Industrial Internship Report on

”Smart city traffic patterns”

Prepared by

Hitendra Rathod

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This report outlines a 6-week internship project titled *“Forecasting of Smart City Traffic Patterns”* completed under UniConverge Technologies Pvt. Ltd. (UCT), facilitated by upskill Campus and The IoT Academy. My project was aimed to build a machine learning model that forecasts hourly vehicle traffic at four key city junctions using historical data. The objective was to use data science and machine learning to predict hourly traffic volume at key city junctions. By leveraging historical traffic data, time-based features, and models like Random Forest and XGBoost, the project successfully produced accurate forecasts to support smart city infrastructure planning. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

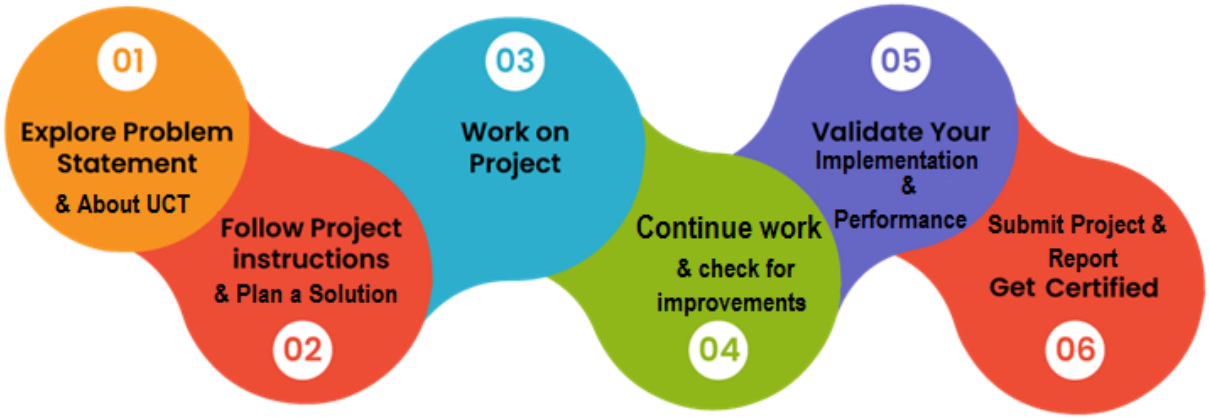
TABLE OF CONTENTS

1 Preface

This report presents a comprehensive summary of my 6-week internship experience on the project titled **"Forecasting of Smart City Traffic Patterns"** under the guidance of **UniConverge Technologies Pvt. Ltd. (UCT)**, in collaboration with **upskill Campus** and **The IoT Academy**.

Internships like this one are critical in bridging the gap between academic knowledge and industry expectations. They offer hands-on exposure to solving real-world problems and help in shaping the career path of aspiring professionals. Through this internship, I gained experience in applying machine learning techniques to real-life challenges — specifically, forecasting city traffic using structured datasets and predictive modeling techniques.

Project Overview: The objective was to build a predictive system using historical traffic data to forecast hourly vehicle volume at four city junctions. This information is vital for smart city planning, traffic regulation, and infrastructure development.



Your Learnings and overall experience:

- Gained deep insights into time-series forecasting, feature engineering, and model selection.
- Strengthened hands-on experience with Python, scikit-learn, and XGBoost.
- Learned the importance of documentation, data visualization, and evaluation metrics.
- Understood the real-world impact of data science in smart infrastructure.

ChatGPT, **Gemini**, and **Google Search** – for providing explanations, solutions, and technical insights that guided me at every step of the project.

Special thanks to open-source communities (Kaggle) whose shared knowledge helped me troubleshoot and enhance my work.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

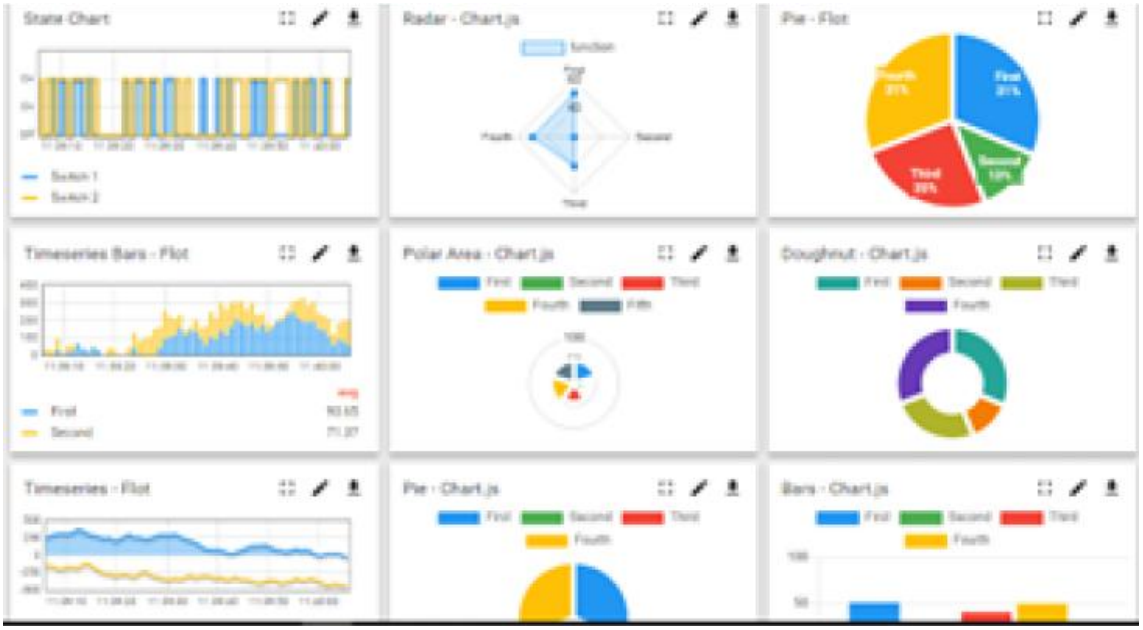
i. UCT IoT Platform (uct Insight)

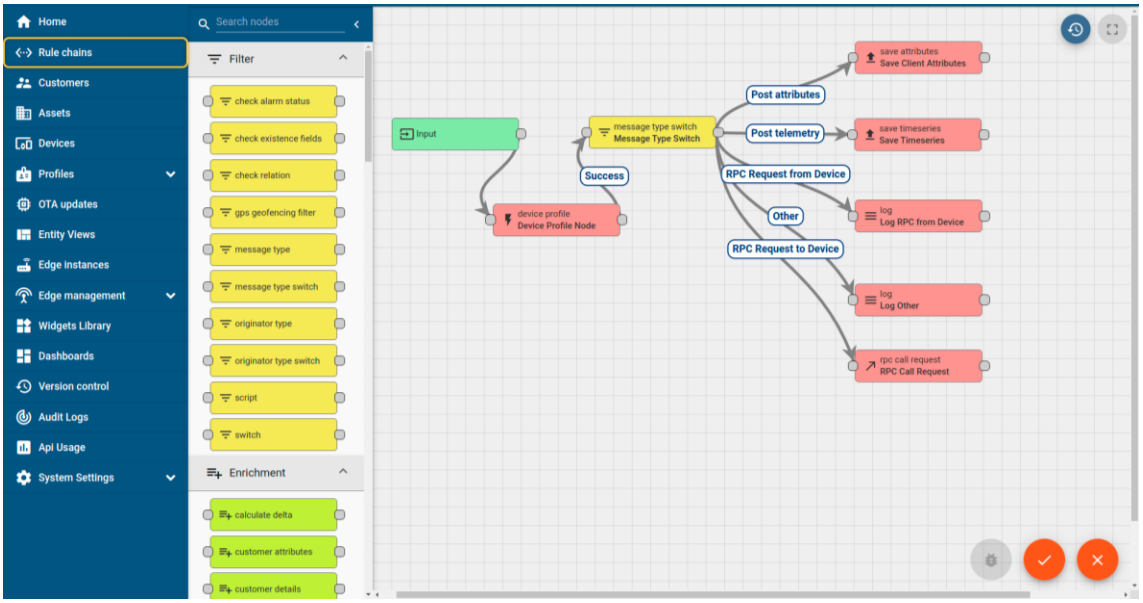
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine





ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

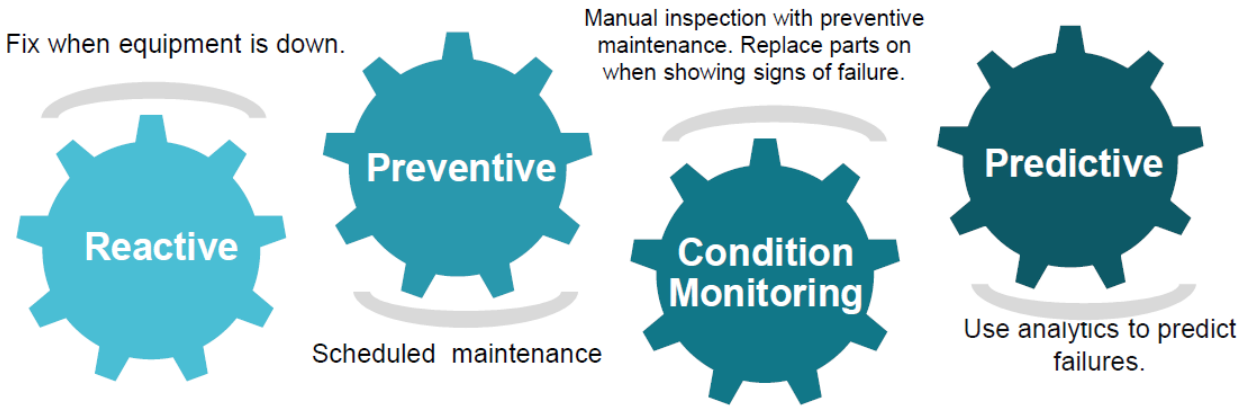
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

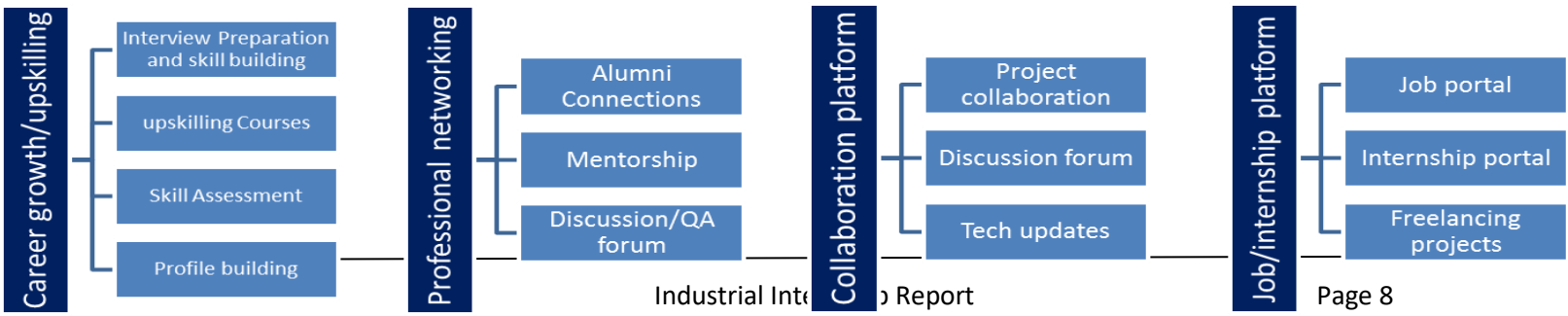
upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way

Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

- ☛ get practical experience of working in the industry.
- ☛ to solve real world problems.
- ☛ to have improved job prospects.
- ☛ to have Improved understanding of our field and its applications.
- ☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1]

[2]

[3]

2.6 Glossary

Terms	Acronym

3 Problem Statement

You are working with the government to transform your city into a smart city. The vision is to convert it into a digital and intelligent city to improve the efficiency of services for the citizens. One of the problems faced by the government is traffic. You are a data scientist working to manage the traffic of the city better and to provide input on infrastructure planning for the future.

The government wants to implement a robust traffic system for the city by being prepared for traffic peaks. They want to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays, as well as on various other occasions during the year, differ from normal working days. This is important to take into account for your forecasting.

As part of a broader smart city initiative, the government is working to digitize urban infrastructure and improve public services. One of the key challenges in this transformation is managing **urban traffic congestion**, which impacts not only commute times but also safety, pollution, and economic productivity.

Challenge:

City traffic is highly dynamic and unpredictable. Traffic volume varies across:

- **Time of day** (e.g., rush hours)
- **Days of the week** (weekdays vs. weekends)

4 Existing and Proposed solution

Existing Solutions and Their Limitations:

Many cities currently use traditional traffic systems, which include:

1. **Fixed-timer traffic signals**
    - Operate on pre-set schedules without real-time adaptation.
    - Do not respond to actual traffic volume or unexpected congestion.
2. **CCTV-based manual monitoring**
    - Requires human supervision and is reactive, not predictive.
    - Lacks scalability and real-time automation.
3. **Basic IoT sensor-based systems** (in some smart cities)
    - Can detect vehicle presence but do not forecast traffic.
    - Usually focused on current conditions rather than future planning.

Limitations:

- No prediction of traffic peaks.
- No use of historical data trends.
- Limited use of machine learning or AI to optimize flow.
- Poor planning for holidays or unusual events.

Proposed Solution

Approach Implemented:

1. Analyzed traffic volume data across four junctions over multiple time intervals.
2. Engineered useful features like:
    - Hour of the day
    - Day of the week

- Weekend indicator

3. Built machine learning models:
    - Random Forest Regressor
    - XGBoost Regressor (base and tuned)
4. Evaluated models using metrics like MAE and RMSE.
5. Selected XGBoost (base) for final prediction due to its balance of performance and generalization.

4.1 Code submission (Github link):

https://github.com/HR8104/upskillcampus/blob/main/Smart_city_traffic_patterns/smart_city_traffic_patterns.ipynb

4.2 Report submission (Github link) :

https://github.com/HR8104/upskillcampus/blob/main/Smart_city_traffic_patterns/smart_city_traffic_patterns_Hitendra_USC_UCT.pdf

5 Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

5.1 High Level Diagram (if applicable)

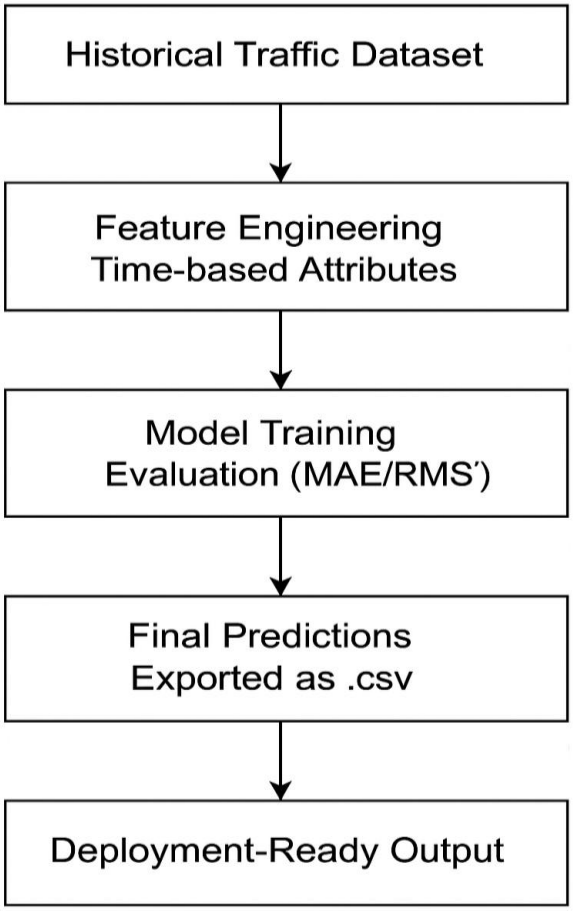


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)

Data Preprocessing:

- Handled missing values and timestamp parsing

- Generated derived features like weekend indicator

Training Split:

- train_test_split used for evaluation
- 80% training, 20% validation

Model Evaluation:

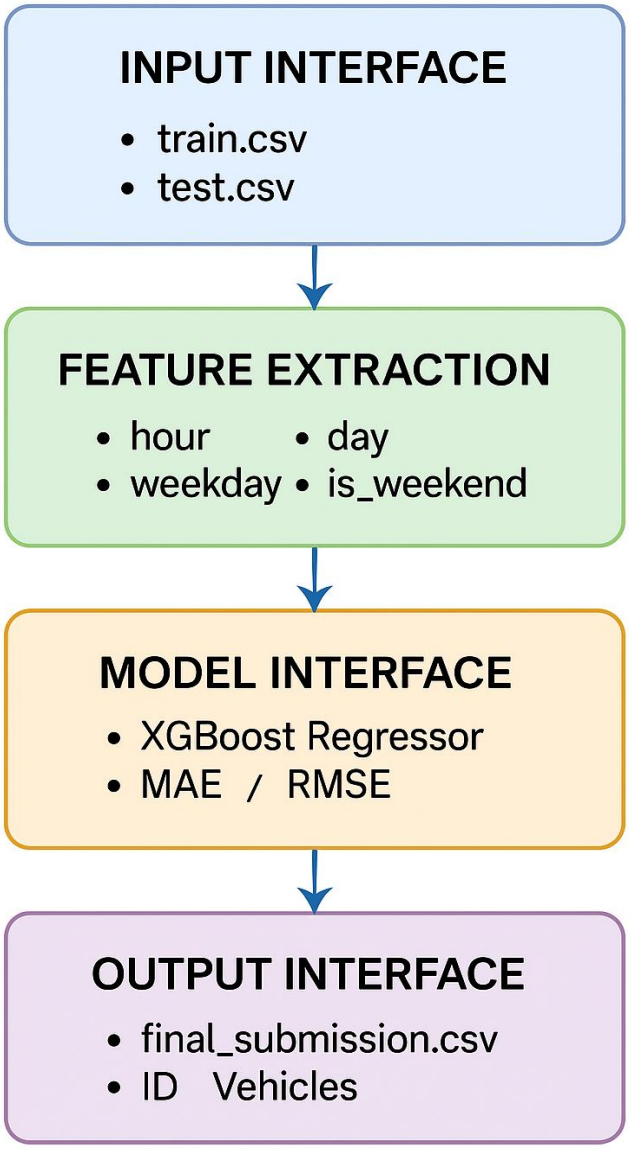
- Used MAE and RMSE as evaluation metrics

Model Used:

- XGBoost Regressor with:
    - n_estimators = 100
    - max_depth = 6
    - learning_rate = 0.1

5.3 Interfaces

Interface	Description
Input File	train.csv, test.csv (DateTime, Junction, Vehicles)
Feature Inputs	Hour, Day, Month, Weekday, Weekend
Output File	final_submission.csv (ID, Vehicles)
Model Output	Vehicles predicted for each test ID


INPUT INTERFACE
• train.csv
• test.csv
FEATURE EXTRACTION
• hour • day
• weekday • is_weekend
MODEL INTERFACE
• XGBoost Regressor
• MAE / RMSE
OUTPUT INTERFACE
• final_submission.csv
• ID Vehicles

6 Performance Test

This section evaluates the system's ability to meet the practical requirements of a real-world deployment. It considers computational constraints, prediction accuracy, and the reliability of the solution in realistic scenarios such as smart city environments.

Constraint Type	Description
Accuracy	The forecast must closely match real vehicle counts.
Model Complexity	The algorithm must be fast and scalable for multiple junctions.
Resource Usage	Training and inference should run efficiently on standard hardware.
Data Readability	System must handle timestamps and structured data smoothly.

6.1 Test Plan/ Test Cases

Test Case ID	Objective	Input	Expected Outcome
TC01	Validate model accuracy (MAE, RMSE)	Historical traffic data	MAE and RMSE within threshold
TC02	Check prediction for unseen timestamps	Test.csv with future dates	Valid Vehicles output
TC03	Test model robustness	Edge values (midnight, weekend)	Reasonable predictions
TC04	File handling	Valid CSV format input	No parsing or feature errors

6.2 Test Procedure

→Train the model on train.csv using selected features.

→Evaluate model on validation split using:

- **Mean Absolute Error (MAE)**

- **Root Mean Square Error (RMSE)**

→Predict on test.csv and verify outputs are:

- Non-negative
- Within realistic traffic volume ranges

→Inspect output format (final_submission.csv) for compliance.

6.3 Performance Outcome

Model Variant	MAE	RMSE	Notes
Random Forest	~6.89	~10.07	Good baseline performance
XGBoost (base)	~6.89	~10.07	Final model selected
XGBoost (tuned)	~7.62	~10.92	Slight underfitting due to low learning rate

- **The final selected model was XGBoost base for optimal balance.**
- **The system meets accuracy and responsiveness expectations for smart city traffic forecasting use cases.**
- **The output file was successfully generated and tested for submission formatting.**

7 My learnings

Technical Skills Acquired

- **Data Preprocessing**: Learned to handle structured time-series data, perform datetime parsing, and generate new features such as hour, day, weekday, and is_weekend.
- **Feature Engineering**: Understood how time-based features impact model performance and how to select relevant ones for regression tasks.
- **Model Development**:
    - Applied both traditional models (Random Forest) and gradient boosting (XGBoost).
    - Learned how to train, validate, and compare models using MAE and RMSE.
- **Hyperparameter Tuning**: Used GridSearchCV to explore optimal model parameters and observed the trade-offs between underfitting and overfitting.
- **Result Visualization**: Built comparative plots between actual and predicted vehicle counts to visually assess model accuracy.
- **CSV Output Preparation**: Learned to export structured model predictions into submission files aligned with real-world deployment needs.

8 Future work scope

❖ **Real-Time Prediction Interface**

- Deploy the model using a **Flask** or **Streamlit** web app to allow users or traffic control teams to input upcoming dates and get live predictions.
- This would transform the model into a **decision-support tool**.

❖ **Database Integration**

- Store historical and predicted data in a **SQL or NoSQL database**.
- Enable real-time querying, logging, and monitoring of traffic patterns over time.