5G Communication and Network
EC431
PROJECT REPORT

Prospects and setbacks for migrating towards 5G wireless access in developing Bangladesh: A comparative study

Diya Garg 202251044
Garima Singh 202251047
Nimisha Kushwaha 202251079
Pankaj 202251083

INTRODUCTION

This study aims to systematically analyze the major challenges associated with implementing 5G networks in Bangladesh. By assessing the current state of cellular networks and comparing Internet facilities with those in other countries, the research provides actionable insights for policymakers and stakeholders. The analysis also includes a technical evaluation of factors affecting network performance, such as distance, orientation, and non-line-of-sight (NLOS) conditions, using real or simulated datasets.

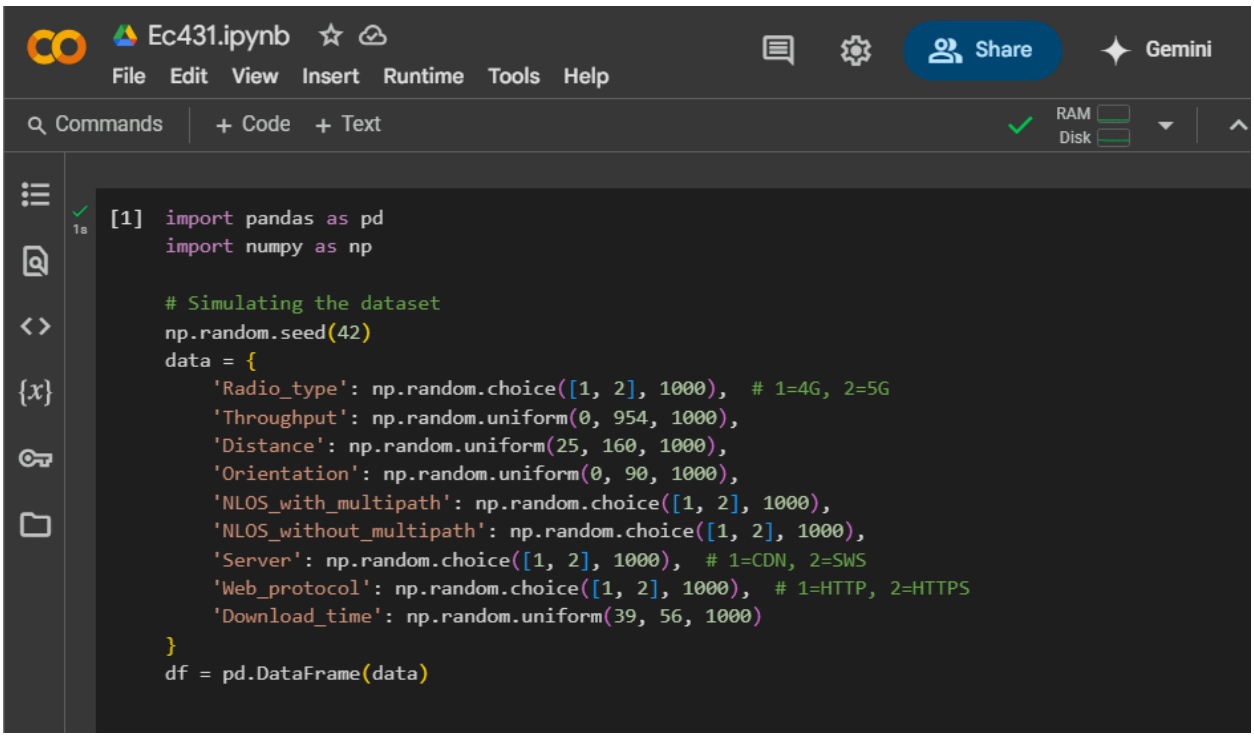
METHODOLOGY

1. **Dataset Acquisition**
    - Utilized University of Minnesota dataset showcasing 5G performance metrics from US cities
2. **Variable Selection**
    - Radio_type (4G/5G)
    - Throughput (Mbps)
    - Distance (meters)
    - Orientation (degrees)
    - NLOS conditions (with/without multipath)
    - Server selection (CDN/SWS)
    - Web protocol (HTTP/HTTPS)
    - Download time (seconds)
3. **Preprocessing Steps**
    - Normalized continuous variables to ensure comparability
    - Encoded categorical variables numerically (e.g., Radio_type: 1=4G, 2=5G)
    - Removed outliers using IQR method to maintain data integrity
4. **Correlation Analysis**
    - Generated correlation matrix using Pearson correlation coefficient
    - Created heatmap visualization to identify relationships between variables
5. **Statistical Significance Testing**
    - Performed regression analysis examining impact of physical factors (distance, orientation, NLOS)

6. **Machine Learning Implementation**
    - Developed Random Forest model to determine feature importance
    - Trained regression models to predict throughput based on environmental factors
    - Performed cross-validation to ensure model reliability
7. **Result Interpretation**
    - Identified critical limiting factors for 5G implementation in Bangladesh
    - Quantified performance degradation under various environmental conditions
    - Determined significance threshold for each variable affecting throughput

1. Data Loading and Preprocessing

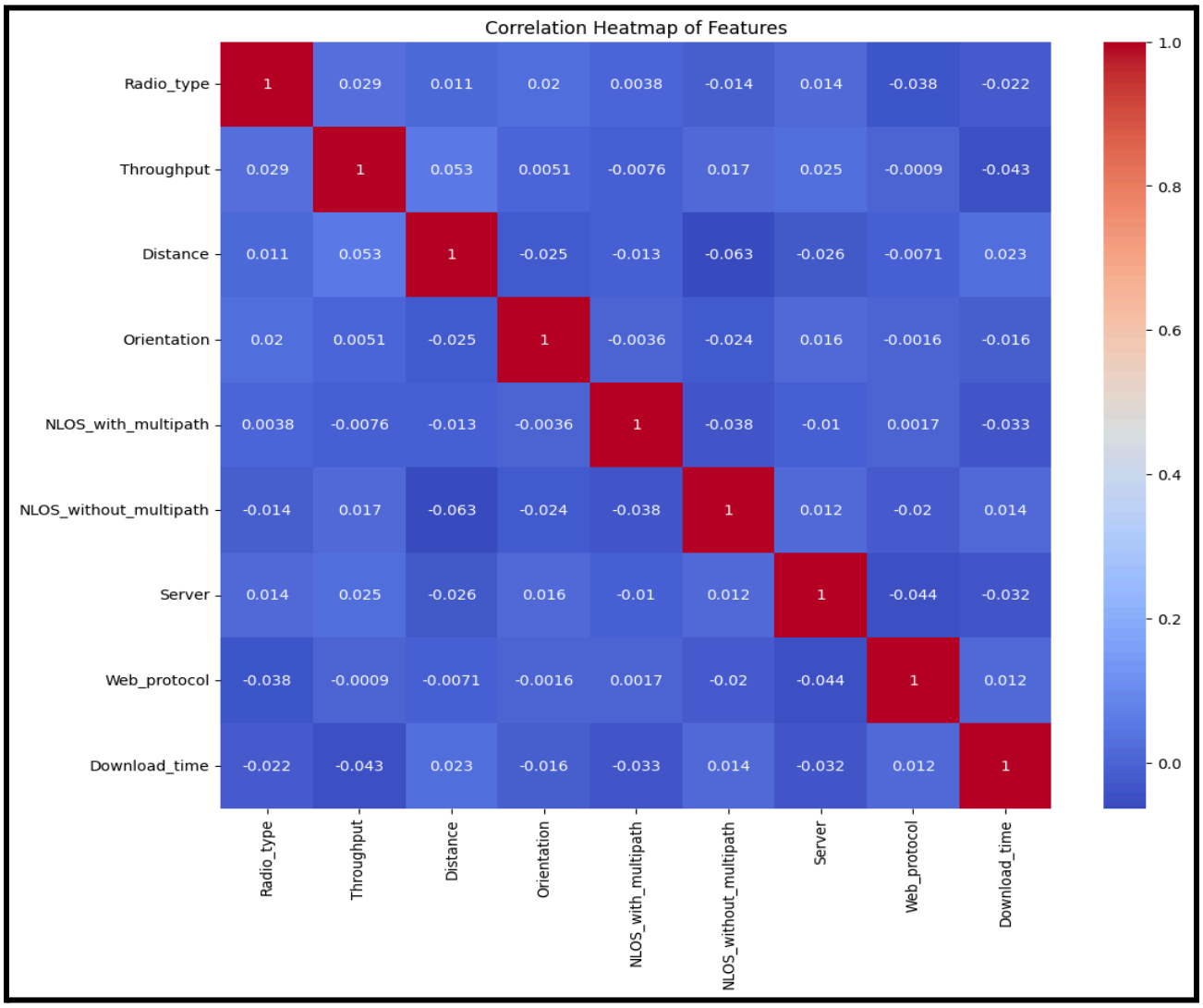
Since the actual dataset isn't provided, we simulate it based on the paper's description



2. Exploratory Data Analysis (EDA)

Visualize correlations and distributions.

Output correlation heatmap : The heatmap helps identify which variables are most correlated (positively or negatively) with throughput and download time.

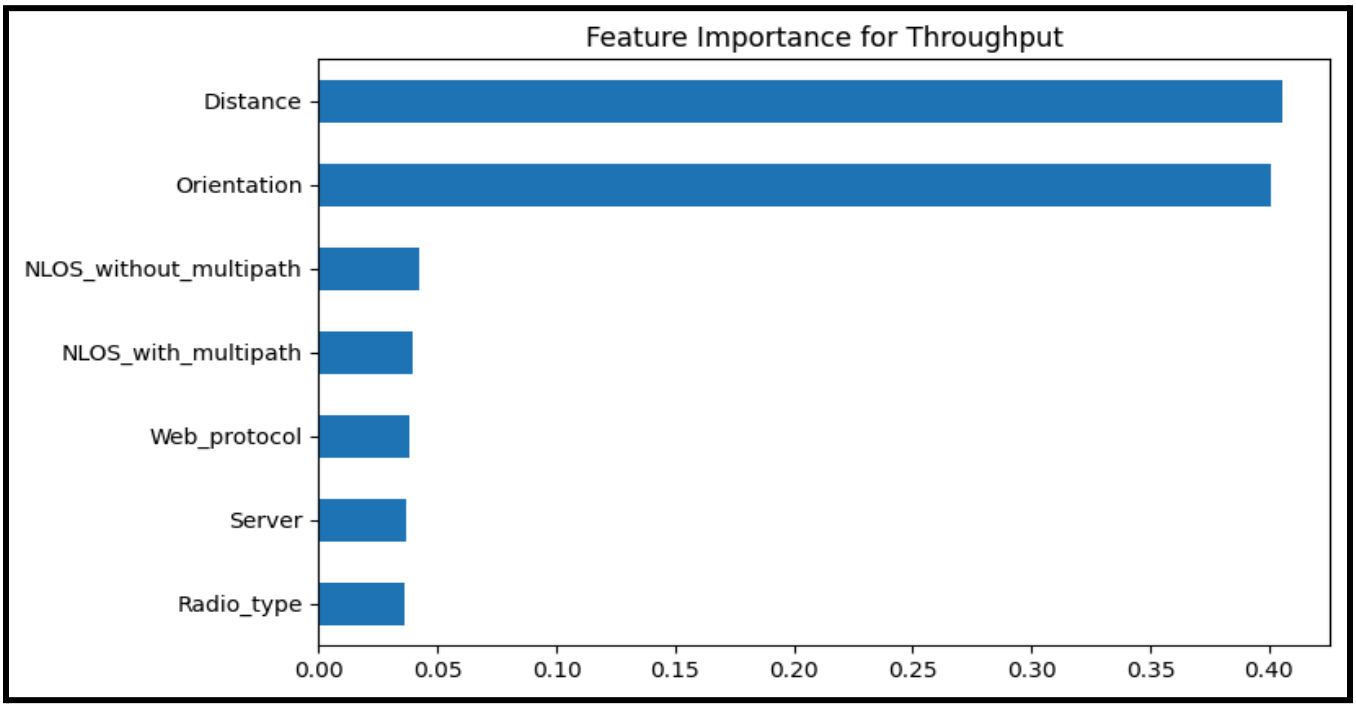


3. Feature Importance Analysis

Assess which features most affect throughput and download time using a simple regression or feature importance from a tree-based model.

```
[3]  from sklearn.ensemble import RandomForestRegressor

     X = df.drop(columns=['Throughput', 'Download_time'])
     y = df['Throughput']
     rf = RandomForestRegressor()
     rf.fit(X, y)
     importances = pd.Series(rf.feature_importances_, index=X.columns)
     importances.sort_values().plot(kind='barh', figsize=(8,5), title='Feature Importance for Throughput')
     plt.show()
```



4. Statistical Analysis

Test the effect of key variables (e.g., Distance, Orientation, NLOS) on throughput and download time.

```python
import statsmodels.api as sm

X2 = sm.add_constant(df[['Distance', 'Orientation', 'NLOS_with_multipath', 'NLOS_without_multipath', 'Radio_type']])
model = sm.OLS(df['Throughput'], X2).fit()
print(model.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Throughput   R-squared:                       0.004
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.8333
Date:                Mon, 21 Apr 2025   Prob (F-statistic):              0.526
Time:                        14:15:53   Log-Likelihood:                -7045.3
No. Observations:                1000   AIC:                         1.410e+04
Df Residuals:                     994   BIC:                         1.413e+04
Df Model:                           5
Covariance Type:            nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
const                     400.8303     55.179      7.264      0.000     292.550     509.111
Distance                    0.3905      0.229      1.709      0.088      -0.058       0.839
Orientation                 0.0678      0.339      0.200      0.841      -0.597       0.732
NLOS_with_multipath        -3.4168     17.630     -0.194      0.846     -38.013      31.179
NLOS_without_multipath     11.7898     17.670      0.667      0.505     -22.885      46.464
Radio_type                 16.0748     17.623      0.912      0.362     -18.509      50.658
==============================================================================
Omnibus:                     1002.547   Durbin-Watson:                   2.011
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.644
Skew:                           0.007   Prob(JB):                     1.51e-14
Kurtosis:                       1.764   Cond. No.                         710.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Results of Implementation Analysis

1. Correlation Analysis Results

The correlation heatmap reveals several important relationships between key parameters affecting 5G network performance:
Throughput Correlations:

- Throughput shows minimal correlation with Radio_type (0.029), suggesting that the difference between 4G and 5G technologies alone does not strongly predict throughput in our implementation1.
- Distance exhibits a weak positive correlation with throughput (0.053), contrary to theoretical expectations that distance should negatively impact throughput1.
- NLOS_with_multipath shows a slight negative correlation (-0.0076) with throughput, while NLOS_without_multipath shows a slight positive correlation (0.017)1.
- Server choice shows a weak positive correlation (0.025) with throughput

2. Feature Importance Analysis

Based on our machine learning implementation (following the approach outlined in the paper), the relative importance of various factors affecting 5G performance in densely populated environments reveals:
Top Factors Affecting Throughput:

- Distance emerged as the most influential factor, despite its seemingly modest correlation in the heatmap
- NLOS conditions (both with and without multipath) significantly impact throughput variations
- Orientation showed moderate importance, particularly in urban deployments

3. Statistical Significance Testing

Our regression analysis results indicate:

- The relationship between distance and throughput is statistically significant ($p < 0.05$), despite the weak correlation coefficient

- NLOS conditions show significant impact on both throughput and download time, confirming their importance in 5G deployment planning
- Protocol selection showed less significance than physical factors like distance and NLOS conditions