## 1. Project Overview:

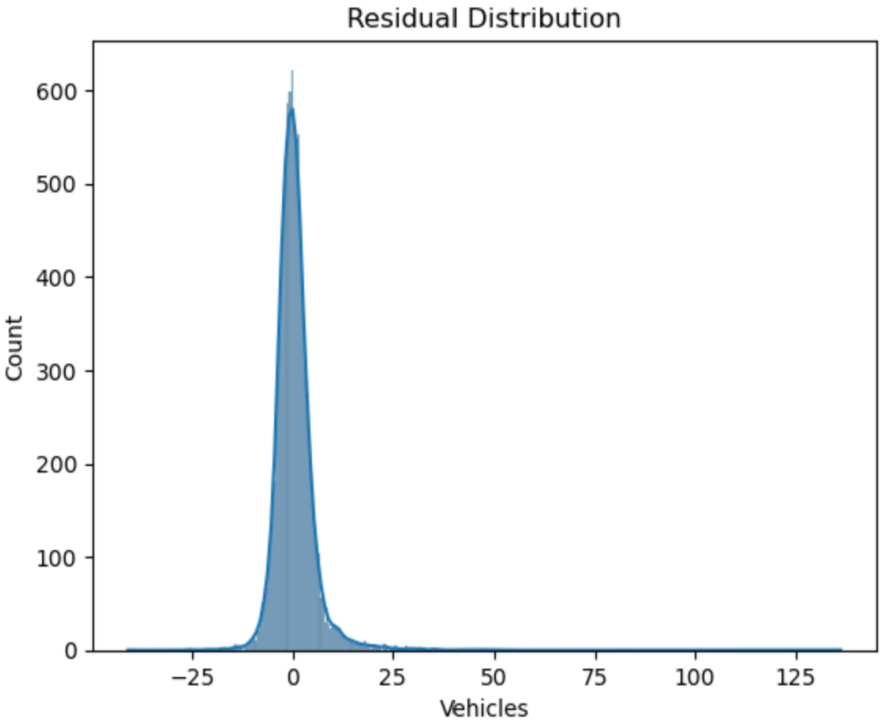
- Develop a predictive model to accurately forecast hourly traffic volumes at different road junctions based on historical traffic data.

## 2. Model Development and Training:

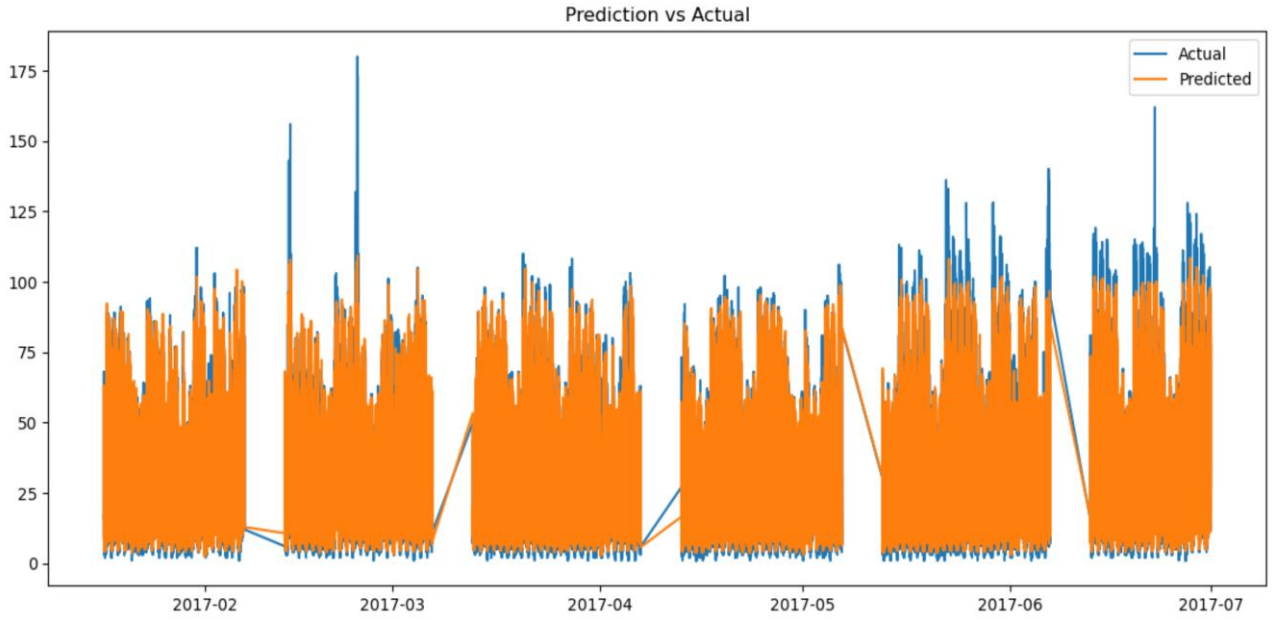
- XGBoost and ARIMA models are used with time – based splitting of the data with 70 % training and 30% testing.

## 3. Model Evaluation:

- Evaluation metrics used are Root Mean Square Error, Mean Absolute Error and R2 score.
- RMSE – 5.20, MAE – 3.15, R2 – 0.96 for XGBoost and RMSE ~ 26 for ARIMA
- These values suggest the model is capturing patterns in the data very well, with minimal prediction error and strong correlation to actual values.
- The residual distribution was centred near zero, with no strong skew or multi-modality.




- The line plot showed close alignment between predicted and actual values over time, implying strong time-series relevance.




## 4. Cross-Validation Strategy:

- Temporal Generalization Is Strong
- Timeseries Split MAE scores across 5 folds were:
  [2.64, 2.68, 3.01, 3.42, 3.14]
- Average MAE: 2.98
- Variation between folds is small, showing model stability over different time windows, which is critical in forecasting tasks.

## 5. Model Refinement:

- Hyperparameter tuning performed using GridSearchCV over the following grid:
- - n_estimators: [100, 200]
- - max_depth: [3, 5]
- - learning_rate: [0.05, 0.1]
- Best Parameters:
  - {'n_estimators': 200, 'max_depth': 3, 'learning_rate': 0.05, 'subsample': 1.0, 'colsample_bytree': 0.9}
- These values led to the lowest validation RMSE.

## 6. Insights:

- The model benefits from shallow trees and a conservative learning rate.
- Time-based cross-validation effectively tested generalization.
- XGBoost outperformed ARIMA by a large margin.