SmartInternz Project Report

TrafficIntelligence: Advanced Traffic Volume Estimation with Machine Learning

Submitted in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

in

CIVIL ENGINEERING

Submitted by

NAMES	ROLL NO
A PRVEEN KUUMAR	22AK5A0128
Y DATTATRAEYA	22AK5A0108
A JAGADEESWAR	22AK5A0116
M PENCHALAIAH	22AK5A0126

DEPARTMENT OF CIVIL ENGINEERING
ANNAMACHARYA INSTITUTE OF TECHNOLOGY AND SCIENCES, TIRUPATI
(AUTONOMOUS)

Approved by AICTE, New Delhi & Permanent Affiliation to JNTUA, Anantapuramu.
Three B . Tech Programmes (CIVIL, ECE & CSE) are accredited by NBA, New Delhi. Accredited by NAAC with ‘A' Grade, Bangalore. Accredited by Institution of Engineers (India) KOLKATA. A-grade awarded by AP Knowledge Mission. Recognised under sections 2(f) & 12(B) of UGC Act 1956.
Venkatapuram (V), Renigunta (M), Tirupati, Tirupati District, Andhra Pradesh

2024-25

Abstract

TrafficTelligence is an advanced machine learning system designed to predict urban traffic volume using environmental and temporal factors. By analyzing historical traffic patterns, weather conditions, time variables, and holidays, the system generates accurate forecasts. The **Random Forest** model, achieving an **R^2 score of 0.84**, was selected for its superior performance in traffic volume estimation. A user-friendly web interface ensures easy access to predictions, aiding city planners, transportation authorities, businesses, and commuters in making data-driven decisions. This system improves urban mobility, reduces congestion, and enhances transportation efficiency.

TrafficIntelligence: Advanced Traffic Volume Estimation with Machine Learning

1. Executive Summary

TrafficTelligence is a comprehensive machine learning solution developed to predict urban traffic volume based on multiple environmental and temporal factors. The system analyses historical traffic patterns alongside weather conditions, time variables, and holiday information to generate accurate forecasts. These predictions enable city planners, transportation authorities, businesses, and commuters to make data-driven decisions that can alleviate congestion, optimize transportation infrastructure, and improve overall urban mobility. Random Forest emerged as the superior model with an R^2 score of 0.84, demonstrating excellent predictive capability in traffic volume estimation. The web-based interface makes these sophisticated predictions accessible to users through an intuitive design, enabling better planning and traffic management.

2. Pre-Requisites

- Python 3.6 or higher
- Knowledge of basic statistical concepts
- Understanding of machine learning fundamentals
- Familiarity with Flask web framework
- Basic HTML/CSS skills

3. Prior Knowledge

- Data analysis and preprocessing techniques
- Regression modeling concepts
- Performance metrics for machine learning models
- Web application development basics
- Data visualization principles

4. Project Objectives

- Develop an accurate machine learning model to predict urban traffic volume
- Identify critical factors influencing traffic patterns
- Create a user-friendly web interface for traffic predictions
- Provide actionable insights for traffic management stakeholders
- Enable data-driven decision making for urban planners and commuters
- Support emergency services in route planning during critical situation

5. Project Flow

- Data Collection
- Data Pre-Processing
- Model Building
- Application Building

6. Project Structure

6.1 Data Collection

Download The Dataset: Historical traffic volume records containing:

- Date and time stamps
- Temperature measurements (°C)
- Precipitation data (rain and snow in mm)
- Holiday indicators
- Weather condition classifications
- Recorded traffic volume (vehicles per time period)

6.2 Data Pre-Processing

- **Import Necessary Libraries**
  - Pandas, NumPy for data manipulation
  - Scikit-learn for preprocessing tools
  - Matplotlib, Seaborn for visualization
  - Flask for web application framework
  - XGBoost for advanced model

- **Importing The Dataset**
  - Load CSV data
  - Initial data inspection
  - Data structure verification

- **Analyse The Data**
  - Statistical summary
  - Correlation analysis
  - Distribution inspection
  - Outlier detection
  - Pattern identification across temporal dimensions

- **Handling Missing Values**
  - Mean imputation for numeric data
  - Mode imputation for categorical data
  - Verification of data completeness

- **Data Visualization**
  - Traffic trends by time of day
  - Traffic variations by weather conditions
  - Holiday impact visualization
  - Correlation heatmaps
  - Seasonal and monthly patterns

• **Splitting The Dataset Into Dependent And Independent Vari**able

- Feature selection
- Target variable isolation (Traffic Volume)
- Feature-target relationship analysis

• **Feature Scaling**

- Standardization of numerical features
- Normalization where appropriate
- Label encoding for categorical variables
- Creation of cyclical features for time variables

• **Splitting The Data Into Train And Test**

- 80/20 train-test split
- Stratification considerations
- Data randomization

6.3 Model Building

• **Training And Testing The Model**

- Linear Regression implementation
- Decision Tree Regressor training
- Random Forest Regressor training
- Support Vector Regressor (SVR) implementation
- XGBoost Regressor training
- K-Nearest Neighbors (KNN) implementation

• **Model Evaluation**

- Mean Absolute Error (MAE) calculation
- Mean Squared Error (MSE) assessment
- Root Mean Squared Error (RMSE) analysis
- R^2 Score determination
- Cross-validation implementation

• Save The Model

- Pickle serialization of best model (Random Forest)
- Feature transformer serialization
- Label encoder preservation
- Version control for model iteration

6.4 Application Building

• Build HTML Code

- Homepage design (index.html)
- Input form creation
- Results display templates (chance.html, noChance.html)
- Responsive styling implementation

• Main Python Script

- Flask application setup
- Route definitions
- Model loading and prediction logic
- Input validation and processing

• Run The App

- Local development server configuration
- Application debugging and testing
- Browser-based interface initialization

• Output

- Traffic volume prediction display
- Traffic category classification (High/Low)
- Confidence metrics presentation
- Visualization of results

7. Key Data Insights

7.1 Temporal Patterns

- Rush hours (7-9 AM and 4-7 PM) consistently show highest congestion levels.
- Weekend traffic follows distinct patterns compared to weekdays (30-40% lower volume).
- Seasonal variations show summer months with higher average traffic volumes.
- Morning and evening peak times vary by day of week.

7.2 Weather Impacts

- Precipitation significantly affects traffic behavior:
  - Rain reduces traffic volume by approximately 15-20%
  - Snow causes more severe reductions (30-40% in heavy conditions)
- Temperature extremes correlate with lower traffic volumes
- Clear weather correlates with higher traffic volumes than adverse conditions
- Gradual weather changes have less impact than sudden weather events

7.3 Holiday Effects

- Major holidays show 40-50% reduction in traffic volume
- Days preceding holidays often show elevated evening traffic
- Regional holidays demonstrate location-specific patterns
- Holiday effects vary by proximity to commercial centres\

8. Model Performance Comparison

Model	MAE	MSE	RMSE	R²Score
Linear Regression	1638.79	3404210	1845.05	0.14
Decision Tree	557.34	1134416	1065.09	0.71
Random Forest	499.59	623648	789.71	0.84
SVM	1507.25	29747474	1724.67	0.25
XGBoost	543.42	661448	813.39	0.83
KNN	609.53	812674	901.48	0.79

Best Model: Random Forest with R² Score: 0.8423 Model and encoders saved successfully!

8.1 Model Selection Rationale

The selection of Random Forest as the production model was based on:

1. Superior Performance Metrics: Highest R² score (0.84) and lowest error metrics (MAE: 499.59, RMSE: 789.71)
2. Ensemble Advantages: As an ensemble method, Random Forest mitigates overfitting issues common in individual decision trees
3. Feature Importance Analysis: Provides valuable insights into which factors most significantly influence traffic patterns
4. Robustness: Shows consistent performance across various input conditions and is less susceptible to outliers
5. Interpretability: Maintains a reasonable level of interpretability compared to black-box models

XGBoost showed comparable performance (R² score of 0.83) and remains a viable alternative model that could potentially outperform Random Forest with additional tuning.

9. Application Usage Guide

1. Launch the Flask application using python app.py
2. Navigate to http://localhost:5000 in your web browser
3. Enter the required prediction parameters:
   - Date and time
   - Temperature (°C)
   - Precipitation (rain and snow in mm)
   - Weather condition
   - Holiday status
4. Submit the form to receive traffic predictions
5. View the predicted traffic volume and classification (High/Low)
6. Use provided visualization to understand traffic patterns

10. Business Impact

10.1 Urban Planning Applications

- Infrastructure development planning based on predicted traffic patterns
- Public transportation schedule optimization around forecasted congestion
- Traffic light timing adjustments during peak congestion periods

- Road expansion prioritization based on consistent high-volume areas

10.2 Commercial Applications

- Delivery route optimization for logistics companies
- Staff scheduling for retail and service businesses
- Location selection for new businesses based on traffic patterns
- Marketing campaign timing based on expected traffic volume

10.3 Individual Benefits

- Improved commute planning and travel time estimation
- Reduced fuel consumption and emissions through congestion avoidance
- Lower stress levels from predictable travel experiences
- Enhanced quality of life through time savings

11. Future Enhancement Roadmap

11.1 Technical Improvements

- Integration with real-time GPS data streams
- Incorporation of traffic camera feeds for validation
- Expansion of the feature set to include road construction and events
- Deployment as a cloud-based service with horizontal scaling
- Further hyperparameter tuning for the Random Forest and XGBoost models
- Implementation of neural network approaches for comparison

11.2 Feature Enhancements

- Mobile application development
- Interactive visualization dashboard
- Route-specific predictions
- Traffic anomaly detection
- Predictive alerts for unusual congestion
- Integration with navigation systems
- Multi-city model adaptation
- Air quality correlation analysis

12. Conclusion

TrafficTelligence effectively showcases the power of machine learning in addressing urban mobility challenges. By leveraging multiple environmental and temporal factors, the system accurately forecasts traffic volumes, providing critical decision-making support for urban transportation stakeholders.

The strong performance of the Random Forest model underscores the potential of data-driven solutions to enhance traffic management. As the system integrates additional data sources and advanced features, its accuracy and effectiveness will continue to improve, leading to more efficient and less congested cities.

With a user-friendly web application, TrafficTelligence makes complex predictions accessible, bridging the gap between advanced machine learning and practical traffic management. This innovation marks a significant step forward in applying data science to real-world urban transportation challenges.