**OVERVIEW**

The objective of this analysis is to investigate the feasibility of using “Wi-Fi fingerprinting” to determine a person’s location in indoor spaces. Alert! Analytics is to analyze the dataset provided from the Universitat Janume I and determine recommended models that is accurate enough to estimate locations. This dataset contains 3 building locations, up to 4 floors, and many location identifiers for the room and position. It was created by more than 20 different users and 25 Android devices. The Wi-Fi signals were captured in WAPs (Wireless Access Points) with a range from -104 (weakest) to 0 (strongest). The actual longitude and latitude were also recorded for each data point. The client is hoping that with this information and machine learning, they can deploy a smartphone app for indoor locationing.

**METHODOLOGY**

First, the dependent variable needed to be identified. In this dataset, there are 3 dependent variables, Longitude, Latitude, and Floor. Longitude and Latitude will be able to tell where a person is located on a map, while the Floor will show the altitude of the person’s location. Hence, with all 3 variables, a person’s exact location can be determined. In order to analyze 3 dependent variables, the problem must be separated and treated as 3 different analyses. Also, there are many extra variables that are not needed since the focus is on how Wi-Fi signals can estimate a person’s location. So in this analysis, all independent variables were removed except the WAP readings.

Both the Longitude and Latitude variables are continuous and numeric, hence they are both treated as regression problems. Floor, however, is an ordinal categorical variable so it was treated as a classification problem. Each different dependent variable problem had different datasets to train models on. For the first problem, Longitude, the dataset only included WAP readings and the Longitude variable. For the second problem, Latitude, the dataset only included WAP readings and the Latitude variable. Finally, for the third problem, Floor, the dataset only included WAP readings and the Floor variable.

For each of these problems, 4 different algorithms were used to train the 70% training set of a sample size of 10,000 observations. Then, further feature selection was tested to see if it could improve the out-of-the-box metrics by removing some extra WAPs. For feature selection, removing zero variance features and PCA were applied here. After determining the best model after tuning parameters on the sample set, the best model was then used against the full dataset with all 19,937 observations. The output metrics of this shows the confidence in the model of predicting actual locations.

**CONFIDENCE**

Note that there were many models that had very similar high-performing metrics, which means many models are just as capable of predicting locations as the best model. Below just highlights one algorithm for each problem.

To determine Longitude, Random Forest without any extra feature selection for WAP was the best performing model (highest adjusted R-squared and lowest RMSE on testing set). When the model was used against the full dataset, the metrics showed R-squared as 0.9949 and RMSE as 8.8648.

To determine Latitude, KNN without any extra feature selection for WAP was the best performing model. When the model was used against the full dataset, the metrics showed R-squared as 0.9952 and RMSE as 8.1297.

Finally, to determine Floor, Random Forest without any extra feature selection for WAP was the best performing model once again (highest accuracy and kappa scores). The accuracy score was 0.99 and the kappa score was 0.9929.

**IMPLICATIONS**

Based on the level of confidence of each problem, it is concluded that Wi-Fi fingerprinting with the use of machine learning is definitely a feasible way of predicting a person’s location for Android users. The client may go into developing the smartphone app with full confidence that the accuracy of the location prediction by machine learning is near to perfect. However, it is important to note that Apple iOS system was not tested in this dataset so the results cannot be implied for iPhones. Further research must be done for iPhones if the client wants to open the smartphone app to the iPhone market.