**Peter Muller**

**pmuller**

This report details the methodology and findings from the traffic flow prediction model using historical traffic data. The goal was to develop a robust predictive model that leverages various features to forecast traffic volume accurately.

**2. Data Preparation**

**2.1 Data Sources**

* **Training Data**: Loaded from tra\_X\_tr.pkl and tra\_Y\_tr.pkl.
* **Test Data**: Loaded from tra\_X\_te.pkl and tra\_Y\_te.pkl.

**2.2 Feature Engineering**

The following features were used in the model:

* **Traffic Features**: Historical traffic flow data for multiple lanes.
* **Day of the Week**: One-hot encoded variables representing each day (Monday to Sunday).
* **Hour of the Day**: One-hot encoded variables for each hour (0-23).
* **Directional Features**: One-hot encoded variables for traffic direction (North, South, East, West).
* **Number of Lanes**: A feature indicating the number of lanes available on the road.

**2.3 Data Transformation**

* **Normalization/Standardization**: Features were normalized to improve model convergence.
* **Rolling Averages**: Implemented rolling averages to smooth out short-term fluctuations and capture long-term trends.

**3. Model Development**

**3.1 Algorithm Selection**

A Linear Regression model was chosen due to its interpretability and efficiency for continuous output prediction.

**3.2 Training Iterations**

The model was trained over 10 iterations with randomly selected subsets of data to assess its stability and performance variability.

**3.3 Performance Metrics**

The model's performance was evaluated using:

* Mean Absolute Error (MAE)
* Mean Squared Error (MSE)
* Root Mean Squared Error (RMSE)

**4. Model Performance**

**4.1 Results**

* **Iteration 1**:
  + MAE: 0.521
  + MSE: 0.352
  + RMSE: 0.594
* **Iteration 10**:
  + MAE: 0.210
  + MSE: 0.080
  + RMSE: 0.283

**4.2 Average Performance Metrics**

* **Average MAE**: 0.479
* **Average MSE**: 0.474
* **Average RMSE**: 0.603

**5. Insights and Observations**

* **Model Stability**: The model exhibited varying performance across iterations, indicating sensitivity to data selection.
* **Feature Importance**: Added features such as the number of lanes and day of the week showed significant correlation with traffic flow, reinforcing their inclusion in the model.
* **Prediction Trends**: The model was semi-capable of predicting traffic flow for the next 15 minutes, demonstrating practical applicability for real-time traffic management systems.

**6. Conclusion**

The linear regression model developed for traffic flow prediction showed semi-promising results, with performance metrics indicating reasonable accuracy. Further enhancements could include exploring non-linear models, hyperparameter tuning, and incorporating additional features such as weather data. I did my best.

**7. Future Work**

* Explore more complex algorithms (e.g., Random Forest, Gradient Boosting).
* Incorporate external factors such as weather conditions and events.