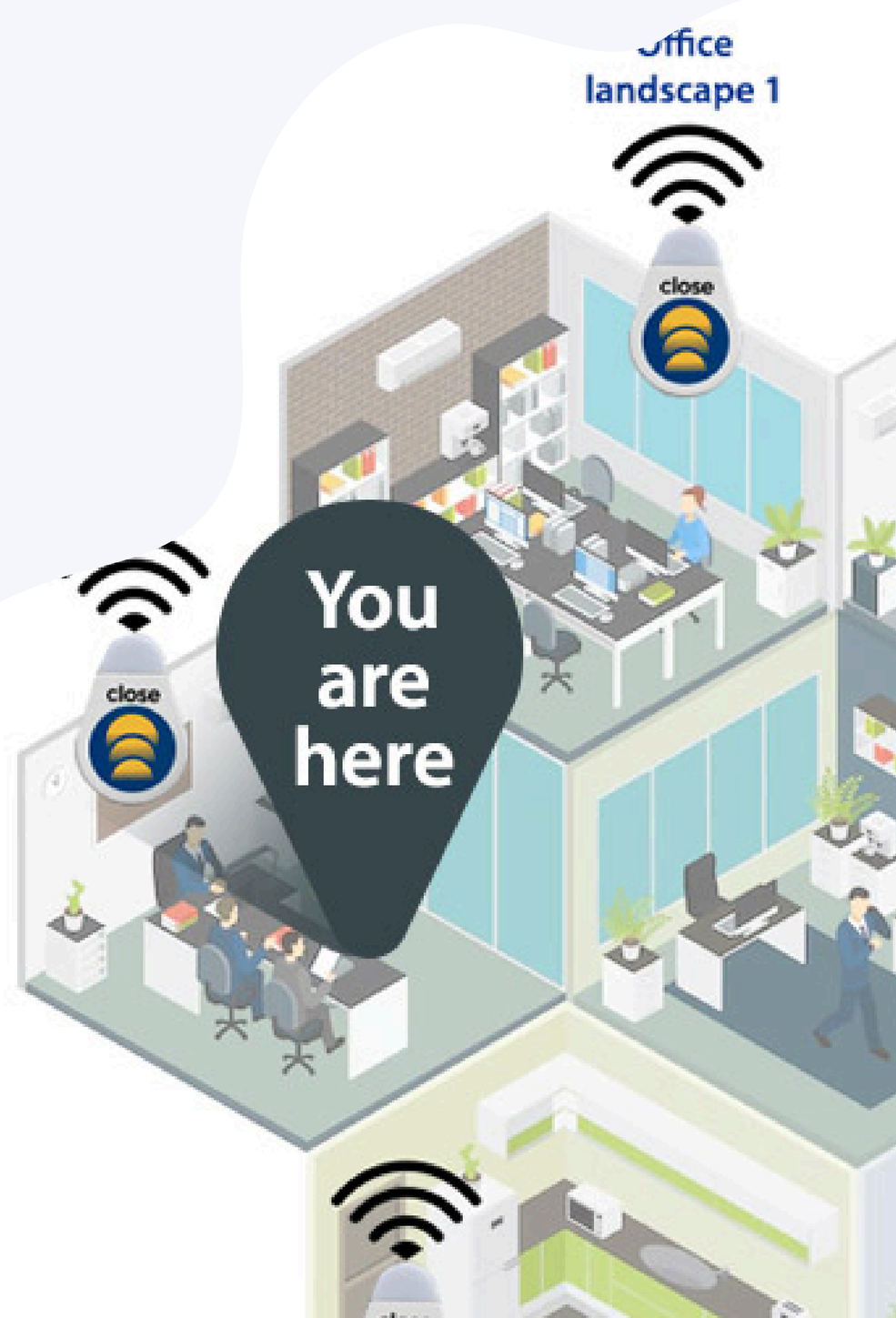
Machine Learning

Indoor Localization Using WLAN Fingerprinting project

Joy Anne P. Dela Cruz - 60301959
Asma AlWan - 60106908



Introduction

Indoor localization aka (IPS) is a technology that determines the position of objects or people within enclosed spaces . There are many ways of applying indoor localization; this project will focus on using Wi-Fi Fingerprinting. There are three components to apply indoor localization in this method

Wi-Fi access points that broadcast signals to be received by

mobile devices; these signals will be input in

an algorithm that processes the data to give an estimated location

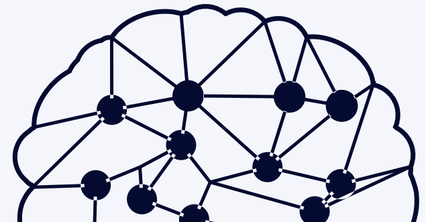
Objectives

01 Develop a Regression Model

02 Create a Classification Model

03 Evaluate Models

04 Simulate Real-World Application



Exploratory Data Analysis (EDA)

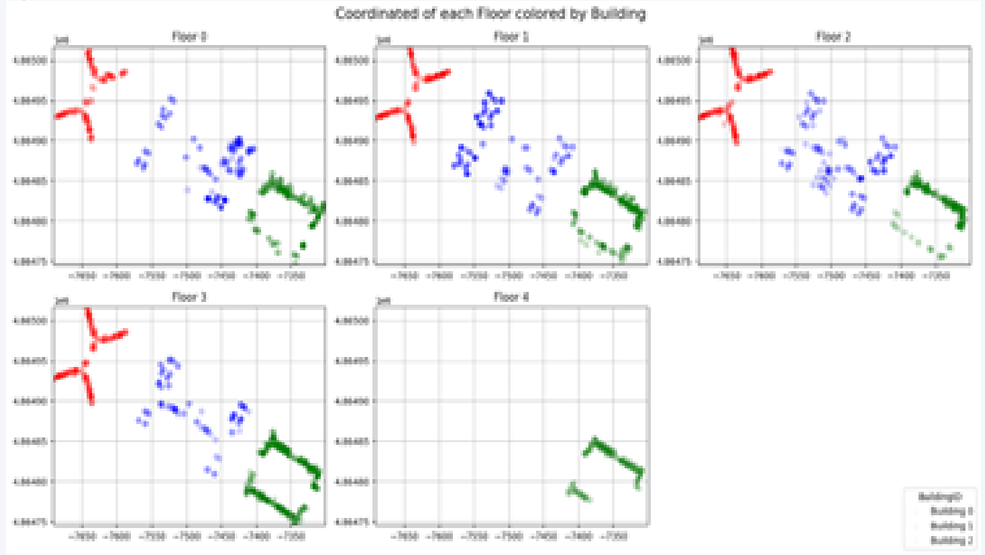
Dataset Overview

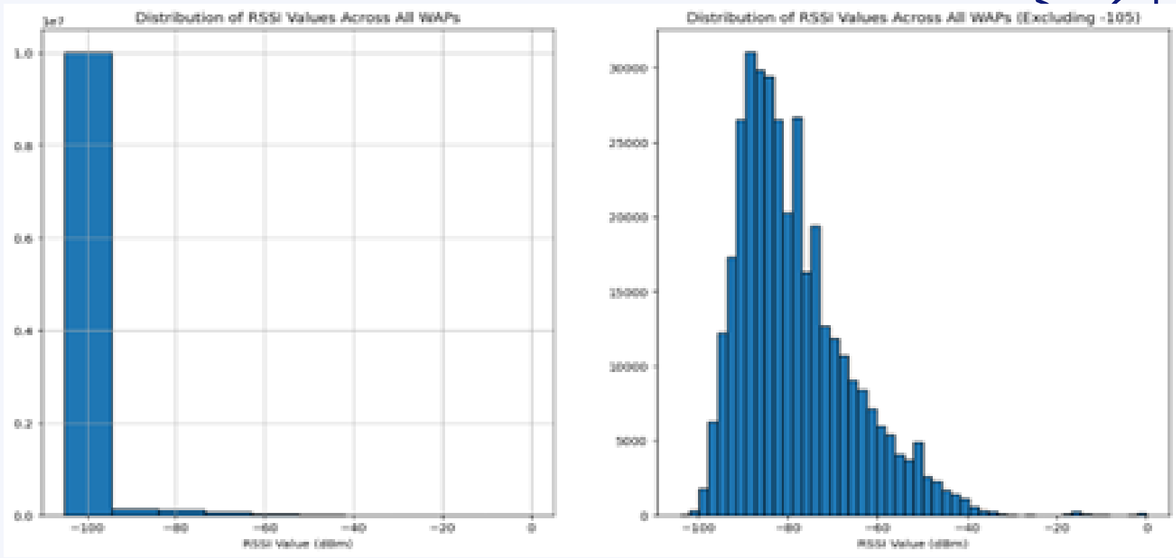
The UJIIndoorLoc dataset is a Multi-Building, Multi-Floor indoor localization database designed to test Indoor Positioning Systems that rely on WLAN/Wi-Fi fingerprinting.

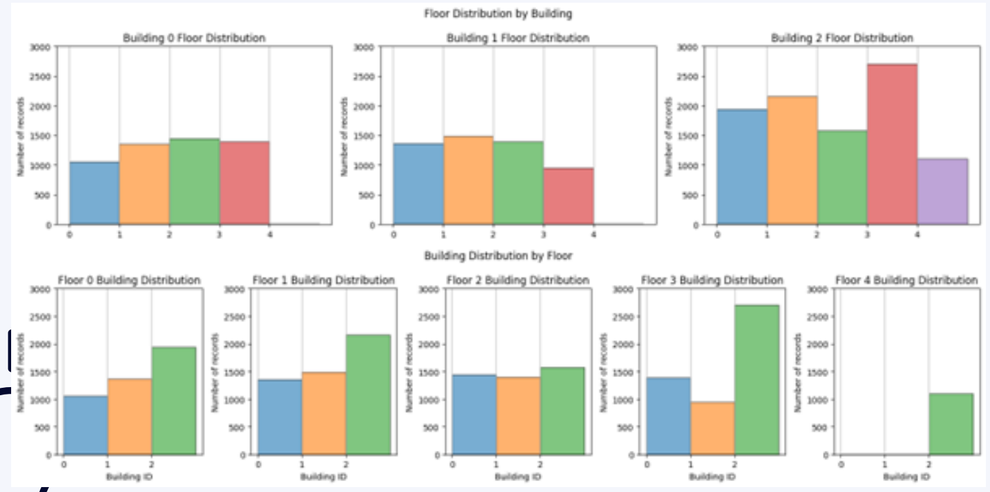
This dataset contains 528 attributes. Of these, 520 are RSSI values, which represent signal intensities from different Wi-Fi Access Points (WAPs).

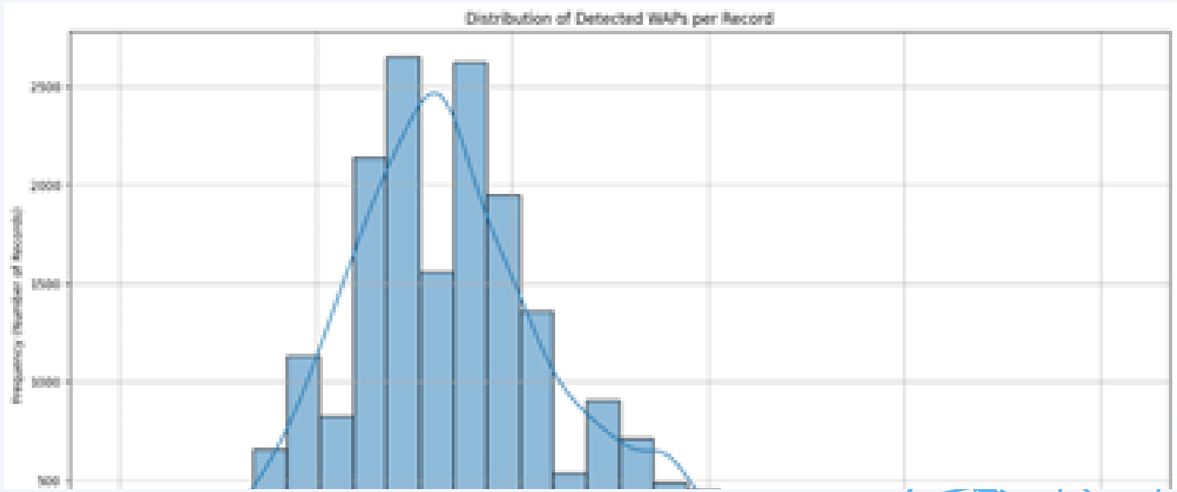
Key Insights from EDA:

The connectionist method, on the other hand, focuses on building neural networks by replicating the biological composition of the human brain.


Coordinated of each Floor colored by Building
Floor 0
Floor 1
Floor 2
Floor 3
Floor 4
BuildingID
Building 0
Building 1
Building 2


Distribution of RSSI Values Across All WAPs
RSSI Value (dBm)
Distribution of RSSI Values Across All WAPs (Excluding -105)
RSSI Value (dBm)


Floor Distribution by Building
Building 0 Floor Distribution
Building 1 Floor Distribution
Building 2 Floor Distribution
Number of records
Building Distribution by Floor
Floor 0 Building Distribution
Floor 1 Building Distribution
Floor 2 Building Distribution
Floor 3 Building Distribution
Floor 4 Building Distribution
Building ID


Distribution of Detected WAPs per Record
Frequency (Number of Records)

Model Development: Regression Model

To predict a user's exact indoor location (coordinates: Longitude , Latitude), we developed a regression model using WLAN signal strengths (RSSI values) as input features.

- Objective: Estimate precise coordinates of a user.
- Basic Regression Model:
  - Linear Regression
  - KNN
  - Decision tree
- Advanced Regression Model:
  - Neural Network
  - Convolutional Neural Network

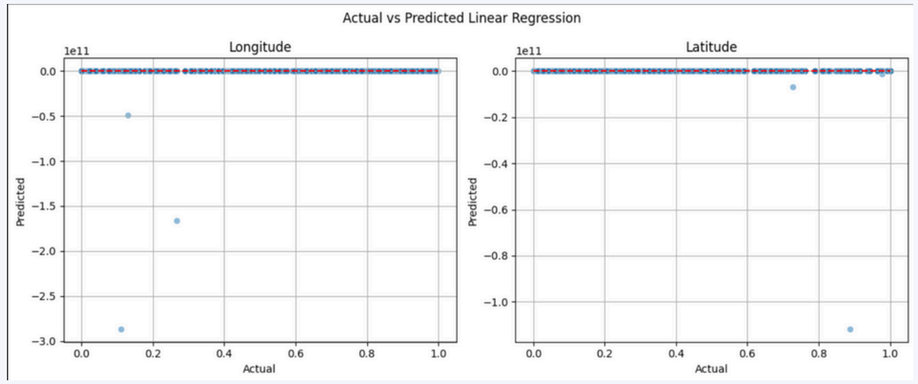
Classification Model

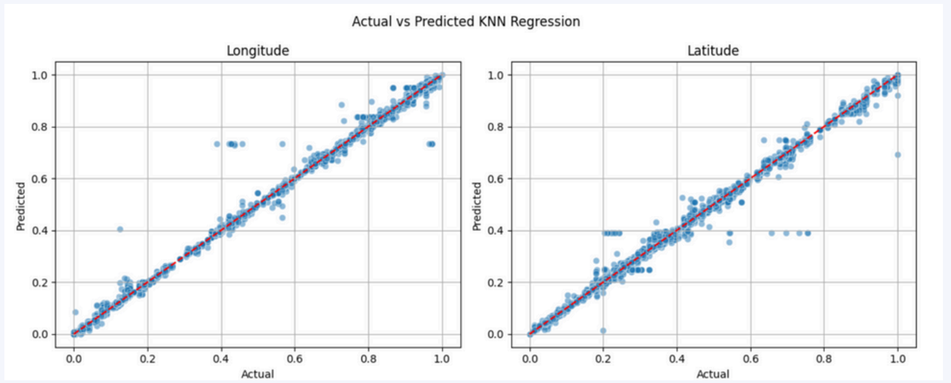
developing a classification model is needed for narrowing down the user's location before predicting exact coordinates

- Objective: Classify which building and floor a user is located on using Wi-Fi signal patterns.
- Advanced Classification Model:
  - Deep learning Neural Network
- Basic Classification : Desicion Tree , Support Vector Machine(SVM) and Random Forest.

These models offers potential applications in asset tracking, navigation systems, and smart environments by providing a basis for predicting user location indoors where GPS is unreliable.

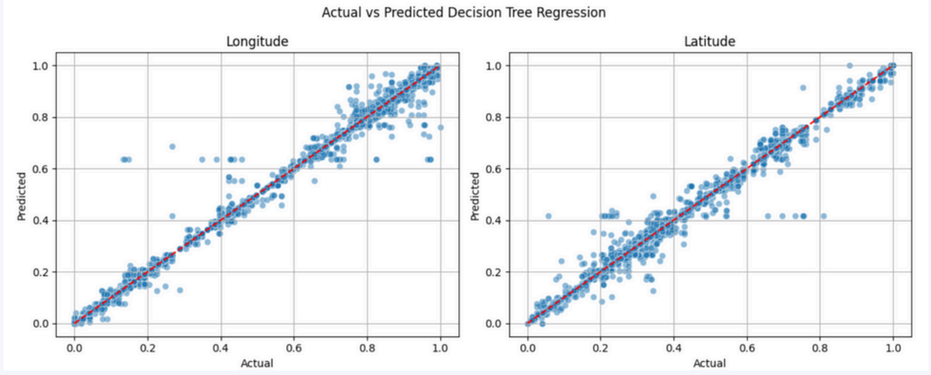
Baisc Regression Model Evaluation





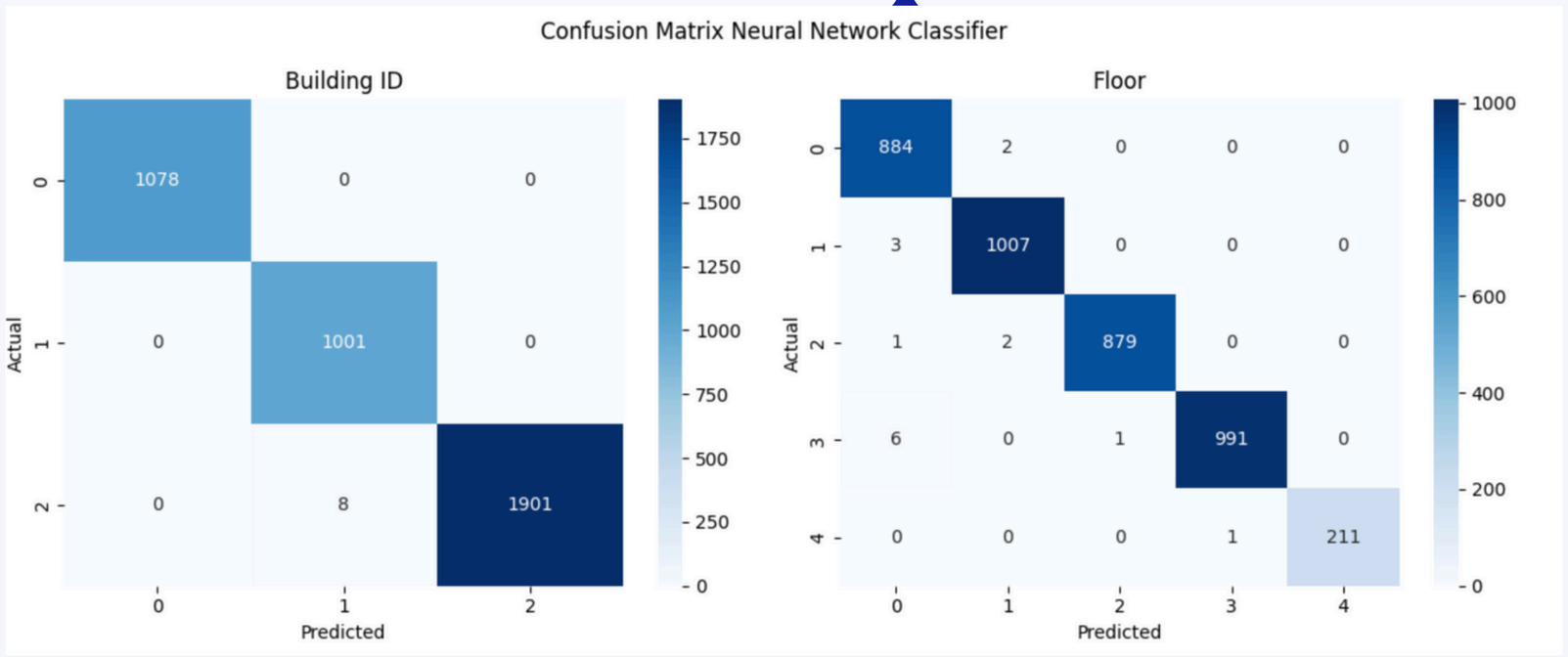

```
============== KNN Regression Longitude =====
>> R2:      0.9962969976754037
>> MSE:     0.0003776058626060708
>> RMSE:    0.01943208333159548
>> MAE:     0.005666715982331163

============== KNN Regression Latitude ======
>> R2:      0.9935700217857931
>> MSE:     0.0003938891273756219 5
>> RMSE:    0.019846640203712615
>> MAE:     0.006936412229373794
```

```
============== Linear Regression Longitude =========
>> R2:      -2.7595046785093947e+20
>> MSE:     2.8139467738725302e+19
>> RMSE:    5304664715.015013
>> MAE:     125972249.93362425

============== Linear Regression Latitude ==========
>> R2:      -5.130627484946384e+19
>> MSE:     3.142931928555845e+18
>> RMSE:    1772831613.1420505
>> MAE:     30061203.08977939
```




```
============== Decision Tree Regression Longitude ========
>> R2:      0.9917097917841592
>> MSE:     0.0008453765215682828
>> RMSE:    0.029075359354069606
>> MAE:     0.008826224407514849

============== Decision Tree Regression Latitude =========
>> R2:      0.9887807561895432
>> MSE:     0.000687271092855533
>> RMSE:    0.02621585575287469
>> MAE:     0.008858853847447709
```

Advanced Classification Model Evalution



...

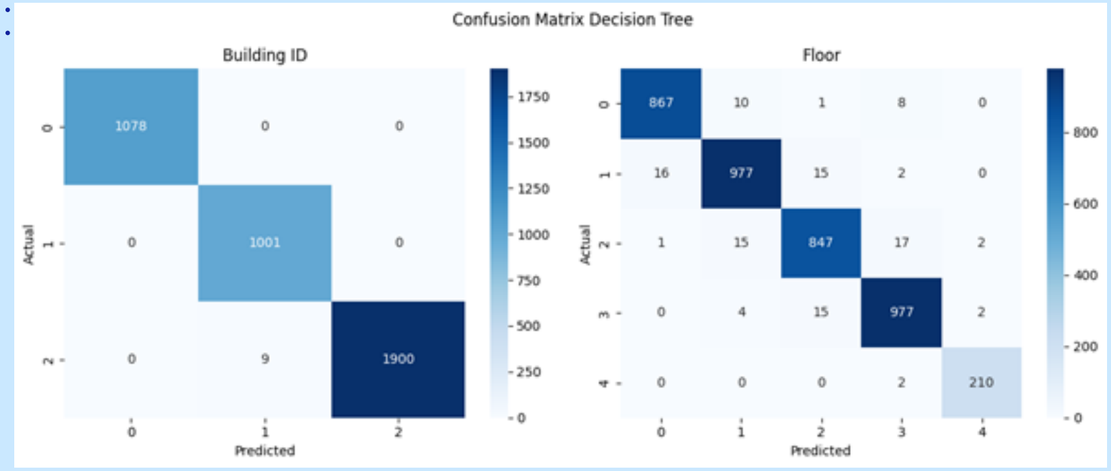
accuracy			0.996	3988
macro avg	0.997	0.996	0.996	3988
weighted avg	0.996	0.996	0.996	3988

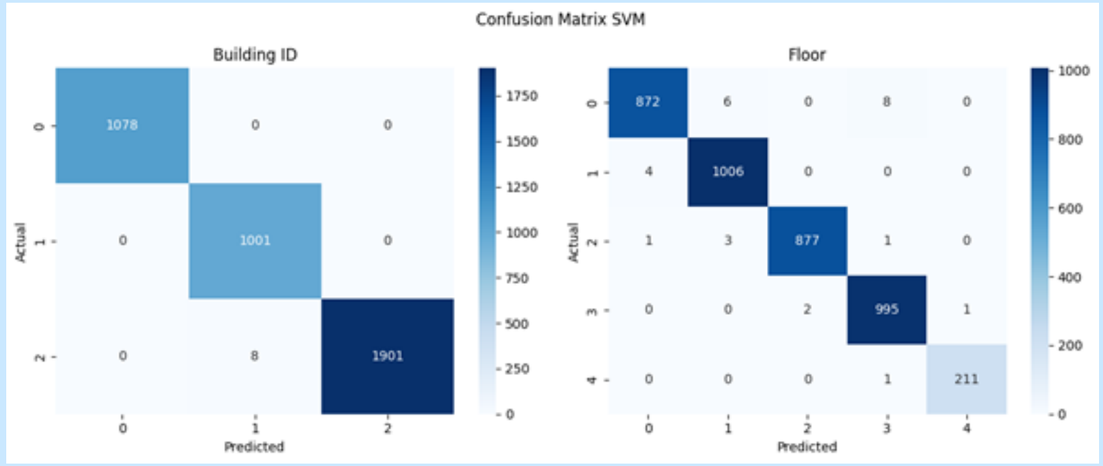
Cross validation for Basic regression & Advanced Classification

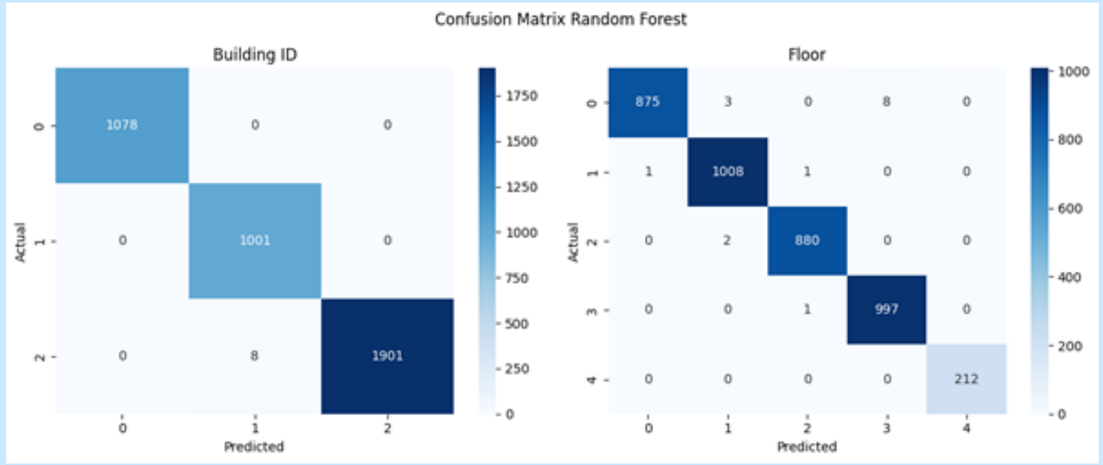
	r2	neg_mse	neg_mae	neg_rmse
knn_latitude	9.920323e-01	4.863379e-04	6.858861e-03	2.178530e-02
knn_longitude	9.945644e-01	5.428723e-04	5.816557e-03	2.292337e-02
tree_latitude	9.890702e-01	6.666593e-04	8.525180e-03	2.573555e-02
tree_longitude	9.931235e-01	6.873599e-04	7.922119e-03	2.609382e-02
linear_longitude	-1.902545e+22	1.929121e+21	7.183938e+08	2.710088e+10
linear_latitude	-3.715878e+22	2.286212e+21	8.234643e+08	2.743071e+10

```
[BUILDINGID] Fold1 Accuracy: 0.9969909729187563 | F1-macro: 0.9970331017902235
[FLOOR] Fold1 Accuracy: 0.9964894684052157 | F1-macro: 0.9970335972042855
[BUILDINGID] Fold2 Accuracy: 0.9989969909729187 | F1-macro: 0.9989601263746349
[FLOOR] Fold2 Accuracy: 0.995987963891675 | F1-macro: 0.995809012786508
```

Basic Classification Model Evaluation





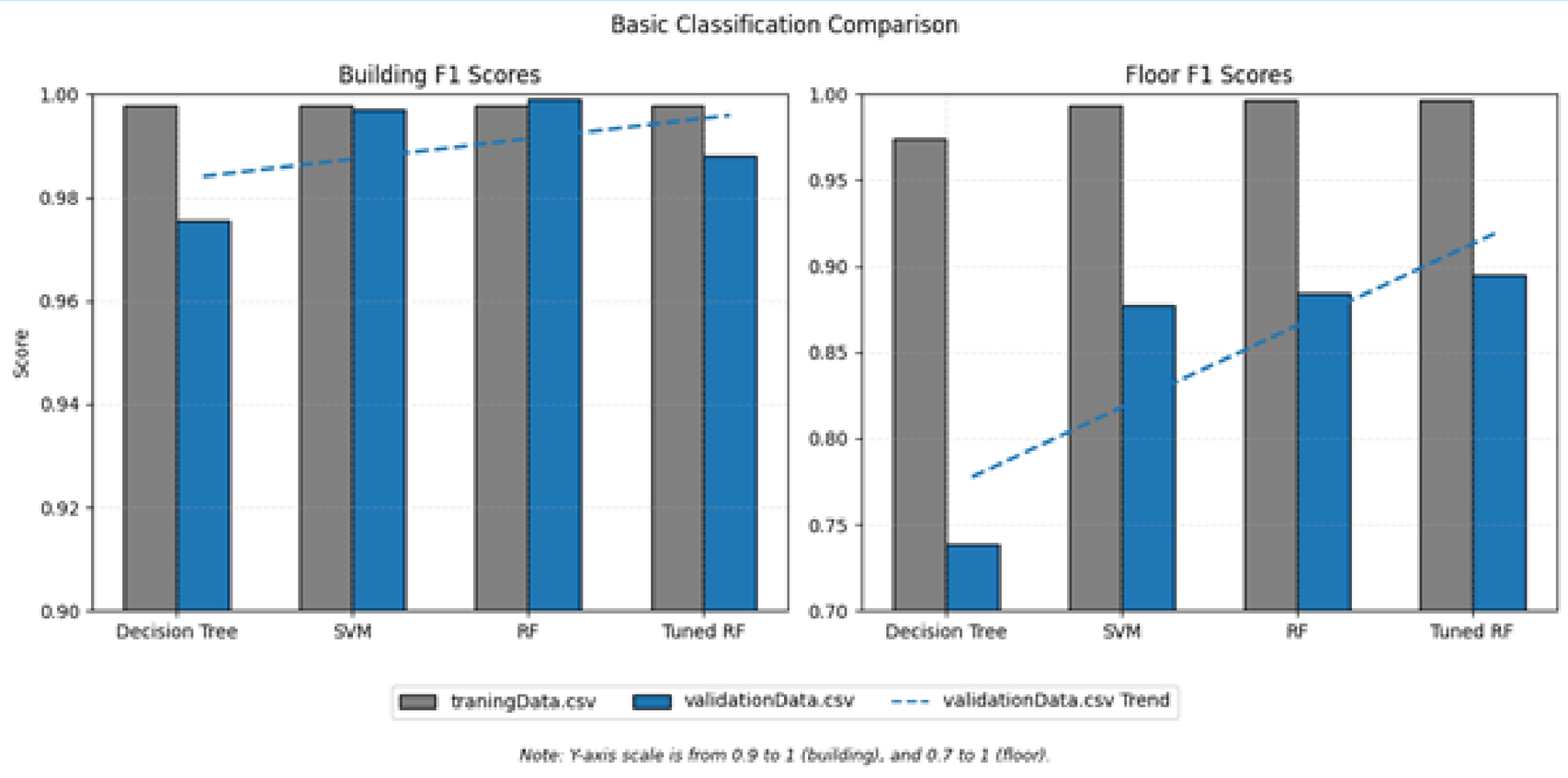


Cross Validation

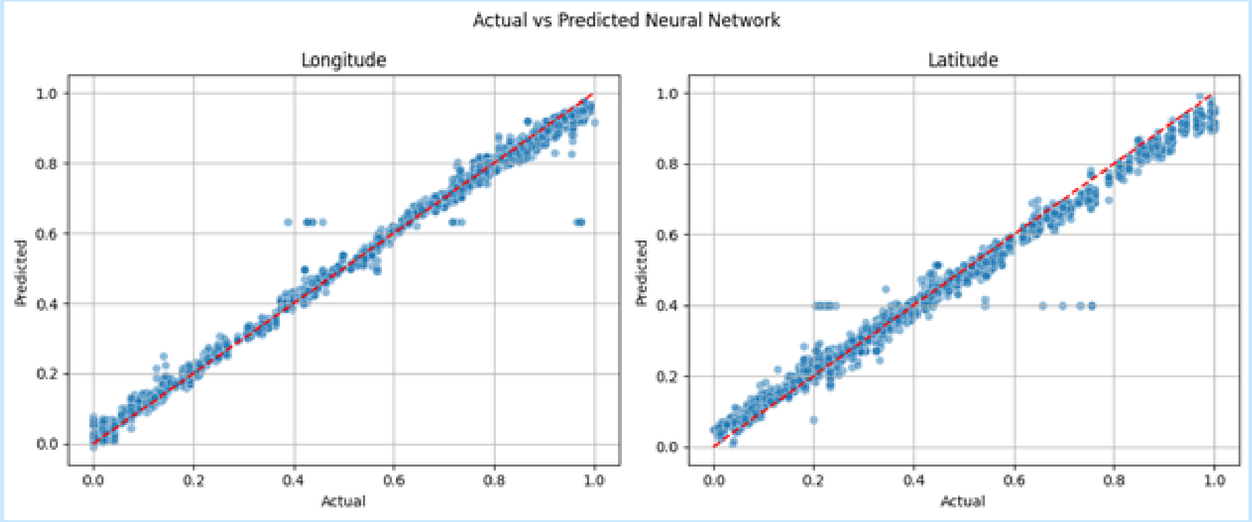
Table 1: Average Cross Validation Metrics on Initial Classification Models

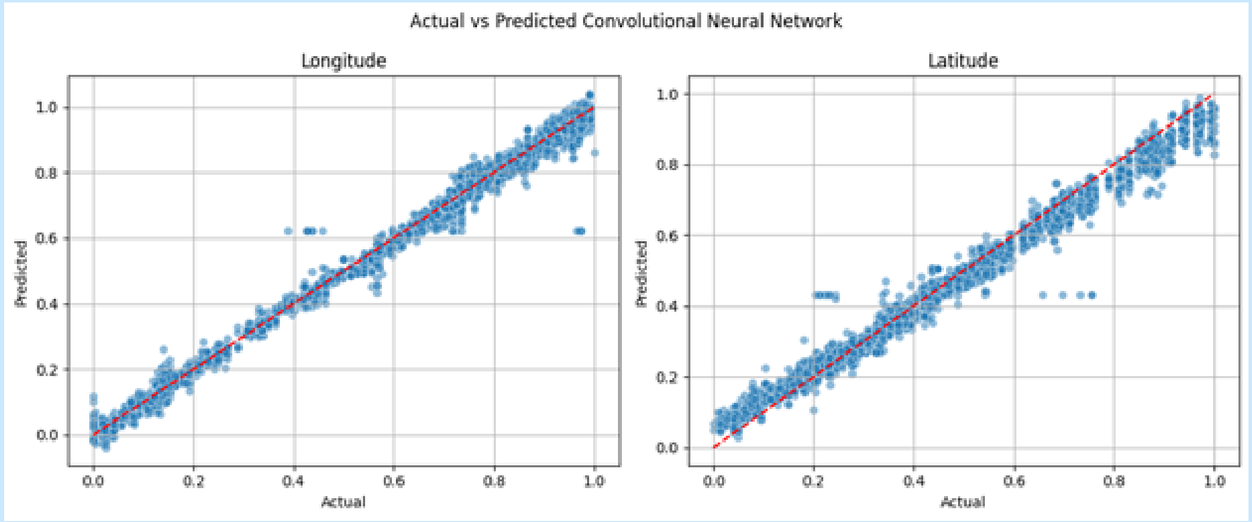
Metric	SVM (Building)	RF (Building)	Tree (Building)	RF (Floor)	SVM (Floor)	Tree (Floor)
Accuracy	0.997793	0.997743	0.996890	0.996740	0.993178	0.974319
Precision Macro	0.997725	0.997962	0.996829	0.997230	0.994047	0.974701
Precision Micro	0.997793	0.997743	0.996890	0.996740	0.993178	0.974319
Precision Weighted	0.997810	0.997758	0.996914	0.996764	0.993203	0.974387
Recall Macro	0.997902	0.997602	0.996874	0.997208	0.993367	0.976212
Recall Micro	0.997793	0.997743	0.996890	0.996740	0.993178	0.974319
Recall Weighted	0.997793	0.997743	0.996890	0.996740	0.993178	0.974319
F1 Macro	0.997805	0.997774	0.996839	0.997209	0.993692	0.975401
F1 Micro	0.997793	0.997743	0.996890	0.996740	0.993178	0.974319
F1 Weighted	0.997793	0.997742	0.996891	0.996741	0.993177	0.974306

Model Comparisons



Advanced Regression Model Evaluation



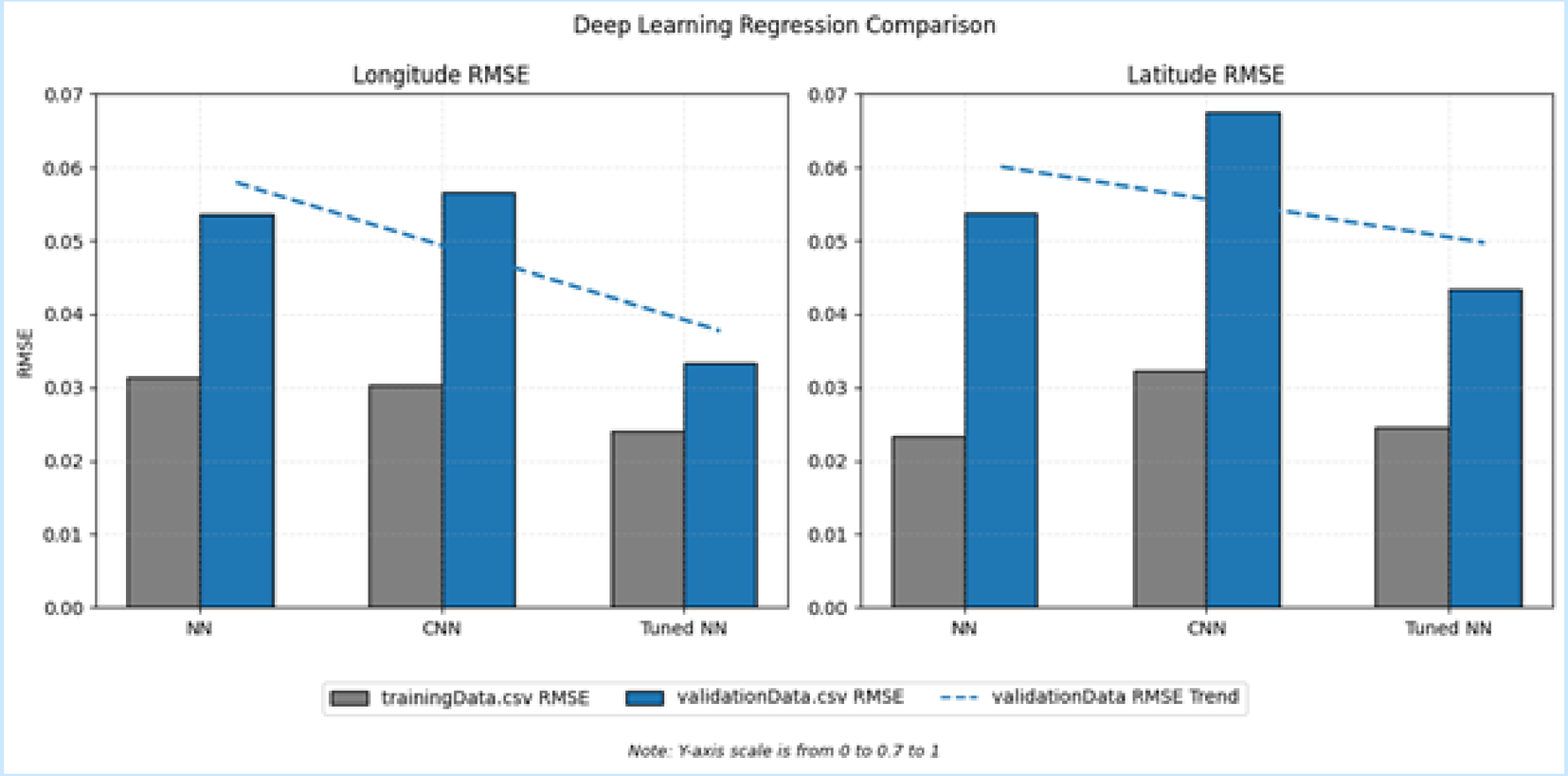


Cross Validation

Table 2: Average Cross Validation Metrics on Initial Regression Models

Metric	NN Longitude	NN Latitude	CNN Longitude	CNN Latitude
R2	0.9921	0.9901	0.9909	0.9842
MSE	0.00078	0.00060	0.00091	0.00097
RMSE	0.02785	0.02418	0.02996	0.03084
MAE	0.01966	0.01587	0.02074	0.02138

Model Comparisons



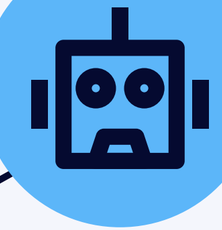
Practical Implications & Applications

Indoor Navigation:

Enhancing navigation systems in malls, airports, museums, etc.

Asset Tracking:

Real-time tracking of equipment, people, and resources inside buildings.



Emergency Services:

Assisting first responders in locating people in large buildings.

Smart Buildings:

Implementing indoor positioning systems in IoT-enabled buildings.

Lessons Learned and Future Improvments

01 the Importance of clean, accurate data for model performance and prediction.

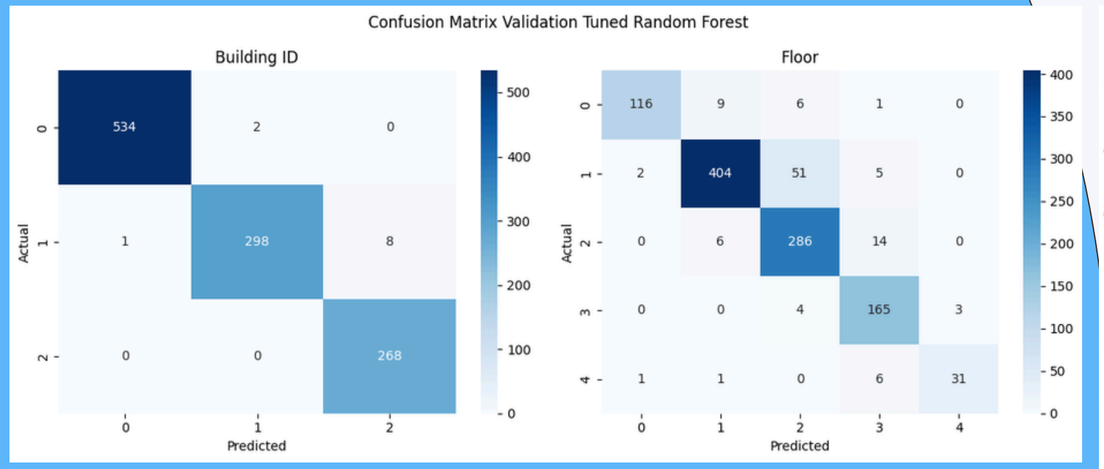
02 new experiment with deep learning models—which I have not tried before.

03 simplicity over complexity:Applying the deep learning models were good practice but the results are not much better the simple models

04 WLAN signal variability, need for more data for better generalization.

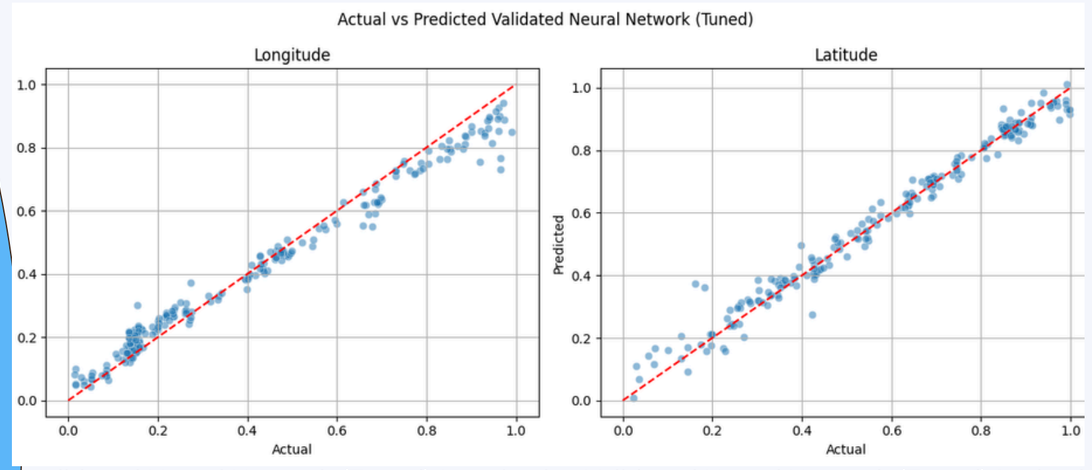
Real-life model deployment simulation

to truly make sure that there is little to no overfitting and the model generalizes well enough for real-world deployment, we now used the validationData.csv which has not been touched or seen by the models yet.



Validation Tuned Random Forest Building
>> Accuracy: 0.9900990099009901
>> Precision macro: 0.9874928890694838
>> Recall macro: 0.9889842319347886
>> F1 Score macro: 0.9881236953869962

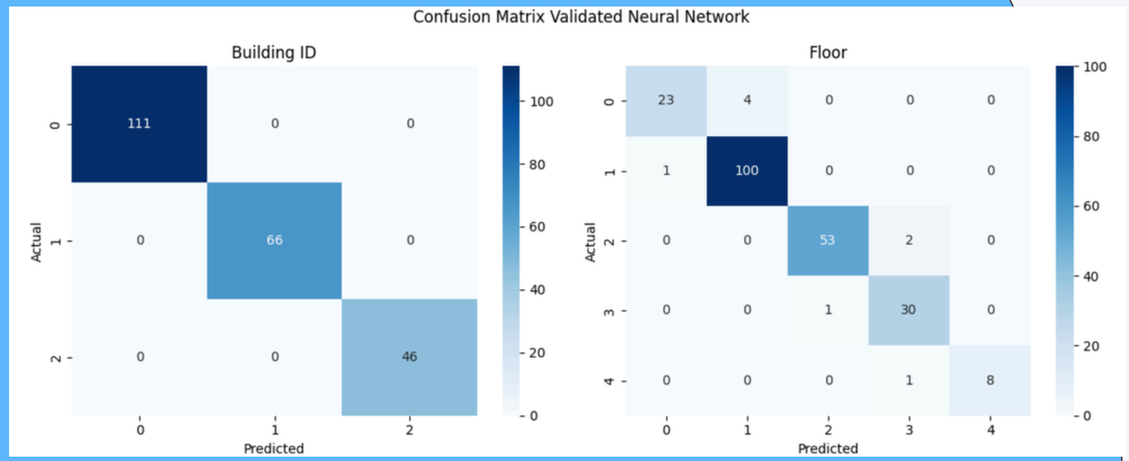
Validation Tuned Random Forest Floor
>> Accuracy: 0.9018901890189019
>> Precision macro: 0.9073082444197258
>> Recall macro: 0.8884122793151521
>> F1 Score macro: 0.8949531321302807

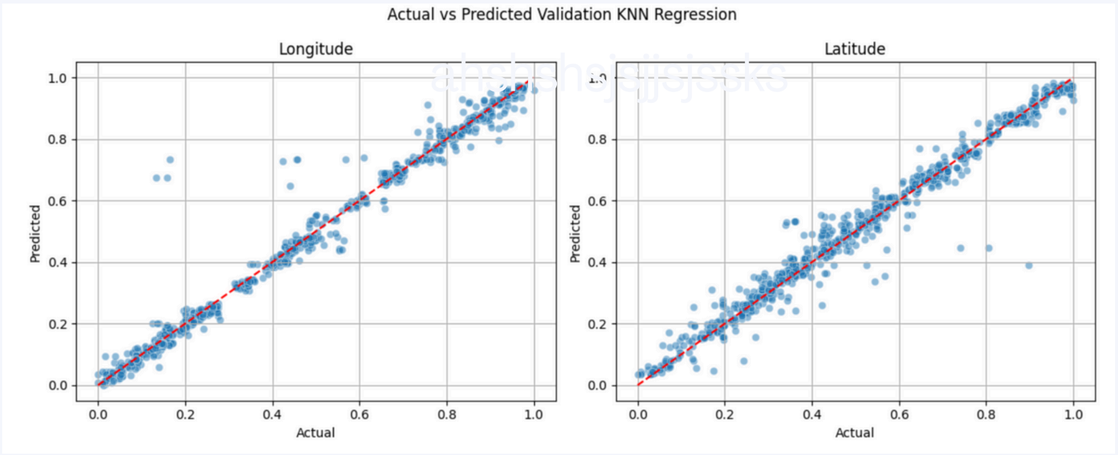


validated Neural Network (Tuned) Longitude
>> R2: 0.9682915558849592
>> MSE: 0.0027582927837203593
>> RMSE: 0.052519451479621906
>> MAE: 0.04028888003605451

Validated Neural Network (Tuned) Latitude
>> R2: 0.9791448045780028
>> MSE: 0.001415959046974701
>> RMSE: 0.037629231283334785
>> MAE: 0.026356536510275575

Real-life model deployment simulation





accuracy			0.960	223
macro avg	0.962	0.932	0.946	223
weighted avg	0.960	0.960	0.959	223

```
============== Validation KNN Regression Longitude ==============
>> R2:      0.9827873337998492
>> MSE:     0.0015834677929710103
>> RMSE:    0.039792810820184724
>> MAE:     0.017425597199181957

============== Validation KNN Regression Latitude ==============
>> R2:      0.9782329000040155
>> MSE:     0.0014803098352321141
>> RMSE:    0.038474794804288615
>> MAE:     0.02114527297659572
```

Thank You

References

[1] Magda Chelly and Nel Samama. 2009. New techniques for indoor positioning, combining deterministic and estimation methods. In Proceedings of the European Navigation Conference - Global Navigation Satellite Systems (ENC-GNSS 2009), Naples, Italy, 1–12. https://hal-01367483.