xGBoost Regression Model

The xGBoost model in Lab 2 was designed to predict bus arrival delays using inputs such as dwell time, scheduled travel time, upstream stop delay, and recurrent delay as independent variables, with arrival delay as the dependent variable. The data was split into 80% for training and 20% for testing to avoid overfitting and ensure better generalization. The model's performance was evaluated using Mean Squared Error (MSE), R-squared (R^2), and Mean Absolute Error (MAE). The model achieved an R-squared of 0.99, indicating a high accuracy (99%) in predicting arrival delays. Additionally, the MAE and MSE were 11.36 and 240.9, respectively, highlighting strong predictive performance.

After performing a grid search, the best hyperparameters were identified as follows: colsample_bytree=0.5, gamma=0, learning_rate=0.1, max_depth=3, n_estimators=200, reg_alpha=0, and reg_lambda=1.5. These hyperparameters were tuned to optimize the model's performance and prevent overfitting. xGBoost could also be applied to other transportation problems, such as travel time prediction, route performance analysis, or traffic congestion forecasting, where accurate prediction is crucial for efficient transportation planning and management.