**Evaluation Techniques for Wifi Locationing in Python**

**By Omar Chavez**

**6/20/2022**

**Overview**

Our client is looking for a solution to identify someone’s location indoors. GPS is a reliable solution for outdoor locationing, but for large indoor spaces such as universities, hospitals, or malls, it is not possible to use GPS satellite signals. The solution proposed in the JLI Indoor Locationing dataset, is the use of WIFI fingerprinting to determine someone’s location inside of the JLI University complex. The location in this scenario is determined by the strength of nearby Wireless Access Points (WAPs). We have taken the JLI dataset and analyzed three different machine learning models to attempt to find a sufficiently accurate model for indoor locationing.

**Data Provided**

We were provided with two different datasets, one with a completed database that was used to train the algorithms and one that was incomplete on which we predicted the location of a particular user.

The first 520 features of the datasets consisted of different WAP signal strengths, where a lower value indicated a higher intensity. The other features consisted of: longitude, latitude, floor, building number (there was a total of 3 buildings), userID, phoneID, timestamp, and relative position, and spaceID. SpaceID which was the dependent variable being predicted in this case. To predict SpaceID, we used different classification models. It was discovered that there was only a certain number of SpaceIDs in each building, but the SpaceIDs might be the same across different floors (for example SpaceID 50 exists on floor 1 and floor 2). For this reason, the SpaceID and Floor variable were combined into one dependent variable entitled LocationID.

Because the dataset is so large, with 19937 observations, the dataset was split up into 3 subsets based on what building the locations were in.

Features that were deemed as noise were removed. Those features were: userID, phoneID, timestamp, and relative position. Relative position (whether a user was inside or outside of a room) was not used in this analysis because for our use case, if a person can get a correct location on where they are in a building using this model, they should be able to figure out themselves if they are inside or outside of a room.

**Model Comparison**

The 3 classification models that were used to predict indoor locations in the incomplete dataset were gradient boosting (GBM) , random forest (RF), and K-nearest neighbor (KNN). The models were each run on the database subsets: buildings 0, 1 and 2. Accuracy scores were calculated for each building and then averaged across all buidings. As can be seen below, the worst algorithm was KNN with an accuracy score of 64%. And the best performing model by far was the Random Forest algorithm. The Random Forest algortihm came in with an accuracy of 93%. This is a very reliable model that should work well for wifi locationing.

**Recommendations**

With time, if the app is put into production, we can fine tune the model and greater increase accuracy. Some things that we will need to also think about is what about sections of the building where there is no nearby wireless access point? Also, how much of an impact did people’s different phone models have on the signal strength? Perhaps the experiment should be conducted with everyone having the same device type. In the real world though this would not be practical, so a better idea may be to put a location monitoring device on students’ ID card, and then their phone could just be the graphical user interface they use to navigate themselves.