from the website cricksheet.org which were in the YAML format. The data files included data of 1165 One-Day International Matches played between May 2006 and February 2016. Each of these 1165 files provided ball-by-ball data for One-Day International for the 1st and 2nd innings. These files had a lot of attributes that had to be filtered out for our logistic regression model. The YAML files were thus converted to CSV files as it allows data to be stored in a table structured format that can be queried or retrieved easily. Every tuple of the Match\_table has meta-data for each of the 1165 matches. Then, we created two tables for every match corresponding to the innings number of the match. Thus, there are 1165\*2 tables in total. This forms our dataset. The chosen attributes for each of these tables are Date, Venue, Balls, Runs, Wickets, Toss, Team\_Playing, RPO (Runs per over), Team1, Team2, Team\_Won. In addition, a team’s form also plays an important role in the game performance. A team’s current form is calculated as a weighted mean of match outcomes over their last five games. Let, yt = 1 if a team won the match played t matches ago, and 0 otherwise. The team’s current form is then defined as,

Where

The difference in form between the two teams playing the particular match is calculated and added as a new attribute ‘fd’ in addition to the above attributes of the table.

To build a logistic regression model the complete ball-by-ball data is better organized one for each ball of each innings (first or second). Letting K be the number of balls bowled and I be the innings number, tuples corresponding to the Ith inning and Kth ball are extracted from our dataset and written to a new table. For ODI cricket there are K=300 balls in each innings, thus our constructed dynamic logistic regression model will have 300 such tables for every innings. The number of observations may not exactly be 300 as not all innings may necessarily have reached up till the Kth ball. This model is dynamic as the parameters vary with the progress of the match. We use separate models for each innings because, the batting teams in each innings play with different strategies. The first team plays with the aim of scoring as many runs as possible while the second one plays with respect to the runs scored by the first one and has the aim of achieving it’s target. It may do so before the pre-allocated number of overs are exhausted.