A logo with blue text

AI-generated content may be incorrect.Module Code: COMP3045

Title of the project: Path Loss at 5G Frequency Range in South Asia

Semester: Spring

Name: Tan Quan Ming

ID: 20475320

Module Convenor: Dr Kweh Yeah Lun

Contents

[1 Introduction 3](#_Toc195893341)

[1.1 Goals & Objectives 3](#_Toc195893342)

[2 Literature review 5](#_Toc195893343)

[2.1 Atmospheric Absorption 5](#_Toc195893344)

[2.2 Vegetation & Foliage, Building Materials and Human Blockage 6](#_Toc195893345)

[3 Methodology 7](#_Toc195893346)

[3.1 Data Acquisition and Preprocessing 7](#_Toc195893347)

[3.2 Interactive Dashboard (Built using Streamlit) 7](#_Toc195893348)

[3.3 Machine Learning Pipeline 8](#_Toc195893349)

[4 Implementation 10](#_Toc195893350)

[4.1 Core Libraries Used 10](#_Toc195893351)

[4.2 Scatter Plot 10](#_Toc195893352)

[4.3 Box Plot 11](#_Toc195893353)

[4.4 Correlation Heatmap 11](#_Toc195893354)

[4.5 Machine Learning Model Evaluation Charts 11](#_Toc195893355)

[4.6 Model Training and Evaluation 12](#_Toc195893356)

[5 Results & discussion 15](#_Toc195893357)

[5.1 Scatter Plot with Regression Lines 15](#_Toc195893358)

[5.2 Seasonal Box Plot 17](#_Toc195893359)

[5.3 Correlation Heatmap 18](#_Toc195893360)

[5.4 Model Performance Evaluation 18](#_Toc195893361)

[5.5 Discussion 20](#_Toc195893362)

[6 Conclusion & Future works 21](#_Toc195893363)

[6.1 Future Works 21](#_Toc195893364)

1. Introduction

Millimeter-wave (mm-wave) communication is an important technology for modern wireless networks, providing rapid data speeds and increased connection [1]. It is an essential part of 5G technology, which operates at very high frequencies of 30 GHz to 300 GHz [9]. However, the propagation properties of mm-wave signals are greatly impacted by atmospheric factors, making it critical to investigate their impact for optimal network design and implementation. Understanding these variances is critical for creating realistic channel models, enhancing communication protocols, and maximizing overall system effectiveness.

We will utilize a dataset created using the NYUSIM 3.0 mm-Wave channel simulator software to meet this purpose [2]. Realistic atmospheric factors including temperature, humidity, barometric pressure, and rain rate are included in this collection. Because the input data was gathered in South Asia over the course of a year, seasonal fluctuations in mm-wave channel characteristics were accurately represented. There are 2,835 records in the collection, and each one has the following parameters:

* T-R Separation Distance (m) – The distance between transmitter and receiver
* Time Delay (ns) – Time taken for signal to travel from transmitter to receiver
* Received Power (dBm) – Power level of the received signal
* Phase (rad) – Phase of the received signal, indicates position of the point in the wave at time of measure.
* Azimuth Angle of Departure (AoD) (degree) – Horizontal angle at which the signal leaves the transmitter
* Elevation Angle of Departure (AoD) (degree) – Vertical angle at which the signal leaves the transmitter
* Azimuth Angle of Arrival (AoA) (degree) – Horizontal angle at which the signal arrives at the transmitter
* Elevation Angle of Arrival (AoA) (degree) – Vertical angle at which the signal arrives at the transmitter
* Root Mean Square (RMS) Delay Spread (ns) – Measure of signal dispersion
* Season – time of the year (spring, winter, fall, summer)
* Frequency (GHz) – operating frequency of the transmitted signal
* Path Loss (dB) – total signal attenuation as it travels from the transmitter to the receiver

The dataset has been cleaned and normalized, ensuring consistency and readiness for analysis. Researchers in networking can use this information to investigate the impact of weather on channel activity. It can also help evaluate communication protocols and signal processing techniques under a variety of atmospheric conditions. Although the dataset is centred around readings from South Asia, these findings could be applied to similar climatic zones. This dataset is an important resource for developing the science of mm-wave communication since it provides insights on real-world propagation characteristics in a variety of environments.

* 1. Goals & Objectives

The major purpose of this work is to examine and estimate route loss in mm-wave communication channels using a variety of environmental factors. Understanding the link between path loss and other elements such as T-R separation distance, atmospheric conditions, and multipath effects is critical for maximizing wireless network reliability and signal dependability in different environments. This study aims to:

1. Investigate the effect of environmental influences on path loss in mm-wave channels.
2. Investigate the link between path loss and spatial factors - how T-R separation distance, azimuth and elevation angles, and delay spread relate to signal attenuation.
3. To get insights about improving mm-wave communication networks, enhancing network architecture, especially in areas with high seasonal fluctuations.
4. Literature review

In applications like 5G and beyond, automotive networks, high-definition video transmission, and next-generation IoT systems, mm-waves provide ultra-fast communication due to their substantially larger bandwidth and data rates as compared to standard sub-6 GHz wireless signals [7]. However, due to the high frequencies of mm-wave propagation, these waves experience higher path loss, as they are more prone to atmospheric variables. For example, the presence of water vapor and oxygen can cause attenuation of these waves [8]. Consequently, mm-wave propagation is extremely directional, necessitating well-designed pathways to ensure reliable communication between the transmitter and receiver.

In this literature review, we will look at existing studies on how environmental factors affect route loss in mm-Wave communications systems.

Several research have been conducted to explore mm-wave communication and how it responds to changing atmospheric conditions. Numerous studies have demonstrated how atmospheric conditions like as rain, humidity, and temperature may influence mm-wave propagation. For example, Rappaport et al. [3] discovered that high-frequency transmissions (especially mm-wave) are attenuated owing to rain and human blocking, stressing the significance of constructing trustworthy communication channel models. Marcus and Pattan [4] have investigated the influence of ambient gases and humidity on mm-wave propagation loss by modifying molecular absorption and scattering effects via diffraction and reflection.

The NYUSIM simulator, which was used to create the dataset, has been thoroughly verified in academic research as an excellent tool for simulating real-world mm-wave communication. Sun et al.’s paper in 2017 [13] unveiled NYUSIM, which can predict channel properties under a variety of environmental circumstances. In the same year, Rappaport et al. found that the NYUSIM provides more realistic simulations of urban mm-wave channels than the 3GPP TR 38.900 Release 14 model.

* 1. Atmospheric Absorption

A study by Rappaport et. al. in 2019 [10] discovered that at mm-wave frequencies (30 GHz – 300 GHz), water (H₂O) and oxygen (O₂) molecules generate large amounts of attenuation due to higher absorption, especially at 60 GHz where oxygen absorption peaks and 183 GHz. In the same study, the atmospheric attenuation at 28 GHz is around 0.1dB/km, but increases to around 0.5dB/km at 73 GHz.

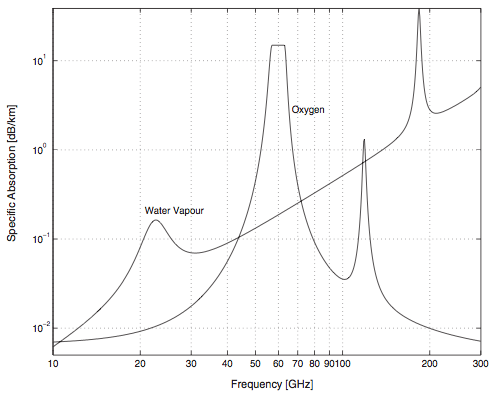


Figure 1. Specific attenuation of water vapor and oxygen. [ITU-P676-9, 11]

* 1. Vegetation & Foliage, Building Materials and Human Blockage

Besides atmospheric absorption, the presence of physical obstructions, such as vegetation and foliage, buildings and human bodies would also introduce a lot of complexity in mm-wave propagation as they can also cause attenuation:

* **Vegetation & Foliage**: Highly dense foliage can cause signal dispersion and absorption. Zhang et al. (2019) [12] measured foliage loss at 28 GHz and reported that there was signal degradation compared to empty space. The attenuation effects vary according to density of leaves, amount of moisture and the existence of branches, which makes mm-wave propagation very dependent on vegetation and seasonal vegetation changes.
* **Building Materials & Urban Structures**: Building penetration loss, which occurs when mm-wave signals are attenuated as they pass through walls, windows, and other buildings, poses a serious challenge for mm-wave propagation in urban environments. It is shown that different building materials, such as glass, concrete, metal can contribute to variable amounts of signal attenuation [6], decreasing the stability of mm-wave communication in densely populated urban cities.
* **Human Blockage:** Human bodies can result in temporary signal shadowing and reflection as they behave like movable obstacles. For example, Chen et al. (2016) conducted measurements on a 28 GHz signal indoors and suggested the basic human body shadowing model that emphasized the influence of human blockage on signal attenuation.

These variables must be considered while developing reliable mm-wave networks, as they can present new signal-blocking barriers that can affect path loss and overall connectivity.

1. Methodology

The goal of this project is to investigate, visualize, and forecast Path Loss in 5G communication settings across South Asia by combining interactive data exploration approaches with machine learning-based regression modeling.

* 1. Data Acquisition and Preprocessing

The initial step was to obtain the dataset 5g-South Asia.csv, which provides signal transmission characteristics and its associated season. The raw dataset has the following fundamental features:

* **Transmission-reception distance** (T-R Separation Distance (m))
* **Received signal power** (Received Power (dBm))
* **Directional characteristics** (Azimuth and elevation angles at both the Angle of Departure and Angle of Arrival, demonstrated in Figure 2)

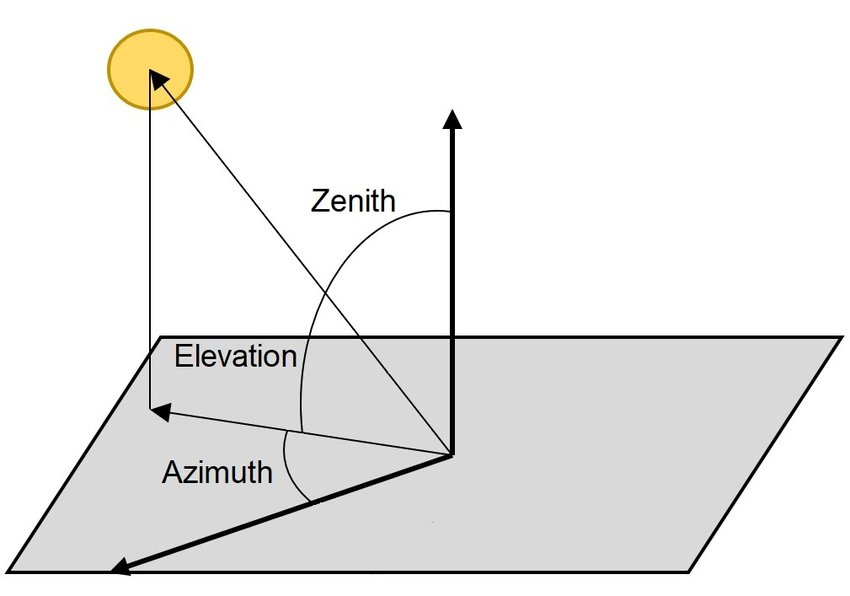


Figure 2. Local coordinates of the sun: elevation, zenith and azimuth angles.

* **RMS Delay Spread (ns)** (How long it takes for different signal paths to arrive at the receiver)
* **Frequency of transmission (GHz)**
* **Seasonal Variation**
* **Target Variable:** Path Loss (dB)

To prepare the data for analysis and modelling, the following preparation processes were used:

* **Seasonal Normalization:** The “Seasonal Variation” column has inconsistencies in its format. Using pandas string operations, a new column called “Normalized Season” can be extracted and normalized, which is in lowercase and separates the values into the 4 different seasons (summer, autumn, spring and winter).
* **Data Cleaning:** Using pandas, the dataset is checked for missing values and removed if found. There were no missing values in the dataset.

* 1. Interactive Dashboard (Built using Streamlit)

The interactive dashboard for this project was created with Streamlit and encased in a structured Python script. To improve modularity, custom component functions were moved to a separate components.py module and the code used to train and test the model is moved into a separate “ml” module, encouraging code reusability and readability.

This dashboard allows users to visually explore the dataset, filter it by key factors, and study feature correlations using a variety of interactive visualizations. It also allows for real-time display of model performance based on machine learning experiment outcomes.

Key functionalities of the dashboard:

* **Sidebar Filters:** The seasonal filter allows users to filter data by choosing between the four seasons: fall, spring, summer, and winter. A distance range slider is also included to allow users to modify the T-R Separation distance.
* **Interactive Scatter Plot:** The scatter plot allows users to choose independent variables to compare against path loss. Users can also choose between different types of trendlines to determine relationships between the variables.
* **Seasonal Box Plot:** Shows how path loss varies across different seasons
* **Correlation Heatmap:** Shows the correlations between different variables in numerical figures.
* **Toggle-based Visualization Control:** The checkboxes on the sidebar allow users to show and hide each of the visualizations as they please.

* 1. Machine Learning Pipeline

The machine learning component was implemented in a separate ml.py module, allowing for reusability and helps to abstract the machine learning pipeline logic from the Streamlit dashboard logic. The initial stage of the pipeline was to preprocess the data to make it possible to train the models. The key features used for training were:

* T-R Separation Distance
* Received Power
* Azimuth and Elevation angles of departure and arrival
* RMS Delay Spread
* Frequency

The seasonal data was also encoded with one-hot encoding through the get\_dummies function of the pandas library, to produce binary columns like Normalised\_Season\_fall, Normalised\_Season\_summer and so on. This would ensure that the data could be used to train the machine learning models.

After selecting the features to train the machine learning models, the dataset was split into training and testing groups with an 80/20 ratio. To improve the performance of scale-sensitive models such as SVR, KNN and MLP neural networks, StandardScaler was used to normalize the numerical values of the dataset. K-Fold cross-validation was also used to improve the accuracy of the model. Using K-Fold guarantees that model performance measures, such Mean Absolute Error (MAE) and R² Score, are not influenced by a single train-test split and accurately reflect the model's capabilities. The algorithms used for this machine learning component is displayed in Table 1.

Table 1: Machine Learning Algorithms Used

| Model | Description |
| --- | --- |
| Linear Regression | Baseline linear model |
| Ridge Regression | Regularized linear model |
| Random Forest Regressor | Decision tree-based ensemble model |
| XGBoost Regressor | Optimized decision tree ensemble model |
| Support Vector Regressor | Support Vector Machine based regression model |
| K-Nearest Neighbors | Classification model |
| Multilayer Perceptron | Feedforward neural network model |
|  |  |

To evaluate the performance of the models, we utilize **Mean Absolute Error (MAE)**, which is a measure of the average prediction error and **R2 score**, which determines how well the model predicts the Path Loss. To visualize the performance, we will pass the evaluation scores to Streamlit, which will display the results in a bar chart through plotly.express.

1. Implementation

The visuals in this project were created with Plotly Express and Streamlit. The usage of these libraries enables dynamic data visualizations in the browser.

* 1. Core Libraries Used
* **Streamlit** – To create web-based user interface and to enable interactivity with the visualization.
* **Plotly Express** **–** Generate all the charts and plots
* **Numpy –** Used for numerical computations, namely the model statistics and the correlation heatmap.
* **Pandas –** Used for data handling and processing, with features such as data loading and encoding variables.
  1. Scatter Plot

The purpose of the scatter plot is to visualize and help users to identify the relationship between Path Loss and other variables. The implementation detail is as such:

* A dropdown box (made using st.selectbox) allows the user to choose between different x-axis variables to compare against Path Loss
* Also using st.selectbox, users can choose between different types of trendlines to be displayed on the scatter plot
* The plot is created using plotly.express.scatter(), which includes dynamic labels when hovered and displayed using st.plotly\_chart

st.subheader("Path Loss Scatter Plot")

    trendline\_type = st.selectbox("Select trendline type:", ["ols", "lowess", None])

    x\_var = st.selectbox("Select a factor to compare with Path Loss:",

                        ["T-R Separation Distance (m)", "Received Power (dBm)", "Azimuth AoD (degree)", "Elevation AoD (degree)", "Azimuth AoA (degree)", "Elevation AoA (degree)", "RMS Delay Spread (ns)", “Frequency (GHz)”])

    scatter\_fig = px.scatter(filtered\_df, x=x\_var, y="Path Loss (dB)",

                           trendline=trendline\_type,

                           title=f"Path Loss vs {x\_var}",

                           trendline\_color\_override="red")

    st.plotly\_chart(scatter\_fig)

* 1. Box Plot

The box plot is used to determine how the variations in seasons affect the path loss, it shows the median, top and bottom quartile, and min max values. Created using plotly.express.box(), the x-axis is the season, and y-axis is the path loss:

    st.subheader("Seasonal Impact on Path Loss")

    season\_fig = px.box(filtered\_df, x="Normalized Season", y="Path Loss (dB)",

                       color="Normalized Season",

                       title="Path Loss Distribution Across Seasons")

    st.plotly\_chart(season\_fig)

* 1. Correlation Heatmap

The correlation heatmap is used to display the relationships between the numerical features of the dataset. Firstly, the numerical features are extracted into a DataFrame using select\_dtypes(include=np.number). Then, pandas.DataFrame.corr() is used to compute the correlation matrix, which is then shown using plotly.express.imshow(). The implementation is as follows:

st.subheader("Correlation Heatmap")

    numeric\_df = filtered\_df.select\_dtypes(include=np.number)

    corr\_matrix = numeric\_df.corr().round(2)

    fig = px.imshow(corr\_matrix,

                   text\_auto=True,

                   color\_continuous\_scale='RdBu\_r',

                   title="Correlation Matrix of Numerical Features",

                   aspect='auto')

    st.plotly\_chart(fig)

* 1. Machine Learning Model Evaluation Charts

The machine learning models are evaluated based on the MAE and R2 scores, which are then displayed on 2 bar charts. The logic of training and testing the machine learning models are abstracted away into the train\_and\_evaluate\_models function, which produces a DataFrame object containing the models’ scores. The results are then computed and used to generate the charts in Steamlit using plotly.express.bar().

st.header("Machine Learning Model Evaluation")

    results\_df = train\_and\_evaluate\_models(df, n\_splits=5)

    best\_model = results\_df.loc[results\_df["R2 Score (Mean)"].idxmax()]

    st.success(f"🏆 \*\*Best Model:\*\* {best\_model['Model']}")

    st.write(f"🔹 R² Score: {best\_model['R2 Score (Mean)']:.4f}")

    st.write(f"🔹 MAE: {best\_model['MAE (Mean)']:.4f}")

    sorted\_mae\_df = results\_df.sort\_values(by="MAE (Mean)", ascending=True)

    sorted\_r2\_df = results\_df.sort\_values(by="R2 Score (Mean)", ascending=False)

    st.plotly\_chart(

        px.bar(

            sorted\_mae\_df,

            x="Model",

            y="MAE (Mean)",

            title="Mean Absolute Error (MAE) by Model (K-Fold Avg)",

            color="MAE (Mean)",

            color\_continuous\_scale="blues"

        ).update\_traces(hovertemplate='Model: %{x}<br>MAE: %{y:.2f}')

    )

    st.plotly\_chart(

        px.bar(

            sorted\_r2\_df,

            x="Model",

            y="R2 Score (Mean)",

            title="R² Score by Model (K-Fold Avg)",

            color="R2 Score (Mean)",

            color\_continuous\_scale="viridis"

        ).update\_traces(hovertemplate='Model: %{x}<br>R² Score: %{y:.2f}')

    )

* 1. Model Training and Evaluation

The training and evaluation process is abstracted into the train\_and\_evaluate\_models function, and the implementation is as follows:

def train\_and\_evaluate\_models(df, n\_splits=5):

    df = pd.get\_dummies(df, columns=["Normalized Season"])

    features = ["T-R Separation Distance (m)", "Received Power (dBm)", "Azimuth AoD (degree)",

                "Elevation AoD (degree)", "Azimuth AoA (degree)", "Elevation AoA (degree)",

                "RMS Delay Spread (ns)", "Frequency",

                "Normalized Season\_fall", "Normalized Season\_spring",

                "Normalized Season\_summer", "Normalized Season\_winter"]

    X = df[features].values

    y = df["Path Loss (dB)"].values

    kf = KFold(n\_splits=n\_splits, shuffle=True, random\_state=42)

    models = {

        "Linear Regression": LinearRegression(),

        "Ridge Regression": Ridge(alpha=1.0),

        "Random Forest": RandomForestRegressor(n\_estimators=100, random\_state=42),

        "XGBoost": xgb.XGBRegressor(n\_estimators=100, learning\_rate=0.1, objective='reg:squarederror'),

        "Support Vector Regressor": SVR(kernel='rbf'),

        "K-Nearest Neighbors": KNeighborsRegressor(n\_neighbors=5),

        "Neural Network (MLP)": MLPRegressor(hidden\_layer\_sizes=(64, 64), max\_iter=500, random\_state=42)

    }

    results = {name: {"MAE": [], "R2": []} for name in models.keys()}

    for train\_index, test\_index in kf.split(X):

        X\_train, X\_test = X[train\_index], X[test\_index]

        y\_train, y\_test = y[train\_index], y[test\_index]

        scaler = StandardScaler()

        X\_train\_scaled = scaler.fit\_transform(X\_train)

        X\_test\_scaled = scaler.transform(X\_test)

        for name, model in models.items():

            if "SVR" in name or "KNN" in name or "MLP" in name:

                model.fit(X\_train\_scaled, y\_train)

                y\_pred = model.predict(X\_test\_scaled)

            else:

                model.fit(X\_train, y\_train)

                y\_pred = model.predict(X\_test)

            results[name]["MAE"].append(mean\_absolute\_error(y\_test, y\_pred))

            results[name]["R2"].append(r2\_score(y\_test, y\_pred))

    summary = []

    for name, scores in results.items():

        summary.append({

            "Model": name,

            "MAE (Mean)": np.mean(scores["MAE"]),

            "MAE (Std)": np.std(scores["MAE"]),

            "R2 Score (Mean)": np.mean(scores["R2"]),

            "R2 Score (Std)": np.std(scores["R2"])

        })

    return pd.DataFrame(summary)

1. Results & discussion

This system uses multiple interactive visuals and predictive modelling tools to evaluate the path loss of 5G communication in the South Asia setting. The findings given here include both exploratory data visualizations and machine learning-based prediction ratings.

* 1. Scatter Plot with Regression Lines

A dynamic scatter plot is used to compare path loss with other factors (T-R Separation Distance, Received Power, Azimuth & Elevation AoD and AoA, or RMS Delay Spread). A trendline is included in the plot to reveal any linear relationships. Based on Figure 3, we can determine that:

* Path loss is higher when **T-R Separation, RMS Delay Spread and Frequency is increased**.
* Path loss is lower when **Received Power is lower.**
* The directional characteristics have insignificant effects on path loss except for **Elevation AOA**, which slightly decreases path loss when the angle is increased, as shown in Figure 4.

|  |  |
| --- | --- |
|  |  |

Figure 3: Scatter Plot Results on channel characteristics against path loss

|  |  |
| --- | --- |
|  |  |

Figure 4: Scatter Plot Results on Elevation/Azimuth AoA & AoDs against path loss

* 1. Seasonal Box Plot

A box plot was used to determine the effects of seasons on the path loss, as shown in Figure 5. It is discovered that summer has the lowest median path loss and spring has the highest median path loss.

A graph of different colored squares

AI-generated content may be incorrect.

Figure 5: Box Plot of Path Loss across different seasons

* 1. Correlation Heatmap

The correlation heatmap shows the correlation between different variables in numerical figures, as shown in Figure 6. Through the heatmap, we can discover that:

* Path loss has a positive correlation with T-R Separation Distance, Time Delay and Frequency.
* Path loss has a negative correlation with Received Power.

A screenshot of a graph

AI-generated content may be incorrect.

Figure 6: Correlation Heatmap

* 1. Model Performance Evaluation

K-Fold Cross Validation was used to test different machine learning models for predicting Path Loss (dB). The two measures utilized for comparison were:

* Mean Absolute Error (MAE): Average absolute difference between predicted and actual values.
* R2 Score: How well the model explains variance in data (closer to 1 is better)

In Figure 7 and 8, we can conclude that XGBoost is the most reliable model for this dataset as it performs the best with the highest prediction accuracy. The scores are displayed in Table 2.

A graph of blue and white bars

AI-generated content may be incorrect.

Figure 7: Mean Absolute Error by model

A graph of different colored bars

AI-generated content may be incorrect.

Figure 8: R2 score by model

Table 2: Evaluation of Machine Learning Models

| Model | MAE | R2 Score |
| --- | --- | --- |
| XGBoost | 3.35 | 0.91 |
| Random Forest | 3.42 | 0.9 |
| Neural Network (MLP) | 4.33 | 0.86 |
| K-Nearest Neighbors | 5.65 | 0.78 |
| Linear Regression | 6.04 | 0.74 |
| Ridge Regression | 6.04 | 0.74 |
| Support Vector Regressor | 8.33 | 0.55 |
|  |  |  |

* 1. Discussion

The Path Loss Analysis Dashboard provides users with a simple and data-driven interface for investigating how atmospheric and channel characteristics can influence path loss in 5G wireless systems. One of the most important conclusions from this implementation is that, although path loss can be predicted easily with some accuracy, it can be affected by an intricate mix of features. Variables such as transmitter-receiver separation distance and received power have been found to be the most strongly associated with path loss, consistent with traditional propagation models. However, adding new characteristics, notably seasonal variation, revealed trends that would otherwise go unnoticed.

The machine learning model comparison allows users to understand the significance of picking the right model in forecasting outcomes. Tree-based models, notably XGBoost and Random Forest, has shown to perform better than linear and kernel-based models by providing higher accuracy and in this dataset. These models succeeded in identifying non-linear correlations in the data. At the same time, although the simpler models such as Linear Regression and Ridge offered quicker calculation, they were less accurate compared to the tree-based models. In contrast, the Support Vector Regressor did poorly across all measures, indicating that the model is not suitable for the current dataset without extra tuning or preprocessing.

From our findings, we can also learn that environmental factors, particularly seasonal variations, has an impact on path loss. The seasonal comparison plots demonstrated that the summer and fall frequently led to a greater overall path loss and volatility. This emphasizes the fact that propagation conditions do not remain static, and successful network planning must take seasonal aspects into consideration. This can result in more resilient system designs that account for changes in foliage, humidity, and other environmental variables over time.

Aside from its ability to analyse data, the dashboard allows for simpler and engaging exploration of data. Although no formal user studies were undertaken, informal observations revealed that users enjoyed being able to filter datasets and examine real-time data visualization via graphical representations. The dashboard is extremely useful for engineers and academics who wanted to convey their discoveries to a larger audience, including stakeholders with less technical knowledge. The correlation matrix, together with the implementation of machine learning for prediction, emphasized the notion that statistical relationships should also be proven by forecasting accuracy rather than just on correlation.

1. Conclusion & Future works

The study of the 5G Path Loss dataset shows numerous critical insights regarding mm-wave signal propagation behaviour under different environmental and seasonal settings. One notable discovery is the substantial positive connection between T-R Separation Distance (m) and Path Loss (dB), which is consistent with traditional propagation models and highlights distance as a major component in path loss. Furthermore, the box plot indicates that seasonal fluctuations have a major impact on route loss, which are most likely caused by environmental variables such as foliage, humidity, and temperature changes. Factors such as Angle of Arrival (AoA), Angle of Departure (AoD), RMS Delay Spread, and Frequency also have different levels of relationship to route loss, indicating intricate relationships influencing signal behavior. Through this study, we found out that XGBoost regularly outperformed other regression models in terms of R² Score and Mean Absolute Error (MAE), demonstrating its ability to represent non-linear connections in the dataset.

* 1. Future Works

There are numerous options for expanding and improving the project. For instance, we could incorporate geographic data to generate terrain-based features or environmental obstructions. We can also make use of real-time weather data instead of broad seasonal labelling to get more insight into the nature of environmental influences on path loss. Aside from collecting different data metrics, increasing the geographic reach of the study by including information from various locations with different climates would enable comparative examinations of mm-wave communication throughout various landscapes and environmental conditions. By incorporating this pipeline into 5G network design tools, we can help to create more dependable systems that are better prepared to tolerate environmental fluctuations.

REFERENCES

1. Jacob, M., Priebe, S., Dickhoff, R., Kleine-Ostmann, T., Schrader, T., & Kurner, T. (2012). Diffraction in mm and sub-mm wave indoor propagation channels. *IEEE Transactions on Microwave Theory and Techniques*, *60*(3), 833-844.
2. Ratul, Rashed Hasan, et al. Atmospheric Influence on the Path Loss at High Frequencies for Deployment of 5G Cellular Communication Networks. arXiv, 2 June 2023. arXiv.org, <https://doi.org/10.48550/arXiv.2306.01894>.
3. Rappaport, T. S., Heath Jr, R. W., Daniels, R. C., & Murdock, J. N. (2015). *Millimeter wave wireless communications*. Pearson Education.
4. Marcus, M., & Pattan, B. (2005). Millimeter wave propagation: spectrum management implications. *IEEE Microwave Magazine*, *6*(2), 54-62.
5. Chen, X., Tian, L., Tang, P., & Zhang, J. (2016). Modelling of Human Body Shadowing Based on 28 GHz Indoor Measurement Results. 2016 IEEE 84th Vehicular Technology Conference (VTC-Fall), 1-5.
6. Asp, A., Sydorov, Y., Valkama, M., & Niemelä, J. (2012, December). Radio signal propagation and attenuation measurements for modern residential buildings. In *2012 IEEE Globecom Workshops* (pp. 580-584). IEEE.
7. Karjalainen, J., Nekovee, M., Benn, H., Kim, W., Park, J., & Sungsoo, H. (2014, June). Challenges and opportunities of mm-wave communication in 5G networks. In *2014 9th international conference on cognitive radio oriented wireless networks and communications (CROWNCOM)* (pp. 372-376). IEEE.
8. Krofli, R. A., & Kelly, R. D. (1996). Meteorological research applications of mm-wave radar. *Meteorology and Atmospheric Physics*, *59*(1), 105-121.
9. Ratul, R. H., Zaman, S. M., Chowdhury, H. A., Sagor, M. Z. H., Kawser, M. T., & Nishat, M. M. (2023, July). Atmospheric influence on the path loss at high frequencies for deployment of 5g cellular communication networks. In *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
10. Rappaport, T. S., Xing, Y., Kanhere, O., Ju, S., Madanayake, A., Mandal, S., ... & Trichopoulos, G. C. (2019). Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond. *IEEE access*, *7*, 78729-78757.
11. Modification of Earth-Space Rain Attenuation Model for EarthSpace Link - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Specific-attenuation-of-water-vapor-and-oxygen-ITU-P676-9-12\_fig1\_270445506 [accessed 13 Apr 2025]
12. Zhang, Y., Anderson, C. R., Michelusi, N., Love, D. J., Baker, K. R., & Krogmeier, J. V. (2019). Propagation modeling through foliage in a coniferous forest at 28 GHz. *IEEE Wireless Communications Letters*, *8*(3), 901-904.
13. Sun, Shu & Maccartney, George & Rappaport, T.S.. (2017). A novel millimeter-wave channel simulator and applications for 5G wireless communications. 1-7. 10.1109/ICC.2017.7996792.
14. Rappaport, T. S., Sun, S., & Shafi, M. (2017, September). Investigation and comparison of 3GPP and NYUSIM channel models for 5G wireless communications. In *2017 IEEE 86th vehicular technology conference (VTC-Fall)* (pp. 1-5). IEEE.