**World cup 2018 match result prediction using machine learning**

**Introduction** – With the ongoing 2018 FIFA World Cup fast approaching, every soccer fan in the world is dying to know: Who will capture the coveted trophy?

The document presents the prediction methodology and approach using Machine Learning Algorithm to predict matches outcome at every stage in the FIFA World Cup.

**Data Used** – Collected Data Sources:

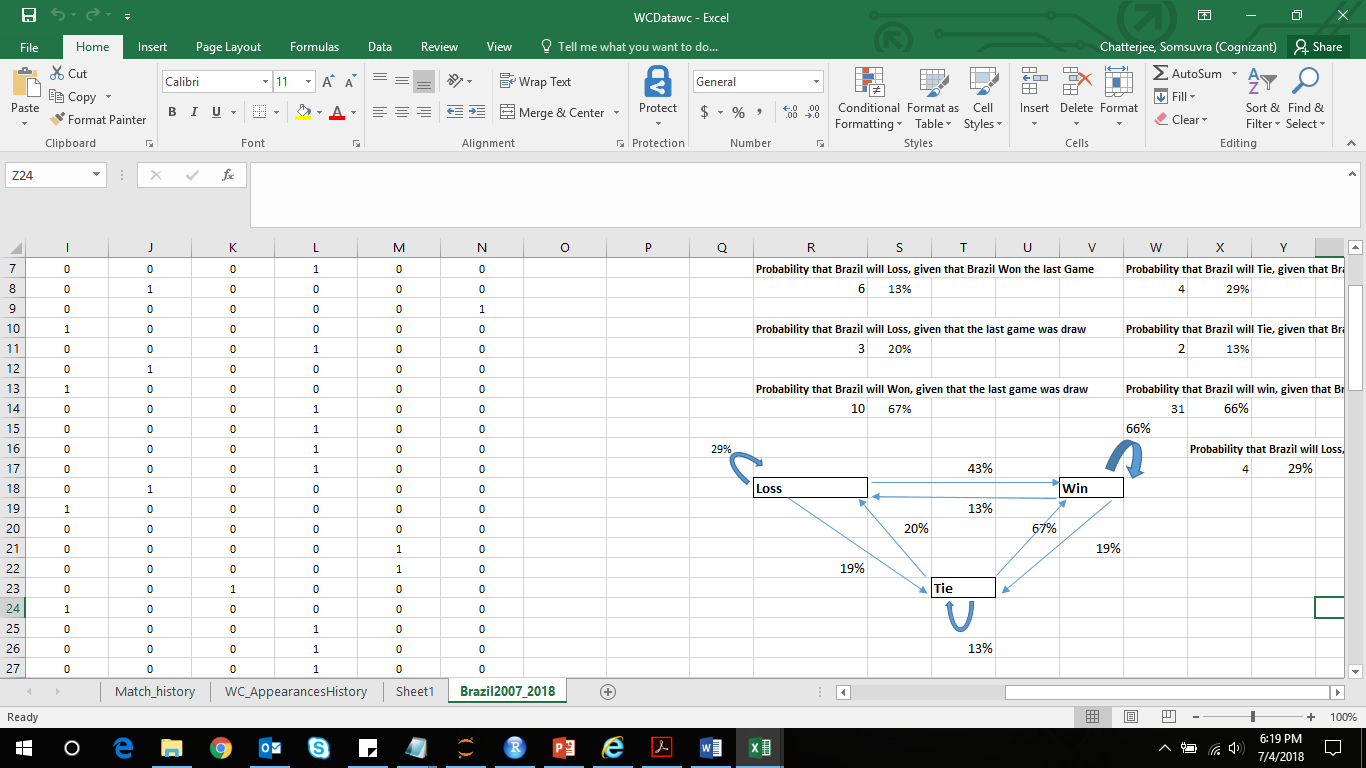
1. Result of all world cup matches downloaded from Kaggle
2. Historical match results starting from 1925 – 2018
3. FIFA ranking for each month from 1993-08 – 2018-05
4. Team wise Players’ data
5. Current playing position of teams

The above data-sources were integrated into a single master-dataset. The attributes covers following categories

* **Performance Attributes:**
  + Rank/Point/World Cup Appearance Differences etc
  + Free-Kick Accuracy, Heading Accuracy,
* **Team Attributes:** 
  + Median Age Difference between the teams
  + Attributes on Team Potential, Acceleration, Aggression, Agility
  + Team Strength, Stamina
  + Attributes measured on Ball Control, Balance
* **Strategic Attributes:**
  + Positioning of Players
  + Attributes measured on Short Passing, Long Passing
  + Attributes on player positioning
* **Derived Markov Chain Probabilities:**
  + Based on historical matches, Markov chain probability columns on

last sequence.

Example below:



**Note**: *Due to data limitation, its assumed current Team strength/weakness characteristics remain same as in history.*

**Algorithm used** –Machine learning classification algorithm (random forest) to predict probabilities of Team1 Win/Team1 Loss/Draw chances. Higher probabilities are to be concluded in favour of respective classes. Model accuracy in validation sample is close to 50%.

**Prediction** – The prediction is to be implemented at each stages. The results will be out in

phases accordingly. In current scenario, the predictions are done to arrive at probable semi-finalists. Thereafter further predictions are done to arrive at finalists and henceforth the final winner.

**Challenges & Scope of Improvement** –

* Players’ information based on past match occurrences is unavailable, hence results can be improved further if proper history can be accumulated.
* Unstructured history like expert comments, blogs, online newspaper feeds on team strategies in favour of win/loss odds if fed as an input, the result can be further improved.
* Betting odds in football history if acquired can be helpful further

**Results of Quarter-Finals:**

Quarter-final1: Uruguay vs. France: Winner- **France**

Quarter-final2: Brazil vs. Belgium: Winner- **Belgium**

Quarter-final3: Russia vs. Croatia: Winner- **Croatia**

Quarter-final4: Sweden vs. England: Winner – **Sweden**

**Results of Semi-Finals:**

Semi-Final1: France vs. Belgium: Winner- **France**

Semi-Final2: Croatia vs. Sweden: Winner: **Croatia**

**Final:**

France vs. Croatia: **France**

Script:

