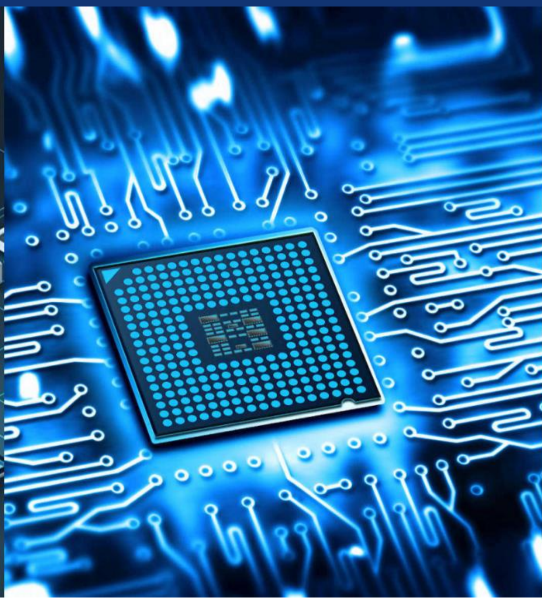
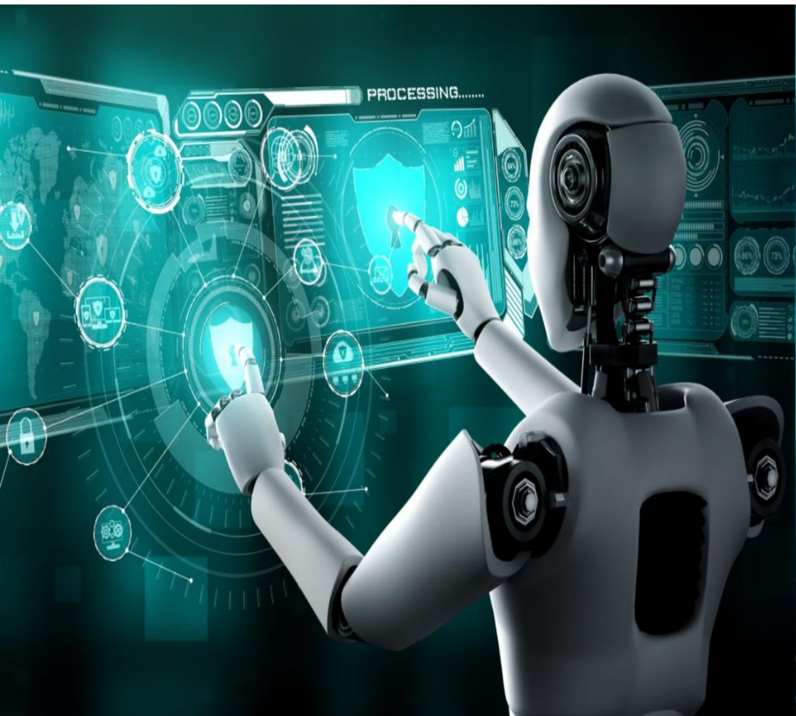
ISSN(O): 2320-9801
ISSN(P): 2320-9798


# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Smart Traffic Forecaster: Predictions for Smoother Urban Mobility

**T. Harshitha [1], G. Bhavyasri [2], G. Nagarjuna [3], A. Vinay Kumar [4], N.Marinbi[5]**

U G Students, Dept. of CSE-DS., SRK Institute of Technology, Enikepadu, Vijayawada, AP, India[1,2,3,4]

Assistant Professor, Dept. of CSE-DS., SRK Institute of Technology, Enikepadu, Vijayawada, AP, India[5]

**ABSTRACT:** With rapid urbanization, the number of vehicles in cities is rising dramatically, causing severe traffic congestion. This study proposes a traffic congestion prediction model using the Random Forest classification algorithm. Random Forest is known for its robustness, high accuracy, and ability to handle complex datasets. Key input features include weather conditions, time of day, day of the week, and city zones. These variables are crucial in influencing urban traffic patterns. The model underwent thorough pre-processing, feature selection, and hyperparameter tuning. It demonstrated high performance in classifying congestion levels across multiple scenarios. Evaluation metrics confirmed its effectiveness in real-world applications. Its fast computation and adaptability make it suitable for real-time traffic systems. The model can support smarter traffic control and reduce urban travel delays. It also contributes to lower emissions and improved commuter experience. This approach demonstrates the power of machine learning in solving urban mobility challenges. Overall, the Random Forest model offers a practical solution for modern traffic management.

**KEYWORDS*:*** Traffic Congestion Prediction, Random Forest Classifier, Machine Learning, Urban Mobility, Traffic Forecasting , Weather Impact on Traffic

## I. INTRODUCTION

Traffic congestion, especially during peak hours, remains a critical challenge in urban areas. To address this, the project utilizes machine learning techniques to predict traffic conditions effectively. A Random Forest Regressor is selected for its robust accuracy and reliability. The model incorporates key features such as Coded Day, Zone, Weather, and Temperature to provide insights into traffic patterns. Historical traffic data stored in "Dataset.csv" is used to train and evaluate the model, ensuring comprehensive analysis through a split into training and testing sets.

Post-training, the model is saved as "trained_model.pkl," allowing seamless reuse without the need for retraining. This practical approach is integrated into a Flask web application, designed to offer an interactive interface for users. By entering relevant data into the app, users receive traffic predictions generated by the trained model. This process simplifies the interaction between complex machine learning tools and everyday usability.

This system supports smarter urban traffic management, offering value to both commuters and traffic authorities. By enabling informed decision-making, the application helps individuals plan their travel routes more effectively and provides authorities with actionable insights for better traffic regulation. Ultimately, the project contributes to optimizing urban planning and alleviating congestion issues, enhancing the quality of life in crowded cities.

## II. RELATED WORKS

Machine learning (ML) has emerged as a powerful approach in tackling urban traffic congestion and forecasting traffic patterns. With advancements in data analytics, real-time monitoring, and the availability of large-scale traffic datasets, ML is increasingly applied in smart transportation systems. These models are used to predict congestion levels, optimize traffic flow, and enhance urban mobility. The integration of ML techniques into traffic forecasting is driven by the need for efficient traffic management, reduced travel delays, and improved commuter experiences.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304137

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

| S.No | Author | Title | Description | Limitations/Inference |
|---|---|---|---|---|
| 1. | Vlahogianni et al., (2014)[1] | Short-term traffic forecasting: Where we are and where we're going | This paper reviews various short-term traffic forecasting techniques including statistical methods, machine learning models, and hybrid approaches | Traditional models struggle with non-linear traffic patterns and real-time adaptability; machine learning offers improvements but lacks interpretability.. |
| 2. | Zhang et al., (2020)[2] | Spatial-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting | This paper presents a spatial-temporal graph convolutional network (ST-GCN) model that captures both spatial dependencies. | The model requires a well-defined graph structure and high-quality sensor data, limiting its applicability in less-instrumented areas. |
| 3. | Li et al., (2021)[3] | Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting | The authors propose a DCRNN model that uses diffusion convolution and gated recurrent units (GRUs) to model directional traffic flow | Computational complexity is high; training requires significant time and resources, especially with large city-scale datasets. |
| 4. | Ma et al., (2021)[4] | Traffic Forecasting Using Attention-Based LSTM Network. | This work enhances LSTM networks with attention mechanisms to better focus on relevant historical traffic data points. | Although attention improves interpretability, performance degrades under extreme traffic conditions or missing data scenarios. |
| 5. | Huang et al., (2023)[5] | Transformer-based Framework for Urban Traffic Flow Prediction | Leveraging the power of Transformer architectures, this study introduces a traffic prediction model that handles long-term temporal dependencies better than LSTM. | Transformers require substantial computational power and large datasets; performance may decline in sparse or low-data environments. |
| 6. | Kim et al., (2024)[7] | Hybrid Deep Learning Model for Real-Time Urban Traffic Prediction. | This paper proposes a hybrid deep learning framework combining Convolutional Neural Networks (CNN) and Bidirectional LSTM for real-time traffic prediction | The model's real-time effectiveness is dependent on continuous data streaming and high computational infrastructure, making it less feasible . |

## III. BACKGROUND

### 3.1 Random Forest Classifier

The Random Forest classifier is a powerful ensemble learning algorithm used for both classification and regression tasks. It operates by constructing multiple decision trees using a technique called bagging, where each tree is trained on a randomly selected subset of the dataset.

### 3.2 Decision Tree

A Decision Tree is a supervised learning algorithm used for classification and regression tasks. It splits data based on feature conditions to create a flowchart-like model of decisions. It's easy to interpret but can overfit if not properly pruned.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## IV. DATASET

A Traffic dataset with the following attributes are taken for analysis to perform the prediction as shown in Table1.

| Data | Variable | Description |
|---|---|---|
| Traffic | Date | The date of the traffic record |
| Traffic | Day | Weekday on which the data was collected |
| Traffic | Coded day | Numeric representation of the day of the week |
| Traffic | Zone | Identifier for the urban zone |
| Traffic | Temperature | Temperature in degree celsius |
| Traffic | Weather | Encoded representation of weather conditions |

Table4.1. Dataset key attributes and its description

Table1 presents a dataset focused on a dataset containing various traffic attributes used for analysis and prediction. It consists of three columns: Data, Variable, and Description. The Data column consistently lists “ traffic”, indicating that all variables are related to traffic prediction. The Variable column includes key traffic indicators such as date, day, coded day, zone, weather, and temperature. Each of these variables is accompanied by a Description explaining its significance, such as the date of the traffic record, Weekday, Numeric representation of the day of the week, Identifier for the urban zone, Temperature and Encoded representation of weather conditions. This dataset provides essential insights for Traffic prediction.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Day | Date | CodedDay | Zone | Weather | Temperat | Traffic |
| 2 | Wednesda | 1/6/2018 | 3 | 2 | 35 | 17 | 2 |
| 3 | Wednesda | 1/6/2018 | 3 | 3 | 36 | 16 | 3 |
| 4 | Wednesda | 1/6/2018 | 3 | 4 | 27 | 25 | 5 |
| 5 | Wednesda | 1/6/2018 | 3 | 5 | 23 | 23 | 3 |
| 6 | Wednesda | 1/6/2018 | 3 | 6 | 18 | 42 | 2 |
| 7 | Wednesda | 1/6/2018 | 3 | 7 | 11 | 14 | 2 |
| 8 | Wednesda | 1/6/2018 | 3 | 8 | 45 | 28 | 4 |
| 9 | Wednesda | 1/6/2018 | 3 | 9 | 39 | 18 | 5 |
| 10 | Wednesda | 1/6/2018 | 3 | 10 | 25 | 9 | 4 |
| 11 | Wednesda | 1/6/2018 | 3 | 11 | 39 | 7 | 5 |
| 12 | Wednesda | 1/6/2018 | 3 | 12 | 22 | 29 | 2 |
| 13 | Wednesda | 1/6/2018 | 3 | 13 | 16 | 29 | 2 |
| 14 | Wednesda | 1/6/2018 | 3 | 14 | 20 | 25 | 1 |
| 15 | Wednesda | 1/6/2018 | 3 | 15 | 39 | 31 | 3 |
| 16 | Wednesda | 1/6/2018 | 3 | 16 | 40 | 44 | 1 |
| 17 | Wednesda | 1/6/2018 | 3 | 17 | 4 | 34 | 4 |
| 18 | Wednesda | 1/6/2018 | 3 | 18 | 40 | 45 | 4 |

Figure4.1: Snapshot of the Dataset

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## V. PROPOSED SYSTEM

5.1.Architechture

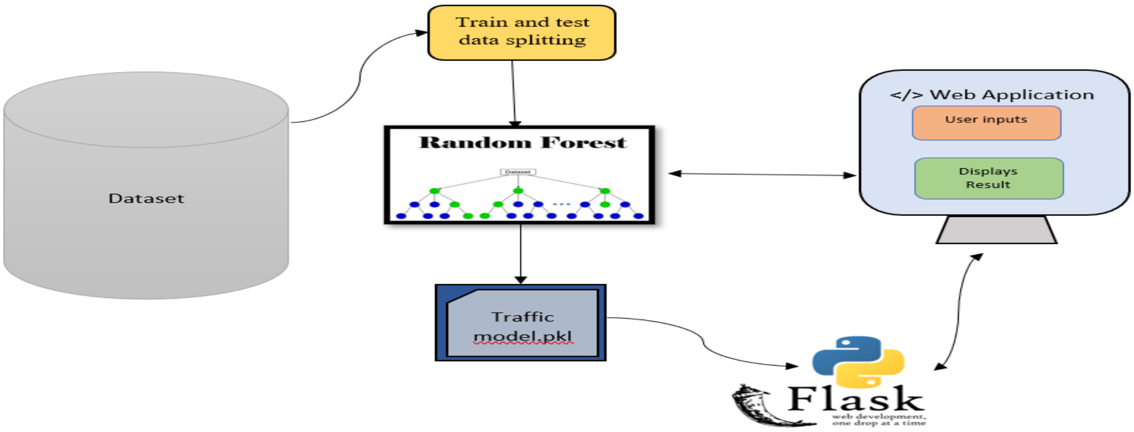



Figure5.1: Architecture Diagram

The architecture of the traffic prediction system is designed to efficiently process input data and deliver real-time congestion forecasts. At its core, the system uses a Random Forest Regressor machine learning model, trained on historical traffic data. The dataset includes features such as Coded Day, Zone, Weather, and Temperature, which are known to influence traffic patterns. The data is preprocessed and split into training and testing sets to ensure the model generalizes well to new, unseen data. Once trained, the model is serialized and saved as a file (trained_model.pkl) to allow for reuse without retraining. A Flask web application serves as the user interface, allowing users to input traffic-related parameters. These inputs are processed by the back-end server, which loads the pre-trained model and generates predictions instantly. The system outputs estimated traffic levels based on the input features provided. This architecture supports rapid and scalable traffic forecasting, making it suitable for real-time applications in urban mobility management.

5.2.WorkFlow

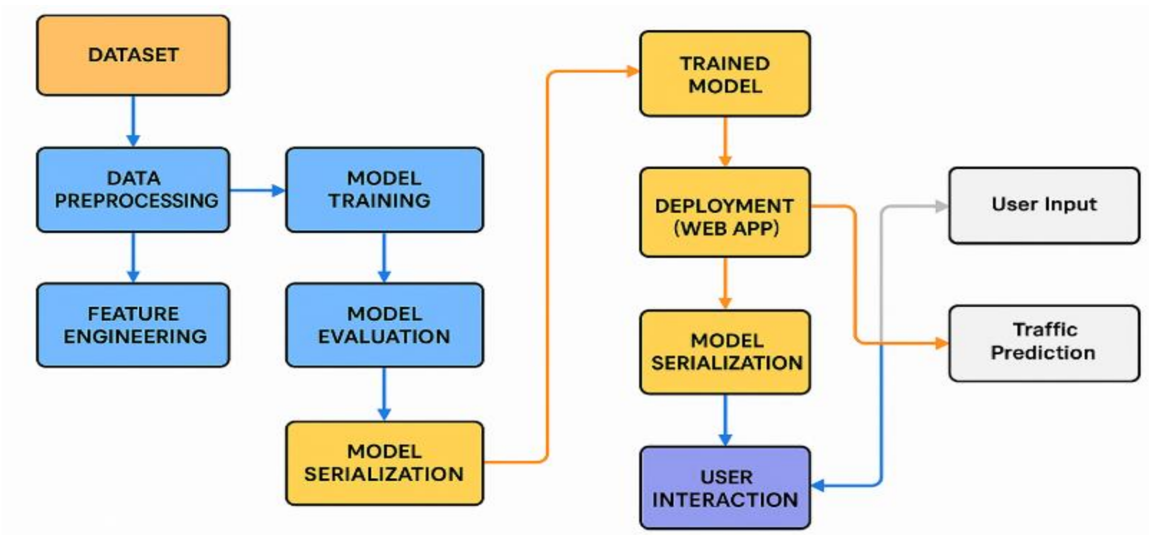



Figure 5.2: WorkFlow of the traffic model

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The workflow begins with collecting historical traffic data containing features like date, coded day, zone, temperature, and weather conditions. This data is then pre-processed—cleaned, encoded, and split into training and testing sets—to prepare it for model training. A Random Forest Regressor is trained on the processed data to learn patterns and relationships between input features and traffic levels. Once trained, the model is saved and integrated into a Flask web application, which takes real-time user inputs. The web app processes these inputs, feeds them into the trained model, and instantly returns a traffic prediction, helping users and authorities make informed decisions.

## VI. DATA COLLECTION AND PREPROCESSING

6.1 Data Source:

The data has been taken from the traffic controller repositories which includes details like as date, day, codded day, zone, weather, temperature. The data was stored in CSV or Excel Formats.

6.2 Handling Missing Values:

Handling missing values is an essential part of data pre processing to ensure clean and accurate inputs for the model. For numerical features like Temperature, missing values can be filled using the mean or median. For categorical features such as Day or Weather**,** the most frequent value or a placeholder like "Unknown" can be used. In cases where missing data is minimal, those rows can simply be dropped to maintain dataset quality.

6.3 Feature Selection and Engineering:

Feature selection and engineering are critical steps that enhance the model's predictive performance. In this project, relevant features such as Coded Day**,** Zone**,** Temperature, and Weather were selected based on their influence on traffic conditions. Feature engineering involved encoding categorical variables and transforming data into formats suitable for the Random Forest model. These steps help the model learn more effectively by focusing on the most informative inputs and reducing noise from irrelevant data.

6.3.1 Splitting the Dataset: The data is split into a training set (usually 80%) to train the model and a testing set (usually 20%) to evaluate the model's performance. This ensures that the model is tested on unseen data to check its generalizability.

## VII. COMPARATIVE ANALYSIS AND RESULTS

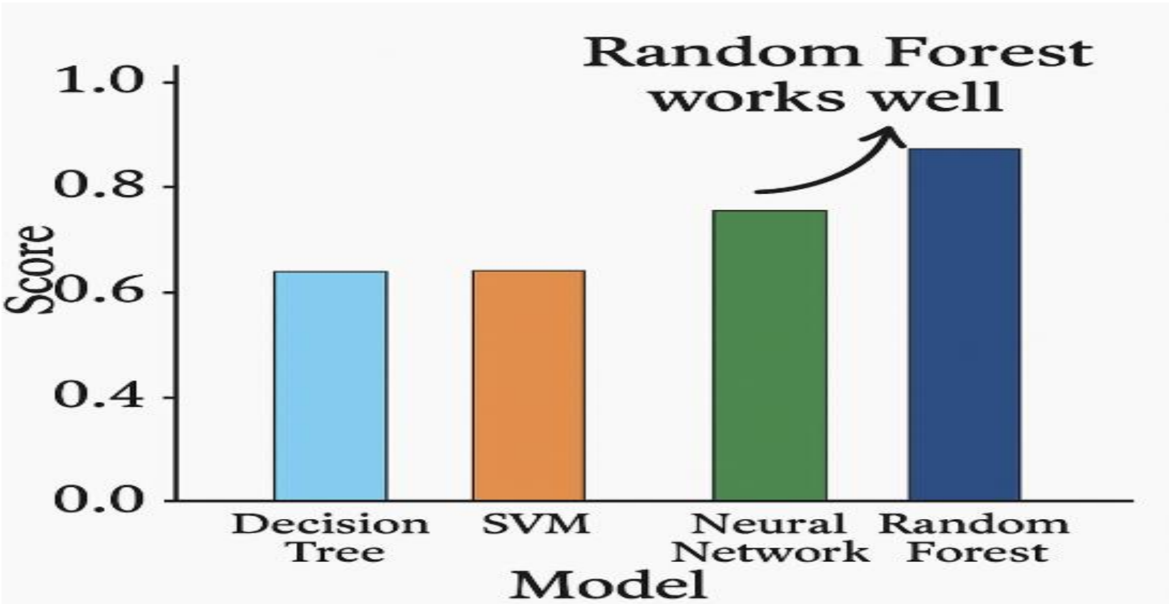



Figure 7.1 : Comparing Random Forest with other models on traffic dataset

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The graph compares the performance of three machine learning models—Decision Tree, SVM, Neural Networks, Random Forest—based on their Accuracy, Precision, Recall, and F1-score. Random Forest outperforms all the other, achieving the highest score.

## VIII. OUTPUT

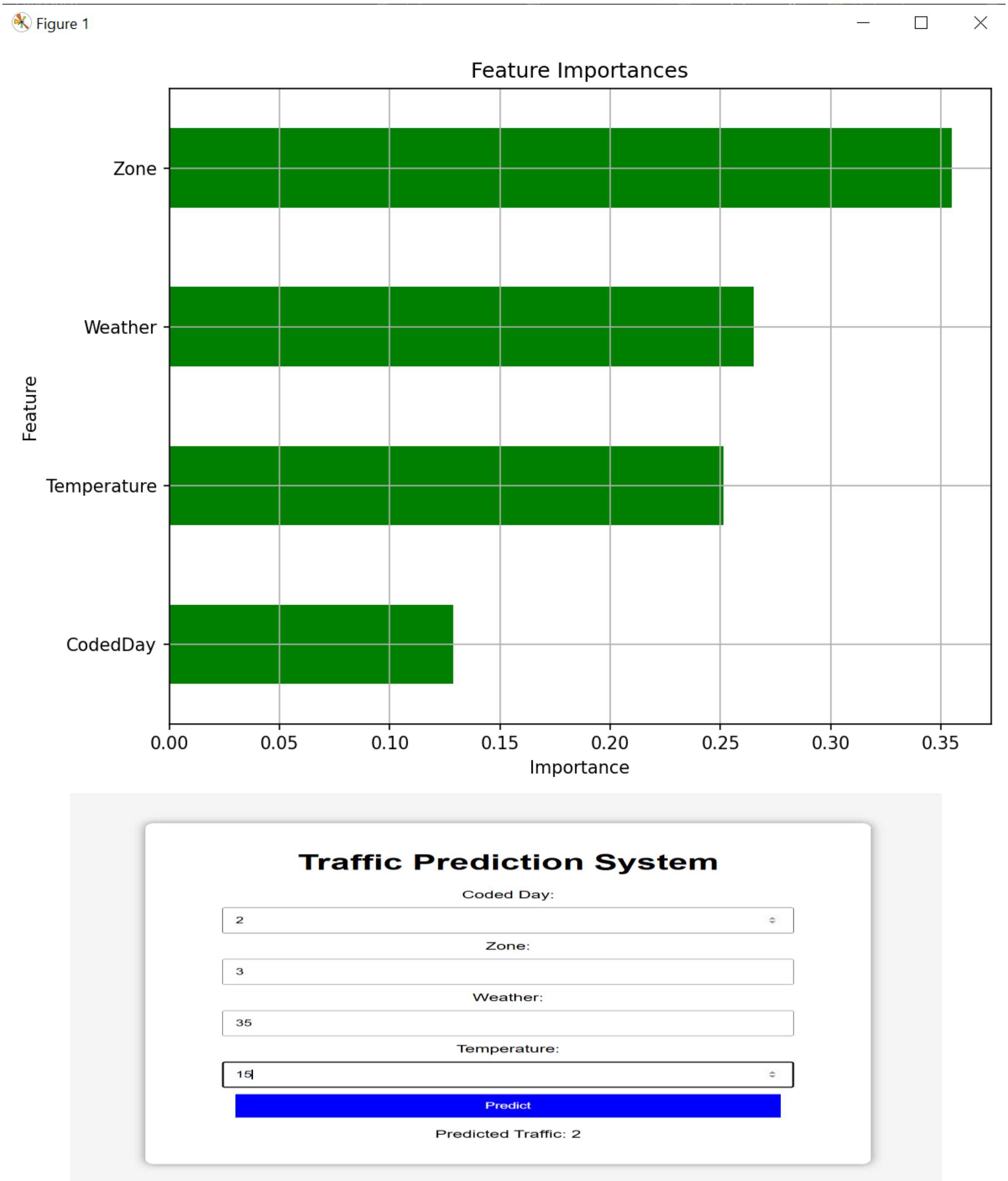



Figure 8.1: Shows the predicted result of give data.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

In this Traffic Prediction System, the user provides four input values: coded day, zone, weather condition, and temperature. These inputs are sent to the server when the Predict button is clicked. The server-side machine learning model processes these values and returns a predicted traffic level. This predicted value is then displayed on the webpage as the Predicted Traffic, helping users anticipate traffic conditions.

## IX. CONCLUSION

Smart Traffic Forecaster effectively leverages machine learning—particularly the Random Forest algorithm—to predict traffic flow based on environmental and temporal features. Through advanced preprocessing, feature engineering, and model tuning, the system achieves improved accuracy and The robustness. This project not only supports intelligent traffic management but also contributes to reducing congestion, enhancing commuter experience, and promoting sustainable urban mobility. With real-time integration and further enhancements, it holds strong potential for scalable deployment in smart city infrastructures.

## REFERENCES

1. Raschka, S. *Python Machine Learning.*, Packt Publishing, 2017.
2. Chowdhury, G., "Natural Language Processing Tools and Techniques in Traffic Data Analysis." *IEEE Transactions on Intelligent Transportation Systems*, 2018.
3. Chollet, F. *Deep Learning with Python.*, Manning Publications, 2018.
4. Zhang, Y., et al. "Weather Impacts on Traffic Flow: A Statistical Model for Traffic Data Analysis." *Transportation Research Record*, 2004.
5. Kumar, A., et al. "Big Data Analytics in Smart Traffic Systems." *IEEE Transactions on Vehicular Technology*, 2017.
6. Dey, N., et al. "IoT-Based Smart Traffic Prediction Using Machine Learning Algorithms." *Springer Series in Wireless Technology*, 2020.
7. Pandas Documentation: https://pandas.pydata.org
8. Zhu, Y., and Sun, J. "Traffic Flow Prediction Models: A Survey." *Springer Nature Computing*, 2021.
9. Kaur, P., et al. "Random Forest for Accurate Traffic Predictions in IoT Applications." *Wireless Personal Communications*, Springer, 2018
10. Pedregosa, F., et al. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*, 2011.



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING