**HTU Upskilling Program: Data Science Track**

**Capstone Project: Predict location from Wi-Fi signal strength**

**Student Name: Majd Qutaish**

**Date: 15th Jan 2023**

* **Introduction**

The objective of this project was to analyze the data collected by a smartphone about the strength of Wi-Fi signals in an indoor space. The collected data was then used to develop a model that can predict the location of an individual based on their measurements.

The data will be preprocessed and then several classification models such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Neural Networks (NN), Gradient Boosting Machine (GBM) will be applied to the dataset and their accuracy will be evaluated to select the best model.

The results of this project can have practical applications in various fields such as smart buildings, emergency response, and personal navigation. By accurately predicting a person's location within an indoor space, it is possible to improve building efficiency, emergency response times, and personal navigation experiences.

* **Problem**

## The problem that this project aims to address is the accurate prediction of a person's location within an indoor space using Wi-Fi signal strength data. The problem is formulated as a supervised learning task, where the goal is to develop a machine learning model that can predict the room label (location) of a person based on their Wi-Fi signal strength measurements.

## The challenge of this problem is due to the high dimensionality and variability of the data, as well as the factors that can affect the strength of Wi-Fi signals, such as building materials and interference. Another issue is the need to accurately identify the location of an individual in several rooms, rather than just a single location. To solve this issue, the data must be analyzed and understood to develop a strategy that can accurately predict an individual's location based on their measurements. The use of machine learning techniques should also be considered.

* **Data set**

The dataset used for this project contains measurements of Wi-Fi signal strength observed from 7 different Wi-Fi devices, as well as the corresponding room label for each measurement. These measurements were collected in an indoor space, and the goal is to use this data to train a model that can accurately predict the location of a person based on their Wi-Fi signal strength measurements.

The data set has 2000 rows and 8 columns. The data set contains Wi-Fi signal strength observed from 7 Wi-Fi devices on a smartphone collected in indoor space. The data could be used to estimate the location in one of the four rooms.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Position | Attribute Name | Definition | Data Type | % Null Ratio |
| 1 | wifi 1 | Signal strength of wifi 1 | Quantitative | 0 |
| 2 | wifi 2 | Signal strength of wifi 2 | Quantitative | 0 |
| 3 | wifi 3 | Signal strength of wifi 3 | Quantitative | 0 |
| 4 | wifi 4 | Signal strength of wifi 4 | Quantitative | 0 |
| 5 | wifi 5 | Signal strength of wifi 5 | Quantitative | 0 |
| 6 | wifi 6 | Signal strength of wifi 6 | Quantitative | 0 |
| 7 | wifi 7 | Signal strength of wifi 7 | Quantitative | 0 |
| 8 | Room | One of the four rooms (1,2,3,4) | Quantitative | 0 |

**Source:**

This data set has been sourced from the Machine Learning Repository of University of California, Irvine Wireless Indoor Localization Data Set (UC Irvine).

The UCI page mentions the following 2 publications as the original source of the data set:

1. Rajen Bhatt, 'Fuzzy-Rough Approaches for Pattern Classification: Hybrid measures, Mathematical analysis, Feature selection algorithms, Decision tree algorithms, Neural learning, and Applications', Amazon Books.
2. Jayant G Rohra, Boominathan Perumal, Swathi Jamjala Narayanan, Priya Thakur, and Rajen B Bhatt, 'User Localization in an Indoor Environment Using Fuzzy Hybrid of Particle Swarm Optimization & Gravitational Search Algorithm with Neural Networks', in Proceedings of Sixth International Conference on Soft Computing for Problem Solving,2017, pp. 286-295.

* **Data visualization:**

**Text

Description automatically generated with medium confidence**

**Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedA picture containing graphical user interface

Description automatically generated**

* **Models:**

1. **The K-Nearest Neighbors model - KNN:**

**Text

Description automatically generated**

1. **The Support Vector Machine model – SVM**

**Text

Description automatically generated**

1. **Gradient Boosting Machine model – GBM**

**Text

Description automatically generated**

1. **Neural Networks model - NN**

**Graphical user interface, text, application

Description automatically generated**

Git-hub link: <https://github.com/MajdQ93/wifi_localization>

* **Results:**

The results of the project revealed that the various models performed well and had high accuracy scores. The K-Nearest Neighbors model had the highest score with an accuracy of 0.985, followed by the SVM model with an accuracy of 0.9825, and the GBM and NN models with an accuracy of 0.9725 and 0.9825, respectively.

The results of the project revealed that the various models performed well and had high accuracy scores. The use of machine learning techniques allowed them to accurately identify the patterns in the data and make predictions about the location of an individual based on their measurements.

Although the accuracy of the different models varied, the K-Nearest Neighbors model performed well and was the best choice for this task. The performance of the models is not always the same across different scenarios and data sets. Therefore, it is important to thoroughly evaluate the models and make sure that they can generalize well to new data.