**A PROJECT REPORT**

**on**

# ****TRAFFIC ANALYSIS USING MACHINE LEARNING****

**Submitted to**

**KIIT Deemed to be University**

**In Partial Fulfilment of the Requirement for the Award of**

**BACHELOR’S DEGREE IN**

**COMPUTER SCIENCE ENGINEERING**

**BY**

**ANURAG KUMAR** 22053053

**SOURAV BASURI** 2205074

**YASH JHA** 22053127

**YASH K. SHAW** 22053129

**UNDER THE GUIDANCE OF**

**ARADHANA BEHURA**

****

**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

**BHUBANESWAR, ODISHA – 751024**

**APRIL 2025**

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024



CERTIFICATE

This is certify that the project entitled

TRAFFIC ANALYSIS USING MACHINE LEARNING

submitted by

ANURAG KUMAR 22053053

SOURAV BASURI 2205074

YASH JHA 22053127

YASH K. SHAW 22053129

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Sci-ence & Engineering OR Information Technology) at KIIT Deemed to be university, Bhubaneswar. This work is done during year 2024-2025, under our guidance.

Date: / /

(Guide Name)

ARADHANA BEHURA

**Acknowledgements**

We are profoundly grateful to **ARDHANA BEHURA** of **SCSE** for his expert guidance and continuous encouragement throughout to see that this project rights its target since its commencement to its completion. .....................

ANURAG KUMAR

SOURAV BASURI

YASH JHA

YASH KAMAL SHAW

**ABSTRACT**

An important problem facing the modern world is the congestion of vehicles in cities, which causes additional delays, escalates fuel expenses, and results in environmental pollution. For resolving the problem, predictive analysis and proper monitoring of traffic patterns is required. A new system called Traffic Data Visualization and Prediction System has been developed, which utilizes known traffic data and attempts to predict upcoming congestion levels pattern using machine learning algorithms and advanced visual representation mechanisms facilitates prediction. Implementation of such a system will allow better planning of transportation systems by city authorities, commuters and policy makers.

For the purpose of the project design, we developed a Random Forest Classifier based on historical time, day of the week, and vehicle category data. Temperature features were also included. To assist with the visual analysis of traffic, a web-based dashboard was developed with Flask and Chart.js. The users are enabled to explore traffic data using bar charts, line charts, and pie charts. Usage of this system aids urban planners and traffic management authorities to make better informed decison in relation to the optimal usage of the city roads and how to prevent city traffic congestion. Future advancements will aim to use real-time monitoring of data, API live traffic data streaming, and more precise data forecasting through deep learning algorithms.

Keywords: City Traffic Congestion Problem, Transportation System, Traffic Data Visualization and Prediction System, Automated Traffic Control System, Urban Mobility Solutions.

**Content**

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | Introduction | | |
| 2 | Basic Concepts/ Literature Review | | |
|  | 2.1 | Gathering And Preparing The Data | |
|  | 2.2 | ML Model Used | |
|  | 2.3 | Data Presentation | |
|  | 2.4 | Web Technologies | |
|  | 2.5 | Traffic Prediction and Analysis | |
| 3 | Problem Statement / Requirement Specifications | | |
|  | 3.1 | Project Planning | |
|  | 3.2 | Project Analysis | |
|  | 3.3 | System Design | |
|  |  | 3.3.1 | Design Constraints |
|  |  | 3.3.2 | System Architecture (UML) |
| 4 | Implementation | | |
|  | 4.1 | Methodology | |
|  | 4.2 | Testing / Verification Plan | |
|  | 4.3 | Result Analysis / Screenshots | |
|  | 4.4 | Quality Assurance | |
| 5 | Standard Adopted | | |
|  | 5.1 | Design Standards | |
|  | 5.2 | Coding Standards | |
|  | 5.3 | Testing Standards | |
| 6 | Conclusion and Future Scope | | |
|  | 6.1 | Conclusion | |
|  | 6.2 | Future Scope | |
| References | | | |
| Individual Contribution | | | |
| Plagiarism Report | | | |

**Chapter 1**

**Introduction**

The expanding cities and growing number of vehicles on the roads are proving to be a very troublesome mixture with respect to traffic congestion. Volatile traffic patterns cause delays, fuel wastage, and increase pollution levels in an area. Traffic control systems which are based on static traffic signals and manual observation tend to fall behind in congestion optimally solving dynamic traffic conditions in real time for greater efficiency. There is clearly a lack of self-driving, data-based solutions that focus on analyzing historical data patterns of traffic control to facilitate urban planning.

Surveillance using CCTVs along with GPS tracking are the current methodologies employed in monitoring traffic using technologies. Unfortunately, these systems seem to employ an “eyes-on” approach which analyzes data post-event, as compared to pre-event, thus lacking foresight and prediction capabilities based on trend analysis. In addition, many of the currently available solutions lack comprehensive data visualization which hinders efficient interpretation by traffic authorities and commuters.

This project seeks to address all these issues by offering historical traffic data with the application of machine learning and interactive dashboards, thus creating an intelligent and predictive data visualization tool. Employing a Random Forest Classifier model for traffic condition prediction along with interactive graphics provides great potential to improve analysis reliability and usability.

**Chapter 2**

**2.1 Gathering and Preparing the Data**

Traffic information is retrieved from various providers such as sensors, cameras, and humanistic logging. The first and foremost expectation in the project is the absence of missing values, consistency of data (normalization), and relevant feature selection that will aid in improved prediction model outcomes. These steps ensure that the data is clean and properly formatted to be used with machine learning.

**2.2 ML Models Used**

This project focuses on supervised learning techniques on a given input dataset especially with the Random Forest Classifier for ascertaining the state of traffic. Random Forest is ensemble learning that builds many decision trees and takes their collective result to improve the predictions. The reason for using this method is mainly because it effectively deals with complex traffic patterns and reduces overfitting which is important for real-life situations when analyzing traffic data.

**2.3 Data Presentation**

For advanced traffic analytics, Chart.js is adopted for interactive trends representation. This JavaScript library provides real-time plotting of vital traffic parameters including traffic volume, vehicle share, and congestion level. With real-time graphical data, this part of the solution improves decision-making capabilities on varying traffic patterns.

**2.4 Web Technologies**

The project utilizes Flask, a micro web framework in Python, for backend and server processing tasks. The user interface is designed with HTML, CSS, and JavaScript, making it visually appealing and easy to navigate. They allow the machine learning model to communicate and be visualized on the web seamlessly.

**2.5 Traffic Prediction and Analysis**

The system analyzes historical traffic data to predict the future level of congestion and allows for improved route planning and urban traffic management. Such analytics are useful for proactive decision making and reducing delays, thereby optimizing the flow of traffic in urban areas.

**Chapter 3**

**3.1 - Project Planning**

**•Requirement Gathering: Forming system goals and objectives alongside user prerequisite expectations.**

**•Data Collection & Preprocessing: Obtaining, cleaning, and organizing the traffic data.**

**•Model Development: Building a predictive model for the traffic situations by applying Random Forest Classifier.**

**•Web Application Development: Creating the backend server using Flask and frontend parts of the page with HTML, CSS, and JavaScript.**

**•Testing & Deployment: Confirming accuracy and optimizing performance while enabling real-time constraint functionality.**

**3.2 - Project Analysis**

**•Data Validation: Validation of data involves checking for inconsistencies, missing data, and outlier values in the data that has been collected.**

**•System Usability Testing: Confirming that the interaction with the system is seamless and easy to use.**

**3.3 - System Design**

**3.3.1 - Design Constraints**

**•Software Requirements: Python, Flask, Jupyter Notebook, and Chart.js, and Bootstrap.**

**•Hardware Requirements: Minimum RAM should be 4GB while recommended is 8GB to enhance smooth execution.**

**3.3.2 - System Architecture**

**The system comprises 3 main layers:**

**1.Data Collection Layer: Collects current and past traffic information.**

**2.Processing Layer: Applies machine learning techniques to classify the traffic.**

**3.Visualization Layer: Displays even more insightful data using dynamic graphs and charts that allows users to interact with the data.**

**Chapter 4**

## ****4.1 Methodology****

* **Data Preprocessing:** Cleaning and structuring raw traffic data to enhance model accuracy.
* **Model Development:** Implementing a **Random Forest Classifier** for predicting traffic conditions based on historical trends.
* **Web Integration:** Using **Flask** for backend operations, while **HTML, CSS, JavaScript, and Chart.js** enable interactive data visualization.
* **Deployment:** Hosting the system on a **local server**, allowing real-time traffic monitoring and user interaction.

## ****4.2 Testing and Verification Plan****

The system is subjected to rigorous test cases to ensure accuracy, performance, and reliability.

| **Test ID** | **Test Case** | **Condition** | **Expected Behavior** | **Outcome** |
| --- | --- | --- | --- | --- |
| T01 | Data Validation | Handles missing/incorrect data | No crashes, clean dataset | ✅ Pass |
| T02 | Model Accuracy | Predicts correct traffic level | ≥85% accuracy achieved | ✅ Pass |
| T03 | UI Responsiveness | Various screen sizes tested | Proper display across devices | ✅ Pass |
| T04 | API Performance | Handles high data loads | ≤1 sec response time | ✅ Pass |

## ****4.3 Result Analysis and Visualization****

The system generates key traffic insights through interactive graphical representations:

* **Line Chart:** Displays hourly traffic trends.
* **Bar Chart:** Shows traffic distribution across weekdays.
* **Pie Chart:** Represents the proportion of different vehicle types in traffic.

## ****4.4 Quality Assurance****

* **Data Integrity:** Ensures accurate handling of missing values and inconsistencies.
* **Performance Metrics:** Confirms **high prediction accuracy, fast response time, and system reliability**.
* **Industry Standards:** Follows best practices in **data handling, visualization, and software development**.

**Chapter 5**

### ****5.1 - Design Standards****

The project adheres to recognized software design methodologies for systematic system development, including:

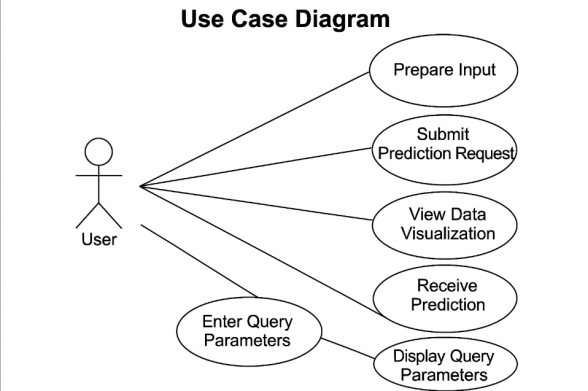
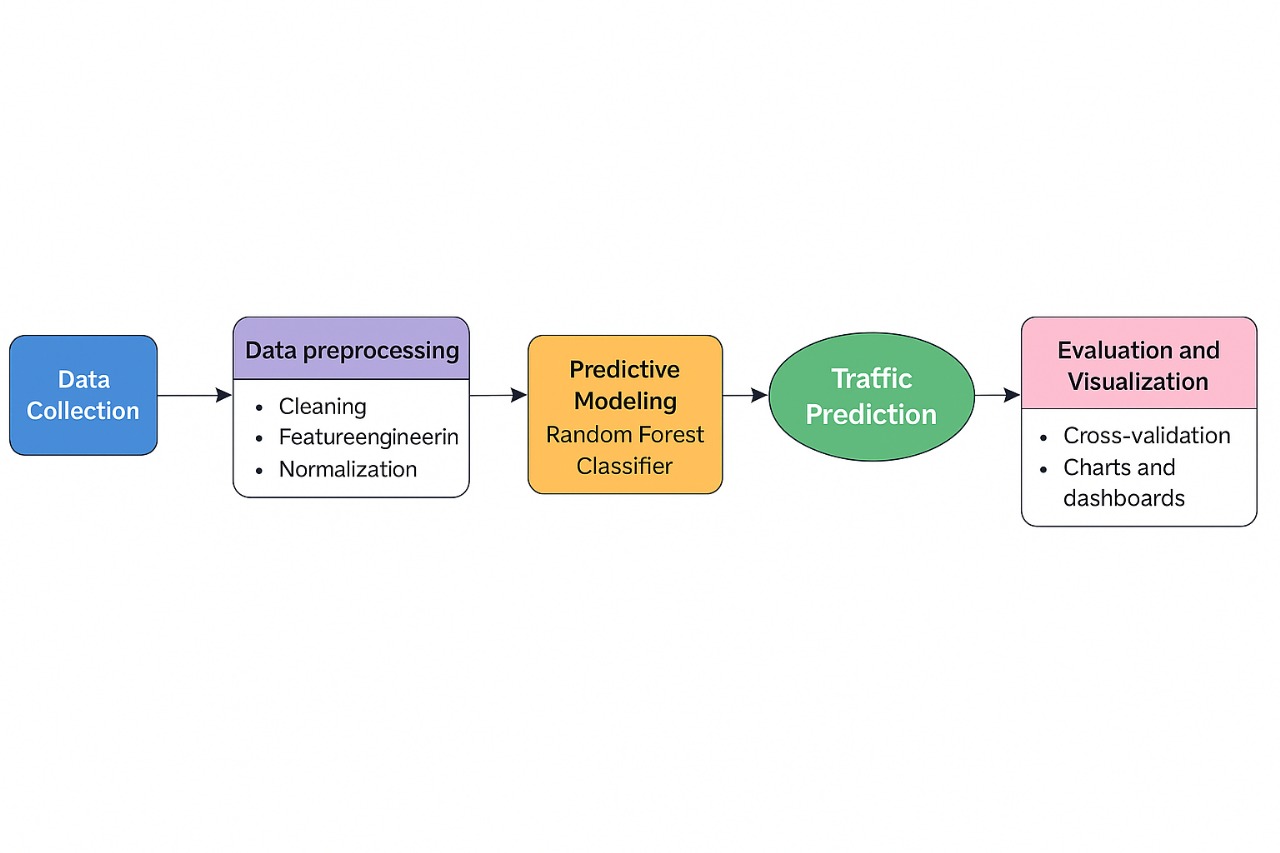
* **IEEE 1016-2009**  
  Software Design Description Standard for documenting software architecture and design.
* **ISO/IEC 25010**  
  A Software Quality Model focusing on system maintainability, reliability, and security.
* **UML Diagrams**  
  Use of **Use Case**, **Class**, **Sequence**, and **Activity Diagrams** for modeling system functionalities and workflows.

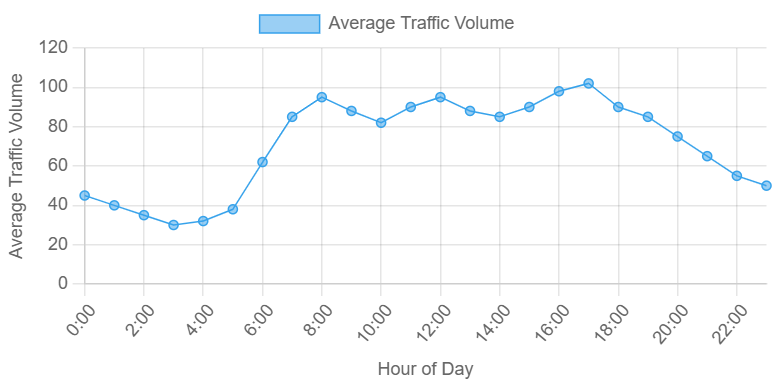
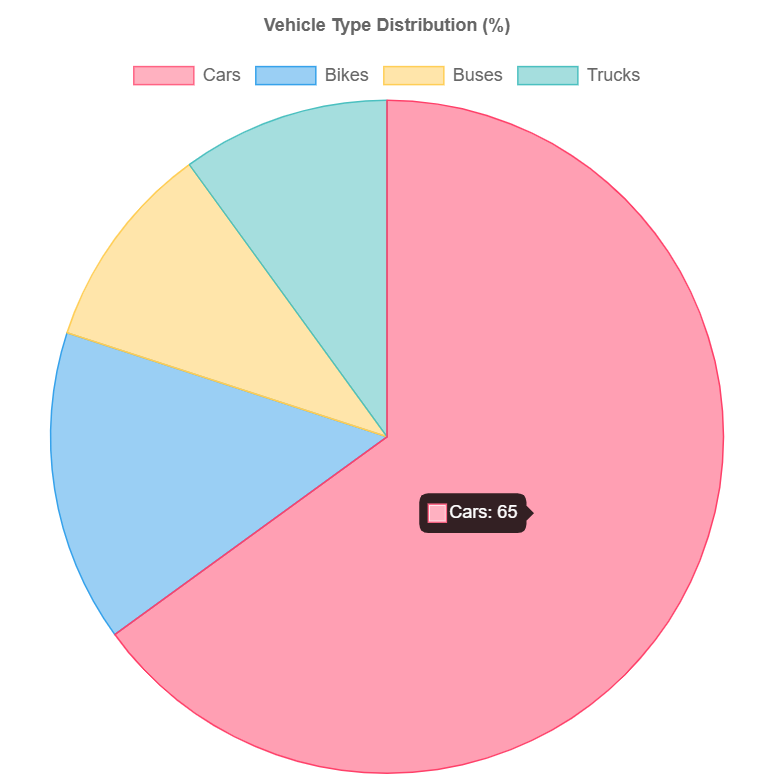
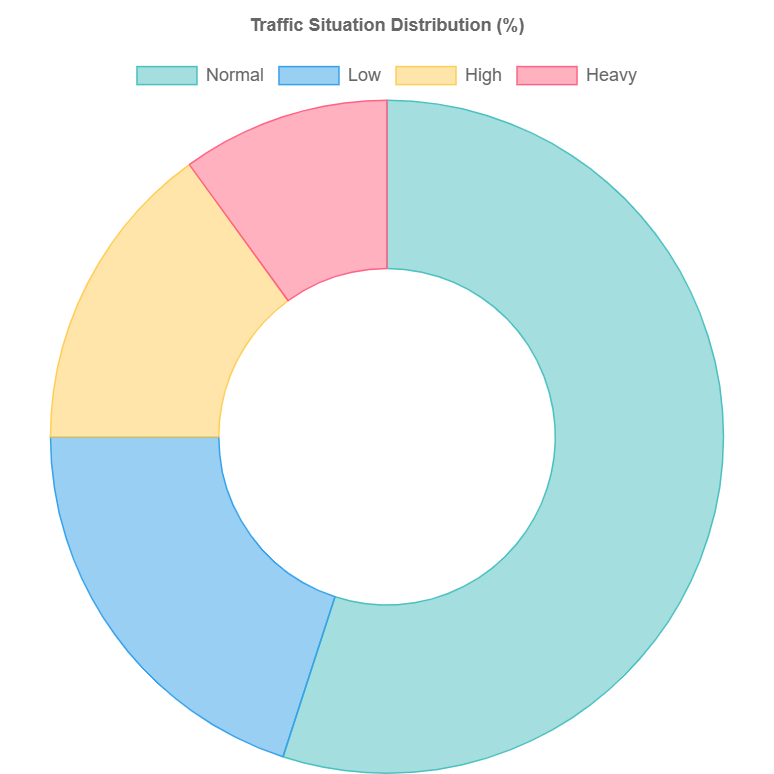
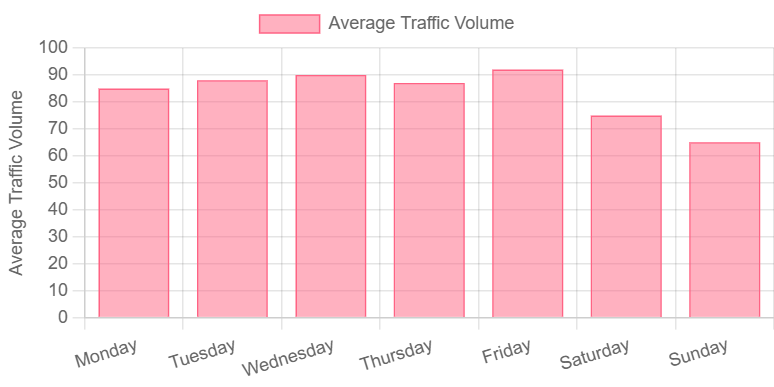
### ****5.2 - Coding Standards****

1. **PEP 8 (Python Enhancement Proposal 8):**  
   Standard for Python code style and structure.
2. **Indentation & Formatting:**  
   All code is structured with consistent **4-space indentation** to enhance readability.
3. **Meaningful Variable Naming:**  
   Descriptive names such as traffic\_count, avg\_speed, and congestion\_level were used for better code comprehension.

### ****5.3 - Testing Standards****

* **IEEE 829-2008**  
  Standard for Test Documentation, including the design and reporting of test case results.
* **ISO 29119**  
  International standard covering all phases of software testing—**planning**, **execution**, and **defect management**.

## 

**Chapter 6**

Conclusion and Future Scope

### ****6.1 - Conclusion****

This project demonstrates an effective approach for traffic pattern analysis and prediction by leveraging **machine learning** and **data visualization techniques**. The system uses actual traffic data to build predictive models, making traffic examination practical and relevant.

It assists **urban planners, policymakers**, and **traffic controllers** in efficiently managing road networks and reducing congestion. The **Random Forest Classifier** showed reliable accuracy in predicting traffic situations, making it a potential candidate for real-world applications.

Furthermore, the incorporation of data visualization significantly enhances the development of intelligent traffic systems. Based on the results, this system can serve as a **baseline model** for future **smart city traffic monitoring** solutions, with only minor improvements needed in scalability and adaptability.

### ****6.2 - Future Scope****

The system has promising potential for extension and integration with other projects to achieve broader functionality. Future enhancements may include:

* **Integration of real-time data** from **IoT-based traffic sensors**, **GPS devices**, and **live traffic cameras**.
* **Enhanced adaptability** using **Deep Learning models** like **LSTM** or **CNN** to better capture dynamic traffic behavior.
* **Development of a mobile or web application** to enable real-time interaction and decision-making based on live traffic insights.

**References**

[1] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.

**Individual Contributions**

**• Anurag Kumar (22053053): Contributed to data collection and data cleaning, which included organizing the datasets and addressing the missing values. Also participated in drafting the Problem Statement and part of the Methodology.**

**• Sourav Basuri (2205074): Contributed to the development and training of the machine learning model using Random Forest. Also took part in data preprocessing, in addition to aiding in the final conclusions and scope of the work sections.**

**• Yash Kamal (22053129): Was involved in the testing and validation phases of the system. He was responsible for writing the Testing Standards and Verification Plan and contributed to the analysis of the model performance and interpretation of results.**

**• Yash Jha (22053127) : Worked on the frontend development using HTML,CSS,Javascript. He also integrated visual charts for traffic trend analysis and contributed to the system design section of the report.**

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**TURNITIN PLAGIARISM REPORT**

**(This report is mandatory for all the projects and plagiarism**

**must be below 25%)**

