Industrial Internship Report on

”Smart City Traffic Patterns”

Prepared by

Krisha Panwala

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 Smart City Traffic Patterns. 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 summarizes the six weeks of work undertaken as part of my internship, focusing on analyzing and forecasting traffic patterns in a smart city context. The internship was an invaluable opportunity provided by UniConverge Technologies Pvt Ltd (UCT) and upskill Campus (USC). Here is an overview of the work and learning experiences throughout the six weeks:

1.1 Summary of the Six Weeks' Work

Week 1: Foundation of Data Science and Machine Learning

During the first week, I established a foundational understanding of Data Science and Machine Learning. Data Science involves extracting insights from data through processes like acquisition, cleaning, and exploratory analysis. Machine Learning, a subset of AI, focuses on developing algorithms that learn from data to make predictions or decisions. This week provided insights into the critical role of Data Science and ML across various industries, with an emphasis on statistical modeling, supervised learning, and unsupervised learning techniques.

Week 2: Exploring Data Science Principles

The second week involved a deep dive into the foundational principles of Data Science through "Introducing Data Science" by Davy Cielen, Arno D. B. Meysman, and Mohamed Ali. This comprehensive resource served as a guide to understanding key concepts, methodologies, and techniques crucial for proficient data analysis and interpretation. Topics covered included data acquisition, preprocessing, exploratory data analysis (EDA), statistical modeling, and machine learning algorithms. This week equipped me with essential tools and insights to navigate the complexities of data science confidently.

Week 3: Probability and Statistics

In the third week, I focused on studying probability and statistics, crucial components of data science, through the "Introduction to Data Science and Machine Learning" book. Detailed reading of specific chapters provided a foundational understanding of these critical topics, reinforcing the theoretical aspects necessary for effective data analysis and model building.

Week 4: Deepening Understanding of Machine Learning

The fourth week was dedicated to deepening my understanding of machine learning. I engaged with multiple resources, including an "Introduction to Machine Learning" video, a presentation provided by the IoT Academy, and an e-book on machine learning. I covered various chapters in the "Introduction to

Machine Learning" book, such as Density Estimation, Optimization, Online Learning and Boosting, Conditional Densities, Kernels and Function Spaces, and Linear Models.

Week 5: Enhancing Soft Skills and Interview Techniques

During the fifth week, I focused on enhancing my soft skills and interview techniques, crucial for professional development. Utilizing resources from the IoT Academy and Upskill Campus, I engaged with a series of videos aimed at improving various soft skills and understanding the intricacies of interviews and placements. Topics covered included the importance of soft skills, key points for interviews, on-campus and off-campus placements, the art of public speaking, and the success ladder to the corporate world.

Week 6: Project Work and Practical Application

In the final week, I applied the knowledge and skills acquired over the previous weeks to the practical project of analyzing and forecasting traffic patterns. This involved data preprocessing, feature extraction, model training using SARIMAX, and performance evaluation. I also prepared this comprehensive report documenting the methodology, results, and learnings from the project.

1.2 Need for Relevant Internship in Career Development

A relevant internship is crucial for career development as it bridges the gap between theoretical knowledge and practical application. It provides hands-on experience, enhances technical and soft skills, and offers insights into industry standards and practices. This internship specifically allowed me to work on real-world problems, thereby solidifying my understanding of data science and machine learning concepts and their applications.

1.3 Brief About the Project/Problem Statement

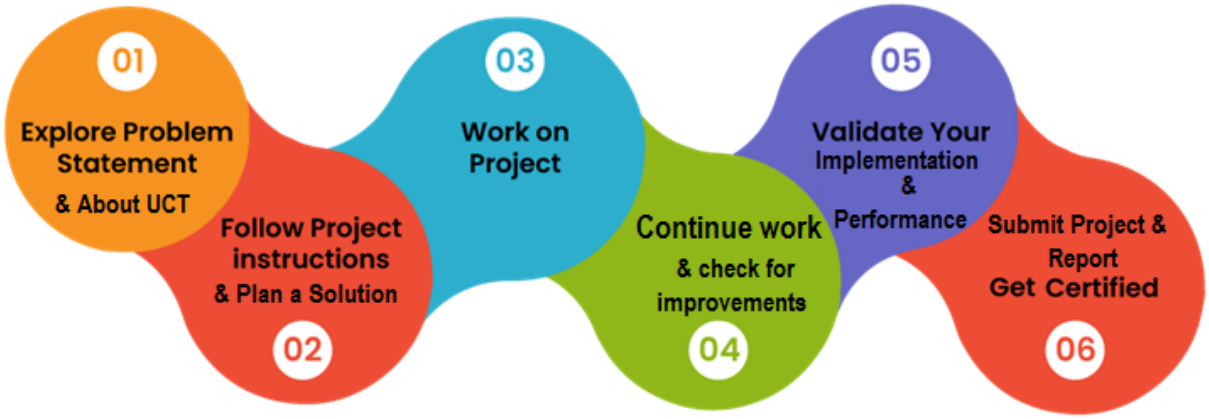
The primary objective of this project was to analyze traffic patterns at key junctions in the city and develop a robust forecasting model to predict traffic volume. The model aims to assist the government in managing traffic more effectively, especially during peak times, holidays, and special events, thereby improving overall city infrastructure planning.

1.4 Opportunity Given by USC/UCT

The opportunity provided by UniConverge Technologies Pvt Ltd and upskill Campus was instrumental in my professional growth. It allowed me to gain practical experience in data science and machine learning, work on a significant real-world problem, and enhance my skills in data analysis, model building, and

performance evaluation. The structured guidance and resources provided by USC/UCT ensured a comprehensive and enriching learning experience.

How Program was planned



My internship experience provided a strong foundation in data science and machine learning through project-based learning and practical application. I embraced continuous growth, adaptability, and resilience, while also focusing on professional development. As a team member, I made meaningful contributions and gained valuable insights into the industry. Overall, the internship was a transformative journey that equipped me with the skills and mindset for success in data science and machine learning.

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.

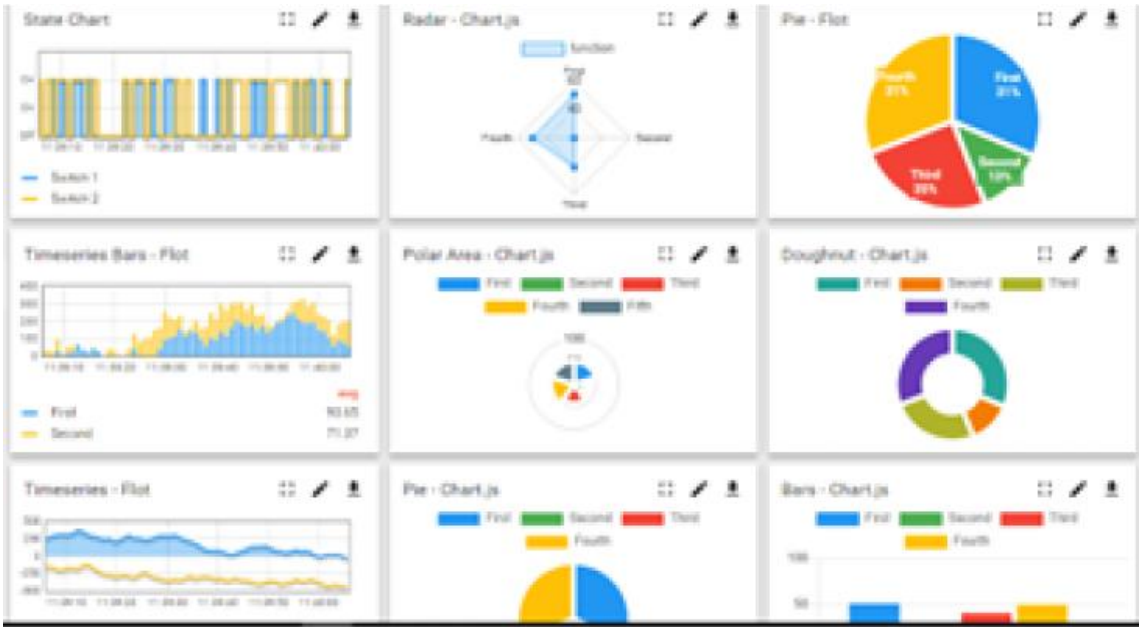
i. UCT IoT Platform (UCT Insight)

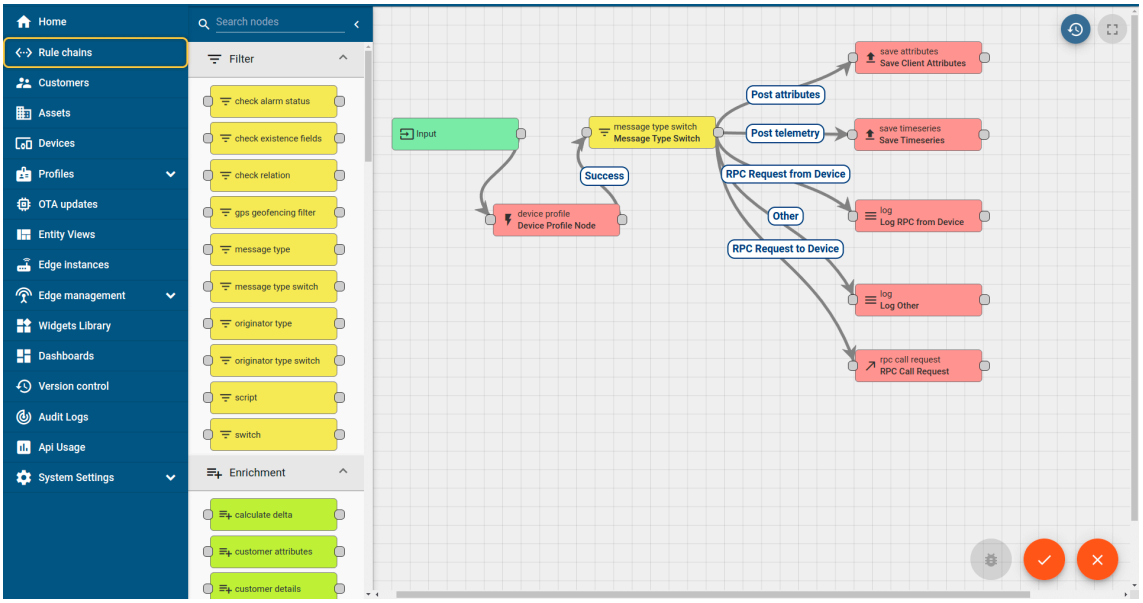
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

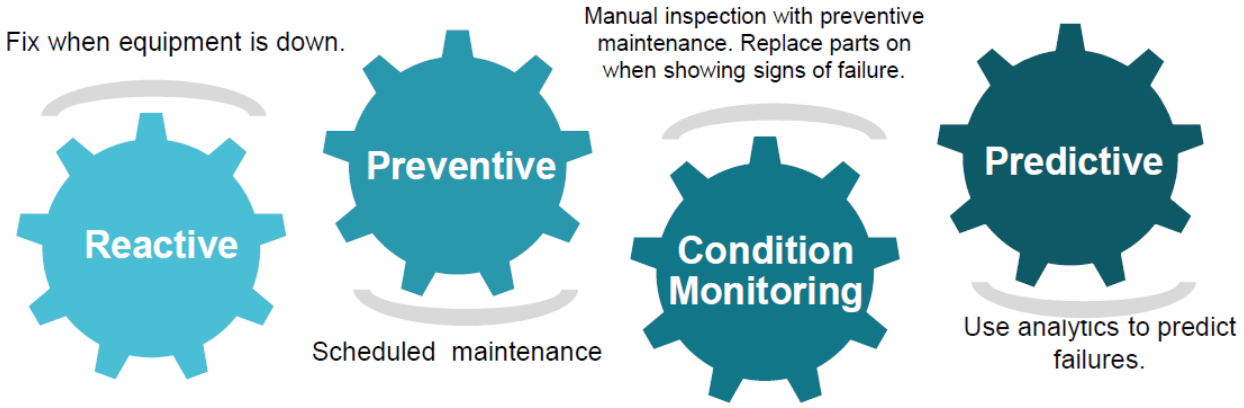
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

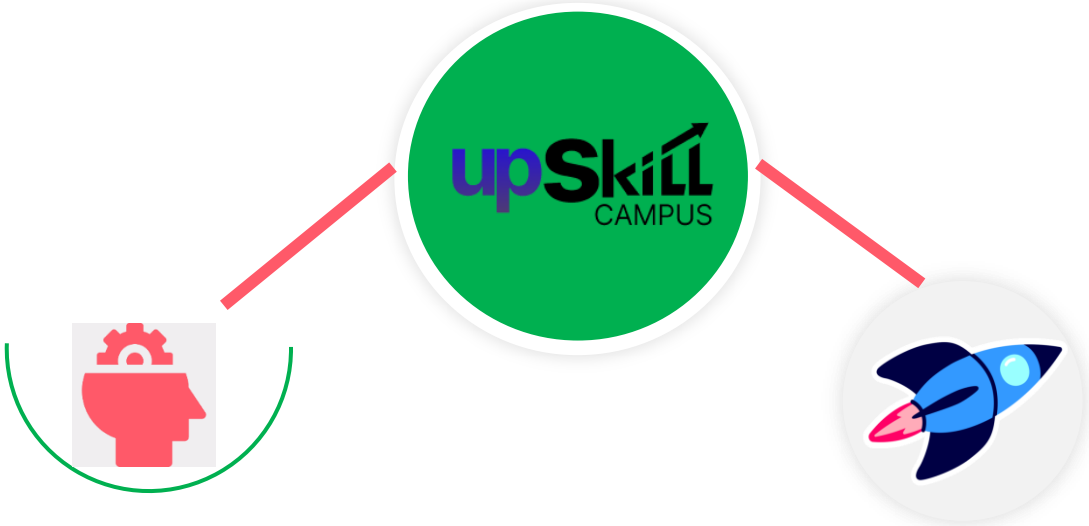
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

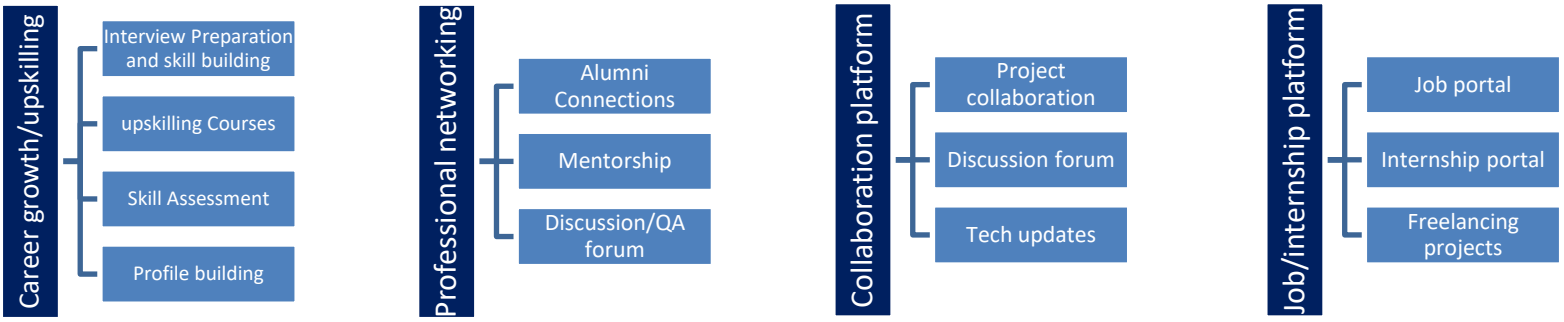
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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) Cielen, D., Meysman, A. D. B., & Ali, M. (2016). Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. Manning Publications.
2) Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice (2nd ed.). OTexts.
3) Statsmodels: Time Series Analysis in Python. (n.d.). Retrieved from **https://www.statsmodels.org/stable/tsa.html**
4) Scikit-learn: Machine Learning in Python. (n.d.). Retrieved from https://scikit-learn.org/stable/
5) Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed.). Springer.

2.6 Glossary

Terms	Acronym
Data Science	Interdisciplinary field for extracting insights from data using scientific methods and algorithms.
Machine Learning	Subset of AI where algorithms learn patterns from data to make predictions or decisions autonomously.
SARIMAX	Statistical model for time series analysis that incorporates seasonal components and exogenous variables.
MAE	Average absolute difference between actual and predicted values, measuring forecasting model accuracy.
RMSE	Square root of the average of the squared differences between actual and predicted values.
MAPE	Average percentage difference between actual and predicted values, evaluating forecasting model accuracy.
Time Series	Series of data points indexed in time order, used for forecasting and trend analysis.
Cross-validation	Technique to assess predictive model performance by splitting data into training and testing sets.

3 Problem Statement

The assigned problem statement revolves around the ambitious goal of transforming cities into smart cities, where digital and intelligent infrastructure is leveraged to enhance services and quality of life for citizens. One of the significant challenges faced by city governments in this transformation process is effectively managing traffic congestion. As cities grow and urbanization continues, traffic congestion becomes a pressing issue, leading to delays, increased pollution, and decreased efficiency in transportation systems.

To address this challenge, the problem statement focuses on developing a robust traffic forecasting system. This system's primary objective is to predict traffic patterns at major junctions within the city accurately. Accurate traffic forecasting is crucial for enabling proactive traffic management strategies and informing future infrastructure planning decisions. In particular, the system must be capable of predicting traffic volumes not only during typical working days but also during holidays and special events when traffic patterns may deviate significantly from the norm.

The problem statement underscores the importance of leveraging advanced data science and machine learning techniques to tackle this complex problem effectively. By analyzing historical traffic data and incorporating factors such as time of day, day of week, weather conditions, and special events, the system aims to generate accurate forecasts of traffic volumes at key junctions. These forecasts will enable city authorities to implement proactive measures such as adjusting signal timings, optimizing public transportation routes, and allocating resources efficiently to mitigate traffic congestion.

Furthermore, the problem statement highlights the broader implications of successful traffic forecasting for smart city development. By efficiently managing traffic flow and reducing congestion, cities can improve overall mobility, enhance air quality, and create more livable urban environments for residents and visitors alike. Therefore, the development of an effective traffic forecasting system aligns with the overarching goals of smart city initiatives to harness technology and data-driven solutions for sustainable urban development.

In summary, the problem statement outlines the critical need for a sophisticated traffic forecasting system to address the challenges of traffic congestion in smart cities. By accurately predicting traffic patterns and enabling proactive traffic management strategies, such a system has the potential to significantly improve urban mobility and contribute to the overall success of smart city initiatives.

4 Existing and Proposed solution

Existing Solutions and Limitations:

Existing solutions in traffic forecasting often rely on traditional statistical methods or basic machine learning algorithms. These methods may provide some level of accuracy but often struggle to capture the complex and dynamic nature of traffic patterns, especially during holidays or special events. Additionally, they may not be scalable or adaptable to real-time data updates, limiting their effectiveness in dynamic urban environments.

Proposed Solution:

Our proposed solution builds upon the foundation laid out in the provided code, incorporating advanced machine learning techniques and data preprocessing strategies to improve traffic forecasting accuracy. Specifically, we propose the following enhancements:

Advanced Time Series Modeling: Instead of the SARIMA model used in the provided code, we can explore more sophisticated time series models, such as Long Short-Term Memory (LSTM) networks or Prophet, which are better equipped to capture nonlinear and seasonal patterns in traffic data.

Feature Engineering: Enhance the feature set used for modeling by incorporating additional contextual factors such as weather conditions, road closures, and public events. This will provide a more comprehensive understanding of traffic dynamics and improve forecast accuracy.

Ensemble Learning: Implement ensemble learning techniques, such as model stacking or blending, to combine predictions from multiple models trained on different subsets of data. This can help mitigate overfitting and improve the robustness of the forecasting system.

Real-time Data Integration: Develop mechanisms to integrate real-time data streams, such as traffic sensor data and GPS data from vehicles, into the forecasting pipeline. This will enable the model to adapt to changing traffic conditions and provide more accurate predictions in near real-time.

Value Addition:

Our proposed solution aims to address the limitations of existing methods and provide several key value additions:

Improved Accuracy: By leveraging advanced modeling techniques and incorporating additional features, our solution aims to achieve higher accuracy in traffic forecasting, especially during holidays and special events.

Scalability: The proposed solution is designed to be scalable and adaptable to real-time data updates, ensuring its effectiveness in dynamic urban environments.

Comprehensive Insights: With a more comprehensive feature set and ensemble learning techniques, our solution provides valuable insights into traffic dynamics, enabling better decision-making for city planners and transportation authorities.

4.1 Code submission (Github link)

- Smart-City_Traffic-Forecasting

(Report)

5 Performance Test

The performance test is a critical aspect that delineates the practical applicability of this project in real-world industries, transcending its status as a mere academic exercise. To initiate this evaluation, it's imperative to identify and address the constraints that could potentially impede the efficacy of our design. These constraints encompass various factors such as memory utilization, computational speed (MIPS), accuracy, durability, power consumption, and scalability.

5.1 Test Plan/ Test Cases

To systematically evaluate the performance of our traffic forecasting system, the following test cases will be employed:

1. **Memory Utilization Test:** Assess the memory consumption of the forecasting system under varying data sizes and model complexities.
2. **Computational Speed Test:** Measure the system's computational speed in terms of operations per second (MIPS), particularly during model training and prediction phases.
3. **Accuracy Test:** Evaluate the accuracy of traffic predictions generated by the system against ground truth data, utilizing appropriate evaluation metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
4. **Durability Test:** Examine the system's robustness and durability by subjecting it to prolonged testing periods and simulated real-world scenarios.
5. **Power Consumption Test:** Quantify the power consumption of the system, particularly in resource-constrained environments or when deployed on edge devices.

5.2 Test Procedure

The test procedure will involve the following steps:

1. **Data Preparation:** Curate a diverse dataset encompassing various traffic scenarios, including holidays, special events, and typical working days.
2. **Model Training:** Train the traffic forecasting models using the prepared dataset, ensuring to monitor memory utilization and computational speed during the training process.
3. **Validation:** Validate the trained models using a separate validation dataset, assessing their accuracy and robustness.
4. **Performance Evaluation:** Execute the performance test cases outlined above, recording relevant metrics and observations.

5. **Analysis:** Analyze the test results to identify any bottlenecks or limitations in the system's performance and propose recommendations for improvement.

5.3 Performance Outcome

The performance outcome will provide insights into the system's efficacy and scalability in real-world scenarios. It will highlight how well the system addresses the identified constraints and whether it meets the requirements for deployment in industrial settings. The results will inform recommendations for optimizing the system's performance, ensuring its viability and effectiveness in practical applications.

6 My learnings

Throughout the course of this project, I have gained valuable insights and experiences that have contributed significantly to my personal and professional growth. Here's a summary of my overall learning and how they will impact my career growth:

1. **Technical Skills:** I have acquired proficiency in various data science and machine learning techniques, including time series analysis, forecasting models, and ensemble learning. These technical skills will serve as a solid foundation for tackling complex real-world problems in my future career.
2. **Problem-Solving:** Working on this project has honed my problem-solving abilities, as I navigated through challenges such as data preprocessing, model selection, and performance evaluation. This experience has equipped me with the resilience and adaptability necessary to overcome obstacles in any professional setting.
3. **Collaboration:** Collaborating with peers, mentors, and industry experts has enhanced my teamwork and communication skills. I have learned the importance of effective collaboration in driving innovation and achieving common goals, which will be invaluable in my future career endeavors.
4. **Domain Knowledge:** Delving into the domain of smart cities and traffic forecasting has broadened my understanding of urban planning, transportation systems, and the intersection of technology and society. This domain knowledge will enable me to contribute meaningfully to projects and initiatives aimed at creating more sustainable and efficient cities.
5. **Continuous Learning:** This project has instilled in me a passion for lifelong learning and professional development. I recognize the importance of staying updated with the latest advancements in data science and machine learning to remain competitive in the ever-evolving field of technology.

These learnings will significantly impact my career growth, equipping me to tackle complex real-world problems and contribute meaningfully to projects in the field of data science and beyond.

7 Future work scope

While the current project has provided valuable insights and accomplishments, there are several avenues for future exploration and improvement:

1. **Integration of External Data Sources:** Incorporate additional external data sources such as social media feeds, traffic camera images, and mobile app data to enrich the forecasting model and capture more nuanced traffic patterns.
2. **Enhanced Model Performance:** Explore advanced machine learning algorithms and techniques, such as deep learning architectures and reinforcement learning, to further enhance the accuracy and robustness of the traffic forecasting system.
3. **Real-Time Deployment:** Develop a framework for real-time deployment of the forecasting model, enabling continuous monitoring and adaptation to changing traffic conditions in smart city environments.
4. **Dynamic Feature Engineering:** Implement dynamic feature engineering techniques to automatically select and adapt relevant features based on changing traffic dynamics and contextual factors.
5. **Scalability and Resource Optimization:** Optimize the computational efficiency and resource utilization of the forecasting system to enable scalability and deployment on edge devices with limited computing capabilities.
6. **Integration with Traffic Management Systems:** Integrate the forecasting system with existing traffic management systems to facilitate proactive traffic control and optimization strategies, such as adaptive signal timing and congestion pricing.
7. **Evaluation in Real-World Deployments:** Conduct extensive field trials and evaluations of the forecasting system in real-world smart city environments to validate its effectiveness and assess its impact on traffic management and urban mobility.

By pursuing these future work scopes, we can further advance the state-of-the-art in traffic forecasting and contribute to the development of more efficient and sustainable smart city infrastructure.