1. Selection of algorithms/research papers which would be part of the study.

So far, we have selected 3 algorithms which we will be training and evaluating on our selected dataset. Detail is as follows:

- Polynomial Regression: In this algorithm, we will be mainly concerned with fine-tuning the degree of the polynomial, such as to avoid under and over-fitting. We are hoping that this algorithm will be giving acceptable results.
- Artificial Neural Networks: An ANN will be able to learn the relationships in the dataset easily since we have a large dataset. The main problems faced in training an ANN are to select an architecture of the ANN and to fine-tune the hyper-parameters.
- Regression Tree: We will apply the concept of decision trees for regression. We anticipate that this algorithm might be slow and will have a hard time to learn the relationships in data due to overfitting.

2. Detail of dataset

The dataset we are using is a database of the player characteristics and attributes of all the players in FIFA 18. It contains almost 185 fields for each player. Some of those attributes are personal information of the player, which are not useful for our application of position based rating prediction.

Also, non-goalkeeper (non-GK) players do not have ratings for GK position and visa versa. Therefore, we have divided the data-set into 2 part, the GK and non-GK datasets. We will be handling their regressors separately. Dataset Link

3. Progress as against the timeline.

We have completed the phases of Data Gathering, Visualization, Preprocessing, Feature Extraction and Algorithm Selection.
The data-set had anomalous values for some columns which have been fixed. For feature extraction, we removed personal information of players, which does not contribute towards the player rating. We have also completed our ground work, regarding the working and trade-offs of different regression algorithms. The algorithm selection details have already been provided in section 1.

4. Updated timeline.

15 May 2018 - 21 May 2018 (Implementation)
22 May 2018 - 27 May 2018 (Fine-Tuning & Report writing)

5. Issues/difficulties/ pitfalls faced so far.

- The dataset contains anomalous and incorrect values for some fields which have been a little hard to trace.
- Backpropagation for regressive ANNs has been a little tough for us to figure out, as the gradients associated with the final layer change when the non-linear activation function is removed.