Performance Evaluation of RSSI Prediction Methods in Wireless Communication Networks

Mhd Ikhsan Rifki[1], Ali Ikhwan[2], Faisal Muhammad[3]

[1]Department of Computer Science, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[2]Department of Computer engineering, University Malaysia Perlis, Perlis, Malaysia
[3]Department of Mathematics, IPB University, Bogor, Indonesia

Article Info	ABSTRACT
	Continuous communication services that ensure user connectivity with the communication network are important. The communication network should be able to accommodate erratic user movements with high mobility. This research studies the performance of three different RSSI prediction methods: decision trees, random forests, and linear regression. Evaluation is carried out using statistical metrics, including mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and percentage accuracy. This analysis was carried out to understand how well each model predicted RSSI values based on distance. The research results show that Decision Tree Performance has an accuracy of 83.333%, Random Forest has a high accuracy of 97.2545%, and the Linear Regression Model provides quite good predictions with an accuracy of 91.6667% in predicting RSSI values.

Corresponding Author:

Mhd Ikhsan Rifki,
Department of Cmputer Science,
Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: rifki.mhdikhsan@uinsu.ac.id

1. INTRODUCTION

Wireless communication systems provide options for users to be able to carry out the communication process flexibly. The random movement of communication network users provides its own benefits for users without having to be constrained by the cable connections of the communication devices used. However, the quality of a wireless communication network is greatly influenced by the propagation environment. Random user movements have enormous potential to disrupt user connectivity to the communications network. Linearly, the distance between communication network users will be directly proportional to the quality of the communication signal received by the user. The closer the user is to the Access Point (AP) that serves the user, the better the quality of the received signal obtained by the user, and vice versa.

Apart from that, user movement also allows users to move away from communication service coverage from the AP, so this has the potential to interrupt communication services received by users. The user's position outside the coverage area means that the signal received is very small or even lost. The relative quality of signals received by communication devices is often referred to as the Received Signal Strength Indication (RSSI). RSSI in dbm units will weaken as the distance between the user and the AP increases

[1]. In addition to the distance between the user and the AP, the propagation environment has a significant impact on the RSSI power level. Line of Sight (LOS) propagation ensures a high RSSI power level when received by the user. This is because the communication service signal will propagate in a straight line without hitting objects or even humans in the propagation environment.

Meanwhile, a non-LoS propagation environment describes a propagation environment whose surroundings have many obstacles. Propagation: environmental conditions that are crowded with obstacles will affect the amount of RSSI received by the user. The communication signal emitted by the AP will experience several events, including reflection, difraction, and scattering. Reflection describes the event of the reflection of a signal wave, which forms a difference with dimensions greater than the wavelength. When an AP's signal wave encounters an obstruction, difraction occurs, which results in the wave's duplication. Meanwhile, scattering illustrates the event that the signal waves emitted by the AP will hit an obstacle object with a rough surface so that the signal will be scattered in various directions.

All of these conditions will cause RSSI to experience a decrease in level, which can have a direct impact on the user's communication experience. In addition, non-Los propagation environmental conditions will trigger multipath events. The term "multipath" refers to a variety of propagation paths that will form, allowing the signal that the AP emits to travel through paths that may be farther than the Los path before finally reaching the user. Multipath impacts will trigger fading. Fading is a gradual weakening of the signal wave emission due to the signal passing through a number of long-distance transmission paths. Apart from that, there is also the term path loss, which is one of the main components of the link budget and is related to the power level received by the user.

Based on this description, So RSSI has an important role in a wireless communications network. RSSI can be used as a reference for measuring the quality of communication network signals. RSSI can describe information related to the signal level of communication services received by the recipient. So this can be a basis for evaluating the quality of the communication network between the sender and receiver. RSSI information can also be used as a basis for network optimization to optimize the performance of wireless communication networks. This can be done by adjusting the transmission power level, adjusting the antenna direction, or even redesigning the placement of communication devices. The RSSI representation can also be used as an illustration of the distance between the sender and receiver. Variations in RSSI received values can provide an idea of the distance between the AP and the user or even interference that occurs in the propagation environment.

The urgency of measuring and predicting the RSSI received value is one of the important things that can be done, especially with regard to communication network performance. based on [2]'s research which carries out RSSI measurements using trilateration technology to determine the distance of the transmitter to the desired receiver point. The aim of the research is to provide an illustration of determining the location of a room. The KNN implementation is used to determine and adjust test points.

Meanwhile, research conducted by describes using an LCX antenna device, which has better prediction accuracy in multipath environments compared to conventional monopole antennas, which are more popularly used in many linear cell-based wireless communication networks. In addition, research by applied Delaunay Triangulation (DT) for Point of Intersection (POI) on all groups of partitioned nodes. RSSI and Predicted Received Signal Strength Indicator (PRSSI) are used to determine connectivity between node groups and POIs. The prediction concept is used to monitor the location of sensor nodes that have the potential to move when a disaster occurs. [5]. also conducted research on determining RSSI predictions in determining location. The research outlines the importance of predicting RSSI levels in location determination, wireless network topology planning, and scheduling. RSSI mapping is presented in a 2D map using IndoorRSSINet. The test results show that IndoorRSSINet is used 21 times more than the ray tracing approach with an MSE implementation of 0.00962 to normalize RSSI. The use of IndoorRSSINet is oriented towards optimizing transmitter positions in RSSI fingerprint-based wireless localization applications.

In the research carried out, predictions were made using the decision tree, random forest, and linear regression methods. The prediction results will be compared with data from measurements carried out at the Faculty of Science and Technology, UIN North Sumatra Medan. Meanwhile, commonly used evaluation methods are oriented towards regression analysis and predictive modeling. The evaluation approach is carried out using mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The propagation environment applied is a friis propagation model, the application of a propagation model based on the conditions of the propagation environment at the measurement location.

2. RESEARCH METHODE
2.1. Received Signal Strength Indication (RSSI)

A fluctuating transmission power number known as RSSI builds up over time due to user movement and the propagation environment's attenuation. The access point on a wireless network directs this transmission power in the user's direction. The quality of communication services that cellular providers offer to their customers will be impacted by the resulting weakening. The propagation environment affects the RSSI, which gradually loses value when the user moves around the room or even goes outside the room away from the covarage access point that supplies it (Goutam & Unnikrishnan, 2019). Aside from that, when the user gets farther away from the access point, the weakening will get worse. The user's receiver device will in this case scan the area for nearby access sites and determine their RSSI value. RSSI values from one or more nearby access points make up the results. The access point with the best RSSI value will be chosen based on the outcomes of the measurements that were made. As a result, the service has been switched to an access point with a higher RSSI value from one with a lower value. Maintaining connectivity and the standard of user-provided services is the goal of this. Equations (1) and (2) represent the received power formulation approach (RSSI) [6]-[8]

$$P_r = P_0 \cdot \left(\frac{d}{d_0}\right)^n \tag{1}$$

$$P_r(dB) = P_0 - 10n \log\frac{d}{d_0} \tag{2}$$

2.2. Friis Propagation Model

The propagation environment model illustrates room conditions or models the mobility trajectory traversed by the user. In the modeling process, there are several parameter settings that are adjusted to the conditions of the propagation path. The aim of implementing the model is to use an empirical approach to describe path conditions that are influenced by the presence of obstacles, path loss, and distance factors that contribute to weakening the RSSI receive level. A free-space loss environment dominates the propagation environment model in the Faculty of Science and Technology at UIN North Sumatra Medan. This condition describes a straight propagation environment, where the wave will propagate in a straight line without touching obstacles around it or along its propagation path. So, in this case, determining the propagation environment model that is considered suitable to be a comparison value for the RSSI received by the user is the first propagation model. The results of the Francis propagation model will be compared with the RSSI values produced at the measurement stage. The frequency propagation model in wireless networks implemented in this research mathematically uses equations (3) and (4)

$$P_r(d) = P_t G_t G_r \left(\frac{c}{4\pi df}\right)^2 \tag{3}$$

$$P_r(d) = P_t G_t G_r \left(\frac{\lambda}{4\pi d}\right)^2 \tag{4}$$

Description:
P_r : Power transmitter
d : Distance
P_t : Power receiver
G_t : Gain antenna transmitter
G_r : Gain antenna receiver

Regarding equality (3) and equality (4) It has a relationship with mathematics that is expressed by equations (5)

$$\lambda = \frac{c}{f} \tag{5}$$

Description:
λ : Wavelength
c : Speed of light (m/s)
f : Frequency (Hz)

2.3. Prediction Method

2.3.1. Decision Tree

A machine learning model for regression and classification is a decision tree. This model depicts a tree structure in which a decision is represented by each branch, a feature (attribute) is represented by each internal node, and the outcome (label or value) of that decision is represented by each leaf. The best features that best define the target class or target variable are chosen during the decision tree building process. These characteristics are chosen using metrics known as impurity measures, including entropy or Gini impurity. At each node, the separating attribute is chosen based on whatever characteristic best separates various classes or outcomes.

The tree is split into sub-trees according to the value of a splitting property once it has been chosen. Every sub-tree in the system goes through this recursive procedure until the termination condition is satisfied. For instance, reaching a given maximum depth level or ensuring that every data instance in a branch has the same label could be termination conditions. Equation 6 can be used to illustrate the decision tree technique approach.

2.3.2. Random Forest

Random Forest is an ensemble learning technique that aggregates the predictions of multiple machine learning models, typically decision trees, to enhance prediction accuracy and reduce the risk of overfitting. Random Forest is a potent and adaptable algorithm commonly utilized in a variety of machine learning applications due to its strong predictive capabilities, resistance to overfitting, and straightforward implementation. The general steps for utilizing Random Forest for prediction purposes are as follows:

1. Data Preprocessing
   a. involves gathering pertinent data for making predictions.
   b. Remove data with missing or invalid values.
   c. Conduct data transformations, such as normalization or encoding, as needed.
2. Data Sharing:
   a. Divide your data into two sections - training data and testing data.
   b. Training data is utilized for model training, whilst testing data is employed to evaluate the model's performance.
3. Constructing a Random Forest Model
   a. Construct a random forest ensemble.
   b. Every tree will handle a portion of the training data through random sampling with replacement.
   c. The final forecast is generated by aggregating the prediction results from all the trees.
4. Model Evaluation:
   a. Assess the model's performance using test data.
   b. Common evaluation criteria are accuracy, precision, recall, and F1-score.
5. Tuning Hyperparameters
   involves optimizing model performance by modifying parameters like the number of trees, tree depth, and sample size.
6. Forecasting
   a. Utilize the newly acquired data to create predictions after training the model.
   b. Forecasting outcomes can inform decision-making or provide insights into patterns.

2.3.3. Linier Regression

Linear regression is a fundamental and widely utilized statistical and machine learning method for representing the connection between independent variables (predictors) and a dependent variable (target) through a straight line. Linear regression aims to determine the optimal straight line that best represents the relationship between input and output variables. Linear regression models are utilized for prediction, estimation, and analyzing relationships between variables. In a simple linear regression with one independent variable, the model predicts a target value based on the value of a single predictor variable. Linear regression is a valuable technique in data analysis and machine learning due to its simplicity, interpretability, and effectiveness when the assumptions of the model are satisfied.

2.4. Evaluation Metrics of Regression Analysis and Predictive Modeling

2.4.1. Mean Squared Error (MSE)

Mean Squared Error (MSE) is a widely used evaluation metric in statistics and machine learning for assessing the effectiveness of regression models. The metric calculates the mean of the squared deviations between the predicted values of the model and the actual values in the dataset. MSE provides a measure of the proximity between the regression line or prediction model and the actual data points. A lower Mean Squared Error (MSE) number indicates a better fit of the regression model to the data. Mean squared error (MSE) amplifies the impact of significant deviations between expected and observed values. Larger errors will have a greater impact on the total Mean Squared Error (MSE) score. The Mean Squared Error (MSE)

method can be implemented using equation (1).Equation 6 can be used to illustrate the decision tree technique approach [9].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_p - y_t)^2 \tag{6}$$

Description:

n   = Number of samples

y_p = Predictive value

y_t =Actual value.

2.4.2. Root Mean Squared Error (RMSE)

RMSE is a widely used evaluation metric in statistics and machine learning to assess the effectiveness of regression models. RMSE quantifies the square root of the average squared difference between the predicted values of the model and the actual values in the dataset. This provides an indication of the magnitude of the model prediction error in the original data units. A lower RMSE value indicates a better fit of the regression model to the data. RMSE eliminates the impact of the square in the MSE, resulting in a more understandable figure in the data's original units. Similar to Mean Squared Error (MSE), Root Mean Squared Error (RMSE) likewise assigns more significance to significant discrepancies between projected and actual values. RMSE readings are more challenging to interpret and more susceptible to outliers due to being the square root of MSE. The RMSE formula is displayed in equation (7).

$$RMSE = \sqrt{MSE} \tag{7}$$

2.4.3. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a regularly used evaluation metric in statistics and machine learning to assess the performance of regression models. Mean Absolute Error (MAE) calculates the average of the absolute discrepancies between the predicted values of the model and the actual values in the dataset. This provides an indication of the model's prediction error in the original data units, irrespective of the discrepancy between the forecasts and the actual values. Mean Absolute Error (MAE) treats all deviations between expected and actual values equally, without considering their direction. MAE is more resilient to outliers compared to MSE or RMSE due to its lack of reliance on squared differences. Mean Absolute Error (MAE) offers a straightforward interpretation in the data's original units, facilitating comprehension. Equation (8) displays the mathematical representation of mean absolute error (MAE).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

Description

n = number of samples in the dataset.

y_i = the actual value of the ith data point.

$\hat{y}_i$ = value predicted by the model for the eighth data point.

3. RESULT AND ANALYSIS

The process is carried out by carrying out design stages, which include path initialization, variable requirements to be used, propagation environment, and initialization of measurement parameters. The stages are carried out by measuring the trajectories of communication network users in the propagation environment. Measurements were made on the length of the user's movement path along 24 meters. Sampling is carried out in stages every one meter. Measurements were carried out repeatedly, up to 27 measurement processes. The measurement results will be averaged to obtain the average RSSI value received by the user while moving on the propagation path. This is done because the propagation environment is sensitive and because of the dynamic nature of the environment, because objects in the propagation environment also change.

The next process is to classify the training data and test data and use them as a basis for the prediction stage using three prediction methods: decision trees, random forests, and linear regression. The stage continues with testing the model that has been established using MSE, MAE, and RMSE. The predicted results will be displayed and compared with the actual data measurement results, as well as with the model testing results as the final output. The stages of the research process can be seen in Figure 1.
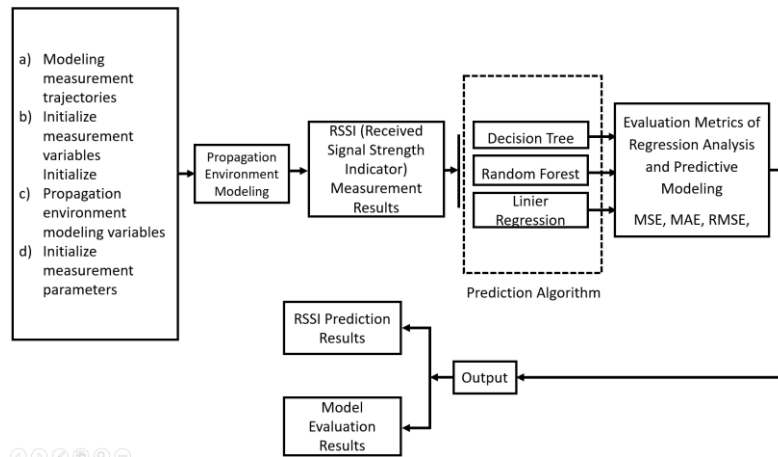
Figure 1. Flowchart for Performance Evaluation of the RSSI Prediction Method

The graph in Figure 2. that shows the original RSSI value along with the predicted values from the three models (Decision Tree, Random Forest, and Linear Regression) shows how well each model predicts RSSI values based on distance. First of all, we can see that the colored dots on the graph represent the original RSSI values observed at a certain distance. For reference, as you move to the right on the x-axis (distance), the RSSI value generally tends to decrease, which corresponds to the general characteristic of signal degradation in wireless communications as the distance increases.
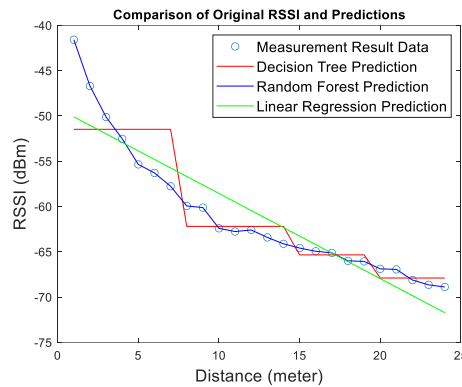

Figure 2. Comparison of prediction results from the three RSSI prediction methods and measurement data

Then, the red (decision tree), blue (random forest), and green (linear regression) lines on the graph represent the prediction results of each model. The difference between these predicted lines and the original RSSI value points gives an idea of how well the models can model the relationship between distance and RSSI.

Further analysis can be done by comparing how well the prediction lines from each model match the general pattern of the original RSSI values. For example, if the predicted line is almost parallel to the original RSSI value points and follows the same decreasing trend, it indicates that the model is quite good at predicting the RSSI value. On the other hand, if the predicted line has a significant deviation from the original RSSI value points, it indicates that the model may not fit the data well.

In addition, the comparison between the prediction lines of the three models is also important. If there is a significant difference between the models' predictions, it might be because each model's prediction strategy differs. The degree to which a model's predicted lines match the original RSSI values can indicate whether it is better at capturing particular patterns in the data in some cases.

3.1. Comparison of RSSI Measurement Results with Decision Tree Prediction Results

3.2. In the comparison graph between the original data RSSI values and the predicted results from the Decision Tree model, we can see how the Decision Tree model performs in predicting RSSI values based on distance. The performance results of the decision tree in predicting can
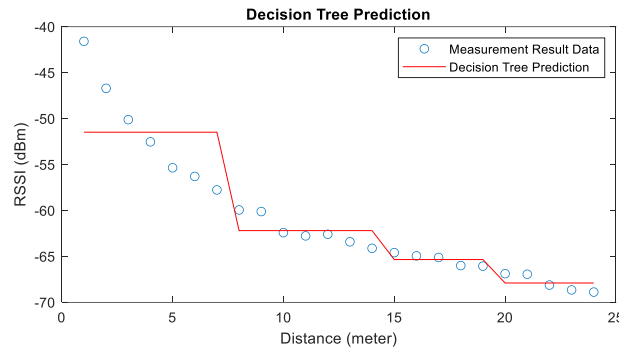
be seen in Figure 3. The comparison results of the RSSI measured against the decision tree.

Figure 3. Comparison of RSSI Measurement Results with Decision Tree Prediction Results

First of all, the colored dots on the graph represent the original RSSI values observed at a certain distance. This is the actual data that we use as a basis for comparison.

Next, the red line on the graph shows the prediction results from the decision tree model. In an ideal world, if the Decision Tree model predictions were perfect, the red line would be above or almost tangent to the original data points.

However, in practice, the decision tree prediction line will show deviations from the original data points because it is impossible to create a perfect model. So, we will see that the red line will have a similar pattern to the original data but may have some discrepancies or deviations at some points.

This comparative analysis is important to evaluate the prediction quality of the Decision Tree model. The closer the prediction line is to the original data points, the better the decision tree model is at predicting RSSI values. Conversely, if there is a significant deviation between the predicted line and the original data points, it indicates that the model may have limitations or needs improvement.

3.3. Comparison of RSSI Measurement Results with Random Forest Prediction Results

In the comparison graph between the RSSI values of the original data and the predicted results from the Random Forest model, it can be observed how the Random Forest model performs in predicting RSSI values based on distance. First of all, the colored dots on the graph represent the original RSSI values observed at a certain distance. This is the actual data used as a reference point to compare the model's prediction performance. The comparison results between RSSI measurement results and Random Forest can be seen in Figure 3.
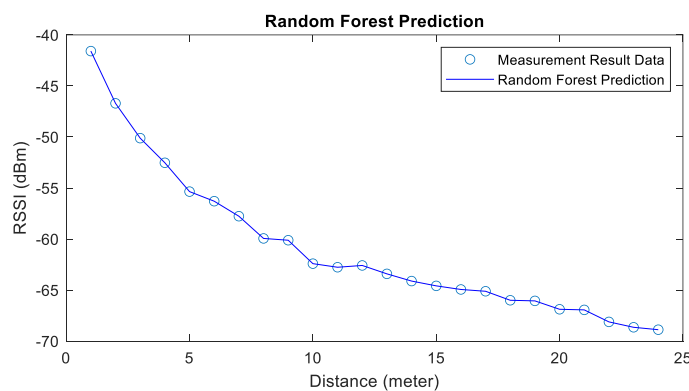

Figure 4. Comparison of RSSI Measurement Results with Random Forest Prediction Results

Next, the blue line in the graph shows the prediction results from the Random Forest model. Under ideal conditions, if the Random Forest model predictions are perfect, the blue lines will cover or almost intersect the original data points. However, in practice, random forest prediction lines will show some deviation from the original data points because it is impossible to create a perfect model. Therefore, we can

observe that the blue line will follow the general pattern of the original data, but it will probably have some irregularities or deviations at some points.

This comparative analysis is important to evaluate the prediction quality of the Random Forest model. The closer the prediction line is to the original data points, the better the Random Forest model is at predicting RSSI values. On the other hand, if there is a significant deviation between the predicted line and the original data points, it indicates that the model may have limitations or needs further improvement. Thus, visual analysis can understand the prediction quality of the Random Forest model and can serve as a basis for further adjustments and improvements to the model.

3.4. Comparison of RSSI Measurement Results with Linier Regression Prediction Results

The linear regression prediction line (green line) on the graph matches the general pattern of the original data (colored dots) fairly well. This shows that the linear regression model is able to capture the general trend of RSSI data based on distance. This can be seen through Figure 6.
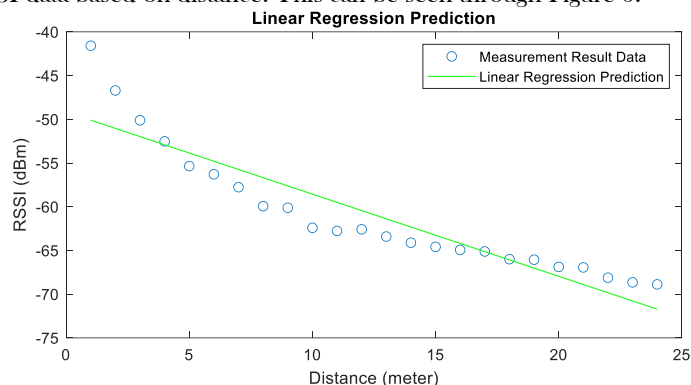

Figure 6. Comparison of RSSI Measurement Results with Linier Regression Prediction Results

Visually, the linear regression prediction line is generally close to the original data points. Although there are some deviations at some points, the linear regression model provides a fairly good estimate of the RSSI value. The linear regression prediction line is relatively consistent in following the original data pattern. This shows that the linear regression model provides stable and consistent prediction results under various conditions. Deviations between the predicted line and the original data points at some points may indicate limitations in the model, especially if the relationship between the input and output variables is not completely linear. In this case, model adjustments or exploration of alternative models may be necessary to improve prediction accuracy.

3.5. Model Evaluation Results with Mean Squared Error (MSE)

In the Mean Squared Error (MSE) analysis of each model, Decision Tree (MSE = 9.2341), A high MSE of a decision tree model indicates that there is a significant error between the model predictions and the actual values of the data. This indicates that the decision tree may have difficulty capturing the correct pattern or relationship between distance and RSSI values. This model tends to provide predictions that are further from the actual value. Random Forest (MSE = 1.1736), The lower MSE of the Random Forest model indicates that this model has a lower error in predicting RSSI values compared to the Decision Tree. Even though there are still errors, the random forest model is able to provide more accurate predictions compared to decision trees. Linear Regression (MSE = 7.5654), Linear regression has a lower MSE compared to decision trees but is higher compared to random forests. This shows that the linear regression model provides better predictions compared to the decision tree, but not as good as the random forest. However, there are still errors in the predictions that can be corrected.

In this context, MSE provides an overview of the error level of each model in predicting RSSI values. The lower the MSE, the better the model is at predicting RSSI values. Of the three models, Random Forest has the lowest MSE, indicating that this model provides the most accurate predictions compared to Decision Tree and Linear Regression. The results of model evaluation using MSE can be seen in Figure 7
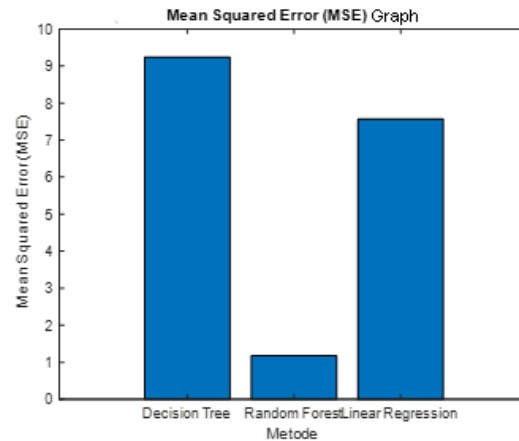

Figure 7. Comparison of Model Evaluation Results using Mean Squared Error (MSE)

3.6. Model Evaluation Results with Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) analysis gives an idea of how big the average error of the model predictions is in the same units as the target variable (in this case, RSSI in dBm). The computer calculations show that when the Decision Tree model's RMSE is high (3,3388), it means that there is a big difference between what the model said would happen and what actually happened. This means that the average prediction error of the Decision Tree model is around 3.0388 dBm. This shows that the decision tree model has a relatively large error in predicting the RSSI value. While Random Forest (RMSE = 0,34258) illustrates a much lower RMSE than the Random Forest model, indicating that this model provides predictions that are much closer to the actual values of the data, The average prediction error of the Random Forest model is only around 0.34258 dBm, which indicates excellent prediction quality. In other computations, linear regression (RMSE = 2.7505) represents linear regression as having a higher RMSE compared to random forests but lower than decision trees. This means that the average prediction error of the linear regression model is around 2.7505 dBm. Although better than a decision tree, there is still a significant error in predictions. The results of the RMSE computation can be seen in Figure 8.
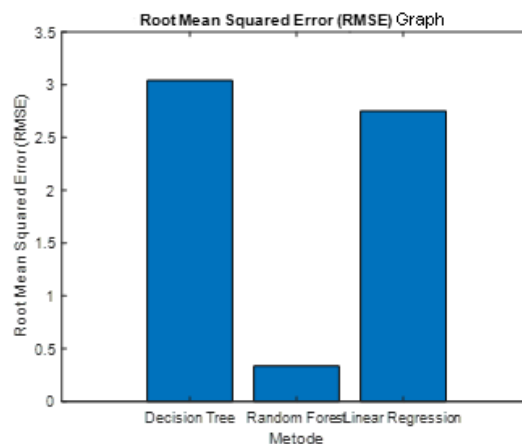

Figure 8. Comparison of Model Evaluation Results using Root Mean Squared Error (RMSE)

In this context, RMSE provides an indication of how well the model is at predicting RSSI values. The lower the RMSE, the better the model is at predicting RSSI values. Of the three models, Random Forest stands out with the lowest RMSE, indicating that it provides the most accurate predictions, followed by Linear Regression and then Decision Tree.

3.7. Model Evaluation Results with Mean Absolute Error (MAE)

Mean Absolute Error (MAE) analysis provides an idea of how close the average model prediction is to the actual value of the observed data. In the context of the MAE value according to the computational results, the following results are obtained:

1. Decision Tree (MAE = 1,9745):

   The relatively high MAE of the Decision Tree model indicates that there is significant variation between the model predictions and the actual values from the data. The average absolute prediction error of the Decision Tree model is approximately 1.9745 dBm. This shows that the decision tree model tends to provide predictions that are quite far from the actual RSSI value.

2. Random Forest (MAE = 0,3088):

   The much lower MAE of the Random Forest model indicates that this model provides predictions that are much closer to the actual values of the data. The average absolute prediction error of the Random Forest model is only about 0.3088 dBm, which indicates excellent prediction quality.

3. Linear Regression (MAE = 2,1595):

   Linear regression has a higher MAE compared to random forests but is lower than decision trees. This means that the average absolute error of predictions from the linear regression model is around 2.1595 dBm. Even though it is better than Decision Tree, there is still quite a large error in predictions.

In this context, MAE provides information about how well the model is at predicting RSSI values. The lower the MAE, the better the model is at predicting RSSI values. Of the three models, random forest stands out with the lowest MAE, indicating that it provides the most accurate predictions, followed by linear regression and then decision trees. The complete results of the MAE computation for the three prediction methods can be seen in Graph Figure 9.
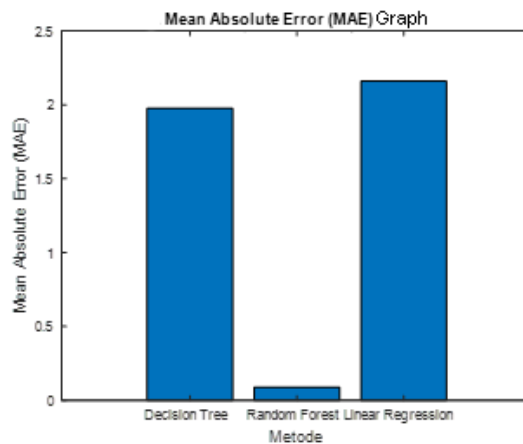

Figure 9. Comparison of Model Evaluation Results using Mean Absolute Error (MAE)

3.8. Results of comparison of percentage accuracy in the RSSI prediction method

The analysis of the results of the percentage accuracy in implementing the RSSI prediction method is as follows:

1. Decision Tree (Accuracy: 83.333%):

   An accuracy of 83.333% indicates that the decision tree model succeeded in predicting around 83.333% of the data correctly. However, around 16,667% of the data has incorrect predictions or does not match the actual values. Although accurate, there is still room for improvement in correcting predictions for this erroneous data.

2. Random Forest (accuracy: 97.2545%):

   The high accuracy of 97.2545% shows that the Random Forest model is very good at predicting RSSI values. This model correctly predicted 97.2545% of the data, which is excellent prediction quality. However, keep in mind that no model is perfect, and it is possible that there is still a small portion of data that is not predicted correctly.

3. Linear Regression (Accuracy: 91.6667%):
    The accuracy of 91.6667% shows that the linear regression model also provides quite good predictions for predicting RSSI values. This model was able to predict about 91.6667% of the data accurately. Even though it is quite high, there is still around 8.3333% of data that has incorrect predictions or does not match the actual values.
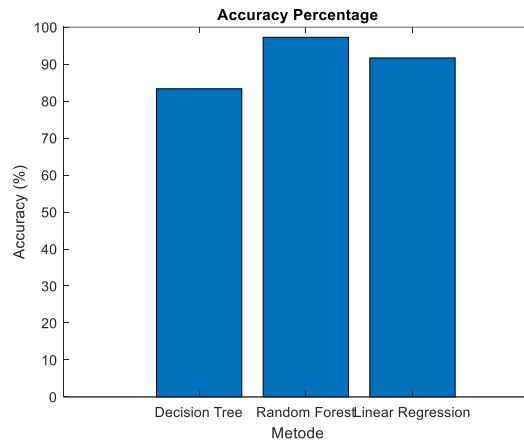

Figure 10. Results of Comparison Percentage Accuracy in The RSSI Prediction Method

4. CONCLUSION

Based on the computational results that have been obtained, there are several conclusions, including:

1. The Decision Tree model has a high MSE, indicating significant errors in predictions. The high RMSE and MAE also confirm these findings, indicating that these models tend to provide predictions that are far from their true values. Meanwhile, the Random Forest Model shows excellent performance with low MSE, RMSE, and MAE. This indicates that the model is able to provide predictions that are very close to the actual values, indicating high prediction quality. Apart from that, linear regression gives quite good results with an MSE that is lower than a decision tree but higher than a random forest. Moderate RMSE and MAE show that this model provides better predictions than Decision Tree but not as good as Random Forest.

2. The decision tree accuracy percentage results have an accuracy percentage of 83,333%; the decision tree model succeeded in predicting around 83,333% of the data correctly. Meanwhile, Random Forest has a high accuracy of 97.2545%, indicating that the Random Forest model is very good at predicting RSSI values. Apart from that, linear regression: The linear regression model provides quite good predictions with an accuracy of 91.6667%.

REFERENCES

[1]   S. Goutam and S. Unnikrishnan, "Handoff decision algorithm in WiFi zone using Fuzzy Logic," *Proc. - 2019 Women Inst. Technol. Conf. Electr. Comput. Eng. WITCON ECE 2019*, pp. 123–128, 2019, doi: 10.1109/WITCONECE48374.2019.9092939.

[2]   X. Zheng, R. Cheng, and Y. Wang, "RSSI-KNN: A RSSI Indoor Localization Approach with KNN," in *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 2023, pp. 600–604.

[3]   P. Hou, J. Zhu, K. Nagayama, Y. Hou, and S. Denno, "Prediction Evaluation for RSSI Data Generated from Leaky Coaxial Cables over Indoor Environment," in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, 2022, pp. 1–5.

[4]   S. L. R. Bashyam and J. Chandra, "An Efficient Node to Node Coverage and Connectivity with RSSI Using Grey-Wolf Prediction Optimization Algorithm in Remote Low Accessibility Area," in *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, 2023, pp. 1–6.

[5]   N. Raj and B. S. Vineeth, "IndoorRSSINet-Deep learning based 2D RSSI map prediction for indoor environments with application to wireless localization," in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2023, pp. 609–616.

[6]   N. Montavont *et al.*, "Handover Triggering in IEEE 802 . 11 Networks To cite this version : HAL Id : hal-01759108 Handover Triggering in IEEE 802 . 11 Networks," 2018.

[7]   T. Manodham, M. Hayasaka, and T. Miki, "A novel handover scheme for improving the performance of WLANs based on IEEE802.11," *2006 Asia-Pacific Conf. Commun. APCC*, vol. 00, 2006, doi: 10.1109/APCC.2006.255784.

[8]   A. H. Ali, M. R. A. Razak, M. Hidayab, S. A. Azman, M. Z. M. Jasmin, and M. A. Zainol, "Investigation of indoor WIFI radio signal propagation," *ISIEA 2010 - 2010 IEEE Symp. Ind. Electron. Appl.*, no. Isiea, pp. 117–119, 2010, doi: 10.1109/ISIEA.2010.5679486.

[9]   Y. Liu, X. Luo, and M. Yang, "Research on passenger flow prediction of bus line based on gradient boosting decision tree," in *2020 Chinese Control And Decision Conference (CCDC)*, 2020, pp. 3212–3218.