Machine Learning-Based RSSI Prediction and Heatmap Generation

Harikeshav Rameshkumar

Dr Venkatesh Ramaiyan

Harikeshav R is a Student at National Public School, Indiranagar. For inquiries or further discussions related to this report, please contact r.harikeshav@icloud.com.

Table of Contents

Abstract 3

Machine Learning-Based RSSI Prediction and Heatmap Generation 4

Dataset Generation 4

  Empirical Formula 5

  Generation of Synthetic Values 5

Training ML Models 5

  Decision Trees 6

  Gaussian Process Regression 6

  K Nearest Neighbor 6

  Linear Regression 6

  Support Vector Regression 6

Testing ML Models 7

Results and Conclusion 7

Bibliography 8

Tables 9

Figures 10

Abstract

The rapid proliferation of indoor wireless networks, driven by the demand for seamless connectivity and the growth of Internet of Things (IoT) devices, has highlighted the need for optimal network performance and coverage. Traditional methods for predicting Received Signal Strength Index (RSSI) values, such as empirical and deterministic models, often fail to consider environmental variables and complexities. This article explores the use of machine learning (ML) models to predict RSSI and generate a heat map. A synthetic dataset was created using a log distance scoring model. Five machine learning algorithms were trained and evaluated, including decision trees, Gaussian process regression, K-nearest neighbor, linear regression, and support vector regression. Gaussian process regression showed the highest R2 score (0.975), indicating superior goodness of fit, while K Nearest Neighbors came in second with an R2 difference of 0.961. Decision trees provided interpretability but with slightly lower accuracy, and support vector regression showed reasonable performance. Linear regression was less effective. Choosing the most appropriate model depends on factors such as complexity, interpretability, and computational requirements. Additional validation on a larger data set or through cross-validation is recommended to evaluate generalization performance. This research contributes to improving the deployment of indoor wireless networks, improving connectivity, and facilitating the growth of IoT devices.

Keywords: Wireless Networks; RSSI Prediction; Machine Learning; Indoor Connectivity; IoT Devices; Network Performance

Machine Learning-Based RSSI Prediction and Heatmap Generation

In recent years, indoor wireless network deployment has surged due to the demand for seamless connectivity and IoT device growth. Achieving optimal network performance and coverage is challenging, but planning wireless device installation considering various factors can save time and costs. Traditional methods, like on-field RSSI measurements, are time-consuming and lack environmental consideration. Conventional models use empirical or deterministic methods to determine the environment-related path loss prediction (Ostlin, Zepernick, & Suzuki, 2010).

Empirical models base signal strength on distance from the transmitter but neglect environmental factors and disturbances, rendering them impractical (Ayadi, Ben Zineb, & Tabbane, 2017). Deterministic models, such as the ray-tracing model, use physical laws for signal strength prediction but are computationally intensive. To address these issues, machine learning can be used for RSSI prediction and heatmap generation. While real data is most accurate, ML approximations are faster and can be sufficiently accurate for practical use cases (Popescu, Nikitopoulos, Constantinou, & Nafornita, 2006). These approximations depend on room shape, transmitter-receiver locations, distances, and other variables to generate signal strength heatmaps.

Dataset Generation

To generate an experimental dataset, the log distance model was chosen for the sake of simplicity and lesser computational load. Real-world data could have been recorded and used as the dataset, but it would have been time-consuming and difficult to tweak certain environment variables to generate different outputs.

Empirical Formula

The following empirical formula was used for the log distance model:

$$RSSI(d) = RSSI(d_0) - 10n \log\left(\frac{d}{d_0}\right) \tag{1}$$

$RSSI(d)$ is the received signal strength by a receiver placed at a distance d from the transmitter, $RSSI(d_0)$ is received signal strength by a receiver placed at a reference distance d_0 from the transmitter, and n is the path loss exponent, which depends upon the propagation environment.

Generation of Synthetic Values

A Python script was created to generate synthetic RSSI values. A rectangular room of length = 6 units and width = 4 units was chosen. The location of the transmitter was randomly chosen. The path loss exponent was taken as 2.5. The reference RSSI was taken as -40 dBm at a reference distance of 1 meter from the transmitter. Thus, 1000 random coordinates were picked within the boundaries of the room, and RSSI values at those locations were generated by substituting values in Equation 1 and obtaining the result. The generated synthetic RSSI values were saved into a CSV file.

Training ML Models

The generated RSSI values were used to train various ML models. The synthetic data was split into an 80/20 split. 80% of the data was used to train these models, whereas the remaining 20% was used to test the accuracy of each model.

Five different machine learning algorithms were trained using the same selection of training data, and their accuracies were compared. They are briefly described for the given use case as follows.

Decision Trees

Decision trees are versatile models for spatial prediction tasks. They split data into branches based on features, making them interpretable and effective for predicting RSSI values. However, they may struggle with capturing complex spatial relationships.

Gaussian Process Regression

Gaussian processes excel in spatial prediction, providing probabilistic estimates of RSSI values. They model spatial correlations well, offering high accuracy, but their computational complexity can be a drawback for large-scale applications.

K Nearest Neighbor

K Nearest Neighbor (KNN) is intuitive for spatial prediction. It relies on nearby data points to estimate RSSI values, which can work well in certain scenarios. However, it might struggle with non-uniformly distributed data or high-dimensional spaces.

Linear Regression

Linear regression is a simple yet effective method for spatial prediction tasks. It works when spatial relationships are approximately linear. However, it may underperform in cases of nonlinear and complex spatial patterns.

Support Vector Regression

Support Vector Regression (SVR) is suitable for spatial prediction when data exhibits non-linearity. SVR seeks to find the best-fitting hyperplane in high-dimensional spaces. It performs well but can be computationally demanding for large datasets.

Testing ML Models

The coordinates of the remaining 20% of the data were fed to the model, and the RSSI value output by the model was compared to the actual RSSI value at that coordinate calculated previously. The results were plotted on a scatter plot. The MAE, MSE, RMSE and R2 values of each model were calculated and compared.

Results and Conclusion

To identify the best RSSI value prediction model, we assess R2 Score, MAE, MSE, and RMSE. In Table 1, model metrics are provided. Gaussian Process Regression scores highest with R2 (0.975), indicating the best fit and lower MAE/RMSE. K Nearest Neighbors follows closely with an R2 of 0.961. Decision Trees are good but slightly less precise. Support Vector Regression is reasonable but has a lower R2. Linear Regression performs poorly. The choice depends on complexity, interpretability, and computational needs. Further evaluation and cross-validation on a larger dataset are recommended.

Bibliography

Ostlin, E., Zepernick, H., & Suzuki, H. (2010). Macrocell path-loss prediction using artificial neural networks. *IEEE Transactions on Vehicular Technology*, 2735–2747.

Ayadi, M., Ben Zineb, A., & Tabbane, S. (2017). A uhf path loss model using learning machine for heterogeneous networks. *EEE Transactions on Antennas and Propagation*, 3675–3683.

Popescu, I., Nikitopoulos, D., Constantinou, P., & Nafornita, I. (2006). Ann prediction models for outdoor environment. *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, 1-5.

Tables

Table 1

Performance Metrics Comparison of RSSI Prediction Models

Model	Best Max Depth (If Applicable)	MAE	MSE	RMSE	R2 Score
Decision Trees	10	0.549	1.729	1.315	0.949
Gaussian Process Regression	N/A	0.282	0.867	0.931	0.975
K Nearest Neighbors	4	0.344	1.333	1.155	0.961
Linear Regression	N/A	3.840	31.677	5.628	0.075
Support Vector Regression	N/A	0.792	9.770	3.126	0.715

Figures

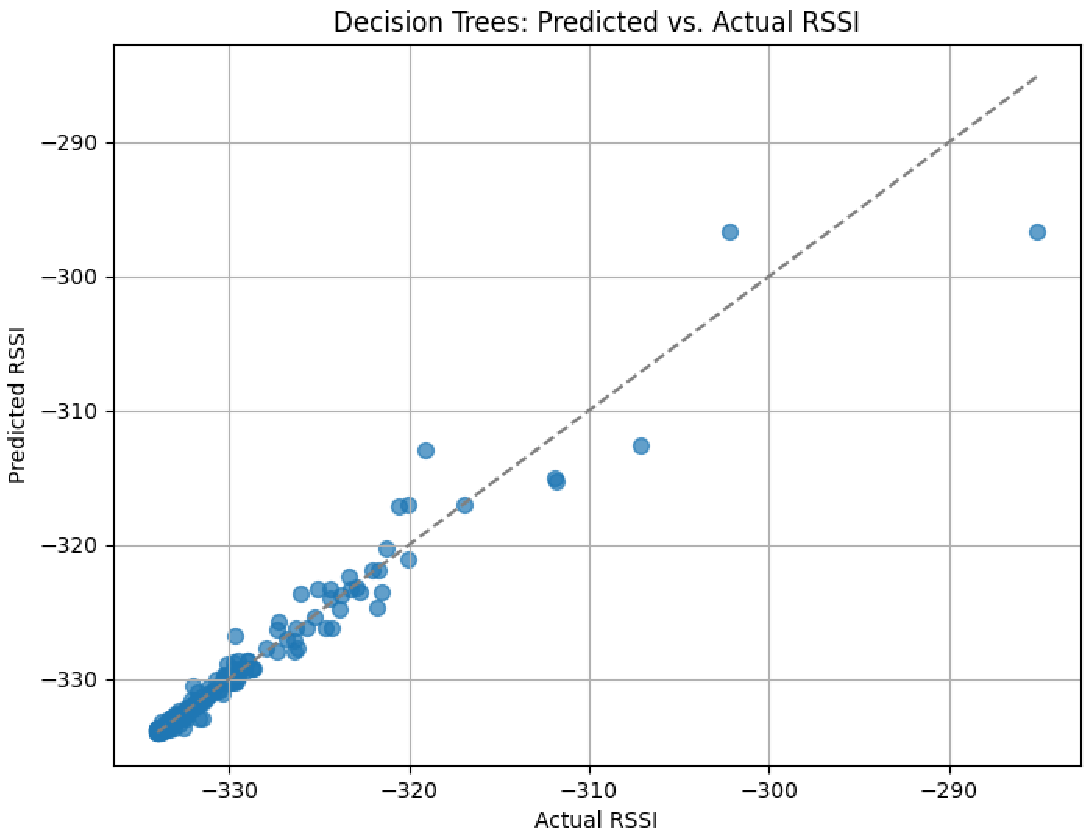

Figure 1: Decision Trees - Predicted vs. Actual RSSI

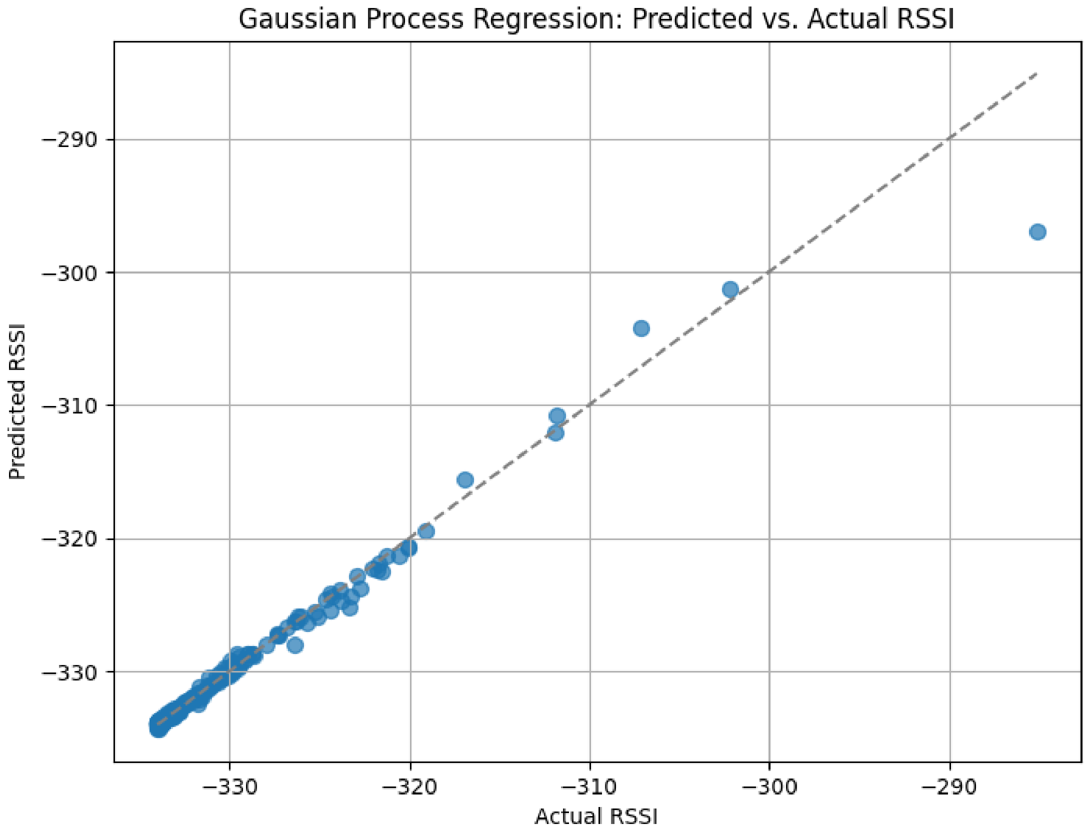

Figure 2: Gaussian Process Regression - Predicted vs. Actual RSSI

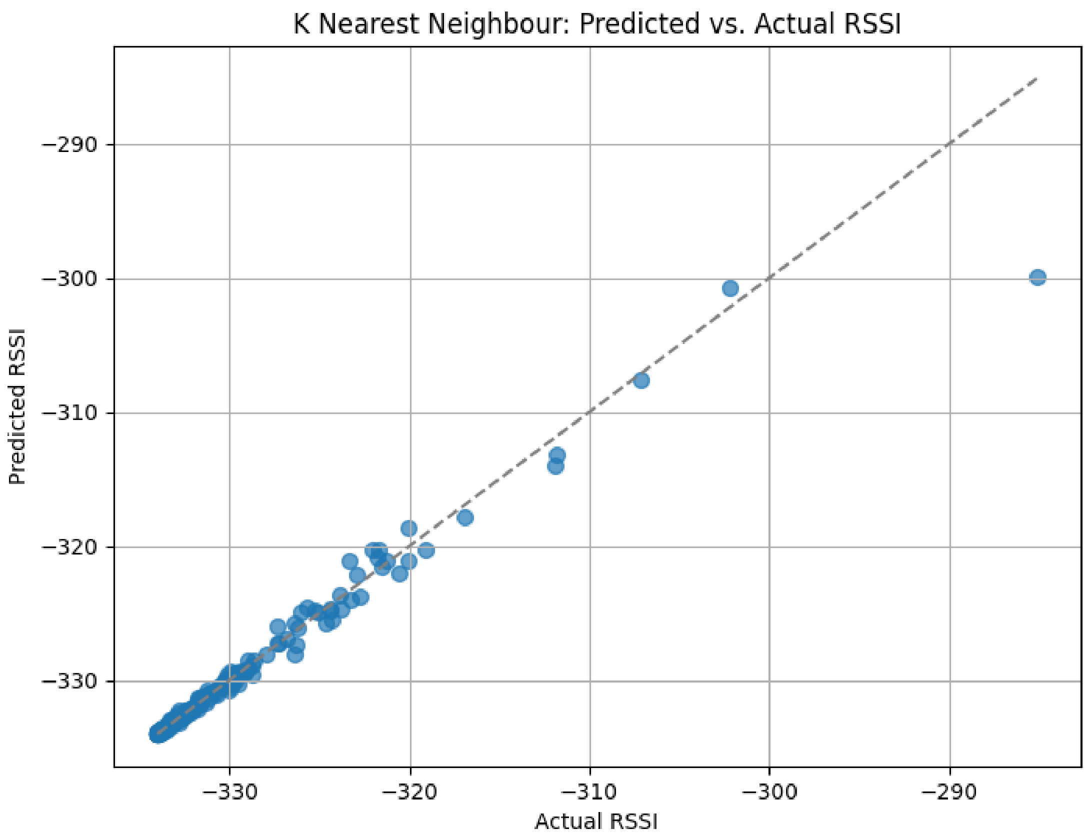

Figure 3: K Nearest Neighbors - Predicted vs. Actual RSSI

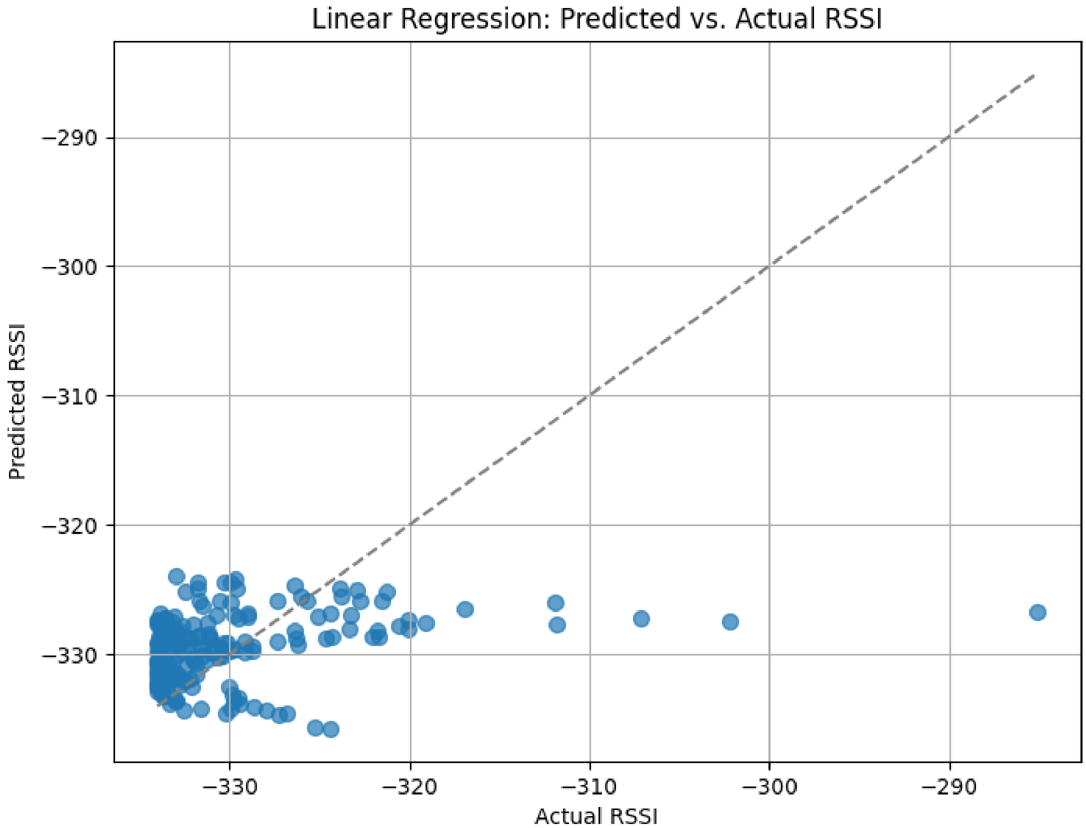

Figure 4: Linear Regression - Predicted vs. Actual RSSI

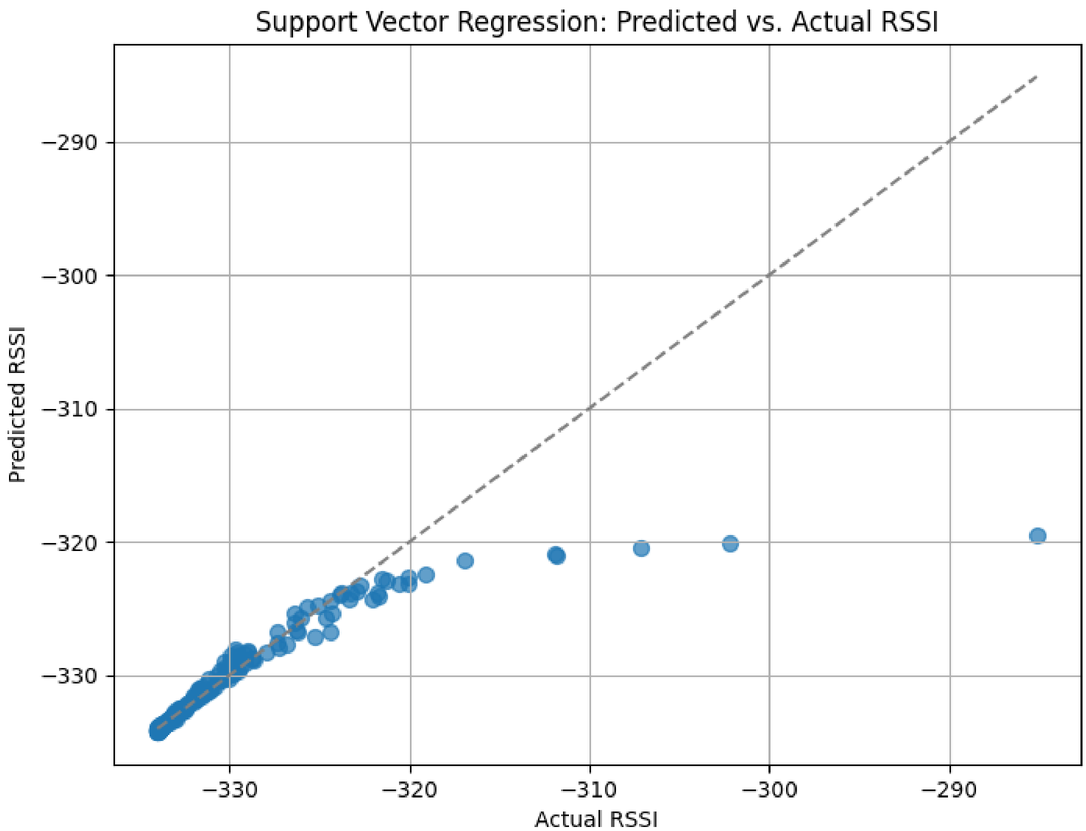

Figure 5: Support Vector Regression - Predicted vs. Actual RSSI