Artificial Intelligence Hackathon

Reliance JIO

Problem Statement

- To build a model to predict the team that will win the match based on input data from the past matches
- We have ball by ball data for each game spanned over 8 years, played by 10 teams across 32 cities and 35 different grounds

Data Set

- Deliveries – This contains ball by ball details for 500 IPL matches
- Matches – For each of IPL matches, this file contains the meta data about the match(venue, city), which team was the winner and other statistics

CHALLENGE – to build a model given the **DELIVERIES** and **MATCHES** dataset to predict the winning team in any given match

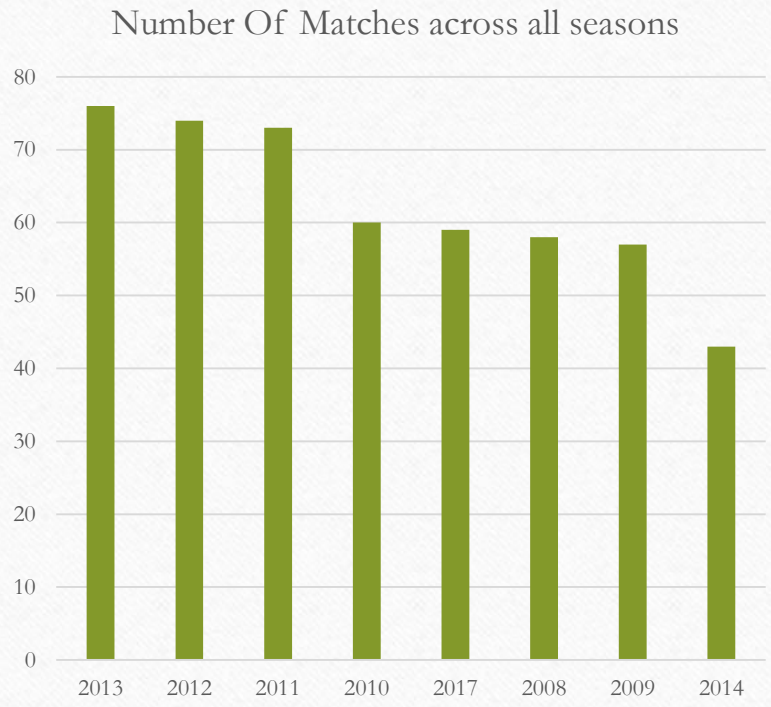
Initial Analysis

Feature Identification

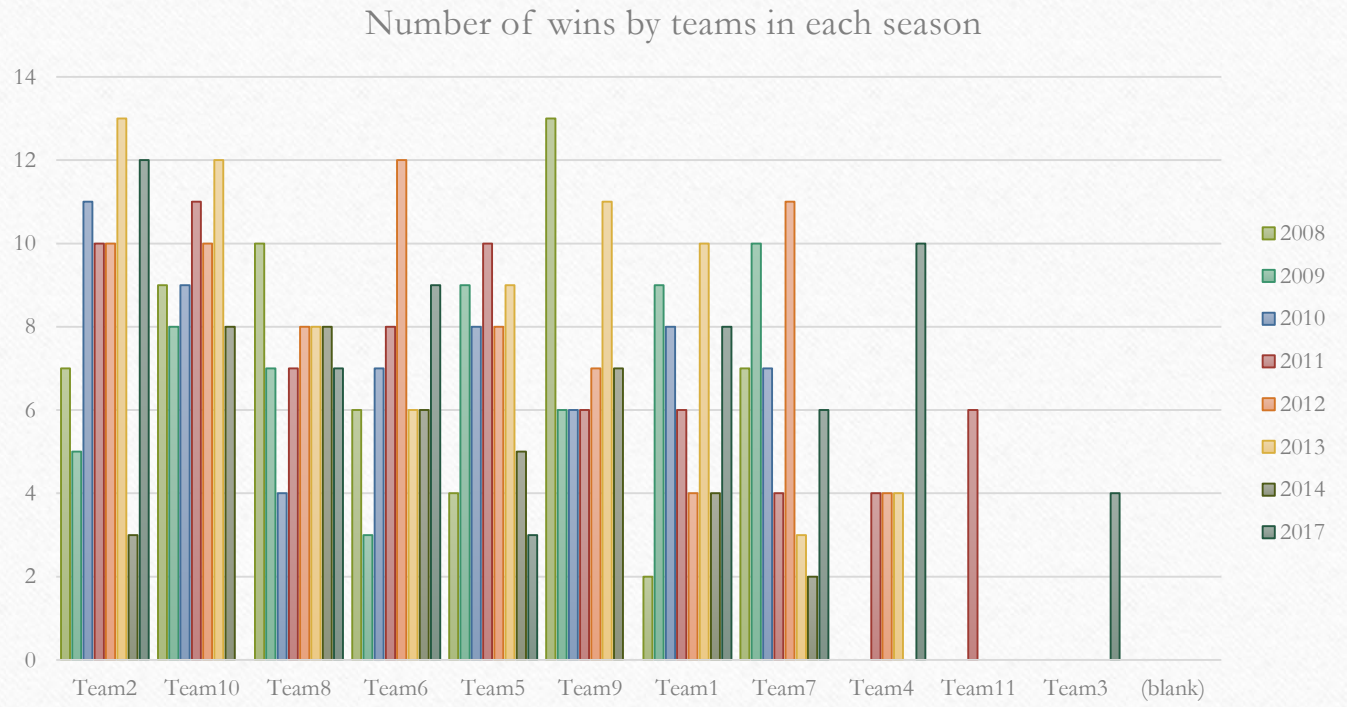
Model Building

Results

1. Most number of matches have been played in 2013(76 matches)



2. Team 2 has won most matches(71 matches in total). Most number of wins were in 2013. Team 10 has given a strong competition with 67 wins in total



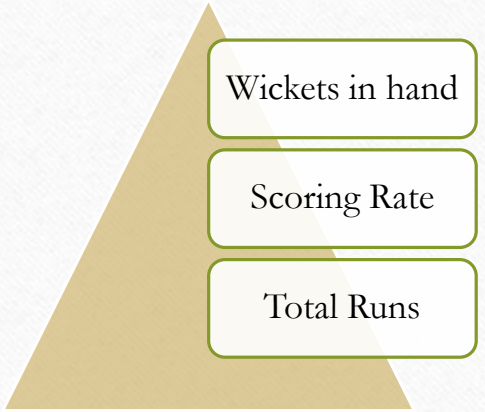
Initial Analysis

Feature Identification

Model Building

Results

- Tool used for analysis is R
- Since the dataset given to us are in 2 separate files – Matches and Deliveries, we have merged them using match ids
- The data granularity is by balls.
- The 3 key features identified to find the winner of each game are –



- We have aggregated the data based on the fields like season, team playing, toss winner, innings, batting team and others to identify the total runs scored by the team in an inning, the net run rate and wickets remaining in hand after the match.
- Scoring Rate and Wickets in hand feature are needed to decide the winner when the D/L method is applied

Initial Analysis

Feature identification

Model Building

Results

The Data Model is created using the following Algorithm

1. If the total number of runs for a team is greater than total number of the runs scored by the opposition, that team stands a winner

2. If D/L method is applied, meaning that the teams have not played all the overs and the match was stopped for some reason or the number of playing overs were reduced, we check the scoring rates of the team. If the scoring rate of a team is greater than opposition, then they are declared as winner

3. In D/L method, along with the scoring rate, we check the number of wickets in hand. If the number of number of wickets are also considerably more than the opposition and we see that the scoring rates are also comparable, we consider the team in the winning bucket

4. In case there is a tie, we have records for the super over. In that case we have compared the runs scored by each team in the super over and then decided the winner of the game

	Predicted Result		
Actual Result		0	1
	blank	1	0
	0	550	0
	1	0	460

- We see that our model is able to predict all 0s and 1s correctly.
- There is one case in which there is no result and winner section is left blank. Since there is not enough information, our model is not able to identify this case

Initial Analysis

Feature identification

Model Building

Results

Result

- The model when run against the training dataset, gave an accuracy of 99.901%
- When the model is run against the test dataset(unseen dataset), it was able to give an accuracy of 99.2647 %