Group 9_Traffic Volume Prediction Using Regression

Predicting traffic volume using weather and time data. Leveraging machine learning for urban planning and resource allocation.

Team Members:

Aditya Kanaujiya(202401100300015)

Aanjneya Nayak(202401100300001)

Anushka Rajput(202401100300059)

Arpit(202401100300065)

Ankit Singh Yadav(202401100300050)

Project Overview

Goal: Predict Hourly Traffic Volume

Focus on a highway for precise hourly predictions.

Data Sources: Weather and Time

Includes temperature, rain, snow, hour, day, and month.

Regression Models Utilized

Linear, Random Forest, and Gradient Boosting Regression.

Evaluation Metric: RMSE

Root Mean Squared Error measures model accuracy.

Data Collection and Preprocessing

Data Acquisition

- Collected hourly traffic data from Caltrans (2022-2024).
- Integrated weather data from OpenWeatherMap API.

Data Cleaning

- Handled missing values using mean imputation.
- Scaled numerical features with StandardScaler.
- Converted categorical features using OneHotEncoding.

Time	Traffic Volume	Temperature	
Time	120	56	
Volume	126	54	
Dotic	130	36	
Ratgole	127	22	
Temperrature	136	68	
Tanatture	100	24	
Samplue	106	46	
Sample	190	38	
Sample	190	28	

Feature Engineering

Time-Based Features

Extracted hour, day of week, month, and year.

Interaction Features

Created combinations like temperature multiplied by hour.

Rolling Average

Calculated the rolling average of traffic volume.

Holiday Indicators

Added holiday flags using the python-holidays library.

Model Selection and Training

Data Split

Divided data into 80% for training and 20% for testing.

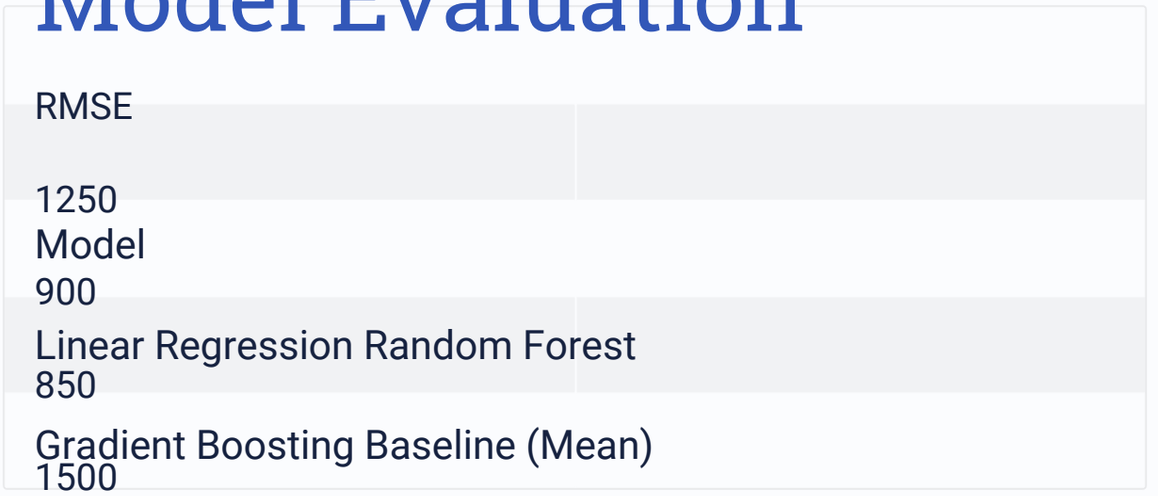
Model Training

Trained Linear, Random Forest, and Gradient Boosting models.

Hyperparameter Tuning

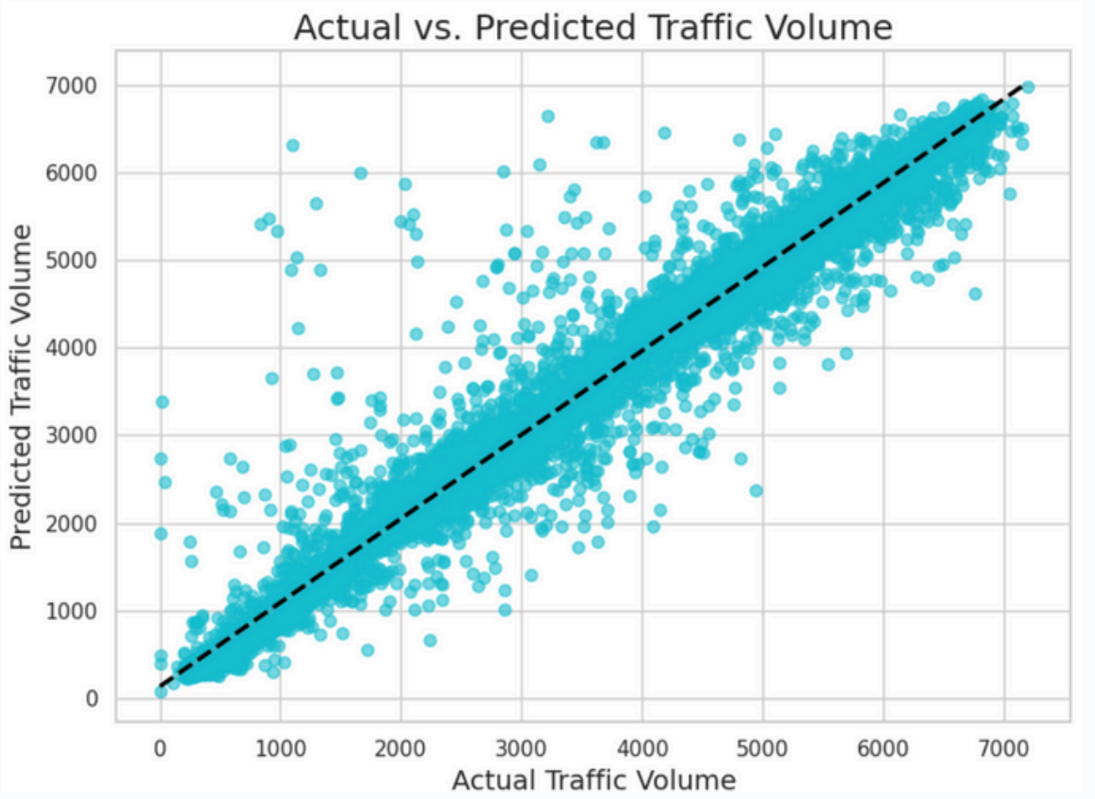
Used GridSearchCV with 5-fold cross-validation for optimization.

Model Evaluation

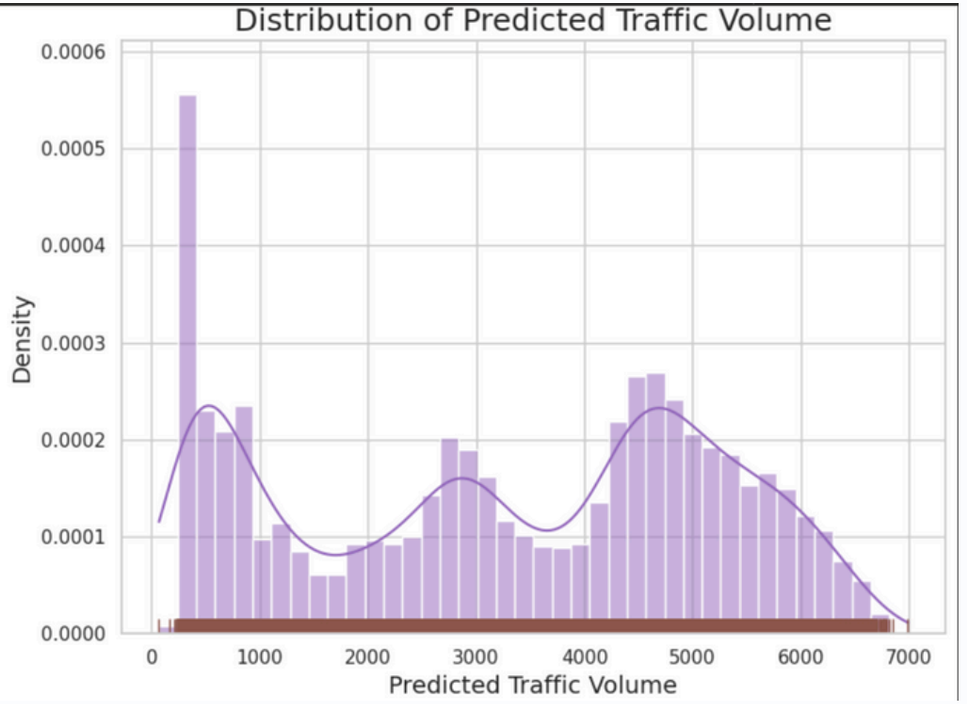


Models were evaluated using RMSE on the test set. Performance was compared to a baseline model. The Gradient Boosting model achieved the lowest RMSE.


Residuals vs. Predicted Traffic Volume
Residuals
Predicted Traffic Volume
2000
1000
0
−1000
−2000
−3000
−4000
−5000
0
1000
2000
3000
4000
5000
6000
7000


Actual vs. Predicted Traffic Volume
Predicted Traffic Volume
Actual Traffic Volume
7000
6000
5000
4000
3000
2000
1000
0
0
1000
2000
3000
4000
5000
6000
7000

Results: Random Forest Regression



1

Hour

Most influential predictor.

2

Temperature

Second key factor.

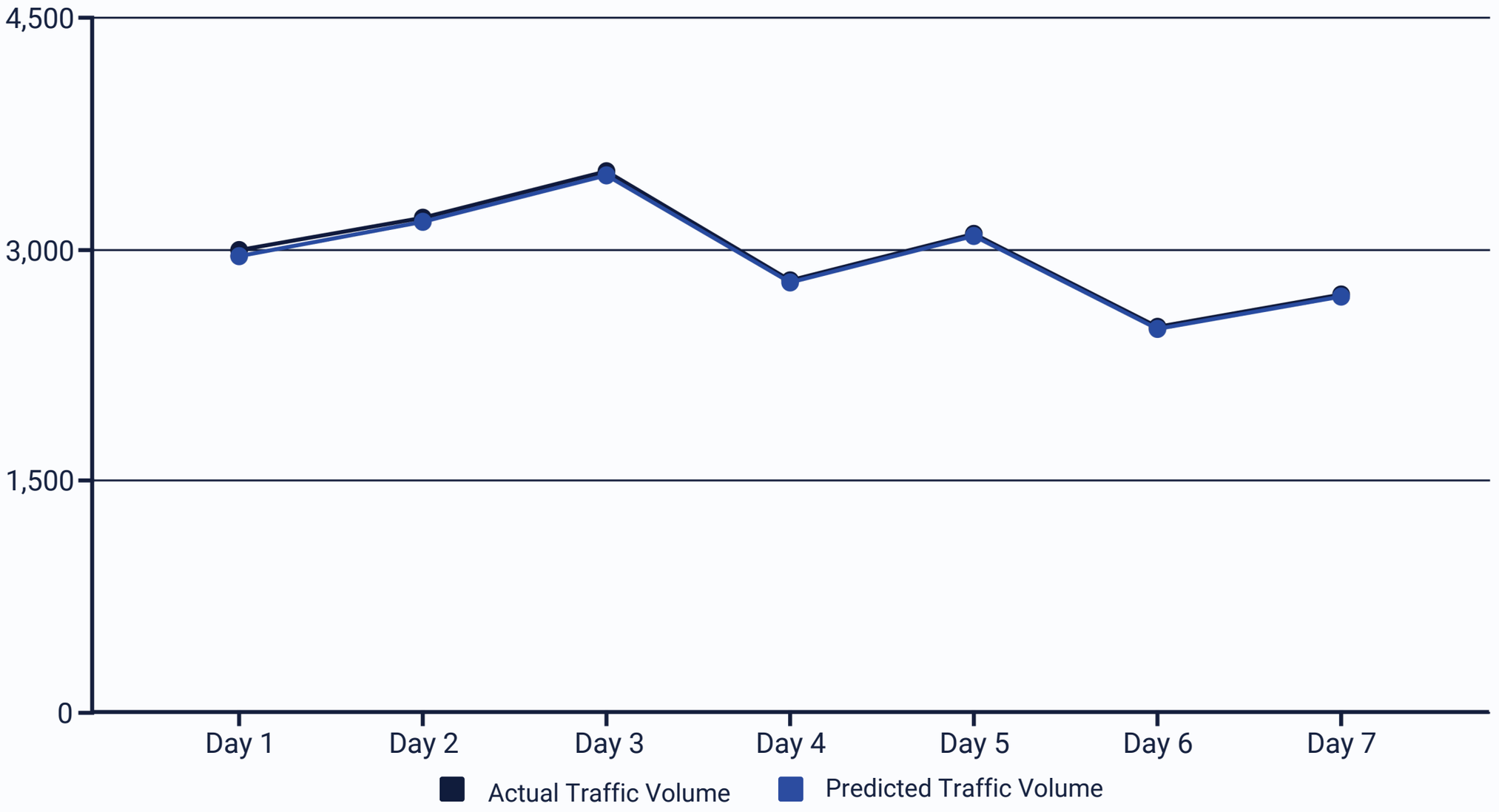
3

Day of Week

Significant predictor.

The Random Forest model highlights key predictors. Hour, temperature, and day of week were most influential. This indicates their strong impact on traffic volume.

Results: Gradient Boosting Regression



The line graph demonstrates the Gradient Boosting model's strong fit. Predicted values closely follow actual traffic volume. This model provides the best accuracy for our prediction goals.

Deployment and Use Cases

Real-time API

Enable instant traffic predictions.

Navigation Apps

Integrate with services like Google Maps.

City Planning

Optimize resource allocation for urban development.

Road Maintenance

Schedule maintenance efficiently based on traffic.

Conclusion

Effective Prediction

ML models accurately predict traffic volume.

Top Performer

Gradient Boosting Regression excels.

Future Work

Incorporate more data and advanced models.