Forecasting of Smart City Traffic Patterns

Introduction

This report presents an analysis of smart city traffic patterns using Python and various machine learning techniques. The dataset used for this analysis includes both training and testing data, and the goal is to predict the number of vehicles at different junctions based on various features.

Data Exploration

Training Data

- The training dataset consists of:
  - Rows: 48120
  - Columns: 8
- Data Types:
  - Numeric Columns: Year, Month, Date_no, Hour
  - Categorical Column: Day
  - Datetime Column: DateTime
- Missing Values:
  - No missing values in the training dataset.

Testing Data

- The testing dataset consists of:
- Rows: [Number of rows in your test dataset] • Columns: [Number of columns in your test dataset]
- Data Types:
- [Specify data types in your testing dataset]
- Missing Values:
- No missing values in the testing dataset.

Data Preprocessing

- Duplicates Removal:
- Duplicates were removed from the training dataset.
- Datetime Conversion:
- The 'DateTime' column was converted from string to datetime format.
- Feature Engineering:
- Additional features were created: Year, Month, Date_no, Hour, Day.

Data Visualization

Time Series Plot

- A time series plot was created to visualize the traffic patterns over time, with different colors representing different junctions.

Traffic Over the Years

- A countplot was generated to display the number of vehicles at different junctions over the years.

Correlation Heatmap

- A correlation heatmap was used to visualize the correlations between numerical features in the dataset.

Model Building and Evaluation

LightGBM Model

- LightGBM regression model was trained with the following parameters:
- Objective: Regression
- Metric: RMSE (Root Mean Squared Error)
- Model Evaluation:
- Mean Squared Error: [MSE value]
- Mean Absolute Error: [MAE value]
- R2 Score: [R2 score value]

Random Forest Model

- Random Forest regression model was trained with 100 estimators and a random seed of 42.
- Model Evaluation:
- Mean Squared Error: [MSE value]
- Mean Absolute Error: [MAE value]
- R2 Score: [R2 score value]

Conclusion

- Both LightGBM and Random Forest models were trained and evaluated for predicting traffic patterns.
- [Include any observations, insights, or comparisons between the models.]
- [Discuss any potential areas for further improvement or analysis.]

Please note that you should replace the placeholders such as [Number of rows in your test dataset], [Number of columns in your test dataset], [Specify data types in your testing dataset], [MSE value], [MAE value], [R2 score value], and any other relevant information with actual values and observations from your analysis.