Traffic Volume
Prediction Using Regression

Group Members

Daksh Singh	202401100400073
Deepanshu Bhardwaj	202401100400077
Devendra Kumar Yadav	202401100400079
Navishka Sharma	202401100400123
Preeti Singh	202401100400146

PROBLEM STATEMENT

Goal: Predict traffic volume based on weather and time data

Importance:

Traffic management

Urban planning

Reducing congestion

DATASET OVERVIEW

- Handled missing values
- Converted time stamps to datetime

Extracted time-based features:

Hour, day, weekday, month, season.

- Categorical encoding for holiday, weather description

METHODOLOGY

- **Data Prep**: Loaded data, checked for missing values, scaled features.
- **EDA**: Explored stats, visualized distributions, checked correlations.
- **Feature Selection:** Used correlation and Random Forest importances.
- **Models:** Tried Linear Regression, Random Forest, XGBoost.
- **Training:** Split data (80/20), trained with default settings.
- **Evaluation:** Metrics used – MAE, RMSE, R^2.
- **Feature Importance:** Visualized top features from Random Forest.

EXPLORATORY DATA ANALYSIS

Visualized traffic volume trends by:

- Hour of day
- Weekday vs weekend
- Season and weather conditions

Found strong rush hour effects and weather influence

Add graphs: Line plots, bar charts, boxplots

»» CODE :

```
# Load and preprocess data
df = pd.read_csv('Metro-Interstate-Traffic-Volume-Encoded.csv')
X = df.drop('traffic_volume', axis=1)
y = df['traffic_volume']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Scale and train models
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
models = {'RF': RandomForestRegressor(), 'XGB': XGBRegressor()}
for name, model in models.items(): model.fit(X_train, y_train)
```

MODELS USED

- **Random Forest Regressor**
  - **Works by building multiple decision trees and averaging their predictions.**
  - **Reduces overfitting and improves accuracy using ensemble learning.**

- **XGBoost Regressor**
  - **Uses boosting technique: builds models sequentially to correct previous errors.**
  - **Known for high speed and performance in structured data problems.**

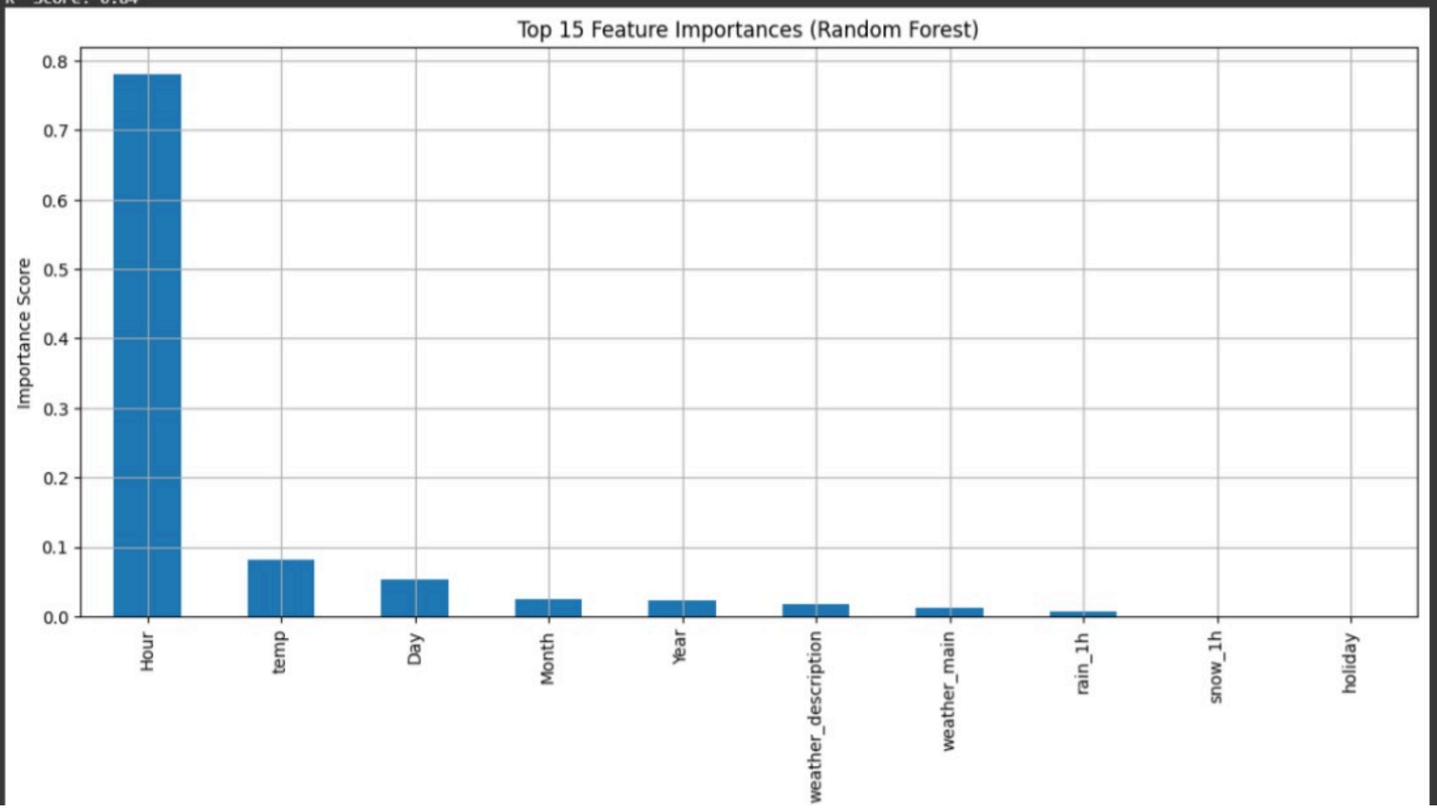
- **Linear Regression**
  - **Linear Regression is a basic machine learning algorithm used to predict a value based on the relationship between variables.It tries to find the best straight line (called a regression line) that fits the data points.**

OUTPUT

```
🔹 Linear Regression Performance:
MAE: 1630.13
RMSE: 1839.22
R² Score: 0.14

🔹 Random Forest Performance:
MAE: 514.15
RMSE: 803.86
R² Score: 0.84

🔹 XGBoost Performance:
MAE: 531.43
RMSE: 798.82
R² Score: 0.84
```



MODEL OUTPUT EXPLANATION

Linear Regression, Random Forest, and XGBoost: we check how well they predicted traffic volume using:

1.MAE (Mean Absolute Error):

- Tells us how far off the predictions are from the real values, on average.
- Smaller MAE means better accuracy

2.(Root Mean Squared Error):

- Similar to MAE but gives more importance to big mistakes.
- Lower RMSE means fewer large errors.

3. R^2 Score (R-squared):

- Shows how well the model explains the changes in traffic volume.
- Closer to 1 means the model fits the data well.

PRESENTATION CONCLUSION

In conclusion, public transportation is an essential component of modern civilization, bringing several benefits such as reducing traffic congestion, lessening environmental impact, and making mobility inexpensive and accessible to all. Despite obstacles such as overpopulation and inadequate infrastructure, continual innovations and investments are paving the way for more efficient, sustainable, and user-friendly systems.

THANK YOU