Industrial Internship Report on

” Forecasting of Smart City Traffic Patterns”

Prepared by

Pushkar.V.Salaskar

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. Forecasting smart city traffic patterns involves predicting future traffic conditions, such as traffic volume, congestion levels, and travel times, in a city using historical data and other relevant factors. The goal is to develop accurate and efficient models that can anticipate traffic patterns and aid in urban planning, traffic management, and resource allocation. Various machine learning (ML) techniques can be employed to address this challenge. Adaptive Traffic Management: ML models can be integrated into traffic management systems to adjust traffic signals, reroute vehicles, and optimize traffic flow based on real-time predictions. Predictive Route Planning: Smart city applications can provide drivers with real-time traffic predictions. By integrating ML techniques into smart city traffic management, cities can improve overall transportation efficiency, reduce congestion, lower emissions, and enhance the overall quality of life for residents and visitors alike. 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

Summary of the whole 6 weeks' work.

During the first week of internship, I choose topic Traffic prediction as my job and I researched and identified the key components and requirements for the project.

In the second week I tried and figured out which all Ml techniques can be applied to the project to have its best fit and gathered a inform about various libraries used in python

Coming on third week I started implementing the techniques on the model and began setting up the necessary libraries. Also worked on other models whether they would be a better fit for the project.

On the fourth- and fifth-week model was run and various error was fixed. Made a note which all model could be the best for my topic and with the output compared three of them.

During the final week of the internship, I focused on removing the unnecessary components elements which would affect the mode. Finally, I came to conclusion which is the best model for my topic.

About need of relevant Internship in career development.

Relevant internships play a crucial role in career development for several reasons.

1. **Potential Job Offers**: Many companies use internships as a recruitment tool to identify potential full-time employees. If an intern performs well during their internship, they may receive a job offer at the end of the program, leading to a direct entry into the job market.
2. **Understanding Workplace Culture**: Internships provide insights into the culture and dynamics of different workplaces. This understanding helps interns assess whether the company's work environment aligns with their values and preferences.
3. **Career Focus**: Internships can help individuals solidify their career goals and aspirations. They allow interns to explore different roles and industries, helping them make informed decisions about their career path.
4. **Industry Exposure**: Internships expose individuals to the industry they are interested in. This exposure helps them gain insights into the workings of the industry, current trends, challenges, and opportunities. It also provides a clearer understanding of the career path they want to pursue.
5. **Networking:** During internships, individuals get a chance to build a professional network within the industry. Connecting with experienced professionals, mentors, and colleagues can open doors to future job opportunities and provide valuable guidance in career advancement.

6. **Building a Resume**: Having relevant internship experiences on a resume enhances its appeal to potential employers. It demonstrates that the candidate has practical experience and a genuine interest in the field, making them more attractive to hiring managers.

Brief about Your project/problem statement.

Forecasting smart city traffic patterns involves predicting future traffic conditions and trends in urban areas using historical traffic data, real-time information, and other relevant factors. The goal is to develop accurate models that can anticipate traffic patterns to assist in urban planning, optimize traffic flow, and improve overall transportation efficiency.

Key steps in forecasting smart city traffic patterns:

1. **Data Collection**: Gathering data is fundamental to forecasting. It includes historical traffic data, traffic counts, weather conditions, special events, road closures, and data from sensors installed throughout the city.
2. **Preprocessing**: Collected data often needs preprocessing to handle missing values, outliers, and noise. Data normalization and feature engineering are done to ensure the data is in a suitable format for modeling.
3. **Time Series Analysis**: Smart city traffic patterns exhibit temporal dependencies. Time series analysis methods, such as Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving-Average (SARIMA), or Prophet, are employed to predict traffic patterns over time.
4. **Machine Learning Models**: Various machine learning algorithms like regression models, neural networks, and ensemble techniques can be used to capture complex relationships between traffic patterns and external factors (e.g., weather, events) for more accurate predictions.
5. **Real-time Data Integration**: To account for dynamic traffic conditions, real-time data feeds from traffic cameras, GPS devices, and other sensors are integrated into the models, enabling continuous updates and adapting to changing situations.
6. **Model Evaluation**: The performance of the forecasting models is assessed using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or R-squared to measure prediction accuracy.
7. **Visualization and Interpretation**: Visualizing forecasted traffic patterns helps decision-makers interpret the results and identify areas that require attention or improvements.

Opportunity given by USC/UCT.

Upskill Campus (USC) or The loT Academy in collaboration with UniConverge Technologies Pvt Ltd (UCT) provides internships and collaborations with industry partners generally offer valuable opportunities for professional growth and development.

Participating in an internship or collaboration program with USC/UCT could provide the following opportunities:

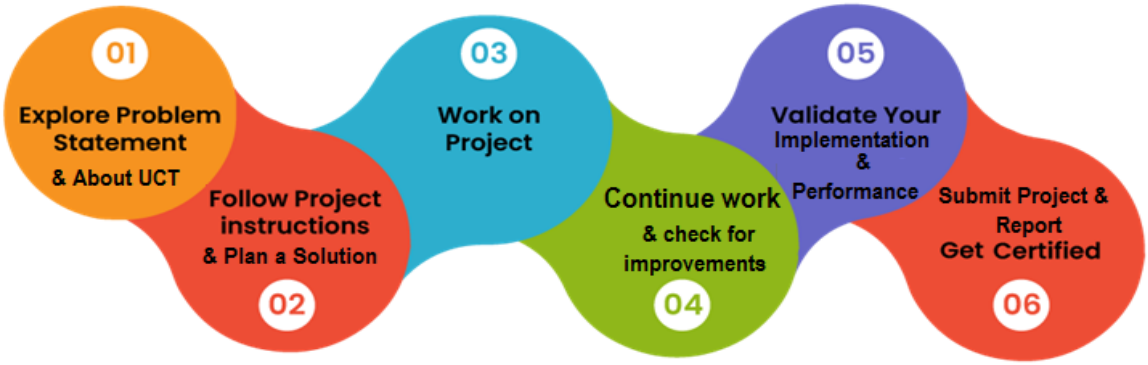
1. **Industry Exposure**: Working with a renowned institution and industry partner gives you exposure to real-world industrial problems, challenges, and practices. It allows you to gain insights into the operations and dynamics of the industry, which can be invaluable for your career

2. **Practical Experience**: Internships and collaborations offer hands-on experience, enabling you to apply your knowledge and skills in a practical setting. Working on projects or problem statements provided by industry partners helps bridge the gap between theory and practice. enhancing your understanding and proficiency in your chosen field.

3. **Networking**: Collaborating with USC, UCT, and their industry partner gives you the opportunity to build a professional network. You can connect with professionals, mentors, and experts in the industry, which can lead to future job prospects, references, and valuable connections.

4.**Skill Development**: Engaging in projects and problem-solving within the context of an internship or collaboration provides an avenue for skill development.

How Program was planned



Your Learnings and overall experience.

1. **Understanding of Traffic Patterns**: Working on this project would deepen your understanding of the complex and dynamic nature of traffic patterns in urban areas. Learned how various factors, such as time of day, weather conditions, events, and road infrastructure, contribute to traffic congestion and flow.
2. **Data Handling and Preprocessing**: Dealing with large and diverse datasets is a significant aspect of traffic pattern forecasting. Gained experience in data cleaning, data normalization, feature engineering, and handling missing values to prepare the data for modeling.
3. **Machine Learning Techniques**: Forecasting traffic patterns involves the application of various machine learning algorithms and time series analysis methods. Implemented Decision tree, Random Forest, SVM and ensemble techniques to achieve accurate predictions.
4. **Model Evaluation and Interpretation**: Evaluating the performance of forecasting models using metrics like MAE, RMSE, or R-squared has been part of the experience. Additionally, interpreting the results and visualizing forecasted traffic patterns would be essential for presenting findings to stakeholders.
5. **Urban Planning and Traffic Management**: Working on a smart city traffic patterns project helped me understand the critical role of data-driven decisions in urban planning and traffic management. Had a look how accurate forecasts can assist in optimizing transportation resources and improving overall city mobility.

   Overall, working on a Forecasting Smart City Traffic Patterns project offers a valuable learning experience that combines technical skills, practical applications, and an understanding of the societal impact of data-driven decision-making in smart cities.

Thanks to Upskill Campus (USC) or the loT Academy in collaboration with UniConverge Technologies Pvt Ltd (UCT) for this Internship Opportunity. Also special thanks to **Brutus , Nitish Sharma and kaushlendra Singh Sir** guiding throughout the internship Program

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)

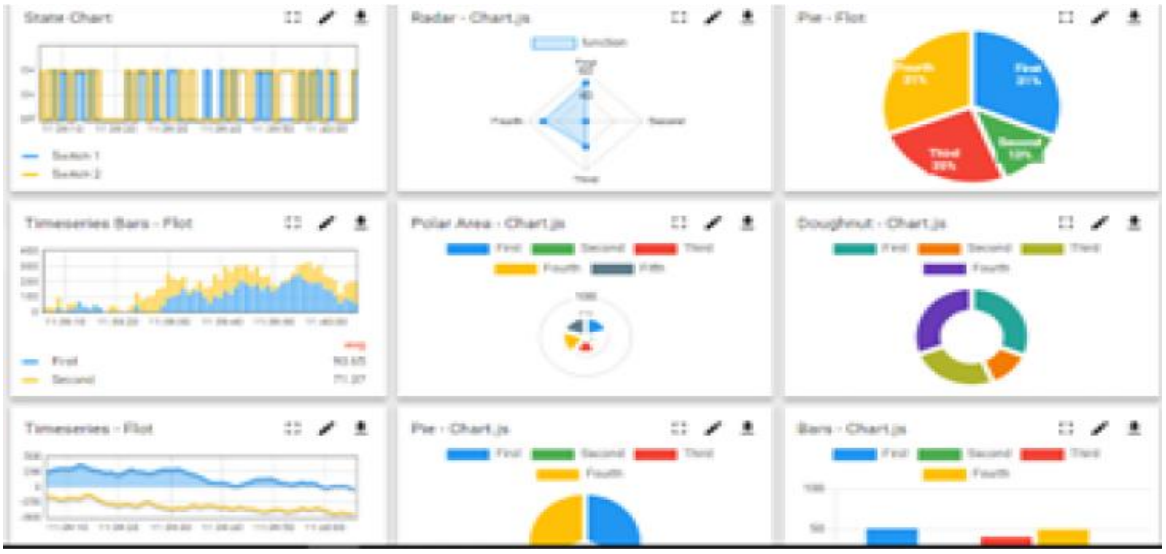
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

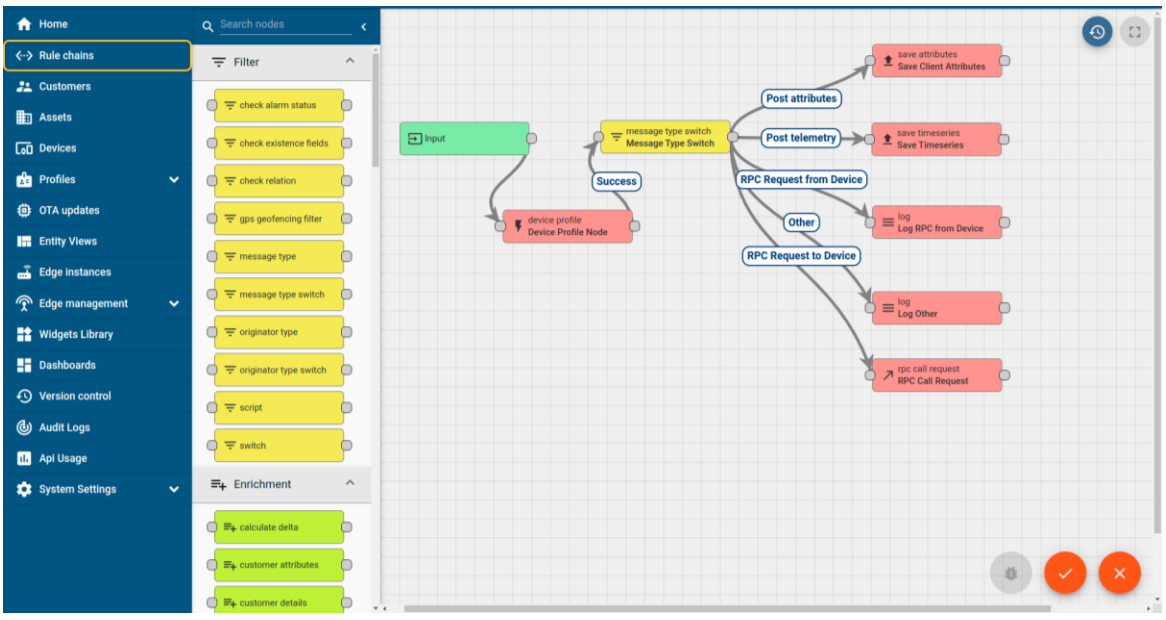
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- 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.

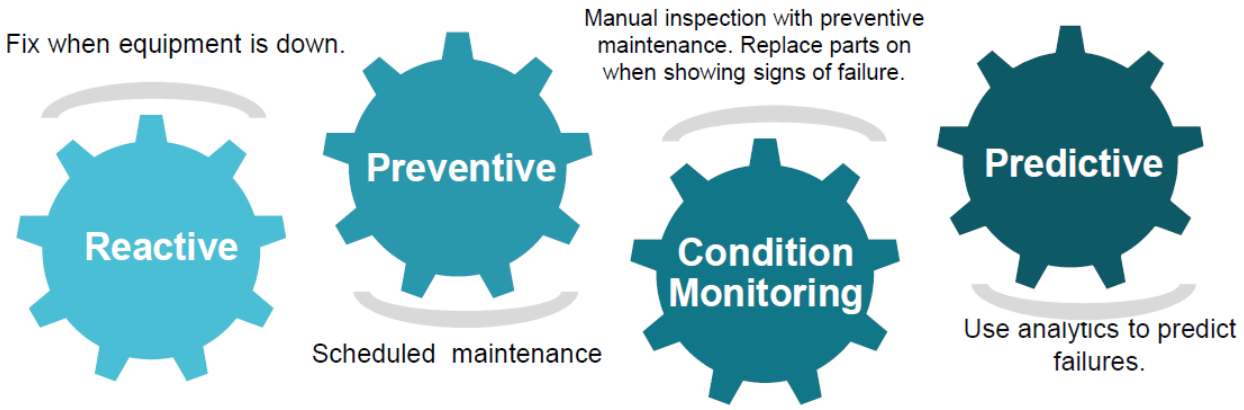
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

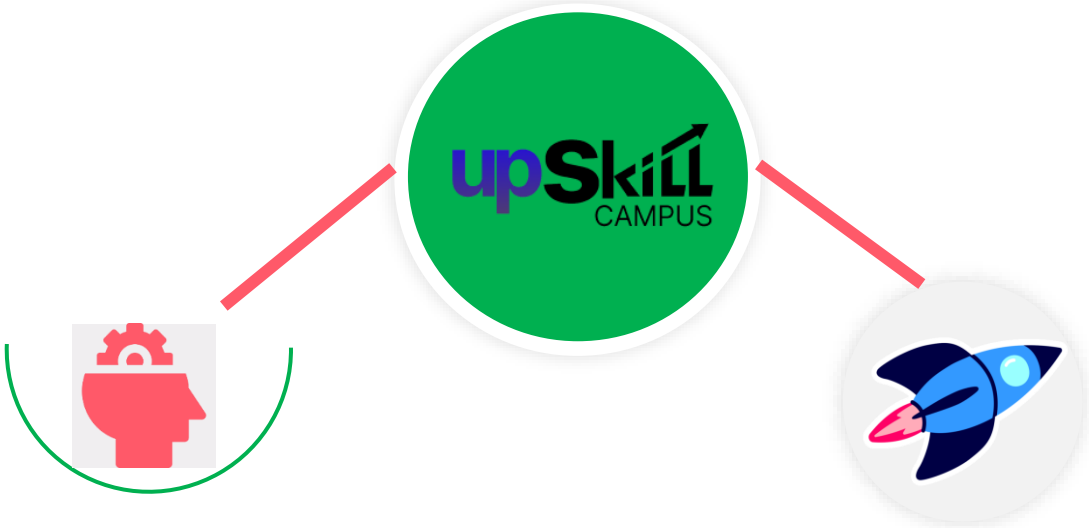
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

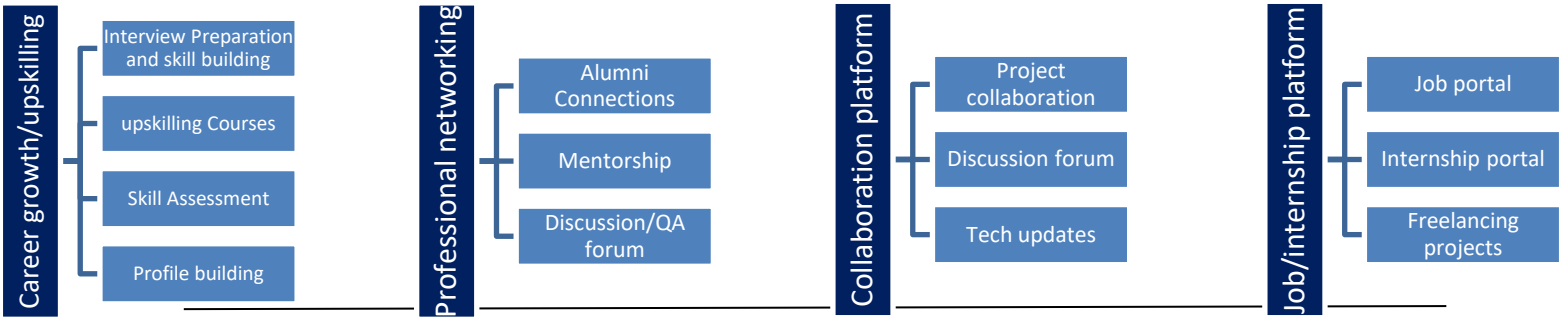
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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]https://www.researchgate.net/publication/363677388_Smart_City_Traffic_Patterns_Prediction_Using_Machine_Learning

[2] https://www.mdpi.com/2071-1050/14/7/4164

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9363163/

2.6 Glossary

Terms	Acronym
Support Vector Machine	SVM
Traffic Management System	TMS
Global Positioning System	GPS
Advanced Traffic Management System	ATMS
Advanced Traveler Information System	ATIS
Random Forest	RF
Decision Tree	DT

3 Problem Statement

In the assigned problem statement

The problem statement for Forecasting Smart City Traffic Patterns involves developing predictive models that can anticipate future traffic conditions, congestion levels, and travel patterns in urban areas. The main objective is to use historical traffic data, real-time information, and other relevant factors to forecast how traffic will behave in the future.

The problem can be further broken down into several key components:

1. Data Collection: Gather relevant data, including historical traffic flow data, weather conditions, special events, road closures, and data from various sensors installed throughout the city.
2. Preprocessing: Clean and preprocess the collected data to handle missing values, outliers, and noise. Data normalization and feature engineering may also be performed to prepare the data for modeling.
3. Machine Learning Models: Various machine learning algorithms and techniques can be employed to capture the relationships between traffic patterns and external factors, such as weather, time of day, and events.
4. Model Evaluation: The performance of the forecasting models should be evaluated using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), or R-squared to assess their accuracy in predicting traffic patterns.
5. Visualization and Interpretation: The forecasted traffic patterns should be visualized and interpreted to gain insights and communicate the results to stakeholders effectively.

The ultimate goal of Forecasting Smart City Traffic Patterns is to create accurate and reliable models that can aid urban planners, traffic management authorities, and policymakers in making data-driven decisions to optimize transportation resources, reduce congestion, and improve overall mobility and efficiency within the city.

4 Existing and Proposed solution

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

Summary of Existing Solutions for Forecasting Smart City Traffic Patterns:

1. **Time Series Analysis**: Some existing solutions rely on traditional time series analysis methods like ARIMA and SARIMA to forecast traffic patterns. These techniques capture temporal dependencies in traffic data and can provide reasonable short-term predictions. However, they might struggle with handling complex and non-linear patterns in traffic data, limiting their accuracy for longer-term forecasts.
2. **Machine Learning Models**: Many solutions incorporate machine learning algorithms, such as regression models, neural networks (e.g., LSTM), and ensemble methods (e.g., Random Forest, Gradient Boosting). These models can capture complex relationships between traffic patterns and various factors. They often outperform traditional time series analysis methods, especially for long-term forecasts.
3. **Real-time Data Integration**: Some solutions integrate real-time data feeds from traffic cameras, GPS devices, and sensors to update traffic forecasts continuously. This real-time approach improves the accuracy of predictions, but it relies heavily on the availability and reliability of data sources. In regions with limited sensor infrastructure or data quality issues, the accuracy of real-time forecasts may be compromised.

Limitations of Existing Solutions for Forecasting Smart City Traffic Patterns:

1. **Data Quality and Availability**: Many existing solutions heavily depend on the availability and quality of data, such as historical traffic data, weather conditions, and event information. Incomplete or inaccurate data can lead to less reliable predictions.
2. **Complexity of Traffic Patterns**: Urban traffic patterns can be highly complex, influenced by numerous interrelated factors. Traditional time series analysis methods might struggle to capture intricate non-linear patterns, resulting in suboptimal predictions.
3. **Sensitivity to Events**: Traffic patterns can drastically change due to unplanned events like accidents, protests, or road closures. While some solutions account for real-time data, sudden events may still disrupt predictions and lead to inaccurate forecasts.

4. **Scalability**: Machine learning models, especially deep learning techniques, might require significant computational resources and time for training, making them less scalable for larger cities or regions with vast traffic data.
5. **Interpretability**: Some machine learning models, particularly neural networks, are often considered black-box models, making it challenging to interpret their predictions. This lack of interpretability can be a limitation in gaining insights into the factors driving traffic forecasts.

What is your proposed solution?

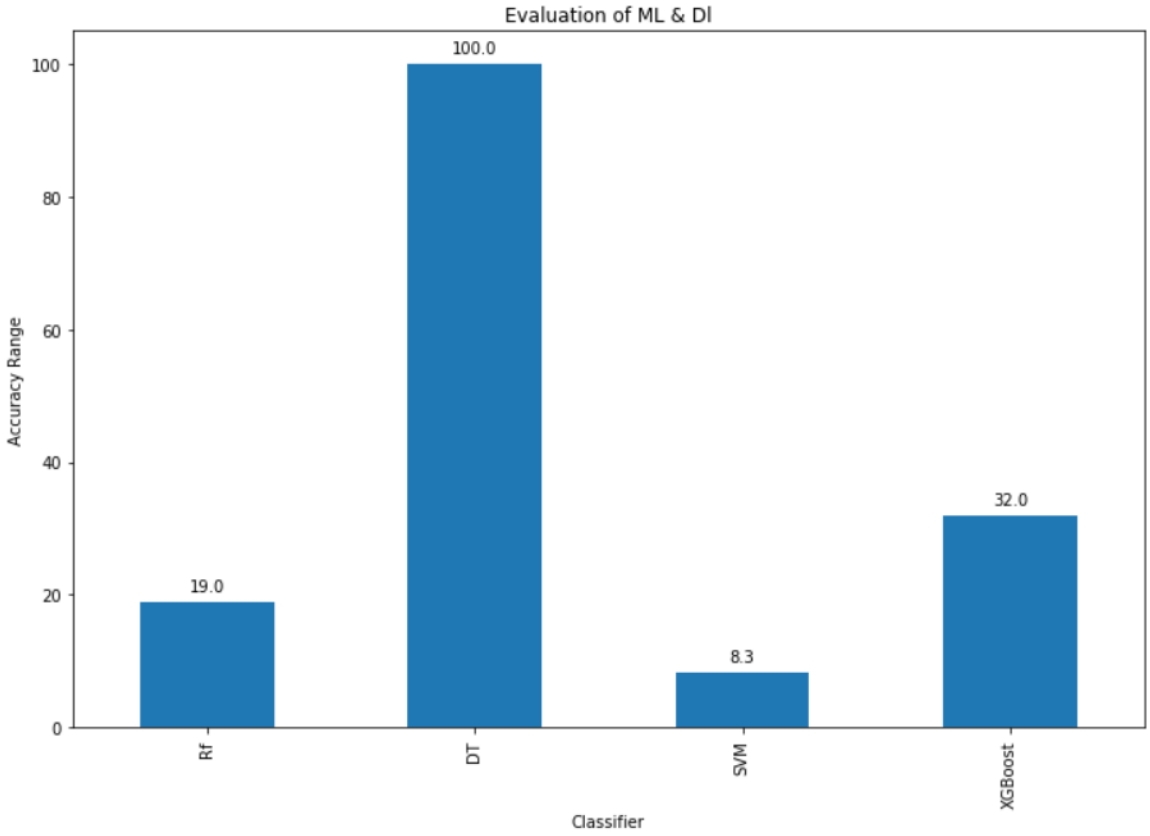
From my findings using the various machine learning techniques the decision tree fits the best for the model. For smart cities traffic flow prediction is the most crucial task. Precise traffic forecasting can help drivers manage their trips effectively. To accurately predict traffic flow, this study initially combined pollution and traffic datasets of Aarhus, Germany. Then different conventional ML approaches were used on the dataset to find out the most accurate approach among them. KNN had the least MAE and RMSE values among them. After observing the results of conventional approaches, the bagging and stacking ensemble was used to improve the MAE and RMSE values. Bootstrapping was used with replacement to split the dataset into samples. The samples were fed into the different number of homogeneous models and their result was aggregated to form a strong bagging ensemble model.

4.1 Code submission (Github link)

https://github.com/Pushkar220901/DS-ML-/blob/main/DS%20%26%20ML%20.py

4.2 **Report submission (Github link) :** first make placeholder, copy the link.

5 Proposed Design/ Model



5.1 High Level Diagram (if applicable)

5.2 Low Level Diagram (if applicable)

5.3 Interfaces (if applicable)

Update with Block Diagrams, Data flow, protocols, FLOW Charts, State Machines, Memory Buffer Management.

6 Performance Test

This is very important part and defines why this work is meant of Real industries, instead of being just academic project.

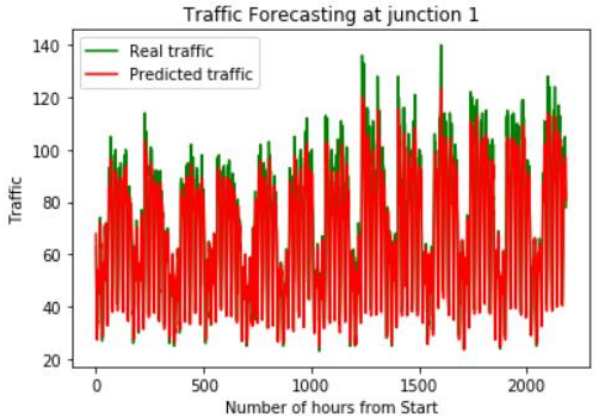
Various model techniques were used to compare the real traffic and the predicted traffic. Model such as DT, RF, SVM were used to compare the traffic forecasting at juunction1 similarly at junction 2 followed by junction 3 and junction 4

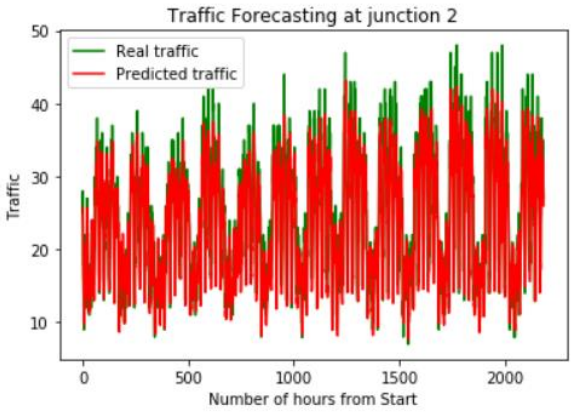
6.1 Test Plan/ Test Cases

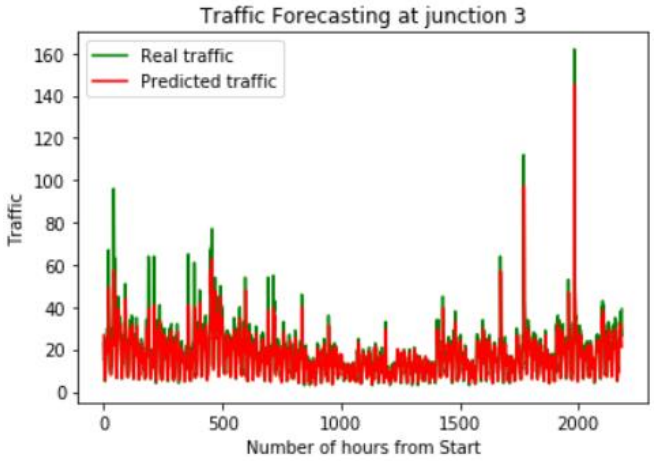
6.2 Test Procedure

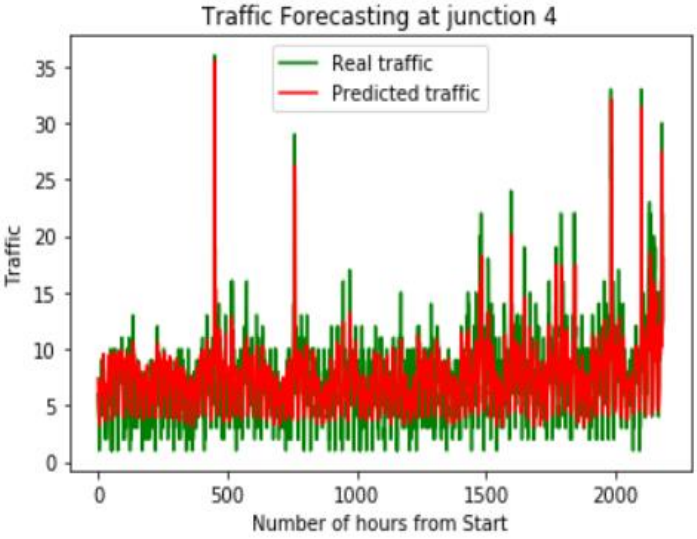
Firstly, data was inputted through csv file using various libraries and framework such as matplotlib etc. Then EDA was performed on the data for what type of data & how many is the length of the vehicles and then we find out the lag-features. So basically, the lag features are basically the target variable but shifted with a period of time, it is used to know the behaviour of our target value in the past, maybe a day before, a week or a month. We found that the data consisted of 14592rowsx4 columns then null and info was performed on the train data . Then we generated the required lag features and made use of Minimaxscaler which shrinks the data within the given range, usually of 0 to 1. Further the data was splited and normalized the data and reshaping of the data was done after that modelling was performed and TensorFlow was used backend and fitting of the model was done after that validating of the data was done followed by visualizing the data.

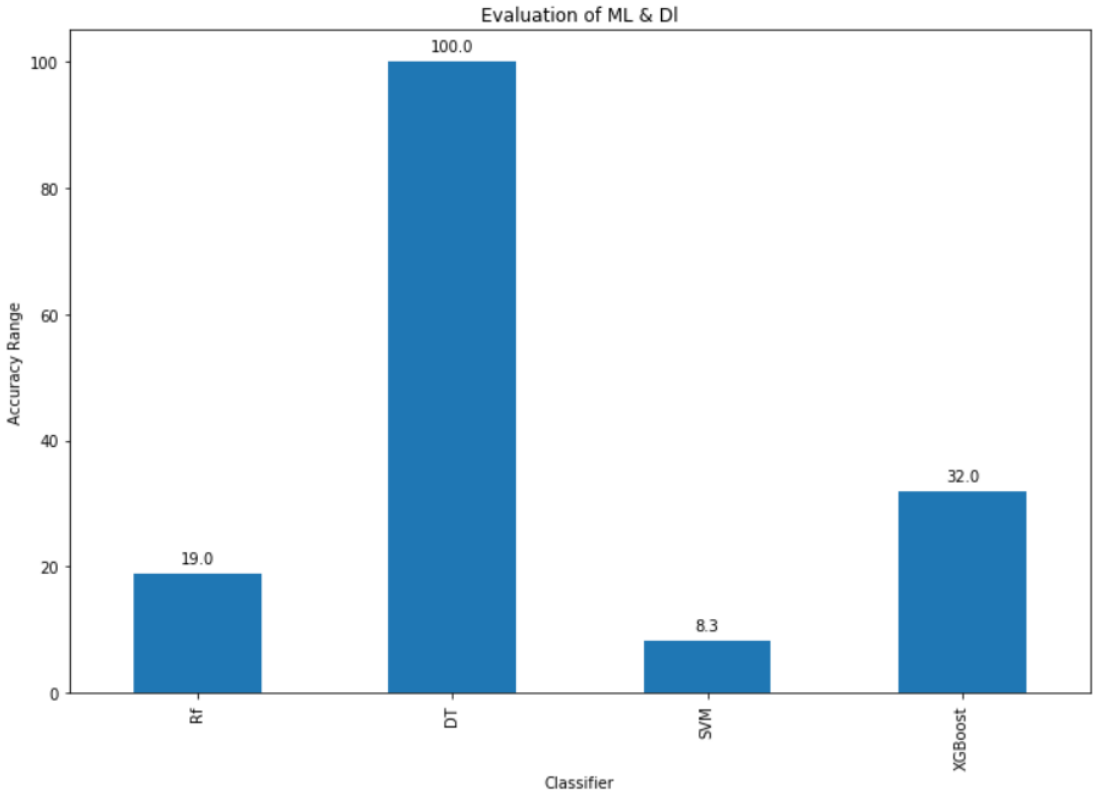
6.3 Performance Outcome


Traffic Forecasting at junction 1
Real traffic
Predicted traffic
Traffic
Number of hours from Start


Traffic Forecasting at junction 2
Real traffic
Predicted traffic
Traffic
Number of hours from Start


Traffic Forecasting at junction 3
Real traffic
Predicted traffic
Traffic
Number of hours from Start


Traffic Forecasting at junction 4
Real traffic
Predicted traffic
Traffic
Number of hours from Start



7 My learnings

All the four-accuracy score was taken out for the model.

Name of the ML techniques	Accuracy Rate	RMSE
Random Forest	94.835827	8.35261
Decision tree classifier	500	2.355
SVM	41.31338320	7.356
XGBoosting	159.83	9.523

8 Future work scope

Forecasting Smart City Traffic Patterns is an evolving field with plenty of opportunities for future work and research. Here are some potential future work scopes to further improve and advance traffic forecasting in smart cities:

1. **Data Augmentation and Fusion**: Explore techniques to augment existing traffic data with simulated data or use data from other sources (e.g., social media, mobile apps) to enhance the dataset's diversity and completeness. Investigate data fusion methods to combine data from various sensors and sources for a more comprehensive view of traffic patterns.
2. **Deep Learning Architectures**: Continue exploring advanced deep learning architectures and novel neural network architectures to improve the accuracy and robustness of traffic forecasting models. Research into attention mechanisms and transformer-based models could offer new insights.
3. **Multimodal Data Integration:** Integrate various types of data, including traffic data, public transportation data, pedestrian flow data, and bicycle usage data, to build comprehensive models that consider all modes of transportation in a smart city.
4. **Adaptive and Online Learning**: Develop forecasting models that can adapt to changing traffic patterns and update in real-time as new data becomes available. Online learning techniques can enable models to continuously learn and adapt from incoming data streams.
5. **Uncertainty Quantification**: Investigate methods for quantifying uncertainty in traffic forecasts, which is crucial for decision-making and managing risk in smart city traffic management systems.