Dhirubhai Ambani Institute of Information and Communication Technology
Gandhinagar, Gujarat, India

Project Report :-

ML based classification and prediction of context using WiFi

RSSI Under the subject of

Advanced Computer Networks
Group 2 (IT531)

By

202311014 - Shrut Shah
202311017 - Dixit Lukhi
202311032 - Abhay Manavadariya

Under the guidance of

Prof. P S Kalyan Sasidhar

Objective

The goal of the project is to classify indoor locations accurately using WiFi RSSI data, while accounting for context variations like device differences and signal fluctuations over time.

Project Execution steps :-

1. **Data Collection with Android App:**
   - The Android app was used to collect WiFi signal strength data at various labeled points across indoor zones. This allowed us to track RSSI variations due to device and environmental factors.

2. **Collaborative Data Gathering:**
   - Partnering with other groups, we ensured the collection of a broad dataset across different times of the day and zones. This helped capture a comprehensive view of signal strength variability.

Sample of the collected Data:

Timestamp	Location	SSID	BSSID	Signal Strength	Frequency
11/14/2024 15:54	LAB_0_3PM_A_1	Attendance	58:b6:33:7a:1c:d8	-68	2437MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_STAFF	58:b6:33:3a:1c:d8	-70	2437MHz
11/14/2024 15:54	LAB_0_3PM_A_1	Lab_211_5G	24:c9:a1:20:e7:4c	-81	5280MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_Student	60:d0:2c:df:63:9c	-56	5220MHz
11/14/2024 15:54	LAB_0_3PM_A_1	TP-Link_13F8_5G_2(	9c:53:22:e7:13:fa	-83	5765MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DA_Public	18:4b:0d:a3:75:38	-67	2412MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_Student	18:4b:0d:e3:75:38	-67	2412MHz
11/14/2024 15:54	LAB_0_3PM_A_1	dasolarpnl	c0:c5:20:24:87:18	-76	2467MHz
11/14/2024 15:54	LAB_0_3PM_A_1		9e:53:22:a7:13:f8	-78	2417MHz
11/14/2024 15:54	LAB_0_3PM_A_1	Dev	b6:16:63:ff:73:29	-86	5745MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DA_Public	58:b6:33:ba:1c:d8	-68	2437MHz
11/14/2024 15:54	LAB_0_3PM_A_1	VM_31	4a:7a:79:08:30:db	-77	2412MHz
11/14/2024 15:54	LAB_0_3PM_A_1		52:91:e3:25:ac:c9	-75	2422MHz
11/14/2024 15:54	LAB_0_3PM_A_1	Dell-Gujcost	78:98:e8:7a:2c:1f	-83	2472MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_Student	44:1e:98:fb:55:d8	-72	2462MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_Student	24:79:2a:34:ea:0d	-43	5745MHz
11/14/2024 15:54	LAB_0_3PM_A_1	DAIICT_STAFF	1c:b9:c4:3d:84:e8	-70	2462MHz
11/14/2024 15:54	LAB_0_3PM_A_1	Attendance	18:4b:0d:63:60:7c	-70	5805MHz
11/14/2024 15:54	LAB_0_3PM_A_1	Lab_211	24:c9:a1:20:e7:48	-72	2457MHz

- In this file, there is data regarding the signal strength of all the access points.

Preprocessing of the Data:

- We will clean and preprocess the data and will work on the specific SSID **(**DAIICT_Student, DA_Public**)**

BSSID	Location	SSID	18:4b:0d:9c:30:98	18:4b:0d:9c:30:9c	18:4b:0d:a3:60:7c	18:4b:0d:a3:75:38	1
0	LAB_0_3PM_A_1	DAIICT_Student	NaN	NaN	NaN	NaN	
1	LAB_0_3PM_A_1	DA_Public	-52.0	-68.0	-69.0	-67.0	
2	LAB_0_3PM_A_10	DAIICT_Student	NaN	NaN	NaN	NaN	
3	LAB_0_3PM_A_10	DA_Public	-70.2	-84.0	NaN	-77.0	
4	LAB_0_3PM_A_11	DAIICT_Student	NaN	NaN	NaN	NaN	
5	LAB_0_3PM_A_11	DA_Public	-76.4	-84.0	NaN	-76.2	
6	LAB_0_3PM_A_12	DAIICT_Student	NaN	NaN	NaN	NaN	
7	LAB_0_3PM_A_12	DA_Public	-78.0	-84.0	NaN	-76.0	
8	LAB_0_3PM_A_2	DAIICT_Student	NaN	NaN	NaN	NaN	
9	LAB_0_3PM_A_2	DA_Public	-57.6	-68.2	NaN	NaN	
10	LAB_0_3PM_A_3	DAIICT_Student	NaN	NaN	NaN	NaN	
11	LAB_0_3PM_A_3	DA_Public	-63.0	-72.8	NaN	NaN	
12	LAB_0_3PM_A_4	DAIICT_Student	NaN	NaN	NaN	NaN	
13	LAB_0_3PM_A_4	DA_Public	-64.0	-78.8	NaN	-82.0	
14	LAB_0_3PM_A_5	DAIICT_Student	NaN	NaN	NaN	NaN	

Step 1: Data Preprocessing

- Ensure the collected WiFi RSSI data is cleaned and normalized.
- Feature scaling is applied for algorithms like SVM that are sensitive to data scale.
- Divide the dataset into training and testing sets (e.g., 80%-20% split).

Step 2: Model Training

1. **K-Nearest Neighbors (KNN):**
    - **Overview:**
        - KNN is a non-parametric, instance-based learning algorithm.
        - It classifies a point based on the majority class of its k nearest neighbors.
    - **Implementation Steps:**
        - Select the value of k (commonly optimized through cross-validation).
        - Use a distance metric such as Euclidean distance to

calculate proximity.
        - Store the training data and compare distances during predictions.
    - **Advantages:**
        - Easy to implement and effective for smaller datasets.
        - Suitable for indoor localization where signal strength correlates spatially.
2. **Support Vector Machine (SVM):**
    - **Overview:**
        - SVM is a supervised learning algorithm that separates classes using a hyperplane.
        - Works effectively for linearly and nonlinearly separable data (via kernels).
    - **Implementation Steps:**
        - Choose a kernel function (e.g., linear, RBF).
        - Train the SVM model using the training data.
        - Optimize hyperparameters (e.g., C, gamma) using techniques like Grid Search or Random Search.
    - **Advantages:**
        - Handles high-dimensional data well.
        - Robust to overfitting in low-noise scenarios.