**Aim:** To predict whether Ronaldo scores a goal or not, depending on the given dataset

**Procedure:**

**Data Pre-processing:** From the dataset given, location\_x, location\_y, power\_to\_goal and distance\_to\_goal data were used. Each of these data columns had some missing NaN elements, which could not be used. Hence first step was to identify the data which can help in prediction and the next step was to remove the rows which had NaN values.

Also from the column is\_goal, the rows which had no value, had to be removed as those rows would be used for prediction testing of the final model.

The other rows would be used to training the model

**EDA:** Once the data was pre-processed and ready to be trained, the data was split into test and train parts (usually 20-80 ratio).

Using this a validation percentage can be taken and used to figure out how good our model is training

**Machine learning Model creation:** Here, there are many algorithms which can be used. The most best suited is logistic regression. This is because the logistic regression uses a sigmoid function which is used for binary classification. Here we need to predict whether goal can be scored or not, which is binary.

Accuracy achieved by different algorithms are as follows-

Logistic regression: 59.9087582093

SVM: 50.5389281596

Random Forest: 61.1971725071

Neural network: 61.0668270918