Predictive Machine Learning (ML) algorithms

**Support Vector Machines (SVM)** (Dataflair Team, September 2018)

It is a supervised classification which separates point of data into different groups, i.e. for the PLePA, this would be home win, lose or draw. This algorithm calculates the vector to optimize the line.

**Naive Bayes** (Dataflair Team, September 2018)

This algorithm uses each feature as independent of each other, e.g. If a team finished 1st last season and on a five loss streak would be evaluated into the algorithm separately.

**Random Forest** (Dataflair Team, September 2018)

A group of decision trees, each tree provides a classification. Classification with most votes is what determines the result.

**K-Nearest Neighbour (KNN)** (The Open University, 2019)

This algorithm is a classification algorithm, it uses data it has already obtained to determine the class of the new data. For PLePA, this would be decided whether new game data is a home win, loss or a draw.

**K-Means** (The Open University, 2019)

This algorithm discovers classes within the data, it groups the data based on parameters which are set, these can be changed each iteration until it groups best for what is trying to be determined.

**Decision**

I have decided on SVM, Naïve Bayes, Random Forest and KNN. The K-Means is not really appropriate for PLePA because it doesn’t use data already known such as whether a match was a home win, loss or draw. The other four give a good variety of styles which will, hopefully, give a good mixture at the end of the project.