Code The Game

Challenge: "Don the selectors cap and pick the "BEST PLAYING 11" for a cricket match season based on their match impact and most recent performances"

Stage 1: From the match reports and ball by ball commentary, Identify top events in the recent matches – "one ball at a time"

Stage 2: Identify top players involved in those top-ranked events and their impact in the match outcome – "one match at a time"

Finally, choose the best playing 11 and walk away with the TROPHIES & PRIZES!!!!

Stage – 1:

From a data analytics perspective participants can look at each ball played in the match as an event. An event in our case would be characterized by the following fields <F1> <F2> .... <FN> <ESPNCommentary>

In this stage of the problem you need build model that can identify the **top-50 major** events in the match.
A major event here could be anything ranging from the fall of a key wicket or 6 sixes, to just taking a single on the last ball of the last over. **In short participants should look for events that had a strong impact on the match.**

Pay special attention to the last field (i.e. ESPN Commentary) that is provided, since this can be quite instrumental in building a good model.

Stage – 2:

Now after Participants identified key events in the match, it's time to look at the contribution of players in these events.
A few points to note regarding this
1) Some of the events shortlisted in the first stage could be more important than others and hence treating them all equally could be quite erroneous
2) The fields from F1 - FN, would help in identifying the contribution of batsmen and bowlers but not of fielders. You could again use the commentary for identifying and significant fielding contributions.

3) Participants need to analyse all matches of IPL and give players a rating in each match and an average rating for the whole tournament, based on which, participants have to make a Dream Team of the tournament.

Description:

1. How do you pick the top batsmen, bowlers and all-rounders from the pool of available players ?
2. Identify the factors you wanted to consider for choosing the player – For an Analytics driven selection?
3. Identify each player's impact based on top events and payer's contribution from Stages 1 &2?
4. How do you rate each player's skills before you pick them?
5. Rank each player as per their most recent match performance data ?
6. Choose the best final playing 11, as per the following rules :
    a. Playing 11 to have a **MAXIMUM** of 3 top order batsman, 2 middle order batsman, 4 bowlers, 2 all-rounders.
    b. All-rounders can be either bowling or batting all-rounders.
    c. Batsman / bowlers can either be left-handed or right handed.
    d. Bowlers can belong to fast, medium, spin bowling categories.

Suggested Approach for the Data & Analytics Challenge :

- You should identify potential factors (data fields or variables impacting match outcomes)
- Finalize information retrieval methods
- Extract and collate the data (for your identified factors) from available data sources
- Build an Analytical model that can score EVENTS and PLAYER's IMPACT
- Train the model with data sourced from recent matches to arrive at top ranked events and impact
- Rank the players and choose the BEST POSSIBLE PLAYING 11.

Data points to address the problem statement :

- Match details, Official Scores and Player overall performance are in official match reports (may be structured, semi-structured and unstructured).
- Structured data in the form of scorecard snapshot including batting scores, bowling figures, extras, fall of wickets etc.
- Ball-by-ball commentaries are in Unstructured data format and they might contain hidden insights about the events and respective player's contributions (bowling, batting and fielding)
- Match histories are in Unstructured text data format and may be sourced from any sources (news reports, newspaper columns, blogs etc.).

- Analytics models may comprise a combination of Association, Classification and Clustering algorithms

Some tips and Match Scenarios to assist you in your approach :

- Tip 1: A combination of information extracted from both unstructured data and structured data helps you get key attributes (fields) to identify major events in the past played matches.
- Tip 2 : Once you have a corpus of major events, corroborate with the commentary to rate the events in DESCENDING ORDER (event ranking based on their impact that changed the course of the match which lead to the "WIN" outcome)
- Tip 3: With the list of final events, identify top players and their involvement, impact in those events. Rate each player based on their type of contribution (bat, bowl, field) – Fall back to the commentary information to fine tune the process of ranking / rating each players.
- Tip 4 : Text analytics / Predictive analytics / Text to Speech / Tone analyzer, Machine learning are some of the key focus areas to be considered for building Analytical models for scoring and picking the best 11 that maximize your winning chances.
- Larger the training data set (from more number of matches) better the Analytical Model accuracy
- Match Scenario 1 : When a batting team need **20 runs of last 3 overs with 8 wickets down**, taking a single run of last ball of the over by a regular batsman CARRIES HIGHER IMPACT than taking a single run of first ball of the over thereby exposing the tail ender.
- Match Scenario 2 : An all-rounder (**a batsman who can bowl**) **to fill in as 5th bowler** is MORE VALUEABLE than a regular 7th batsman against an opponent who have a strong in-depth batting line up.
- Match Scenario 3 : A batsman, also a seasoned **slip fielder who is responsible for 3 great catches** to dismiss top batsman has HIGHER IMPACT than being an average all-rounder.

DATA

The IPL Match Data which is to be analysed for Stage-1 is available at
http://www.techfest.org/resources/MIvsKKR.pdf

Important Guidelines

Language to be used : Any open source language
Platform : IBM Bluemix . https://www.ibm.com/cloud-computing/bluemix/
Watson APIs to be used : Discovery, Alchemy API & ML Tools on IBM BlueMix
: https://www.ibm.com/watson/developercloud/discovery.html
Outcome in the form of : UI where one can upload match data and get the output

Timeline

1. Last date for registration: 1st November
2. Last date for Final Analysis Report of Stage-1 (for all registered teams): 1st November
3. Results and Shortlisting of top 30 teams: 27th November
4. Preparation for Stage-2 (for top 30 teams): 27th November-28th December
5. Presentation of your complete report: 29th-31st December 2017 at Techfest, IIT Bombay

Registration and Submission

The Participants have to register on the official Techfest Website and fill all the necessary details : www.techfest.org > Competitions > Code The Game > Register.

Abstract Submission:

Teams will be required to submit one report to codethegame@techfest.org . This report should contain the idea they are looking forward to work on.

Abstract Format (for Stage 1)-

Everything that is mentioned under the heading “Description”

Submission Format-

The project report should be emailed to codethegame@techfest.org with the subject Code The Game Report: Team Id (For example: Code The Game: CG1234). Teams must follow the following details for the submission:

1. The abstract must be submitted in pdf format only
2. Font: Arial
3. Size: 11
4. Spacing between two lines: 6 pts
5. Spacing between two paragraphs: 10 pts
6. Bottom margin: 1 inch

SHORTLISTING

Top 30 teams will be selected and would get the chance to participate in the Final Round at Techfest, IIT Bombay which is from 29th-31st December, 2017.

TECHFEST OFFICE
STUDENTS’ GYMKHANA, IIT BOMBAY
POWAI, MUMBAI
MAHARASHTRA - 400076

General Rules

1. Every team has to register online on our website for the competition. A Team ID will be allocated to the team on registration which shall be used for future references.
2. A team can register at any point of time before 1st November 2017 and can submit final abstract and video (as mentioned in the structure).
3. The decision of the organizers or judges shall be treated as final and binding on all. Techfest has all the rights to verify the identity and accuracy of the details provided by the participants.
4. No responsibility will be held by Techfest, IIT Bombay for any late, lost or misdirected entries.
5. The idea presented by the teams should be original (not protected by means of patent/copyright/technical publication by anyone).
6. Note that at any point of time the latest information will be that which is on the website. However, registered participants will be informed through mail about any changes.

Eligibility:

All students with a valid identity card of their respective educational institutions are eligible to participate.

Team Specifications:

- One team can have a maximum of 4 members.
- Students from different institutes can be in the same team.

Certificate Policy:

- Certificate of Excellence will be awarded to the top 3 teams
- Certificates of Participation will be given to all the teams who qualify first round of the competition

TUTORIALS:

Primary Watson APIs Link which will help the Participants to solve the problem
https://www.ibm.com/watson/services/discoveryhttps://www.ibm.com/watson/services/natural-language-understanding

FORUM FOR QUERIES:

https://join.slack.com/t/codethegamequeries/shared_invite/MjM1NTg2NjIyMzEwLTE1MDQyMDgwMDMtN2IyMzFjZmIwMQ

https://www.facebook.com/IBMUR/