**Network Traffic Prediction in Telecom**

**Problem Statement:**

Network traffic prediction in telecom is an essential aspect of network management, as it helps predict data usage patterns, optimize resources, plan network upgrades, and improve customer satisfaction. By predicting network traffic volume, telecom operators can efficiently allocate resources, reduce congestion, and ensure smoother operations.

In this case, we'll be using a linear regression model to predict the Traffic Volume (GB) based on several network-related features. These features could include signal strength, latency, packet loss, connection type, congestion level, data throughput, and more. Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables.

**Here’s a detailed explanation of each variable:**

timestamp: The date and time of the data point (hourly).

traffic: Network traffic load (in Mbps or GB), our target variable.

signal\_strength: Signal strength measured in dBm (decibels milliwatts).

latency: The delay in data transmission (in milliseconds).

packet\_loss: The percentage of lost packets during transmission.

connection\_type: The type of connection used (e.g., 4G, 5G, Wi-Fi).

network\_congestion: Simulated congestion level on the network, ranging from 0 (no congestion) to 100 (fully congested).

avg\_throughput: The average data throughput (in Mbps) during the observation period.

num\_users: Number of active users on the network.

network\_availability: The percentage of time the network is available (e.g., 99.5% availability).

snr: Signal-to-Noise Ratio (SNR), a higher value indicates better signal quality.

active\_connections: Number of active connections on the network.

weather\_temp: Temperature (in Celsius), simulating how weather could impact signal strength.

wifi\_interference: A measure of Wi-Fi interference on a scale from 0 to 10.

network\_downtime: The number of hours the network is down in the given period.

customer\_satisfaction: Simulated customer satisfaction score on a scale from 0 to 10.

traffic\_volume: The total data volume (in GB) used in the observation period, calculated as a percentage of traffic.

**Data Encoding:**

* The categorical variable Connection Type is one-hot encoded to convert it into numerical features.

**Model Training:**

* A LinearRegression model is trained using the independent features and the target variable.

**Model Evaluation:**

* The model’s performance is evaluated using Mean Squared Error (MSE) and R-squared (R2) scores, which measure the accuracy and fit of the model.

**Results & Output:**

After training, you will get a model that can predict the traffic volume based on the provided features. We can see in below code file all the above.

**Here’s a code file for the prediction:**

