Real time business use case: Predicting Network Latency for the Indian Telecom Sector

Business Context:

Network latency is a critical performance metric in telecom networks, representing the time delay experienced by data packets traveling between source and destination. High latency can degrade customer experience, affect video streaming quality, increase call drops, and reduce VoIP (Voice over Internet Protocol) efficiency.

Telecom operators in India need to predict network latency based on various infrastructure and network parameters to optimize performance, manage resources efficiently, and enhance customer satisfaction.

Problem Statement:

Develop a regression-based predictive model to estimate network latency in an Indian telecom network based on key influencing factors such as data usage, congestion levels, number of active users per tower, call duration, signal strength, time of day, and geographical factors.

Features in the Dataset:

- Numerical Features:
    - avg_call_duration (in seconds)
    - total_data_consumed (MB per session)
    - num_connected_towers (per user session)
    - network_congestion_level (scale 0-100)
    - signal_strength (dBm, ranging from -120 to -50)
    - num_active_users_per_tower (count)
    - download_speed (Mbps)
    - upload_speed (Mbps)
    - time_spent_on_network (minutes)
    - distance_to_nearest_tower (meters)

- Categorical Features:
    - network_type (2G, 3G, 4G, 5G)
    - tower_type (Macro, Micro, Pico, Small Cell)
    - device_type (Feature Phone, Smartphone, Tablet, IoT Device)

- Datetime-Based Features:
    - timestamp (DateTime)
    - day_of_week (Monday-Sunday)
    - hour_of_day (0-23)

- Target Variable:
  - network_latency (milliseconds)

Expectations from the participants:

1. Data Loading & Initial Inspection

- Load the dataset to check the structure, column names, and data types.
- Display the first few rows to get an overview.
- Check for missing values, duplicate entries, and data inconsistencies.

2. Exploratory Data Analysis (EDA)

- Identify missing values and decide on an appropriate handling method (imputation, deletion, etc.).
- Detect and handle outliers using statistical methods (e.g., Z-score, IQR method).
- Analyze distributions of numerical and categorical features using histograms, boxplots, and count plots.
- Perform correlation analysis using a heatmap to understand feature relationships.

3. Data Preprocessing

- Handle missing values (impute, drop, or fill based on business logic).
- Remove or adjust outliers if necessary.
- Encode categorical variables using label encoding or one-hot encoding.
- Scale numerical features if required (e.g., Min-Max Scaling or Standardization).
- Perform feature engineering if applicable.

4. Data Visualization & Business Insights

- Generate visualizations to extract business-relevant insights.
  - Histograms, scatter plots, boxplots, and line graphs for trends.
  - Heatmaps to analyze correlations between variables.
  - Time-series analysis (if applicable).

- Interpret insights to provide value for decision-making in the telecom industry.

5. Regression Model Building

- Split the dataset into training and testing sets.

- Train multiple regression models:
  - Linear Regression
  - K-Nearest Neighbors (KNN) Regressor

- Decision Tree Regressor
    - Random Forest Regressor
    - AdaBoost Regressor
    - Gradient Boosting Regressor

- Evaluate models using:
    - Mean Squared Error (MSE)
    - Root Mean Squared Error (RMSE)
    - R^2 Score

- Compare the model performances to select the best one.

6. Business Implications & Recommendations

- Interpret the model results and explain their relevance in optimizing telecom operations.
- Provide business recommendations on how telecom companies in India can leverage predictive analytics.
- Highlight challenges in real-world deployment and suggest areas for further improvement.

Aashutosh Kumar Mishra
Corporate Trainer || AI Architect
assimilate.ai03@gmail.com