PUBLIC TRANSPORTATION OPTIMIZATION

Problem Overview :

Our city's public transportation system faces challenges including inefficient routes, unreliable schedules, peak-hour overcrowding, limited accessibility for disabled passengers, environmental concerns, and financial sustainability. To address these issues, we're committed to optimizing the system. This involves redesigning routes, improving schedules, investing in modern vehicles and infrastructure, integrating IoT and real-time tracking technologies, and fostering stakeholder collaboration. Our goal is a user-friendly, eco-friendly, and financially sustainable public transportation system that supports our city's growth and enhances the daily lives of our residents.

Proposed Idea for the Problem :

Our idea is to implement a Machine Learning model with IoT for the prediction of arrival of public transports efficiently .Predicting the arrival of buses helps passengers to plan their journeys efficiently and the time spent for waiting for buses is reduced . We have an idea of implementing **Linear Regression** ML Model along with Internet of Things(IoT) for prediction .

Linear Regression :

In the context of predicting the arrival time of buses or other transportation modes, linear regression can be defined as a statistical modeling technique used to estimate and predict the expected arrival time of a vehicle based on one or more predictor variables such as historical travel time, distance to the destination, time of day, traffic conditions, and other relevant factors.

 The goal of linear regression, in this case, is to establish a linear relationship between the predictor variables and the expected arrival time, allowing for the estimation of arrival times for future trips or journeys based on the values of these predictor variables.

Working of ML model :

1.Data Collection:

        Gather historical data that includes information about bus arrival times and relevant predictor variables. Predictor variables can include factors such as:

> ➤ Time of day, day of the week .

➢ Weather conditions (e.g., temperature, precipitation) .
➢ Traffic conditions (e.g., congestion level) .
➢ distance from previous stops .
➢ Historical arrival times for the same route .

2.Data Preprocessing:

Clean and preprocessing the data. This involves handling missing values, encoding categorical variables (if any), and normalizing or scaling numerical features. Additionally, you may need to convert timestamps into a numerical format that the model can use.

3.Feature Selection:

Select the most relevant predictor variables based on domain knowledge and feature importance analysis. The chosen features will be used as input to the linear regression model. Data Splitting: Split the dataset into a training set and a testing set. The training set is used to train the linear regression model, while the testing set is used to evaluate its performance.

4.Model Training:

Train the Linear Regression model using the training data. The model will learn the linear relationship between the selected predictor variables and the bus arrival times.

5. Model Evaluation:

Use the testing dataset to evaluate the model's performance. Common evaluation metrics for regression tasks include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics quantify how closely the predicted arrival times match the actual arrival times.

6.Prediction:

Once the model is trained and evaluated, you can use it to make predictions on new, unseen data. Provide the selected predictor variables for the current situation, and the model will output a predicted arrival time.

7.Monitoring and Maintenance:

Continuously monitor the model's performance and update it as needed. Factors such as changes in traffic patterns, weather conditions, or infrastructure may require model retraining or adjustments to maintain accuracy.