Ultimate Challenge - Predictive Modelling

1. —
2. We must first decide how to frame this problem given our data and objective.

   **Data:** Labelled with seemingly sufficient features varying in nature.
   **Objective:** Determine whether a new user whose behavior we observe will be retained.

   We choose to model it in a **supervised binary classification** framework as we have sufficient labelled data and the target variable is discrete and further, it is binary (retained vs not retained).

   We then test various classification models - logistic regression, random forest and neural networks.

   With our preliminary study, we use numerous performance metrics (**accuracy, precision, recall, confusion matrices, F1-score, MSE**) to quantify the predictive power of these models and we find that the random forest and neural network classifiers seem to perform better than the logistic regression classifier with the random forest performing marginally better than the neural network model.

   I believe this model is valid assuming the data we are given is an accurate representation of the true population. If we were to extend this solution framework further for analysis of the complete network, we would likely sample the complete population appropriately taking into consideration factors such as location, demographic, gender, age groups, etc.

   To be able to more accurately predict retention, we may also consider further feature engineering taking into account some factors mentioned above such as age, income, gender, etc.

3. A key take-away from our random forest model is to look at **feature importances** to see which of our features have the most predictive power. With this information, Ultimate can make business decisions to improve retention by targeting these factors. For example, we see that the average distance and weekday percent are very important factors in predicting retention and Ultimate could market their product appropriately by introducing promotions related to these factors and focus less marketing resources on things like new-user interaction or promoting Ultimate Black which are comparatively less important.