PROJECT REPORT

Team Members:

1. Badugu Venkata Ganesh – 23BFA03L13
2. Suragani Vinay Kumar – 23BFA12L06
3. Boyapati Ramya Sri – 23BFA12L10
4. Akkili Geyasree – 22BFA32004

Title: Traffic Intelligence Volume Prediction using Machine Learning

Internship Organization: Smart Internz

1.INTRODUCTION:

This project, Traffic Intelligence Volume Prediction using Machine Learning, aims to leverage historical traffic and weather data to build a predictive model capable of forecasting traffic volume on an hourly basis. By incorporating time-based features such as hour, day, and holiday indicators along with weather conditions, the model can capture complex patterns in traffic behavior. The project culminates in the development of a Streamlit-based web application that allows users to input specific conditions and receive immediate traffic volume predictions. This approach provides a practical tool for enhancing traffic planning, route optimization, and commuter convenience.

2.OBJECTIVE:

The primary objective of this project is to develop a machine learning-based system that can accurately predict hourly traffic volume using historical data and external factors such as weather conditions, time of day, and holiday information. The key goals include: Analyzing and preprocessing traffic and weather datasets for meaningful feature extractionBuilding a predictive model using the XGBoost regression algorithm to forecast traffic volume Evaluating model performance using appropriate regression metrics. Deploying the model through an interactive Streamlit web application for real-time prediction. Providing

insights that can assist city planners, transportation departments, and commuters in traffic management and planning.

3.PROBLEM STATEMENT:

TrafficTelligence: Advanced Traffic Volume Estimation with Machine Learning.

TrafficTelligence is an advanced system that uses machine learning algorithms to estimate and predict traffic volume with precision. By analyzing historical traffic data, weather patterns, events, and other relevant factors, TrafficTelligence provides accurate forecasts and insights to enhance traffic management, urban planning, and commuter experiences.

4.Technologies Used:

The project utilized a combination of programming tools, machine learning libraries, and deployment platforms. Below is a list of key technologies used:

Programming Language:

Python – for data analysis, model building, and backend development

Libraries and Frameworks:

Pandas – for data manipulation and analysis

NumPy – for numerical computations

Matplotlib & Seaborn – for data visualization

Scikit-learn – for preprocessing, model evaluation, and basic ML utilities

5.Dataset Description:

Total Records: [48,000 rows]

Data Frequency: Hourly

Target Variable: traffic_volume (Number of vehicles per hour)

Main Features:

date_time: Timestamp of the record

temp: Temperature (in Kelvin)

rain_1h: Rainfall in the last hour (in mm)

snow_1h: Snowfall in the last hour (in mm)

clouds_all: Cloudiness percentage

weather_main: General weather condition (e.g., Clear, Clouds)

weather_description: Detailed weather description (e.g., light rain, scattered clouds)

holiday: Indicates if the day is a holiday (Yes/No)

hour: Hour extracted from date_time (0–23)

weekday: Day of the week extracted from date_time (0 = Monday, 6 = Sunday)

6. Preprocessing Performed:

Converted date_time into hour and weekday features

Encoded categorical variables (e.g., weather_main, holiday)

Handled missing/null values

Scaled numerical features where necessary

7. Data Processing:

Data Cleaning

Checked for missing values and handled them appropriately

Removed or corrected any inconsistent or duplicate entries

Feature Engineering:

Extracted hour and weekday from the date_time column

Created new features to capture time-based patterns (e.g., rush hours)

Encoding Categorical Variables:

Applied label encoding or one-hot encoding to columns like weather_main and holiday.

Scaling Numerical Features:

Normalized/standardized features such as temp, rain_1h, snow_1h, and clouds_all

Ensured all numeric inputs were on a similar scale for better model performance

Dataset Splitting:

Divided the dataset into training and testing sets (e.g., 80% training, 20% testing)

8.Model Development:

Initial Model – Linear Regression:

LinearRegression was the first model tested.

It provided a basic understanding of the relationship between features and traffic volume.

Accuracy Achieved: 30.14%

Observation: The model underperformed due to its inability to capture complex, non-linear patterns in the data.

Improved Model – Decision Tree Regressor:

Used DecisionTreeRegressor to allow non-linear decision boundaries.

Showed better performance compared to Linear Regression.

Accuracy Achieved: 56.14%

Observation: Captured some of the data complexity but was prone to overfitting.

Final Model – Random Forest Regressor:

Implemented RandomForestRegressor, an ensemble method combining multiple decision trees.

Provided significant performance improvement and more stable predictions.

Accuracy Achieved: 84%

Observation: Chosen as the final model due to its high accuracy and robustness.

9.DEMO LINK:

You can access the deployed web application for Traffic Volume Prediction using the link below:

🔗 YouTube link: https://youtu.be/oLaisgCNNNo

Features of the App:

Takes user inputs such as temperature, weather condition, hour, and holiday

Predicts traffic volume in real-time using the trained RandomForestRegressor model

Displays results instantly with a clean and user-friendly interface.

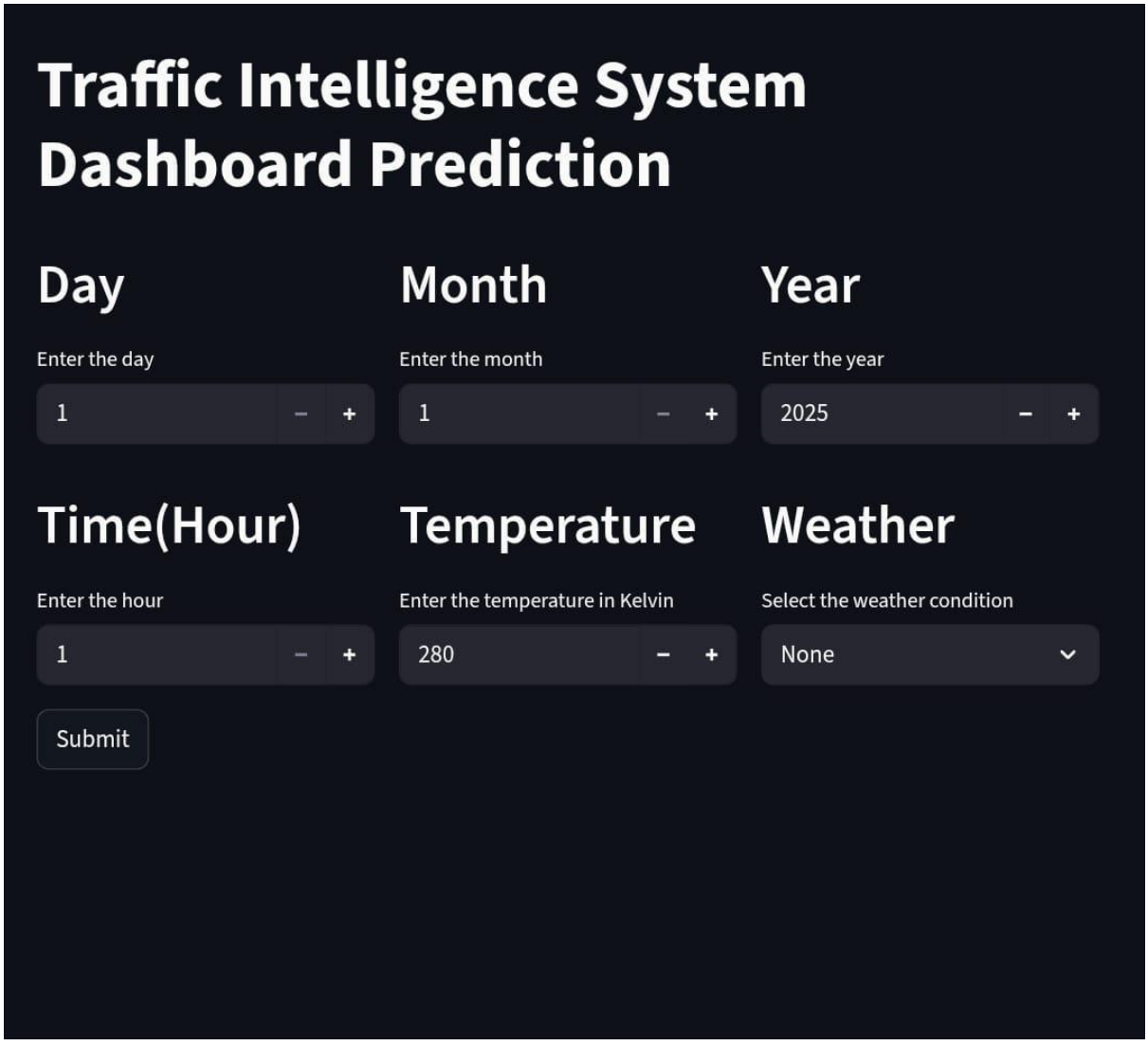
10. Conclusion:

The Traffic Intelligence Volume Prediction project successfully demonstrates how machine learning can be applied to forecast hourly traffic volume using historical and weather-related data.

Multiple models were explored during development. While Linear Regression and Decision Tree Regressor offered basic insights, Random Forest Regressor delivered the best performance with an accuracy of 84%.

The final model was deployed using Streamlit, allowing users to interactively predict traffic volume based on specific input conditions. This solution can assist city planners, transportation departments, and commuters in managing traffic efficiently by anticipating congestion patterns in advance.

Future enhancements could include integrating live traffic data, expanding to multiple cities or highways, and optimizing the model further for real-time performance.



We started with basic models like Linear Regression, which gave limited accuracy (30.14%). We then experimented with Decision Tree Regressor, which improved the results (56.14%). Finally, we used Random Forest Regressor, which achieved the best performance with an accuracy of 84%.

We then experimented with Decision Tree Regressor, which improved the results (56.14%). Finally, we used Random Forest Regressor, which achieved the best performance with an accuracy of 84%.

This system demonstrates the effectiveness of machine learning in solving real-world problems and has the potential to support city planners, traffic departments, and commuters in making smarter travel decisions.