Seamless Indoor-Outdoor Detection Using Machine Learning

Project Report submitted in partial fulfillment of the requirements for the degree of

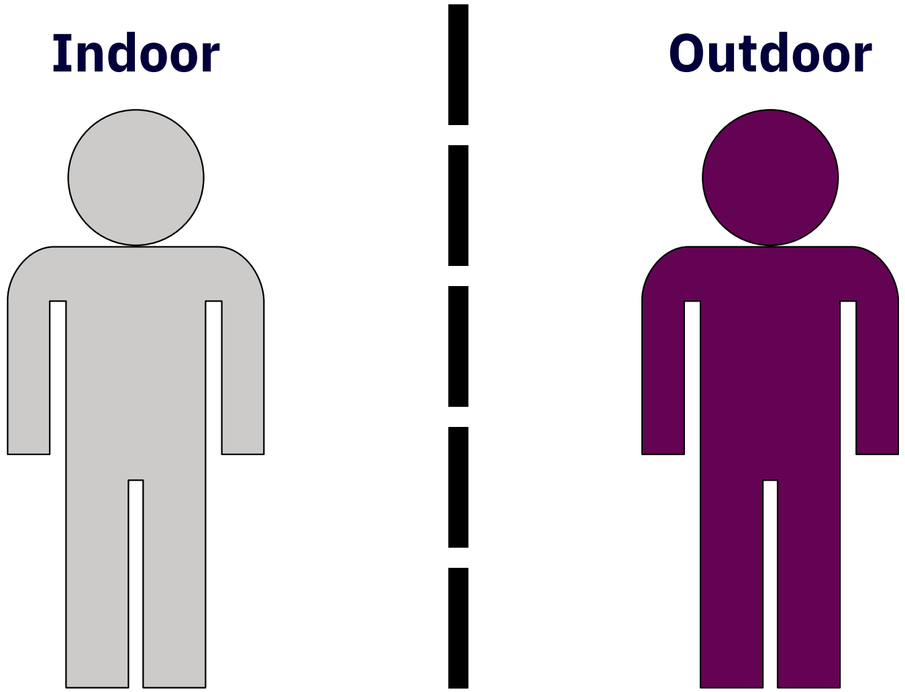
BACHELOR OF TECHNOLOGY IN CSE(AIML)
OF
MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY

BY

Rishi Agarwal, 12020002016023, 201040130810026
Radhika Jalan, 12020002016028, 201040130810007
Pankaj Kumar, 12020002016025, 201040130810019
Keshav Hisariya, 12020002016026, 201040130810018
Srijita Das, 12020002016027, 201040130810005

Under the guidance
of
Prof. Ayan Kr. Panja
DEPARTMENT OF CSE(AIML) & CSBS

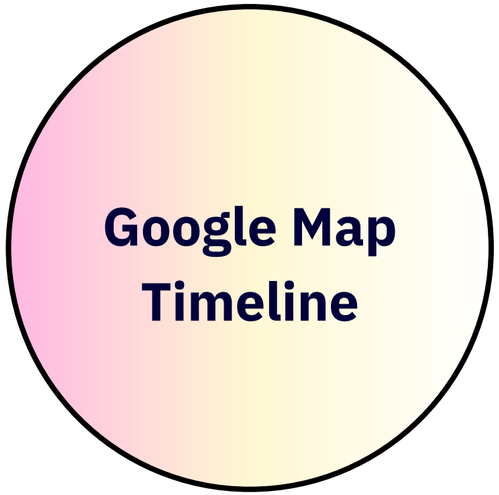
IEM Institute of Engineering & Management



Predict whether someone is *inside* or *outside*..

Application Domain

- Location Based Application
- Security systems
- Smart Home Automation
- Augmented reality gaming



Related Works

1

Smartphone Magnetic Field Data Analysis

- Ashraf et al. explore using smartphone magnetic field data for IO detection.
- Naive Bayes outperforms other classifiers with an accuracy of 83.26%.

2

Diverse Localization Techniques

- StarTrack provides APIs for GPS, GSM, and WiFi-based localization apps.
- Zhou et al. track buses using cell tower sequences, while LANDMARC proposes RFID for indoor localization.

3

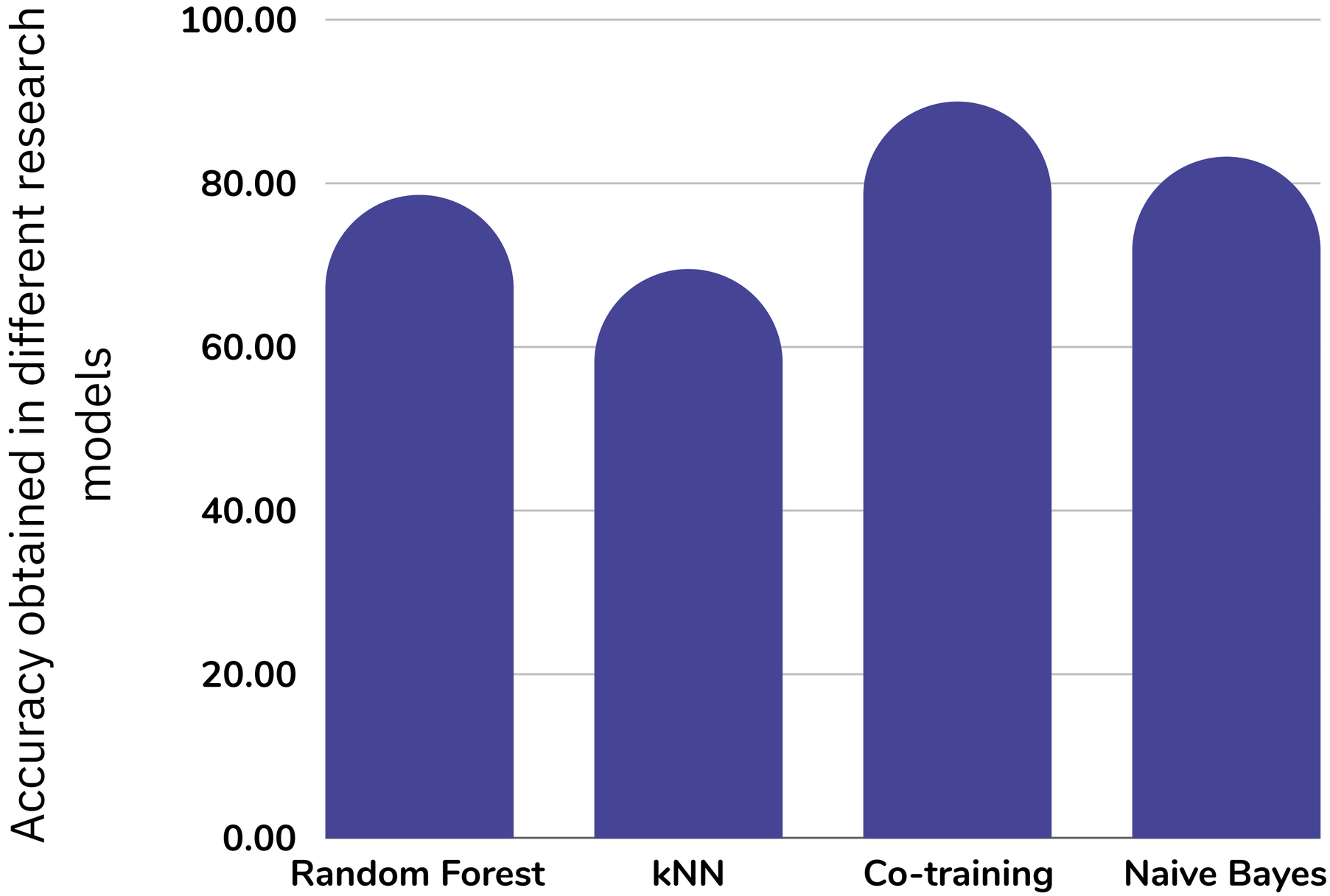
Semi-Supervised Learning for Object Recognition

- Radu et al. employ co-training with "Cluster-then-Label" technique for object recognition.
- Achieves over 90% accuracy even in unknown situations by automatically learning new settings and gadgets.

4

GPS Limitations and SatProbe Solution

- GPS offers positional accuracy but suffers in response time and energy usage.
- SatProbe by Chen et al. improves efficiency by utilizing raw GPS data for IO detection.


Accuracy obtained in different research models
100.00
80.00
60.00
40.00
20.00
0.00
Random Forest
kNN
Co-training
Naive Bayes

Methodology

01 Data Collection

Sensor data was collection from various indoor/outdoor settings, with transitions noted for accurate labeling.

02 Training

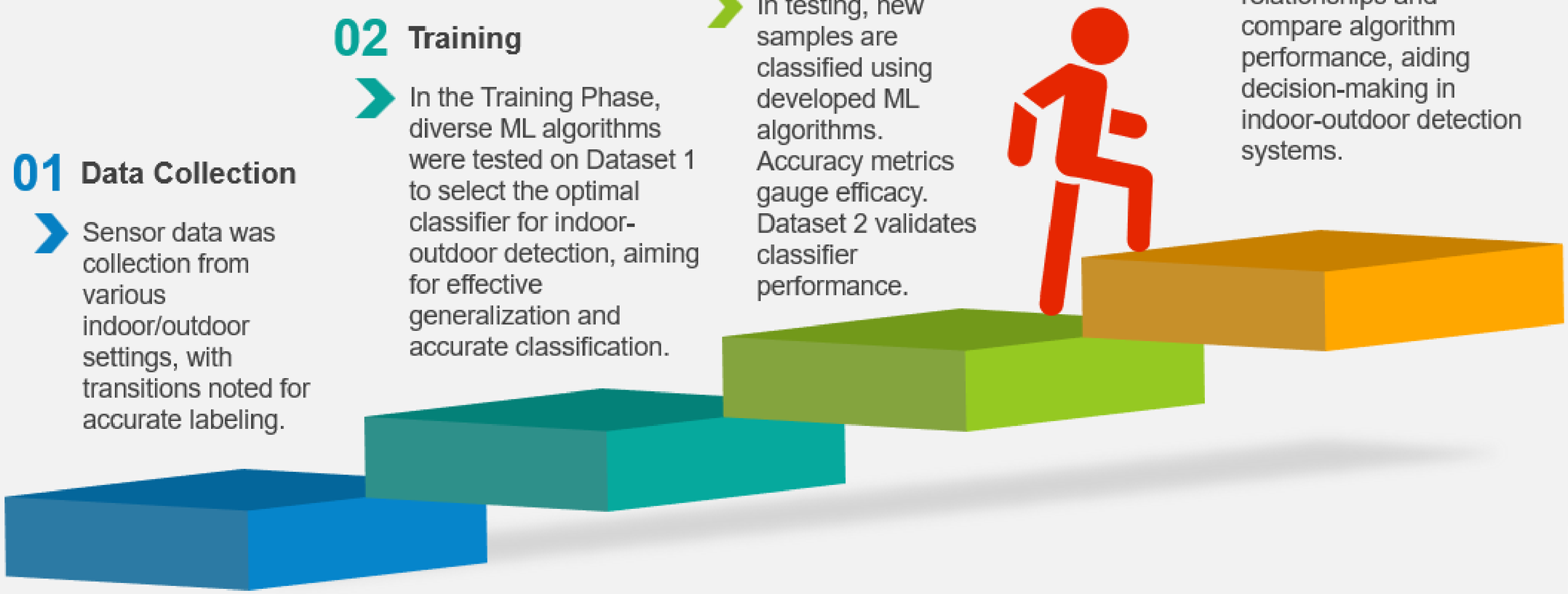
In the Training Phase, diverse ML algorithms were tested on Dataset 1 to select the optimal classifier for indoor-outdoor detection, aiming for effective generalization and accurate classification.

03 Testing

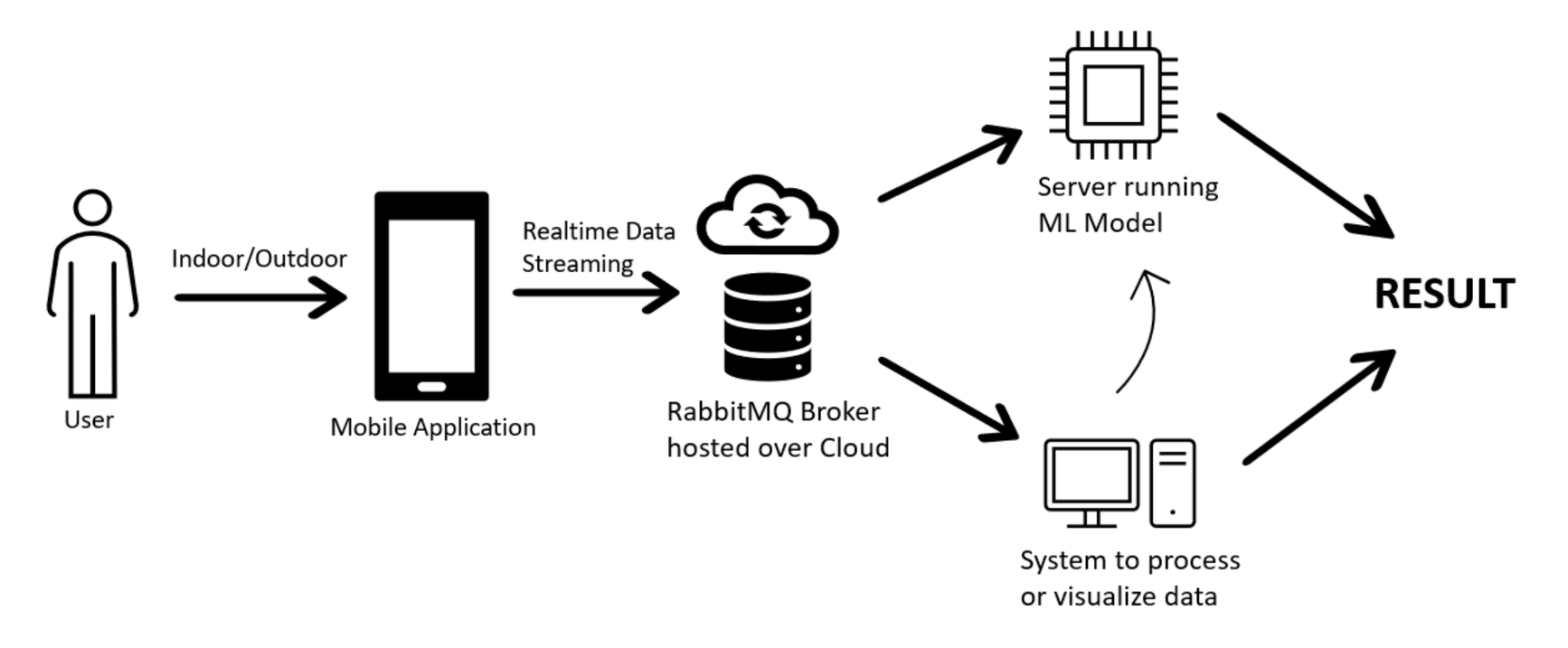
In testing, new samples are classified using developed ML algorithms. Accuracy metrics gauge efficacy. Dataset 2 validates classifier performance.

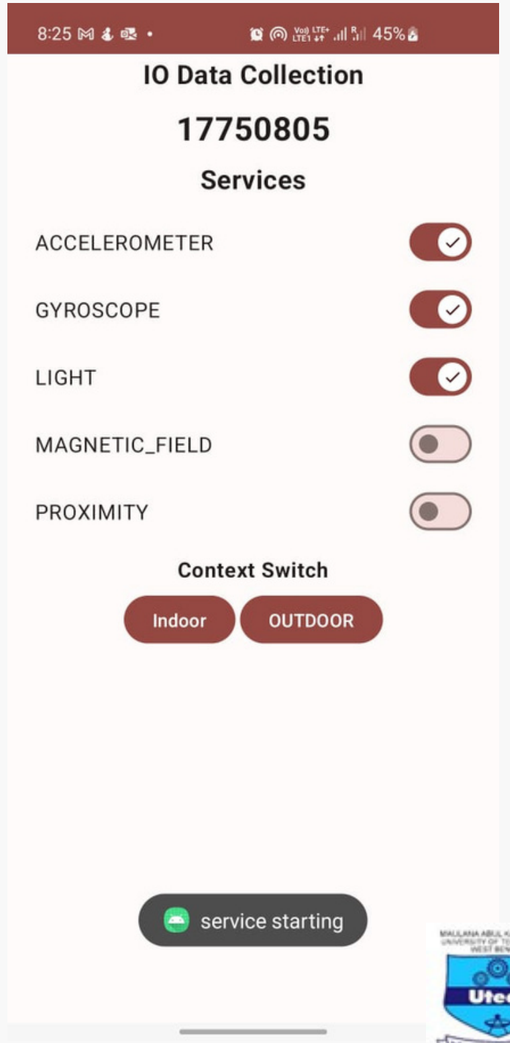
04 Data Visualization

Data plotting techniques helps to reveal sensor relationships and compare algorithm performance, aiding decision-making in indoor-outdoor detection systems.



Methodology

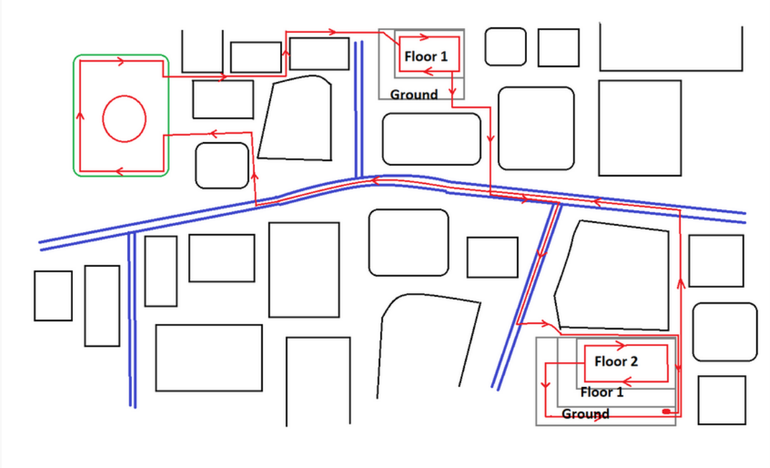




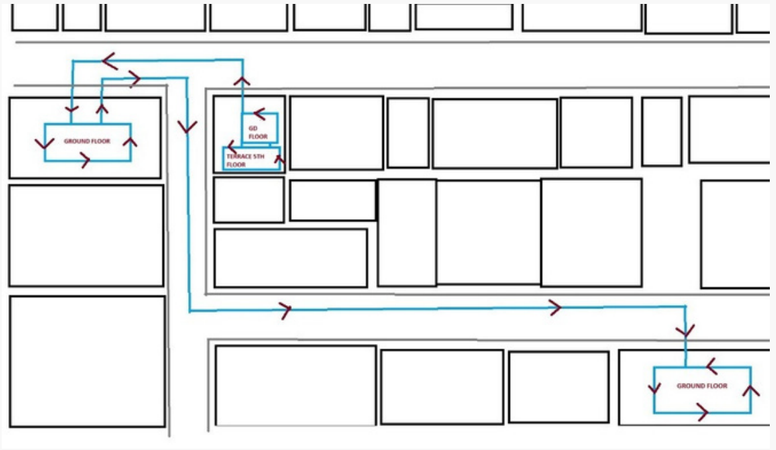
High Level System Design
and
Application Image

Methodology

Particulars	Dataset_1	Dataset_2
Smartphone	Vivo T2X 5G	Redmi Note 10 Pro Max
Android Version	14	13
Data Count	31269	14632
Indoor%	46	33
Outdoor%	54	67



Route 2



Route 1

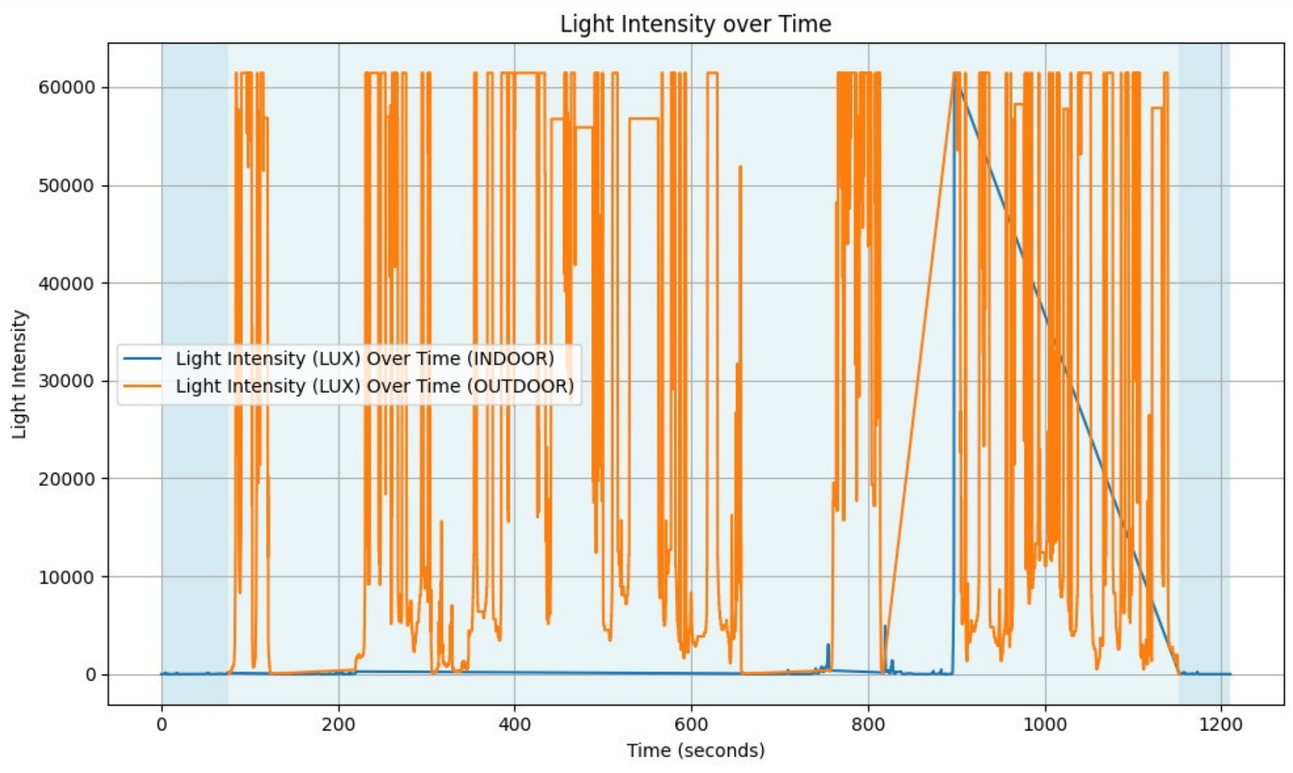
The data sets were gathered in residential environments, albeit in distinct geographical locations, employing varying handsets, all during a 20-minute stroll.

Result and Analysis

1. Two experiments prove that, ML models accurately detect indoor vs. outdoor environments using sensor data.

2. Key metrics and visualizations quantify and illustrate model performance and data nuances.

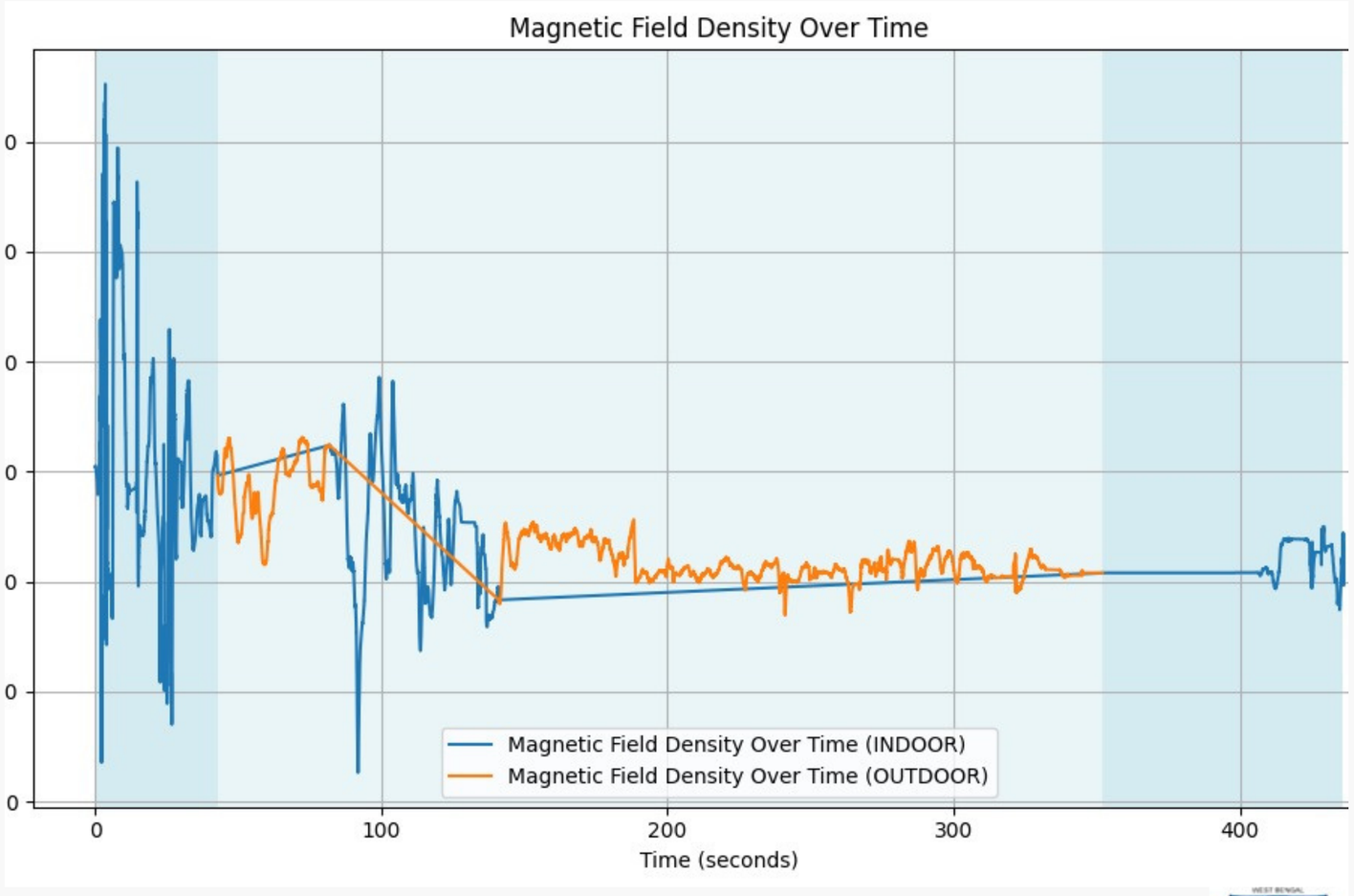
3. Raw sensor data plots, like the given figure, show trends in light intensity, aiding in understanding environmental variability.



Result and Analysis

4 Light sensors detect the intensity of light in their environment. The variation in light sensor readings can be significant due to differences in light sources and environmental conditions.We observe the variations as plotted in the graph.

5 Magnetometers discern between indoor and outdoor environments through variations in electromagnetic interference and the strength of the Earth's magnetic field as shown in the graph.

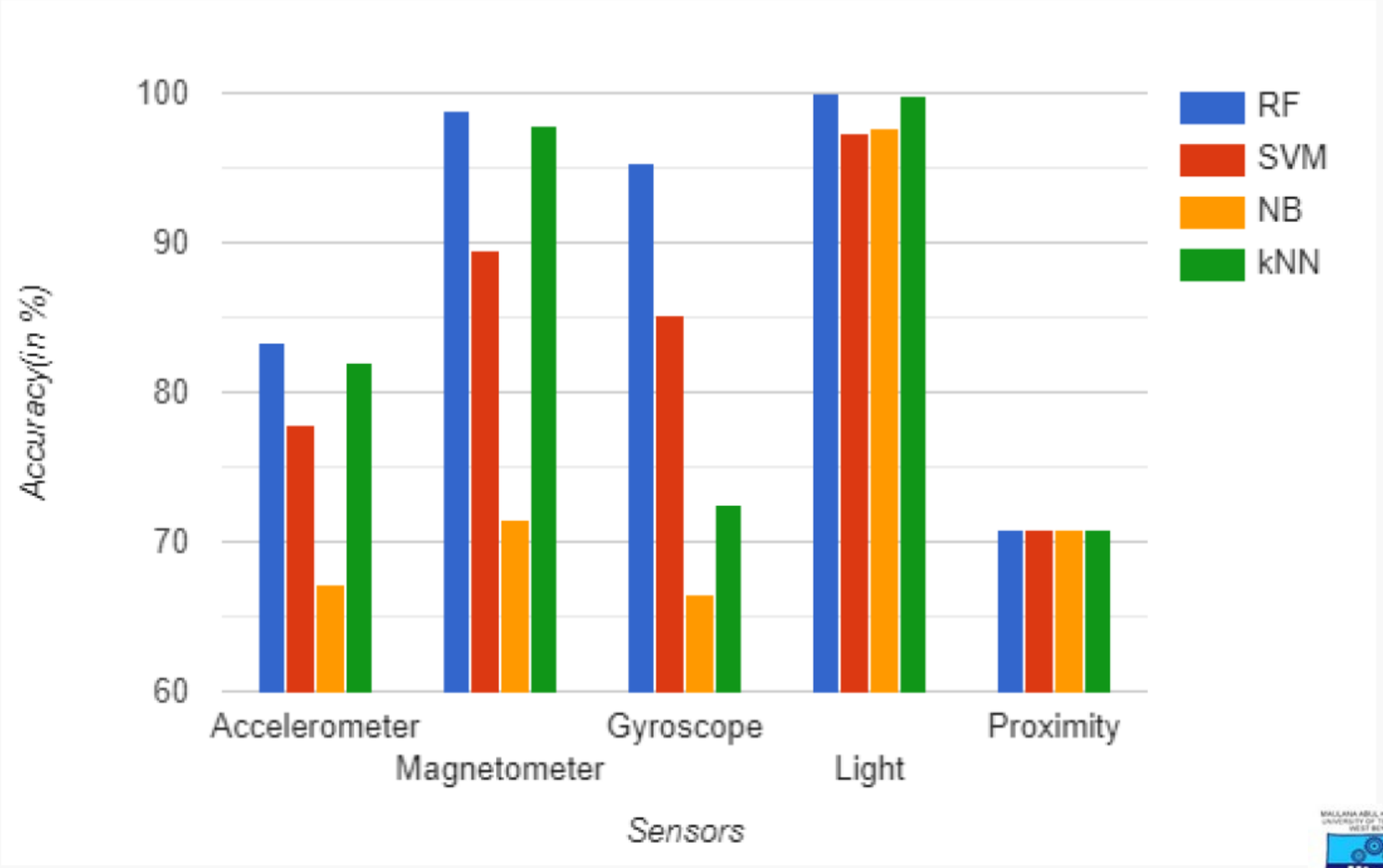


Experiment 1

1. We evaluated each sensor's performance with various machine learning algorithms.

2. Light intensity proves reliable, yet has limitations in low-light or obstructed scenarios.

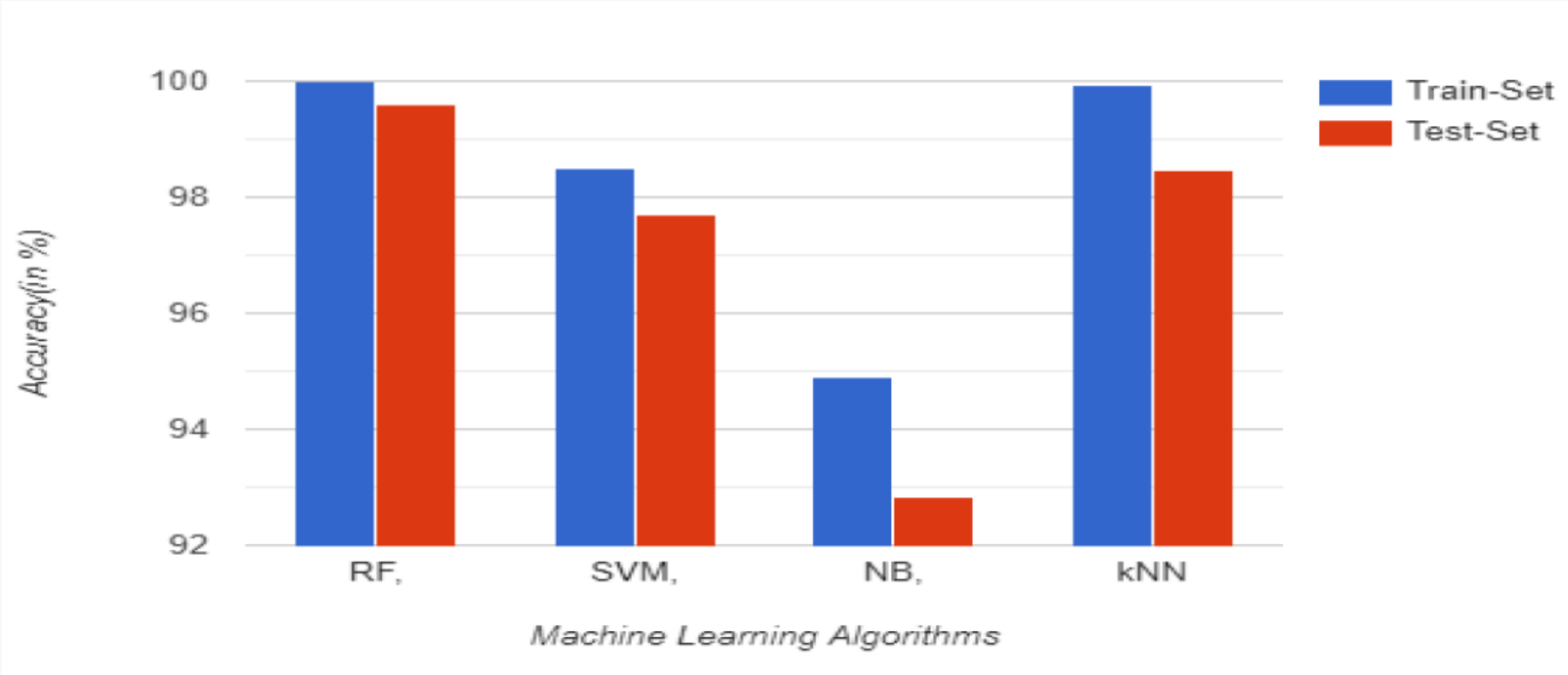
3. Random Forest stands out for its efficiency in handling large datasets without feature selection.



Experiment 2

1. We amalgamated different sensor data, that enhanced the detection accuracy.

2. Random Forest achieved 100% accuracy in the training dataset and 99.59% in the testing dataset.



Summary

This thesis develops algorithms to distinguish indoor/outdoor environments using sensor data and advanced machine learning.

Successfully achieved an exceptional level of accuracy, nearing 100%, in the precise detection of user positions

An Android app was created, released open-source, enabling easier collection of labeled sensor data for research and development.

Future Work

Feature Expansion:

Incorporate additional sensor data like GPS and Wifi for a better understanding of the environment.

Advanced ML Models:

Explore using advanced algorithms like RNNs to improve accuracy.

plore using advanced algorithms like RNNs to improve accuracy.

Integration with Smart Environments:

Integrate with IoT platforms for context-aware applications and personalized recommendations

Thank you