**Team 44 – Spark Machines – Baseball Prediction – Venkatasivanaga Pavan Kumar, Kanumuri**

**Client Name:** Dr. Dennis Wong

**Team roles and responsibilities this semester**

Team roles and responsibilities

•Primary Contact: Niharika Nallappagari

•Client management: Maruthi Chava

•Requirements management: Pavan Kanumuri

•Data management: Vamsi Krishna Pendyala

•Issues management: Gowtham Kumar Karanam

•Quality and testing management: Lahari Surneni

•Communications and documentation management: Jhansi Mora

1.My role in the team for this project is Requirements Management. As we have completed with the requirements, I have cleaned the datasets for a few statistics. I am on the track of schedule and collected the datasets. I would be loading those data sets into the database model this week.

2. I have collected datasets which are required to clean and load them into the database.

3. **Specific:** I will clean the datasets for the Team batting for the years 2000-2015. I want to accomplish the 15years of (team batting) cleaned datasets. This cleaning is done so as to get a non-redundant data. This cleaning is done by an ETL tool, Pentaho.

**Measurable:** Yes, I will clean the datasets for 15(2000-2015) years. When the goal is accomplished, there will be no negative values and null values, and also some values for the labels will not be beyond the extremes after cleaning the data.

**Achievable:** Yes, the goal is reasonable. To train the model, we need to send the cleansed data to model.

**Relevant:** The goal is worthwhile and it will meet our needs. Yes, our goal is consistent.

**Timely:** I will complete this work by end of this week i.e., 09/10/2016. I am going to prepare the slides for the Workshop on Requirements Management.

4. Analysis on data and developing the tables by end of third week i.e., 09/16/2016.

* Developing ETL transformations and jobs to clean, transform and load the data by 09/16/2016.
* Analyze and send the file input/ table data input to Apache spark to build model by 09/23/2016.
* Segregate the data into segments for K-fold cross validation, to train the model by 09/30/2016.
* Implement the initial business logic to train the model by 10/10/2016.
* Implement the machine learning methodology and test the error value by 10/16/2016.
* Implement and reduce the error through appropriate methodology by 10/23/2016.
* For 50% progress, we plan to complete building a model
* For 80% progress, we plan to complete training a model
* Week-12: we will try to meet the client requirements which indicates that progress is 100%

5. No, I don’t have any unresolved challenges for this week.