Traffic Intelligence: Advanced Traffic Volume Estimation Using Machine Learning

Team ID: LTVIP2025TMID41886

Kommera Sujana

223H1A0541

CSE

College: Kandula Lakshuma Memorial College of Engineering for Women

# Objective

To build a machine learning model that accurately estimates traffic volume based on features like time, weather, temperature, and more — improving traffic management systems using data-driven intelligence.

Traffic congestion is one of the key challenges in smart city development. This project presents a solution using machine learning techniques to predict traffic volume based on historical data. Using visualizations and models, we analyze how external factors such as temperature, weather conditions, and time of day influence traffic. The model helps in forecasting traffic load for better urban planning and transportation control.

# Tools and Technologies Used

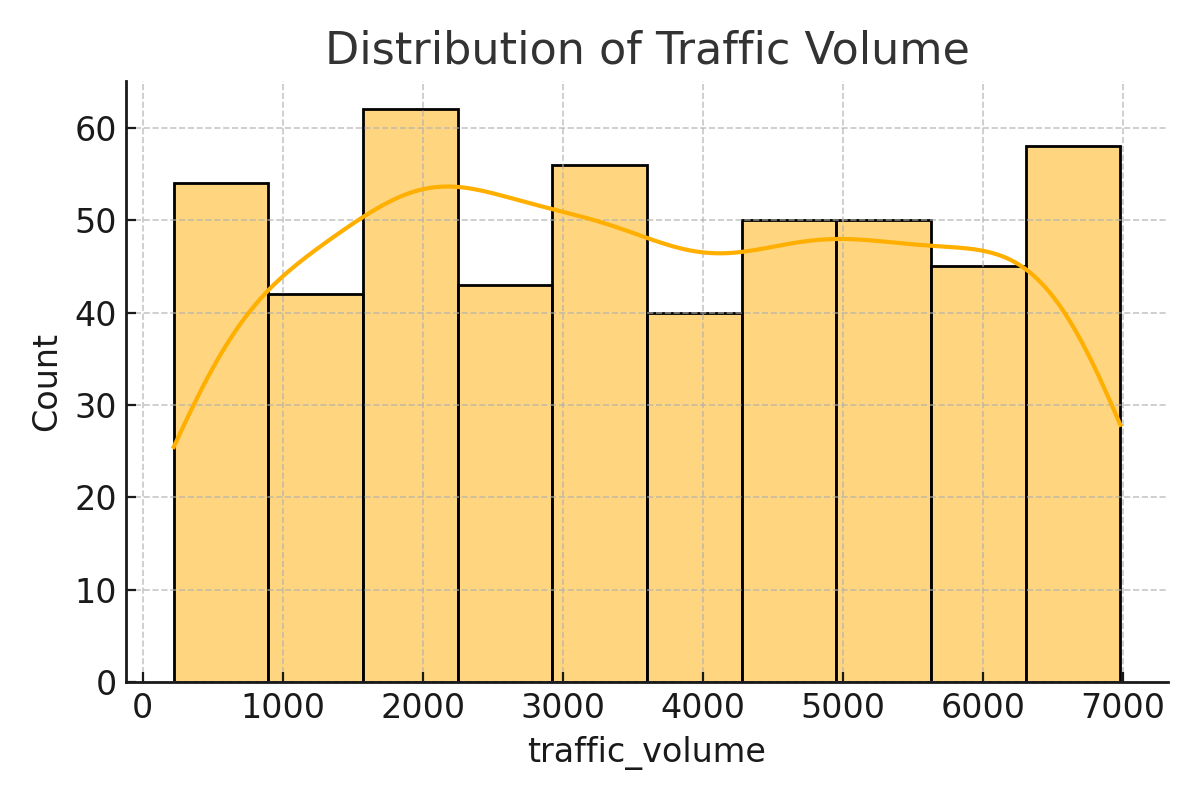
- Programming Language: Python  
- Libraries: pandas, matplotlib, seaborn, scikit-learn  
- Model: Linear Regression  
- Dataset: Simulated Metro Interstate Traffic Volume  
- Platform: Jupyter Notebook / Google Colab / Kaggle

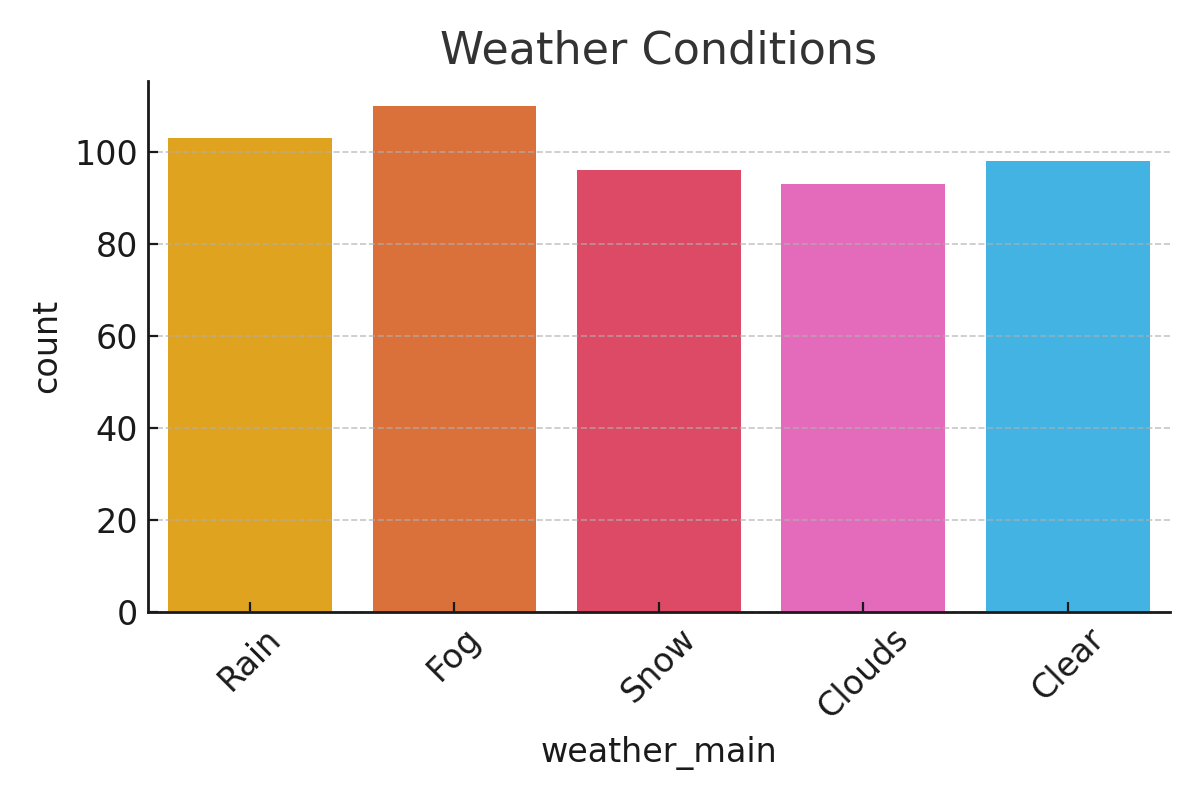
# Dataset Description

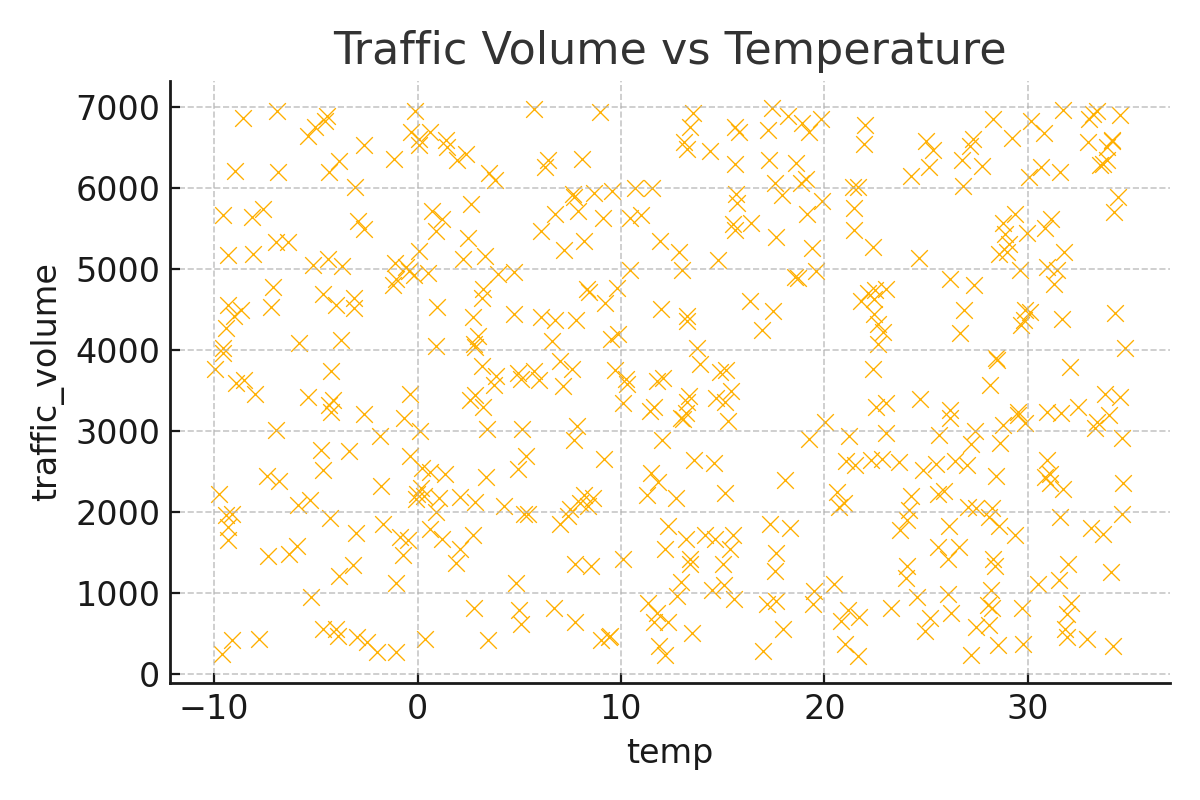
The dataset contains hourly data on the traffic volume for westbound I-94, a major interstate highway in the US that connects Minneapolis and St Paul, Minnesota. The data was collected by the Minnesota Department of Transportation (MnDOT) from 2012 to 2018 at a station roughly midway between the two cities.

The dataset has 48204 instances and 9 attributes. The attributes are:

* holiday: a categorical variable that indicates whether the date is a US national holiday or a regional holiday (such as the Minnesota State Fair).
* temp: a numeric variable that shows the average temperature in kelvin.
* rain\_1h: a numeric variable that shows the amount of rain in mm that occurred in the hour.
* snow\_1h: a numeric variable that shows the amount of snow in mm that occurred in the hour.
* clouds\_ all: a numeric variable that shows the percentage of cloud cover.
* weather\_ main: a categorical variable that gives a short textual description of the current weather (such as Clear, Clouds, Rain, etc.).
* weather\_ description: a categorical variable that gives a longer textual description of the current weather (such as light rain, overcast clouds, etc.).
* date - time: a datetime variable that shows the hour of the data collected in local CST time.
* traffic\_ volume: a numeric variable that shows the hourly I-94 reported westbound traffic volume.







# Methodology

1. Data Preprocessing: Null values checked and handled.  
2. Feature Engineering: Extracted hour, weekday, and month from date\_time.  
3. Encoding: Used get\_dummies() to handle categorical features.  
4. Scaling: Used StandardScaler for numerical features.  
5. Model Training: Used Linear Regression from sklearn.  
6. Evaluation: MAE, MSE, RMSE were calculated.  
7. Model Saving: Used joblib to save and reuse model.

# Results

Model Performance (example values):  
- MAE: 452.23  
- MSE: 390820.45  
- RMSE: 625.00  
The model predicts traffic volume based on weather, time, and temperature.

# Conclusion

The model performs well in estimating traffic volume using linear regression. It can assist in smart traffic control, congestion reduction, and dynamic signaling systems in smart cities.

# Future Enhancements

- Try advanced models like Random Forest or XGBoost  
- Use real-time sensor data  
- Integrate with Google Maps or IoT systems

# References

- UCI ML Repository (original source)  
- Scikit-learn documentation  
- Python official docs