Cairo University Faculty of Engineering

12th Undergraduate Engineering Mathematics Engineering Research Forum

**Data Architects**

Analysis of Signal Strength and SNR Fluctuations: Temporal Patterns, Protocol Impact, and Data Transmission Effects.

## Authors

Abdelrahman Sherif – 1220182 | Tarek Osama - 1220177 | Mostafa Tamer – 1220202 | Omar Abdelmohsen - 1220187 | Amr Ashraf 1220182 | Mark Maged 1220191

##### Under the supervision of:

**Dr. Maha Amin Hassanein**

**MTHS204**

**Spring 2025**

# Abstract

Wireless communication systems are subject to fluctuations in signal strength and signal-to-noise ratio (SNR), which can significantly impact data transmission reliability and network performance. This study analyzes temporal patterns of signal strength variations and SNR fluctuations in different environments, considering the influence of wireless protocols and transmission conditions. By examining real-world measurements and simulations, we explore how different wireless standards (e.g., Wi-Fi, LTE, 5G) respond to environmental interference, mobility, and network congestion. Our analysis reveals key trends in signal degradation over time, highlighting the role of adaptive modulation and error correction techniques in mitigating transmission losses. Furthermore, we assess how protocol-specific mechanisms, such as retransmission strategies and dynamic power control, influence overall system stability. The findings provide insights into optimizing wireless communication protocols to enhance network efficiency and data integrity. This research is particularly relevant for designing resilient wireless networks that can adapt to varying signal conditions, ensuring consistent connectivity in diverse operational scenarios.

# Table of Contents

[Abstract 1](#_bookmark0)

[Table of Contents 2](#_bookmark1)

[List of Figures 3](#_bookmark2)

[Chapter 1: Problem Definition 4](#_bookmark4)

* 1. [Problem Description 4](#_bookmark5)
  2. [Statistical Questions 4](#_bookmark7)

[Chapter 2: Descriptive Statistics 5](#_bookmark8)

* 1. [Definition of Variables 5](#_bookmark9)
  2. [Data Description: 5](#_bookmark15)

2.2.1 Dataset 1: 5

2.2.2 Dataset 2: 9

2.2.3 Dataset 3: 14

[Chapter 3: Methodology 10](#_bookmark18)

* 1. [Data Description 10](#_bookmark19)
     1. [Description of Images 11](#_bookmark22)
     2. [Descriptive Statistical Analyses 11](#_bookmark24)
  2. [Model Architecture and Training Procedure 14](#_bookmark27)
     1. [Layers Parameters 14](#_bookmark28)
     2. [Model Training 15](#_bookmark30)
  3. [Results 15](#_bookmark31)

[Conclusion 17](#_bookmark36)

[References 18](#_bookmark37)

[Appendix 19](#_bookmark38)

# List of Figures

[Figure 1 Dataset 1 Descriptive Statistics Table 5](#_bookmark6)

[Figure 2 Dataset 1 Heatmap 6](#_bookmark10)

[Figure 3 Distribution of Flow Duration 6](#_bookmark11)

[Figure 4 Distribution of Total Length of Forward Packets 7](#_bookmark14)

[Figure 5 Distribution of Total Length of Backward Packets 7](#_bookmark16)

[Figure 6 Distribution of Forward IAT 8](#_bookmark20)

[Figure 7 Distribution of Backward IAT 8](#_bookmark21)

[Figure 8 Distribution of Protocols 9](#_bookmark23)

[Figure 9 Dataset 2 Descriptive Statistics Table 9](#_bookmark25)

[Figure 10 Dataset 2 Heatmap 10](#_bookmark26)

[Figure 11 Distribution of Tx 10](#_bookmark29)

[Figure 12 Distribution of Rx 11](#_bookmark33)

[Figure 13 Distribution of SNR Receiver 11](#_bookmark34)

[Figure 14 Distribution of BER Receiver 12](#_bookmark35)

[Figure 15 Distribution of Transmission Distance 12](#_bookmark25)

[Figure 16 Distribution of Fiber Attenuation 13](#_bookmark26)

[Figure 17 Distribution of Temperature 13](#_bookmark29)

[Figure 18 Distribution of Humidity 14](#_bookmark33)

[Figure 19 Dataset 3 Descriptive Statistics Table 14](#_bookmark34)

[Figure 20 Dataset 3 Heatmap 15](#_bookmark35)

[Figure 21 Distribution of Signal Strength 15](#_bookmark34)

[Figure 22 Distribution of SNR 16](#_bookmark35)

[Figure 23 Distribution of Call Duration 16](#_bookmark25)

[Figure 24 Distribution of Attenuation 17](#_bookmark26)

[Figure 25 Distribution of Distance to Tower 17](#_bookmark29)

[Figure 26 Distribution of Environment 18](#_bookmark33)

[Figure 27 Distribution of Incoming/Outgoing 18](#_bookmark34)

[Figure 27 Distribution of Call Type 19](#_bookmark34)

# 

# Chapter 1: Problem Definition

### Problem Description

Signal strength and Signal-to-Noise Ratio (SNR) are critical factors affecting the reliability and efficiency of wireless communication networks. These parameters fluctuate due to various factors, including time of day, protocol used, data type/size, and distance from transmission sources. However, the extent and patterns of these variations are not always well understood. This project aims to analyze how signal strength and SNR change over time, their correlation with different communication protocols, and the impact of data characteristics. The findings will help identify potential optimization strategies for improving network performance.

*1.2 Statistical Questions*

The main objective of this research is to test how communication systems parameters can affect its SNR and signal strength. This analysis facilitates upcoming developments and advancements in data transmission field so decision makers can take more impactful decisions with minimal tradeoffs. We concluded it into 5-6 main statistical questions we ought to answer.

1. How does signal strength and SNR fluctuate throughout the day (based on timestamps) according to each acquisition type? Are there peak hours of degradation?
2. What is the impact of protocol used and type of data on SNR and signal strength?
3. How is the Performance of protocols on different types/sizes of data throughout the day/year?
4. Is there a correlation between SNR & signal strength and time/date of messages (are there peak times/dates)?
5. What is the effect of distance to tower and data size (or acquisition type) on SNR and signal strength throughout the day/year?
6. What is the effect of environmental conditions with different transmission distances on network performance?

# CHAPTER 2: DESCRIPTIVE STASTISTICS

### 2.1 Definition of Variables

Independent Variables: Timestamp, Total.Length.of.Fwd.Packets, Total.Length.of.Bwd.Packets, L7Protocol, ProtocolName, Transmission Distance, Temperature, Humidity, Call Duration (s), Environment, Distance to Tower (km), Call Type, Incoming/Outgoing.

Dependent Variables: Flow.Duration, Fwd.IAT.Total, Bwd.IAT.Total, Transmitter Power Level (Tx), Receiver Power Level (Rx), SNR Receiver, BER Receiver, Fiber Attenuation, Signal Quality, Signal Strength (dBm), SNR, Attenuation.

### 2.2 Data Description

### 2.2.1 Dataset 1 (to answer Question 3): 8 Columns & 928,676 Rows

### A table with numbers and text AI-generated content may be incorrect.

Fig.1 Dataset 1 Descriptive Statistics Table

### A red and blue squares with white text AI-generated content may be incorrect.

Fig.2 Dataset 1 Heatmap

A graph with blue squares

AI-generated content may be incorrect.

Fig.3 Distribution of Flow Duration

A graph with a number of lines

AI-generated content may be incorrect.

Fig.4 Distribution of Total Length of Forward Packets

A graph with lines in the middle

AI-generated content may be incorrect.

Fig.5 Distribution of Total Length of Backward Packets

A graph with lines in the middle

AI-generated content may be incorrect.

Fig.6 Distribution of Forward IAT

A graph with lines and dots

AI-generated content may be incorrect.

Fig.7 Distribution of Backward IAT

A colorful bars on a white background

AI-generated content may be incorrect.

Fig.8 Protocols Distribution

### 

### 2.2.2 Dataset 2 (to answer Question ): 9 Columns & 586 Rows

### A table with numbers and text AI-generated content may be incorrect.

Fig.9 Descriptive Statistics Table

A screenshot of a computer screen

AI-generated content may be incorrect.

Fig.10 Dataset 2 Heatmap

A graph of blue lines

AI-generated content may be incorrect.

Fig.11 Distribution of Tx

A graph of blue lines

AI-generated content may be incorrect.

Fig.12 Distribution of Rx

A graph with blue lines

AI-generated content may be incorrect.

Fig.13 Distribution of SNR Receiver

A graph of a bar graph

AI-generated content may be incorrect.

Fig.14 Distribution of BER Receiver

A graph with blue lines

AI-generated content may be incorrect.

Fig.15 Distribution of Transmission Distance

A graph with blue lines

AI-generated content may be incorrect.

Fig.16 Distribution of Fiber Attenuation

A graph with blue lines

AI-generated content may be incorrect.

Fig.17 Distribution of Temperature

A graph with blue lines

AI-generated content may be incorrect.

Fig.18 Distribution of Humidity

### 2.2.3 Dataset 3 (to answer Question ): 9 Columns & 463 Rows

### A table with numbers and letters AI-generated content may be incorrect.

Fig.19 Dataset 3 Descriptive Statistics Table

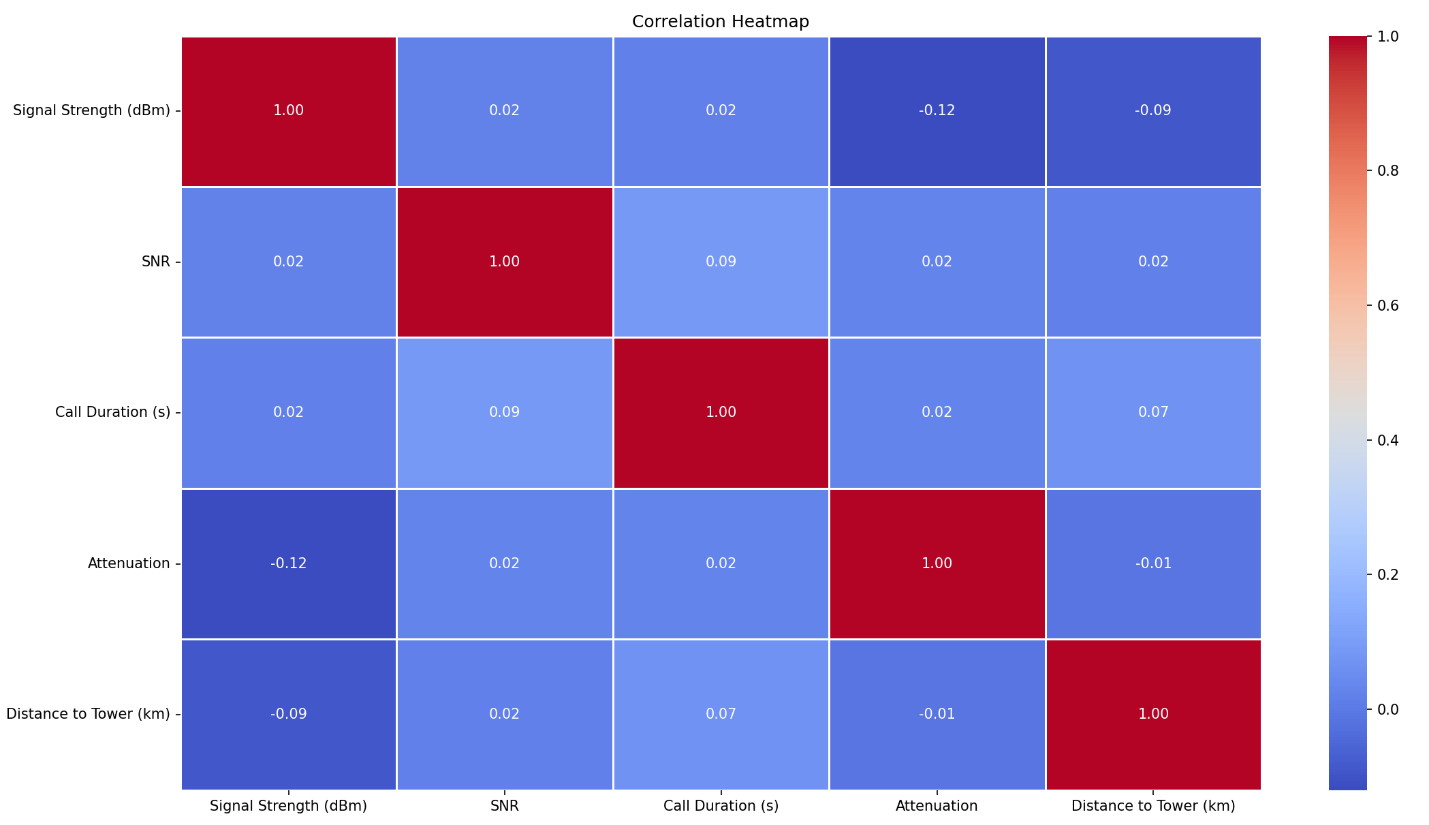


Fig.20 Dataset 3 Heatmap

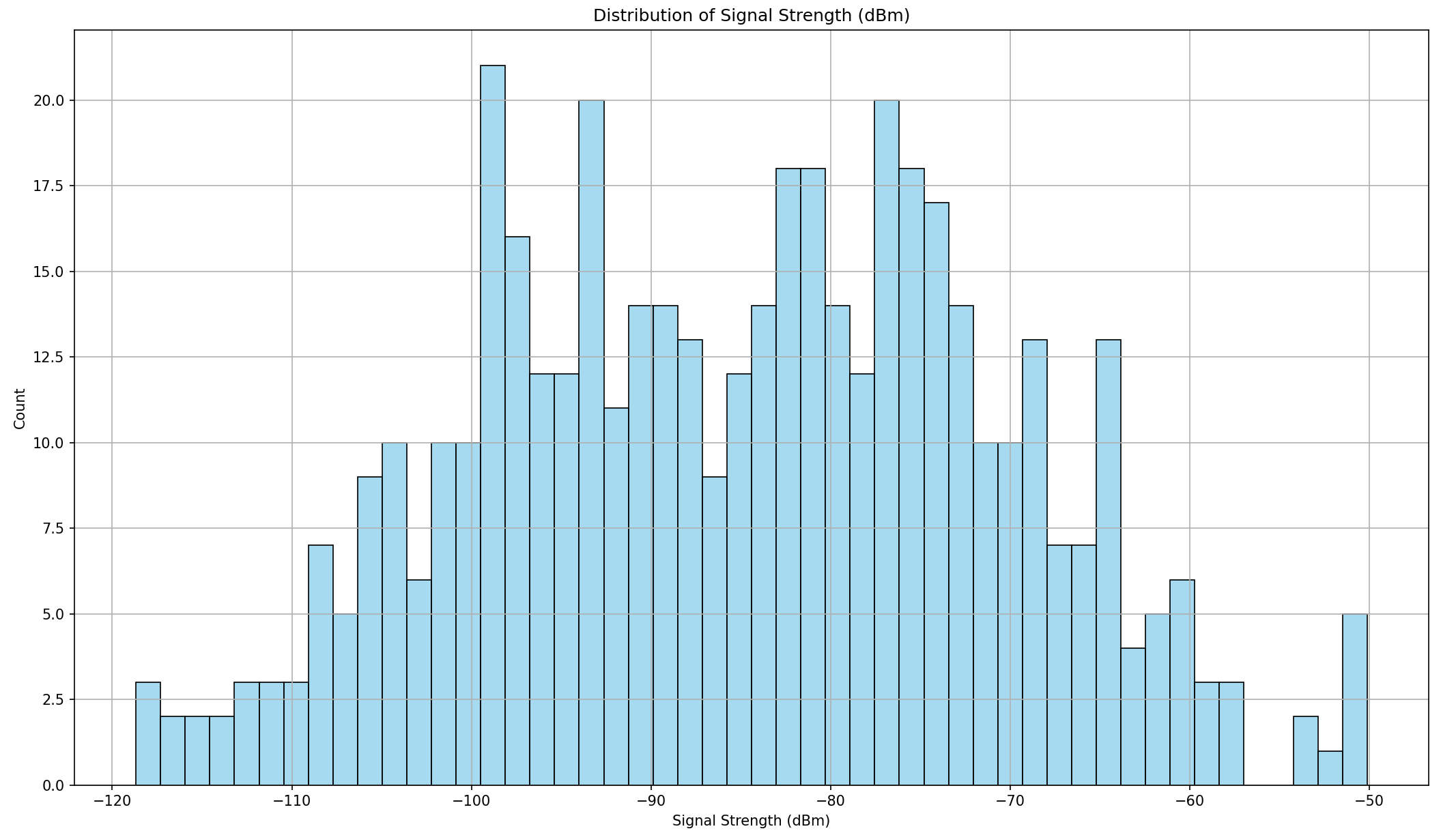


Fig.21 Distribution of Signal Strength

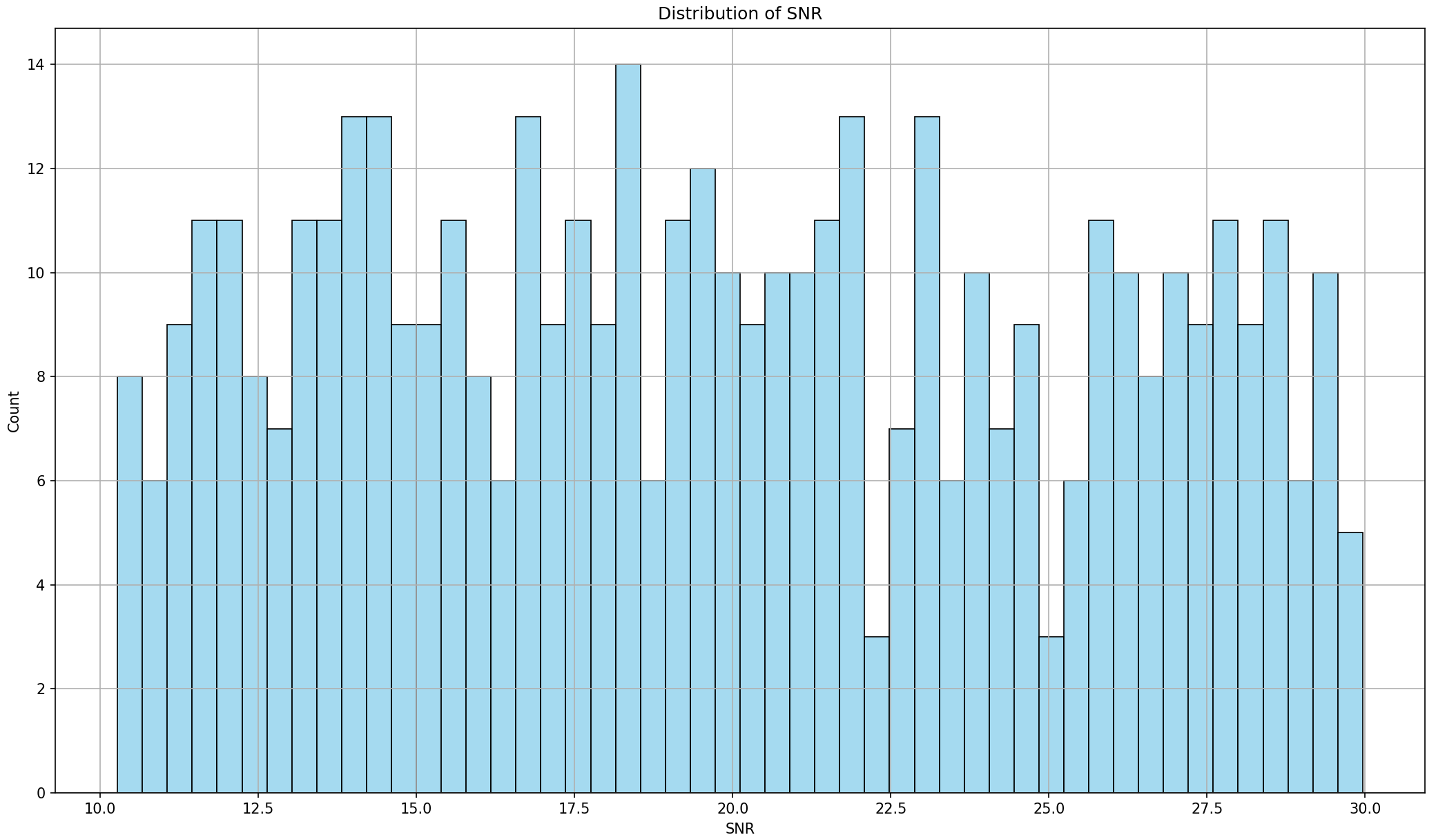


Fig.22 Distribution of SNR

A graph with blue lines

AI-generated content may be incorrect.

Fig.23 Distribution of Call Duration

A graph of a graph

AI-generated content may be incorrect.

Fig.24 Distribution of Attenuation

A graph with blue lines

AI-generated content may be incorrect.

Fig.25 Distribution of Distance to Tower

A colorful rectangular objects with numbers

AI-generated content may be incorrect.

Fig.26 Distribution of Environment

A blue and green rectangles

AI-generated content may be incorrect.

Fig.27 Distribution of Incoming/Outgoing

A screenshot of a computer screen

AI-generated content may be incorrect.

Fig.28 Distribution of Call Type

# Chapter 3: Methodology

The methodology is divided into three main parts: the Artificial Intelligence (AI) model, image preprocessing, and desktop application. The following sections delve deeply into each of these components.

**2. Research Question: What is the impact of protocol used and type of data on SNR and signal strength?**

**2.1 Overview of Analysis**

This study examines how different communication protocols and data types affect **Signal-to-Noise Ratio (SNR) and Signal Strength (dBm)**. By applying **statistical inference and probability models**, we aim to identify patterns and quantify the impact of protocol and data type on wireless communication quality.

**2.2 Statistical Analysis Methods**

**1. Exploratory Data Analysis (EDA)**

Before applying statistical models, an initial exploration of the dataset will include:

* Descriptive statistics (mean, median, variance) for SNR and Signal Strength across different protocols.
* Outlier detection using box plots and Z-score analysis.
* Distribution analysis of SNR and Signal Strength (using histograms and Q-Q plots).
* Correlation matrix to identify relationships between protocol type, data type, and signal metrics.

**2. Comparing SNR and Signal Strength Across Protocols**

This step will determine whether different protocols significantly impact signal metrics.

* **Analysis method:**
  + ANOVA (Analysis of Variance): If SNR and Signal Strength follow a normal distribution, we use ANOVA to compare means across multiple protocols.
  + Kruskal-Wallis Test: A non-parametric alternative if normality assumptions fail.
  + Post-hoc Tukey’s HSD test: If ANOVA detects significant differences, we use this to find which protocols differ significantly.
* **Expected Results:**
  + Some protocols may have **statistically higher SNR or Signal Strength**, indicating better performance.
  + Protocols optimized for real-time communication (e.g., VoIP) may behave differently from data-heavy protocols (e.g., FTP).

**3. Impact of Data Type on SNR and Signal Strength**

Here, we evaluate whether the type of transmitted data affects signal behavior.

* **Analysis method:**
  + T-test (if two data types exist) or ANOVA (if multiple data types exist).
  + Effect Size Calculation: Use Cohen’s d (for two groups) or η² (for multiple groups) to measure the strength of the relationship.
* **Expected Results:**
  + Larger data sizes or streaming applications may correlate with lower SNR due to increased interference.
  + Data-intensive transmissions (e.g., video streaming) may cause fluctuations in signal strength.

**4. Regression Analysis: Predicting SNR & Signal Strength**

We will model SNR and Signal Strength as functions of multiple factors to identify key predictors.

* **Analysis method:**
  + Multiple Linear Regression:
    - Dependent Variables: SNR, Signal Strength.
    - Independent Variables: Protocol, Data Type, Flow Duration, Distance to Tower, and Environmental Factors.
  + Interaction Effects Analysis: Check whether protocol & data type together have a combined impact on signal behavior.
* **Expected Results:**
  + Certain protocols and data types may negatively impact SNR, especially in long-duration transmissions.
  + Environmental conditions (distance to tower, attenuation) may interact with protocol types.

**5. Chi-Square Test for Categorical Dependencies**

To examine relationships between protocol type, call type, and environment, we will use a Chi-Square Test for Independence.

* **Analysis method:**
  + Test whether protocol choice depends on environmental conditions (e.g., urban vs. rural).
  + Determine whether certain call types are associated with higher/lower signal strength.
* **Expected Results:**
  + Protocols may have a preference for certain environments (e.g., some work better in urban areas, others in rural areas).
  + Call types (incoming vs. outgoing) may show signal strength variations due to network load.

**2.3 Additional Considerations**

**Expected Calculations & Outputs**

* Mean & variance of SNR and Signal Strength for each protocol & data type.
* p-values from hypothesis tests (to assess statistical significance).
* Regression coefficients (showing protocol/data type influence on SNR).
* Effect size measures (Cohen’s d, η²) to quantify impact strength.

# Conclusion

pllm k

# References

* Nathaniel Handan. “OptiCom Signal Quality Dataset” [**https://www.kaggle.com/datasets/tinnyrobot/opticom-signal-quality-dataset?select=ocrdataset.csv**](https://www.kaggle.com/datasets/tinnyrobot/opticom-signal-quality-dataset?select=ocrdataset.csv)Feb.15,2024 [March.7,2025]
* Asfand Yar. “Internet Traffic Data Set”

[**https://www.kaggle.com/datasets/asfandyar250/network/data**](https://www.kaggle.com/datasets/asfandyar250/network/data)Feb.23,2023 **[**March.7,2025**].**

* Suraj “cellular-network-performance-data”

[**https://www.kaggle.com/datasets/suraj520/cellular-network-performance-data**](https://www.kaggle.com/datasets/suraj520/cellular-network-performance-data)

Jan.01,2023 [March.7,2025].

# Appendix

Python code to process the data and apply the explained descriptive statistics analysis

import matplotlib.pyplot as plt  
import seaborn as sns  
import pandas as pd  
import numpy as np  
import pandas as pd  
  
# Load dataset  
df = pd.read\_csv("C:/Users/osama/OneDrive/Desktop/Data\_Architects/Datasets/Dataset-Unicauca-Version2-87Atts.csv")  
  
  
# Drop "L7Protocol" column if it exists  
if "L7Protocol" in df.columns:  
 df = df.drop(columns=["L7Protocol"])  
  
# Get descriptive statistics (excluding "count" row)  
desc\_stats = df.describe().drop(index="count", errors="ignore")  
  
# Convert values to 2 decimal places (removes scientific notation)  
desc\_stats = desc\_stats.applymap("{:.2f}".format)  
  
# Convert DataFrame to a string format for visualization  
fig, ax = plt.subplots(figsize=(12, 6)) # Set figure size  
ax.axis('off') # Hide axes  
table = ax.table(cellText=desc\_stats.values,  
 colLabels=desc\_stats.columns,  
 rowLabels=desc\_stats.index,  
 cellLoc='center', loc='center')  
  
# Set table style  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(10)  
table.auto\_set\_column\_width([i for i in range(len(desc\_stats.columns))])  
  
plt.show()  
  
# Plotting correlation heatmap  
plt.figure(figsize=(10, 6))  
#correlation\_matrix = df.corr()  
correlation\_matrix = df.select\_dtypes(include=['number']).corr()  
  
sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt='.2f', linewidths=0.5)  
plt.title('Correlation Heatmap')  
plt.show()  
  
  
# Convert relevant columns to numeric  
numerical\_columns = ['Flow.Duration', 'Total.Length.of.Fwd.Packets',  
 'Total.Length.of.Bwd.Packets', 'Fwd.IAT.Total', 'Bwd.IAT.Total']  
  
# Define numerical columns  
numerical\_columns = ['Flow.Duration', 'Total.Length.of.Fwd.Packets',  
 'Total.Length.of.Bwd.Packets', 'Fwd.IAT.Total', 'Bwd.IAT.Total']  
  
# Convert to numeric and clean data  
for col in numerical\_columns:  
 df[col] = pd.to\_numeric(df[col], errors='coerce') # Convert to numeric  
 df[col] = df[col].replace([np.inf, -np.inf], np.nan) # Remove infinities  
  
# Convert Flow Duration to seconds (assuming it's in microseconds)  
df['Flow.Duration'] = df['Flow.Duration'] / 1\_000\_000  
df['Fwd.IAT.Total'] = df['Fwd.IAT.Total'] / 1\_000\_000  
df['Bwd.IAT.Total'] = df['Bwd.IAT.Total'] / 1\_000\_000  
  
# Plot each numerical column separately  
for col in numerical\_columns:  
 plt.figure(figsize=(8, 5))  
 sns.histplot(df[col].dropna(), bins=50, color='skyblue') # More bins for better detail  
 plt.xlabel(f"{col} ({'10^6 seconds' if col in ['Flow.Duration', 'Fwd.IAT.Total', 'Bwd.IAT.Total'] else 'units'})")  
 plt.ylabel("Count")  
 plt.title(f"Distribution of {col}")  
 plt.ticklabel\_format(style='plain') # Remove scientific notation  
 plt.grid(True)  
 plt.show() #show each plot separately  
  
# Bar plot for ProtocolName (categorical data)  
plt.figure(figsize=(10, 6))  
sns.countplot(y='ProtocolName', data=df, palette='Set2')  
plt.title('Protocol Name Count Distribution')  
plt.xlabel('Count')  
plt.ylabel('Protocol Name')  
plt.show()

# Load dataset

df = pd.read\_csv("C:/Users/osama/OneDrive/Desktop/Data\_Architects/Datasets/ocrdataset.csv")

# Generate descriptive statistics

desc\_stats = df.describe().round(2) # Round for better readability

# Create the figure

fig, ax = plt.subplots(figsize=(12, 6)) # Larger size for clarity

ax.axis('off') # Hide axes

# Create the table with better styling

table = ax.table(cellText=desc\_stats.values,

colLabels=desc\_stats.columns,

rowLabels=desc\_stats.index,

cellLoc='center',

loc='center',

colColours=['lightgray'] \* desc\_stats.shape[1]) # Header row shading

table.auto\_set\_font\_size(False)

table.set\_fontsize(10) # Increase font size

table.auto\_set\_column\_width([i for i in range(len(desc\_stats.columns))]) # Auto-adjust width

# Add title

plt.title("Descriptive Statistics", fontsize=14, fontweight="bold", pad=20)

plt.show()

# Correlation Heatmap

# Convert "Good" to 1 and "Bad" to 0 (Assuming the column name is 'Label')

if 'Label' in df.columns:

df['Label'] = df['Label'].map({'Good': 1, 'Bad': 0})

# Ensure all columns are numeric for correlation

df\_numeric = df.select\_dtypes(include=[np.number])

# Plot the heatmap

plt.figure(figsize=(10, 6))

correlation\_matrix = df\_numeric.corr()

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)

plt.title("Correlation Heatmap")

plt.show()

# Plot distributions for numerical columns

numerical\_columns = df.select\_dtypes(include=[np.number]).columns.tolist()

for col in numerical\_columns:

plt.figure(figsize=(8, 5))

sns.histplot(df[col].dropna(), bins=50, color='skyblue') # Drop NaNs for cleaner plot

plt.xlabel(f"{col}")

plt.ylabel("Count")

plt.title(f"Distribution of {col}")

plt.ticklabel\_format(style='plain') # Avoid scientific notation

plt.grid(True)

plt.show()

# Load dataset  
df = pd.read\_csv("C:/Users/osama/OneDrive/Desktop/Data\_Architects/Datasets/train.csv")  
  
# Generate descriptive statistics  
desc\_stats = df.describe().round(2)  
# Save table as an image  
fig, ax = plt.subplots(figsize=(12, 6))  
ax.axis('tight')  
ax.axis('off')  
table = ax.table(cellText=desc\_stats.values,   
 colLabels=desc\_stats.columns,   
 rowLabels=desc\_stats.index,   
 cellLoc='center',   
 loc='center',  
 colColours=['lightgray'] \* desc\_stats.shape[1])   
  
table.auto\_set\_font\_size(False)  
table.set\_fontsize(10)  
table.auto\_set\_column\_width([i for i in range(len(desc\_stats.columns))])   
  
plt.title("Descriptive Statistics", fontsize=14, fontweight="bold", pad=20)   
plt.show()  
  
  
df\_numeric = df.select\_dtypes(include=[np.number])  
  
# Correlation Heatmap (Numeric Only)  
plt.figure(figsize=(10, 6))  
correlation\_matrix = df\_numeric.corr()  
sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)  
plt.title("Correlation Heatmap")  
plt.show()  
  
# Plot distributions for each numerical column  
for col in df\_numeric.columns:  
 plt.figure(figsize=(8, 5))  
 sns.histplot(df[col].dropna(), bins=50, color='skyblue')  
 plt.xlabel(f"{col}")  
 plt.ylabel("Count")  
 plt.title(f"Distribution of {col}")  
 plt.grid(True)  
 plt.show()  
  
# Identify categorical columns  
categorical\_columns = [col for col in df.select\_dtypes(include=['object']).columns  
 if 'time' not in col.lower() and 'date' not in col.lower()]  
  
# Plot bar charts for each categorical column  
for col in categorical\_columns:  
 plt.figure(figsize=(8, 5))  
 sns.countplot(y=df[col], palette='Set2', order=df[col].value\_counts().index)  
 plt.xlabel("Count")  
 plt.ylabel(col)  
 plt.title(f"Distribution of {col}")  
 plt.show()

Dataset used:

<https://github.com/TarekOsama528/Data_Architects/tree/main/Datasets>