Industrial Internship Report on

“Forecasting of Smart city traffic patterns”

Prepared by

Riya Saxena

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 internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time. My project was "Forecasting Smart City Traffic Patterns" that aimed to transform cities into smart urban cities using digital technologies. One of the problems faced by the government is traffic. To solve this issue and optimize the traffic flow I built a robust traffic system for the city. This model will provide insights for accurate traffic forecasts and helps in understanding the traffic patterns of the four junctions of the city. This project showcases my commitment to urban transformation and data-driven insights. 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 six-week internship journey into the realm of data science and machine learning, focused on the pivotal project of Forecasting Smart City Traffic Patterns.

During this internship, I delved into the intricate process of analyzing vast datasets comprising the historical traffic data over the last few years. Utilizing cutting-edge machine learning algorithms, I sought to develop a model that will provide insights for accurate traffic forecasts and helps in understanding the traffic patterns of the four junctions of the city.

This internship provided by USC/UCT with a dynamic learning environment, for leveraging technology to address real-world challenges. As I reflect upon the culmination of this internship journey, I am profoundly grateful for the opportunity to immerse myself in such a meaningful and impactful project.

Project name - Forecasting Smart City Traffic Patterns

India is the second-largest nation in the world in terms of both population and economic growth. The main cities of India suffer from severe traffic congestion. Sustaining Intelligent Transport Management Systems (ITMS) and improving urban areas are practical but difficult tasks in developed countries. This is one of the biggest problems that India is now experiencing. This is caused by a lack of available space, expensive expenses, and a slower rate of infrastructure expansion than the rapid increase in the number of vehicles. In order to enhance the efficiency of traffic operations and help travelers make better decisions when traveling through crowded regions, traffic flow information is necessary. A clever and sophisticated traffic information system is required to manage traffic congestion and alert vehicles to select other routes to avoid gridlock and save time. For this reason, intelligent traffic systems (ITS) are essential in India's metropolitan areas.

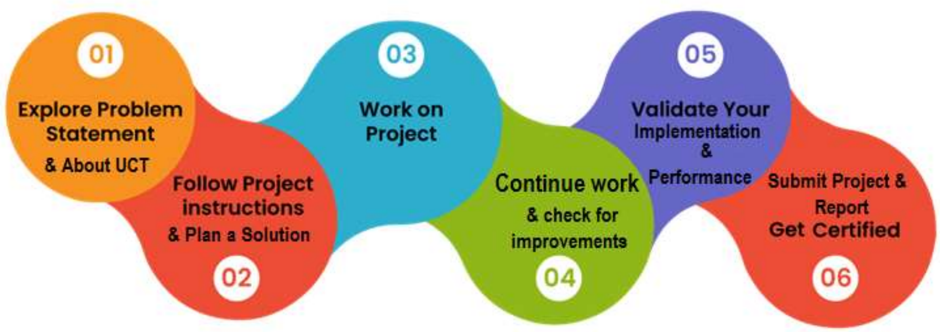
Data science is the study of data with the goal of gaining important business insights. In order to examine enormous amounts of data, this multidisciplinary approach incorporates ideas and methods from the domains of mathematics, statistics, artificial intelligence, and computer engineering. Machine learning is a subset of artificial intelligence (AI) that enables computers to learn from data without being explicitly taught, whereas data science combines tools, methodologies, and technology to derive meaning from data. Because it can find patterns and relationships in enormous volumes of data and base forecasts on these relationships, it is the perfect tool for analyzing the traffic flow of data.

Program Planning:

The program was meticulously planned, offering a structured framework that seamlessly integrated theoretical knowledge with practical application. From foundational concepts in data science to advanced machine learning algorithms, each week was thoughtfully crafted to build upon the previous learning, culminating in a comprehensive understanding of predictive analytics.

Learning:

The journey unfolded with an exploration of data preprocessing techniques, where I learned to cleanse, transform, and prepare raw data for analysis. As I progressed, I delved into the intricacies of regression analysis, time series forecasting, and ensemble learning methods, honing my skills in developing predictive models.



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.

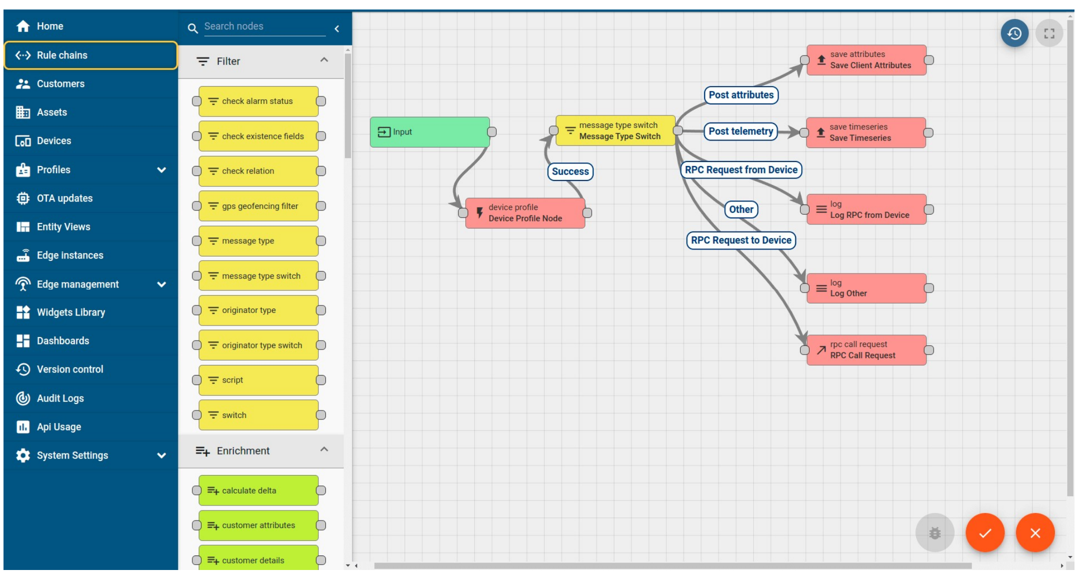
i. UCT IoT Platform( )

UCT Insight 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.

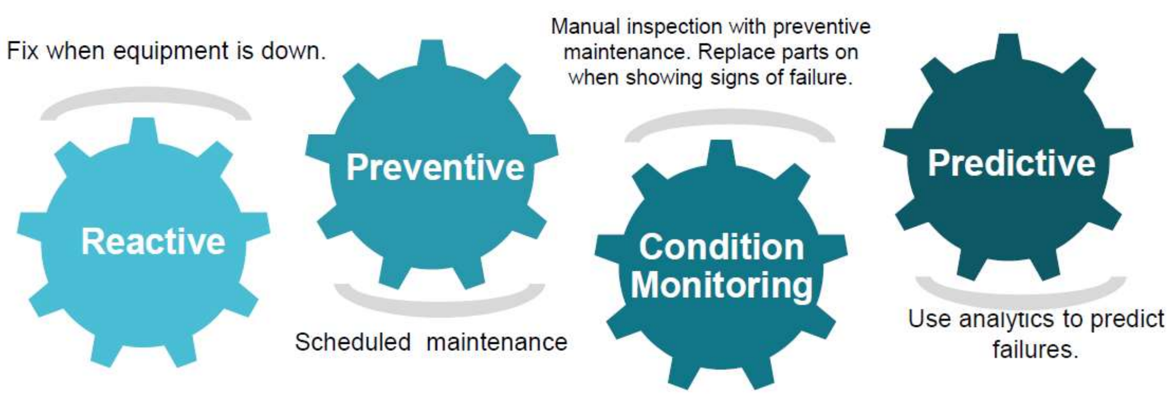
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_57_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_57_81	Operator 1	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 LoRAWANteschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

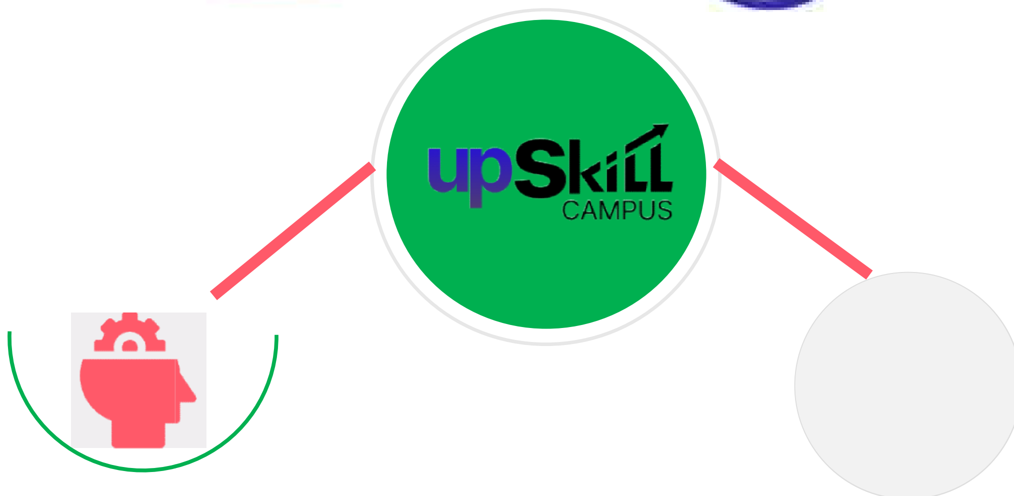
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

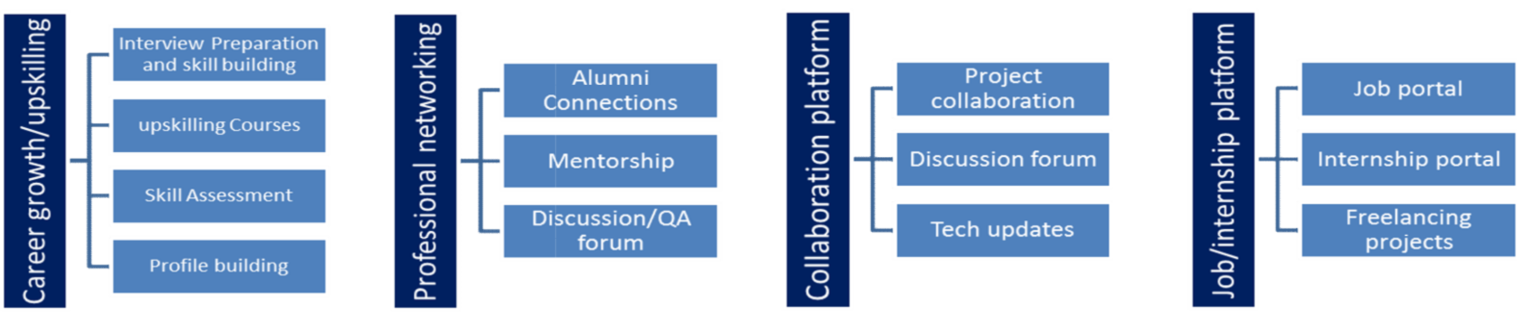
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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]https://learn.upskillcampus.com/s/mycourses

[2] https://www.uniconvergetech.in/

[3] https://upskillcourses.com/

3 Problem Statement

With cities all around the world expanding and changing, traffic congestion is becoming a major problem. The growing number of cars on the road, urbanization, and the quickening pace of population growth all contribute to the worsening of this issue. Smart city projects are being put into action to use technology and data-driven solutions to solve these issues. The capacity to precisely estimate and predict traffic patterns is a critical component of smart city design. Therefore, the problem statement revolves around developing effective forecasting models for smart city traffic patterns. T he objective is to develop a system that can predict traffic patterns in a smart city setting with accuracy. Predicting traffic congestion levels, journey durations, and traffic flow along different routes, junctions, and road networks are all part of this process.

4 Existing and Proposed solution

Provide summary of existing solutions provided by others, what are their limitations?

1. Traditional Statistical Models:

   Existing solutions often employ traditional statistical models such as autoregressive integrated moving average (ARIMA) or exponential smoothing techniques.

   Limitations: These models typically assume linear relationships and fail to capture the complexities of traffic patterns affected by various factors. They may struggle to handle large-scale data and real-time updates, limiting their accuracy and scalability.

2. Machine Learning (ML) Approaches:
   To forecast traffic patterns, machine learning (ML)-based systems use methods like gradient boosting, random forests, and support vector machines (SVM).

Limitations: These methods frequently rely significantly on feature engineering, which may be time-consuming and needs domain knowledge. Predictions may be less than ideal as a result of their inability to manage the dynamic and nonlinear interactions seen in traffic data.

Proposed Solution:

My approach aims to determine the traffic patterns by utilizing cutting-edge techniques and taking into account relevant aspects of traffic dataset. This proactive method visualizes and analyzes the traffic pattern at all the four junctions.

I go through many machine algorithms that are good for forecasting smart city traffic patterns and at the end I observed that the Decision Tree algorithm works great for the forecasting the traffic patterns.

A well-liked machine learning algorithm for classification and regression is the decision tree algorithm. The result is a model that takes the shape of a tree, with each internal node standing in for a feature or attribute, each branch standing in for a decision rule, and each leaf node standing in for an outcome or a class label.

The decision tree algorithm works in following steps:

First it prepares the data by creating a labeled dataset with target labels that correlate to the input features (the dependent variable) as well as independent variables. Then it selects the attributes to maximize the homogeneity of the target labels within each subset. After that algorithm divides the data recursively according to the chosen feature, starting at the root node. Until a halting requirement is satisfied. It also handles the missing values and at last the prediction and analyzing of data is done.

The decision tree algorithm offers a flexible and understandable method for predicting traffic patterns, allowing you to obtain knowledge of the underlying variables affecting traffic and anticipate future behavior using data from the past.

4.1 Code submission (Github link)

https://github.com/riyagh/upskillcampus/blob/main/Forecasting_of_smart_city_traffic_patterns.ipynb

4.2 Report submission (Github link)

https://github.com/riyagh/upskillcampus/blob/main/Forecasting_of_smart_city_traffic_patterns_Riya_USC_UCT.pdf

5 Proposed Design/ Model

Designing a solution/model for forecasting traffic patterns using the machine learning techniques involves several steps. Here's a proposed design flow to help you get started:

1. Data Collection: Gather historical traffic data including features like time of day , day of the week, weather conditions, holidays, road construction, etc.

2. Data Preprocessing: Clean the data by removing duplicates, handling missing values, and addressing outliers. Perform feature engineering to extract meaningful features.

3. Splitting the Data: Divide the preprocessed data into training and testing sets.

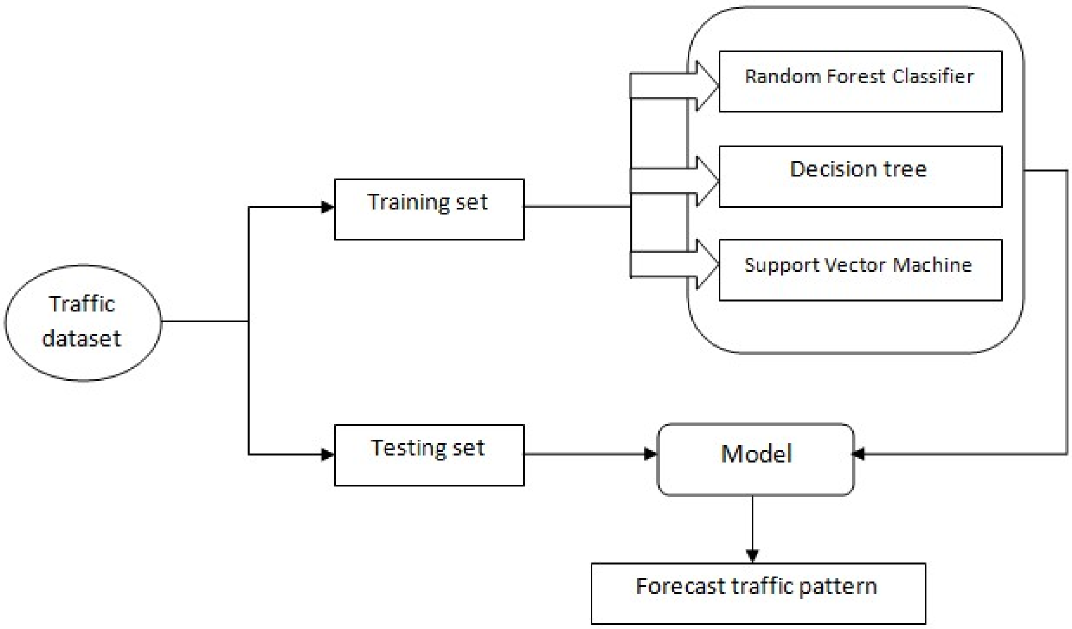
4. Training the Model: Apply the machine learning algorithms to the training data. Select the best feature to split the data at each internal node based on criteria like information gain. Tune hyperparameters to avoid overfitting or underfitting.

5. Model Evaluation: Evaluate the model using the testing data. Calculate metrics such as accuracy, precision, recall, or mean absolute error (MAE).

6. Model Deployment: Deploy the model for traffic pattern forecasting using the entire dataset for training. Consider retraining or updating the model periodically as new data becomes available.

7. Prediction and Monitoring: Use the deployed model to forecast traffic patterns based on new input data. Continuously monitor the performance of the model and consider necessary adjustments.

5.1 Interface



6 Performance Test

The Forecasting Smart City Traffic Patterns model cannot be successfully predicted without the use of predictive models. With the use of cutting-edge machine learning methods and resources, we set out to build a solid model that could examine the traffic patterns. Among the methods employed, the 'train_test_split' function from'sklearn.model_selection' was essential to our prediction models' performance.

The 'train_test_split' allowed us to split our dataset more easily into two subsets: one for model training and another for performance testing. Through random data splitting and assigning a larger percentage to the training set, we ensured that the model identified patterns from a significant amount of data while preserving unseen samples for assessment.

The 'sklearn preprocessing' features made it easy for us to apply a range of preprocessing approaches, such as one-hot encoding and normalization, to get our data ready for modeling.

Using rigorous experimental and cross-validation methods, we assessed these models' performance on the test set using metrics including mean squared error (MSE), mean absolute error (MAE), and R-squared (R2) score.

6.1 Test Plan/ Test Cases

Test Plan for Forecasting Traffic Patterns model:

1. Test Objective: Verify the precision and dependability of different machine learning algorithms for traffic pattern forecasting model.
2. Test Environment:
   - Programming language: Python
   - Libraries: scikit-learn, Pandas, NumPy
   - Traffic data set: Historical traffic data containing features such as date, time, weather conditions, etc.
3. Test Data:
   - Prepare a dataset with historical traffic data, including known patterns and corresponding outcomes.
   - Split the dataset into training and testing sets (e.g., 70% training, 30% testing).

4. Test Cases:

   a. Data Preprocessing:

   i. Verify that the dataset is loaded correctly, and all required features are present.

   ii. Check for missing or invalid values and ensure proper handling or imputation.

   iii. Validate the normalization or scaling of numerical features if applicable.

   b. Model Training:

   i. Train the model using the training dataset.

   ii. Verify that the model has been trained successfully without any errors.

   iii. Validate that the model has learned patterns and relationships from the training data.

c. Model Evaluation:

i. Apply the trained model to the testing dataset.

ii. Compare the predicted traffic patterns with the actual patterns in the testing dataset.

iii. Calculate evaluation metrics such as accuracy, precision, recall, and F1-score.

iv. Ensure that the model performance meets the defined acceptance criteria.

d. Model Validation:

i. Use cross-validation techniques (e.g., k-fold cross-validation) to assess the model's generalization capabilities.

ii. Verify that the model performs consistently across different subsets of the data.

iii. Ensure that the model does not overfit or underfit the training data.

e. Performance Testing:

i. Measure the training and prediction times to ensure they are within acceptable limits.

ii. Evaluate the model's performance on large datasets to validate scalability.

f. Boundary and Edge Cases:

i. Test the model's behavior with extreme or outlier values in the input features.

ii. Verify that the model handles unexpected or novel traffic patterns gracefully.

g. Integration Testing:

i. Validate the integration of the Decision Tree algorithm with other components or systems.

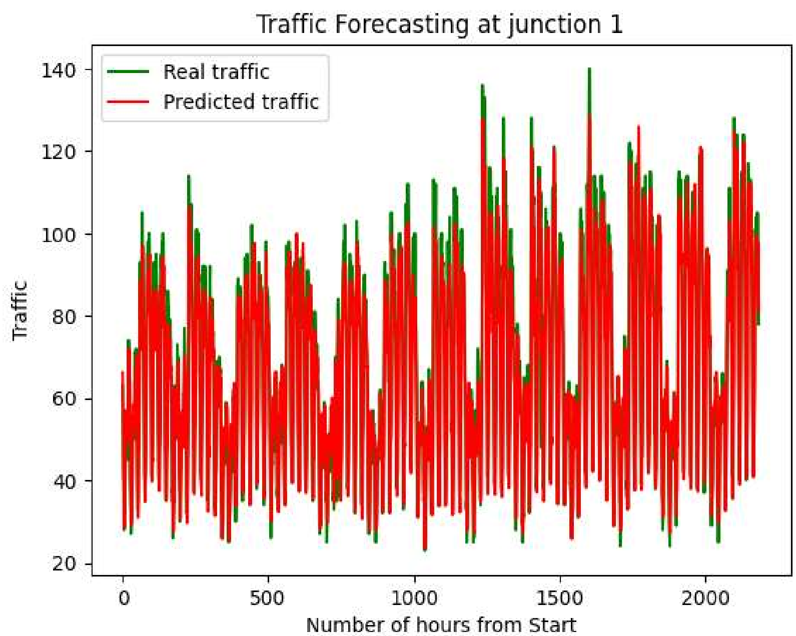
ii. Verify the compatibility and data exchange between the traffic forecasting model and external systems.

6.2 Performance Outcome

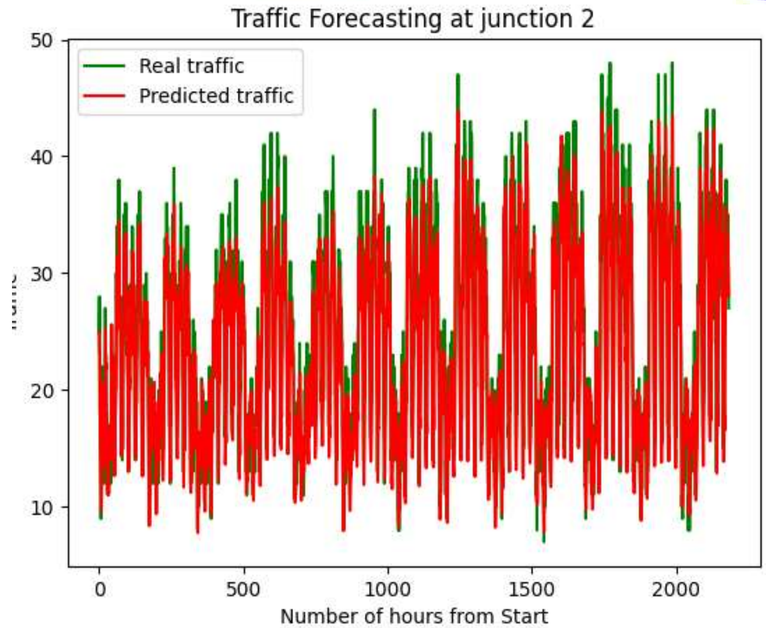
The performance outcome of my forecasting traffic patterns model can vary depending on several factors, including the quality of the data, the complexity of the traffic patterns, the choice of features, and the tuning of the model parameters.

Visualizing Prediction of traffic at different Junction:

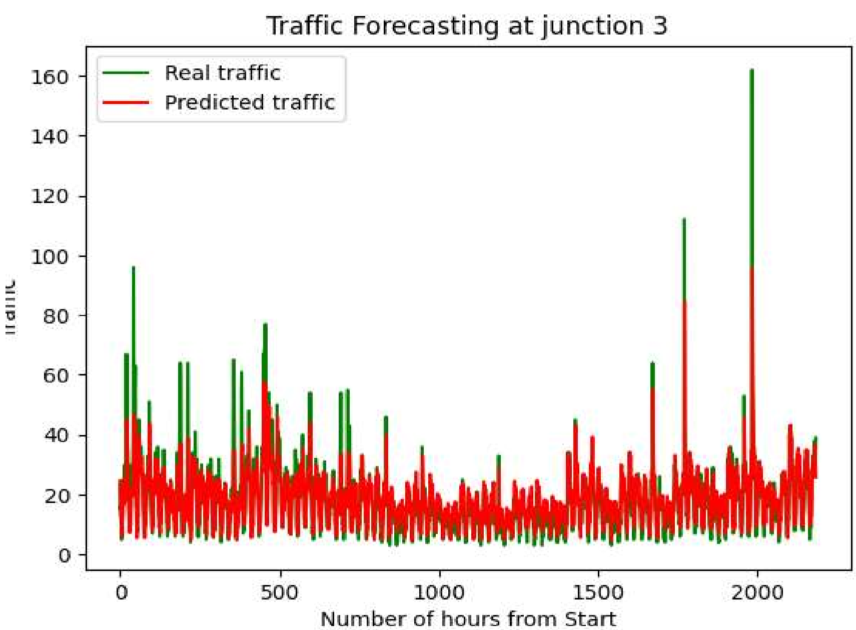
- Junction 1:



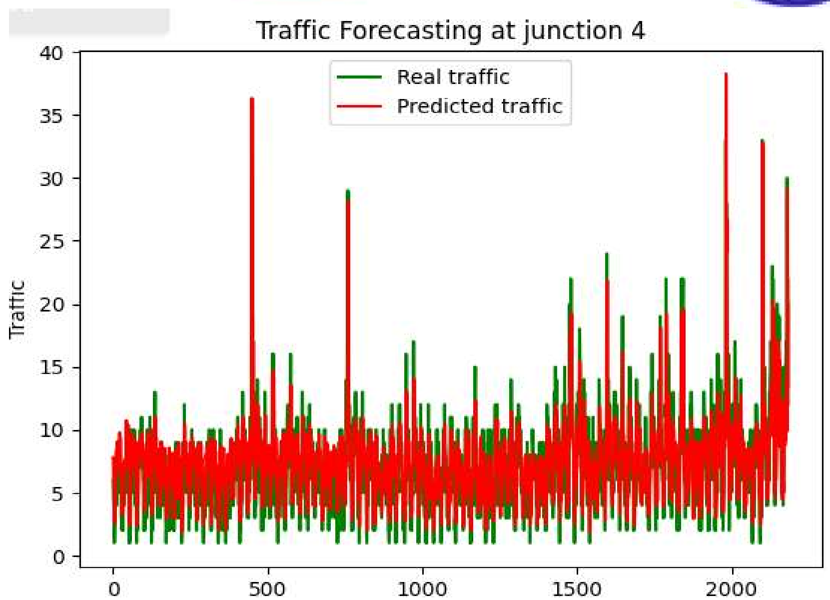
- Junction 2:



- Junction 3:



- Junction 4:



Accuracy of different algorithms:

- Random forest classifier:

  *Accuracy score for RF: 81.21363258520367

- Decision tree classifier:

  *Accuracy score for DT: 81.21363258520367

- Support Vector Machine (SVM):

  *Accuracy score for SVM: 97.45635910224439

7 My Learnings

During my six-week internship in data science and machine learning, I had the opportunity to work on a project focused on Forecasting of Smart city traffic patterns. This project not only exposed me to the intricate idea of Smart city but also allowed me to apply various data science techniques and machine learning algorithms to solve real-world problems. By applying data science techniques and machine learning algorithms to real-world problems, I not only enhanced my technical skills but also gained a deeper appreciation for the role of data-driven solutions in addressing societal challenges. This internship has inspired me to continue exploring the intersection of data science.

8 Future work scope

The Decision Tree Classifier has shown promising results in improving algorithmic efficiency and accuracy. However, there is room for improvement in fine-tuning hyperparameters and exploring ensemble techniques. Integrating real-time data sources, such as live GPS feeds, weather conditions, and special events, can enhance the predictive capabilities of the models. AI-based anomaly detection can help identify unusual traffic patterns or incidents, contributing to proactive traffic management and emergency response systems. Expanding the model to multiple cities and integrating it with smart traffic signal systems can lead to automated adjustments in signal timings, reducing congestion and optimizing traffic flow.

Limitations include data quality and availability, static feature set, model complexity, external factors, and human behaviour variability. Accurate traffic prediction relies on high-quality and diverse datasets, but the model's limitations highlight the need for ongoing development to create a comprehensive and adaptable traffic management solution

As we look towards the future, the potential for enhancing predictive models for forecasting traffic patterns is vast. Building upon the foundation laid by our existing model development efforts, several avenues for further research and innovation emerge. Additionally, the integration of a user interface (UI) adds a layer of accessibility and usability.