upSkill CAMPUS

[Your College Logo]

Industrial Internship Report on

Forecasting of Smart city traffic patterns

Prepared by

Rohit Lahori

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 to implement a robust traffic system for the city and analyze the traffic patterns of four junctions of the city on holidays and other occasions during the year as well, and how they differ from normal working days. 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

In the starting week of my internship, I had read through different projects and chose to work on the smart city. The problem statement of the project excited me and kept me thinking about how to go about it. The next week, I surfed the web to get some ideas and started implementation such as importing the necessary libraries and dataset. In the third week to get some idea about the dataset, I performed data analysis and came up with conclusions which helped me get an idea on how to go forward with it.

In the fourth week, I performed data transformation and preprocessing. The objectives this week were to make unique frames for each junction and plot them. Plotting the series and changing it and making training and test sets. After performing data processing and transformation, this week I started working on the building of the model. I have decided to employ a Gated Recurrent Unit for our project (GRU). We are developing a function in this part that the neural network may use to access and fit the data frames for all four junctions. Trained this model on each junction training set and found root mean squared error for each.

About need of relevant Internship in career development.

A data science and machine learning internship provides practical, real-world experience in applying data science and machine learning techniques. It allows you to work on actual projects, analyze real data, and solve industry-specific problems. This hands-on experience enhances your skills and understanding of data science concepts, tools, and methodologies. It exposes you to the workflows, best practices, and tools commonly used in the industry. You gain insights into how data science and machine learning are applied in real-world scenarios, understand project lifecycles, and learn about the challenges and considerations specific to different industries and domains.

Working on real-world data science projects during an internship hones your problem-solving and critical thinking abilities. You learn to approach complex problems, formulate hypotheses, design experiments, analyze data, and derive actionable insights. These skills are highly valued in data-driven decision-making and problem-solving across industries.

An internship in data science and machine learning adds significant value to your resume. It demonstrates your practical experience, industry exposure, and ability to apply theoretical knowledge in a professional setting. Employers often prioritize candidates with internship experience as it showcases your commitment, adaptability, and potential contributions to their organization.

Brief about Your project/problem statement.

Traffic prediction using machine learning involves using historical and real-time data to forecast traffic patterns, congestion levels, and travel times in each area or road network. By leveraging machine learning algorithms, traffic prediction models can provide valuable insights to improve traffic management, optimize routes, and enhance transportation systems. Traffic prediction models require comprehensive data, including historical traffic data, road network information, weather conditions, and other relevant factors. This data is collected from various sources, such as traffic sensors, surveillance cameras, GPS data, and weather APIs.

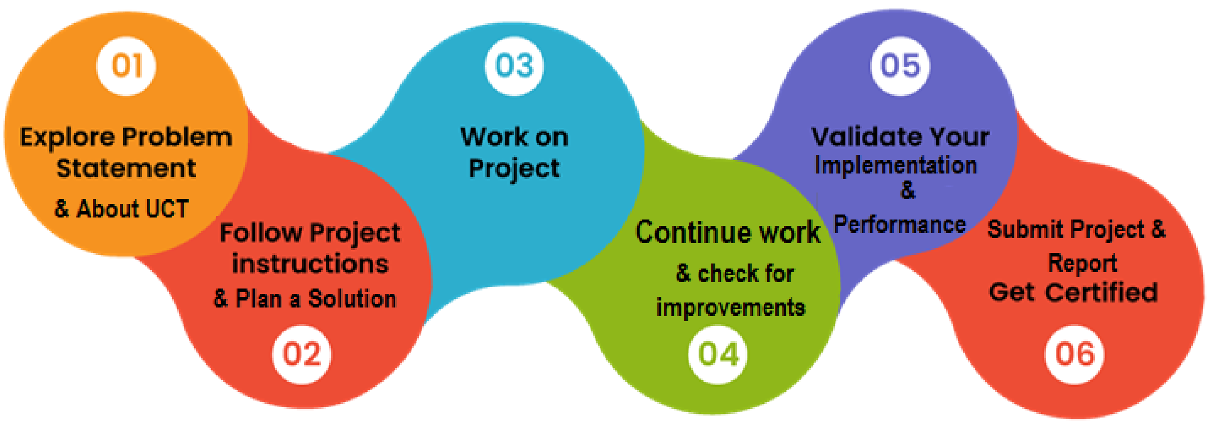
The collected data undergoes preprocessing steps to handle missing values, outliers, and inconsistencies. This includes data cleaning, normalization, and feature engineering. Features such as time of day, day of the week, seasonality, road characteristics, and historical traffic patterns are extracted to capture important information for prediction. Various machine learning algorithms are employed to train the traffic prediction model. These can include regression algorithms, time series forecasting methods, or more advanced techniques like neural networks. The model is trained using the preprocessed data, with the target variable being the traffic condition, such as traffic volume or travel time.

The predicted traffic patterns can be visualized through maps, graphs, or dashboards to provide intuitive insights for traffic management and decision-making. These visualizations help transportation authorities, traffic engineers, and commuters understand current and predicted traffic conditions.

Opportunity given by USC/UCT.

I would like to thank upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process. The opportunity given, provided me with a good knowledge and helped me enhance my machine learning and data science skills.

How Program was planned



Your Learnings and overall experience.

Traffic prediction projects involve working with real-world data, which often requires extensive cleaning, pre-processing, and feature engineering. Learning these skills equips you with valuable data handling techniques applicable to a wide range of data science projects.

Exploring and comparing different machine learning models for traffic prediction allows you to understand their strengths, weaknesses, and applicability to specific scenarios. Evaluating models using appropriate metrics helps you refine your understanding of model performance and make informed decisions.

By acquiring these skills and experiences, you enhance your expertise in machine learning, data analysis, and problem-solving within the specific context of traffic prediction. This knowledge can be transferable to other domains facing similar challenges, such as demand forecasting, supply chain management, and predictive maintenance. It also demonstrates your ability to apply advanced analytics techniques to real-world problems, making you an attractive candidate for data science roles and offering opportunities for career growth in areas related to intelligent transportation systems, smart city initiatives, and data-driven decision-making.

My message to junior and peers

It has been an incredible journey filled with learning, growth, and valuable experiences. I wanted to take a moment to express my gratitude and share some reflections with all of you. To my junior colleagues, I encourage you to embrace every opportunity that comes your way, especially internships. They provide a unique chance to apply your knowledge, gain practical skills, and explore different aspects of your field. Approach your internship with enthusiasm, curiosity, and a willingness to learn. Be proactive, ask questions, seek feedback, and make the most out of the resources available to you. Remember that mistakes are part of the learning process, so don't be afraid to step out of your comfort zone and take on new challenges. This internship is a stepping stone towards building a successful career, and your dedication and hard work will pay off.

To my peers, I want to extend my heartfelt appreciation for the support, collaboration, and camaraderie throughout this internship journey. The opportunity to work alongside talented individuals like you has been inspiring. Together, we have tackled complex projects, shared knowledge, and supported each other's growth. Let's celebrate our achievements and the milestones we have reached during this internship. As we move forward in our careers, I encourage us to stay connected, continue to collaborate, and be each other's support system. We have the potential to accomplish great things, and by supporting one another, we can achieve even more.

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 ( )

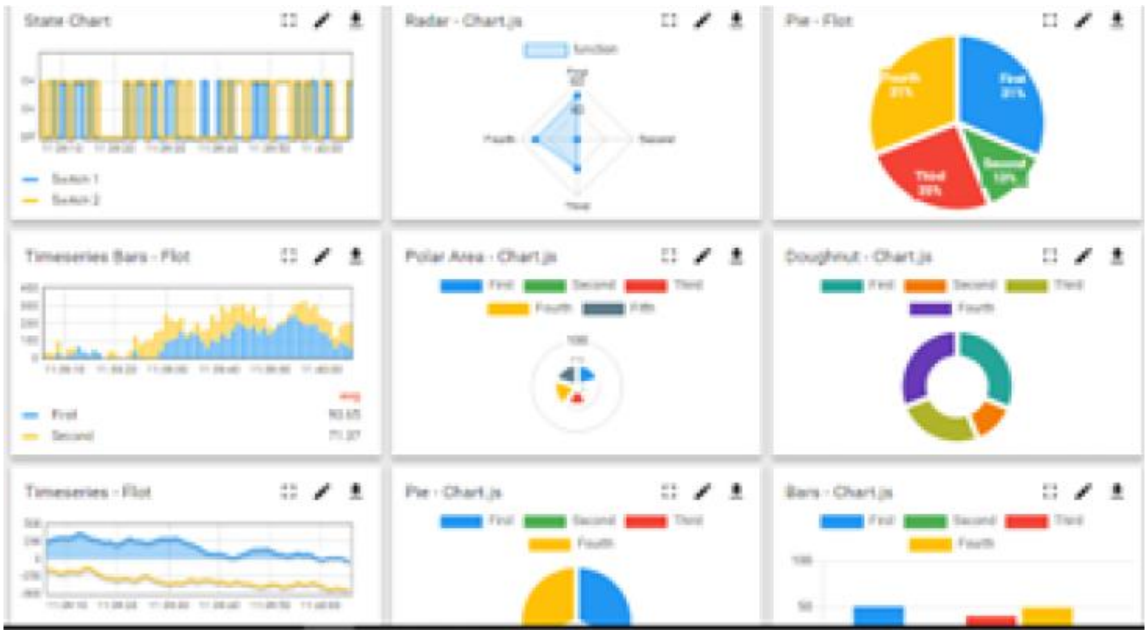
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

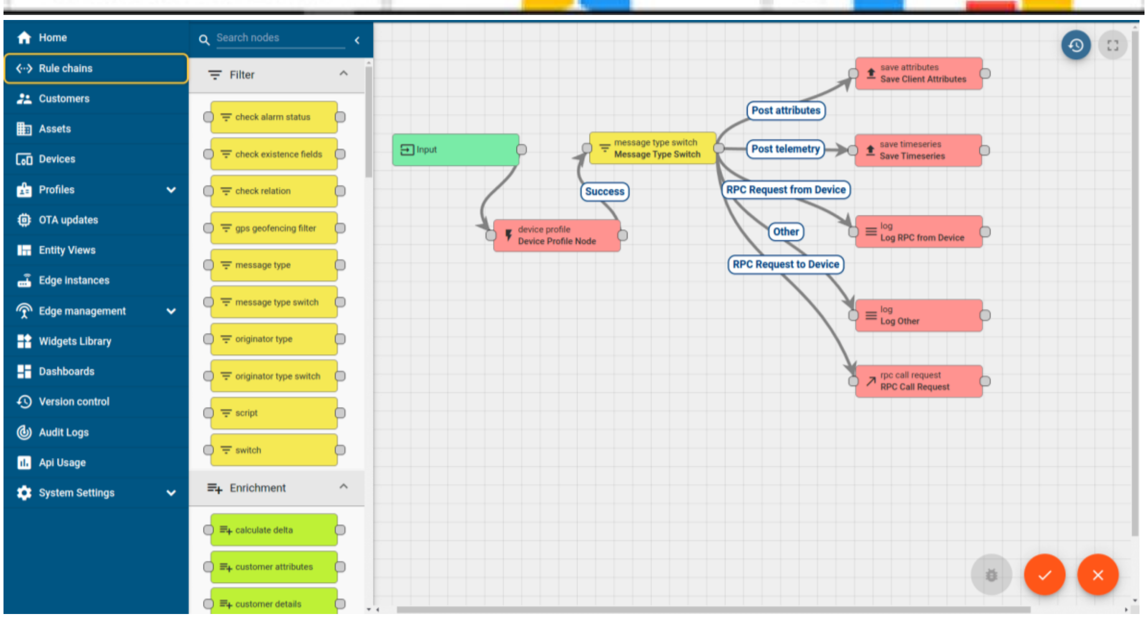
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- 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)





ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

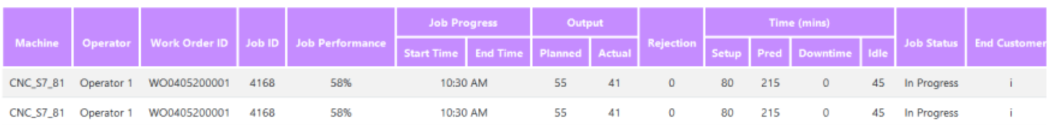
It provides Users/ Factory

- 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.

upSkill CAMPUS

[Your College Logo]



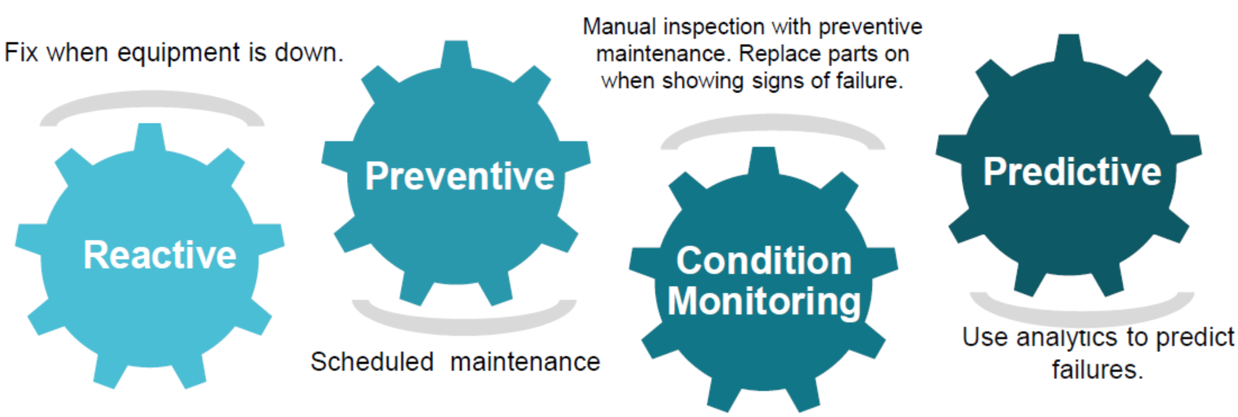
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

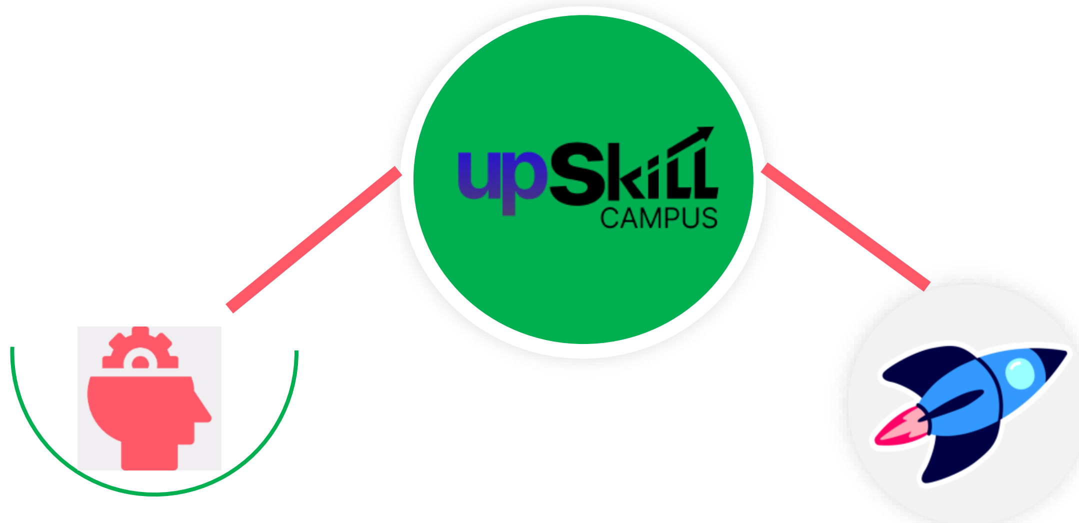
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

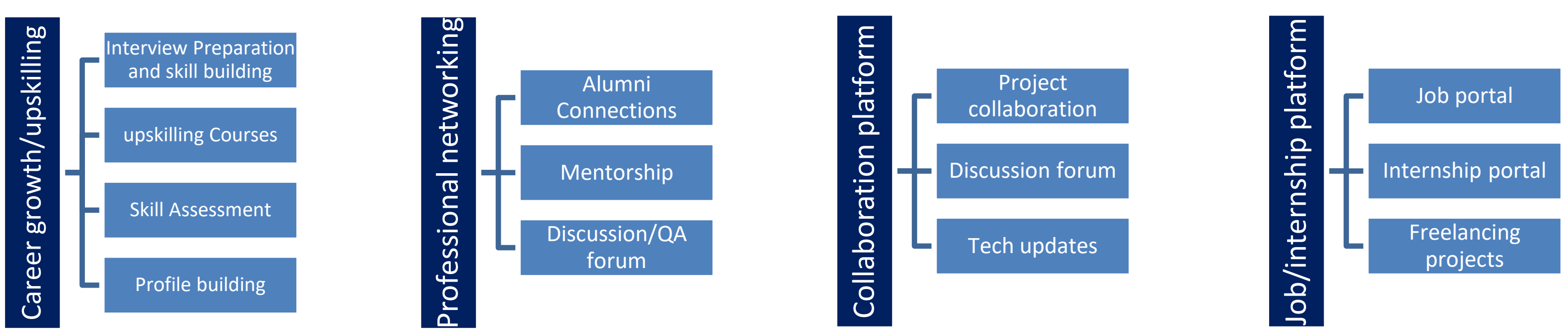
USC is a career development platform that delivers personalized executive coaching in a more affordable, scalable and measurable way.



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 years.

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.javatpoint.com/traffic-prediction-using-machine-learning#:~:text=There%20are%20several%20types%20of,conditions%20based%20on%20past%20trends.

[2] https://www.altexsoft.com/blog/traffic-prediction/

2.6 Glossary

Terms	Acronym
Machine Learning	ML
Data Science	DS
Gated Recurrent Unit	GRU

3 Problem Statement

The vision of transforming a city into a smart and intelligent entity has gained prominence worldwide, driven by the need to enhance citizen services and improve efficiency. Among the challenges faced by governments, traffic congestion stands out as a significant issue impacting the daily lives of citizens. In this scenario, data scientists play a crucial role in managing traffic better and providing insights for infrastructure planning.

This problem discusses the efforts of a data scientist working with the government to implement a robust traffic system that considers traffic patterns during peak times and special occasions.

Efficiency through Data-driven Insights: To effectively manage traffic, data collection serves as the foundation. Multiple sources such as traffic sensors, surveillance cameras, and GPS data are utilized to gather both historical and real-time information. This comprehensive dataset allows data scientists to analyze traffic volume, congestion, and speed, among other relevant metrics. Moreover, data scientists obtain specific data pertaining to holidays, festivals, and special events that impact traffic patterns in the city.

The transformation of a city into a smart and intelligent entity requires data-driven solutions, and traffic management is a critical aspect of this process. Data scientists armed with comprehensive and accurate data, along with advanced analytical techniques, play a pivotal role in understanding traffic patterns, forecasting future trends, and providing valuable insights for infrastructure planning. By implementing a robust traffic system that accounts for traffic peaks during holidays and other special occasions, governments can significantly improve traffic management and enhance the overall efficiency of the city's services.

4 Existing and Proposed solution

Existing solutions and their limitations

Existing solutions for traffic prediction using machine learning vary in their approaches and limitations. Some solutions employ traffic flow models, such as the Cell Transmission Model (CTM) or the Lighthill-Whitham-Richards (LWR) model. These models simulate traffic flow dynamics based on fundamental principles. However, they require detailed information about road networks, traffic states, and calibration parameters, making them computationally intensive and less suitable for real-time prediction.

Another approach utilizes GPS and sensor data from vehicles to estimate traffic conditions. This can involve techniques like trajectory analysis, clustering, or anomaly detection. However, these solutions often face challenges with data availability, coverage, and privacy concerns. They may not be applicable in areas with limited data or where GPS/sensor coverage is sparse.

What is your proposed solution?

My proposed solution for traffic prediction aims to leverage a combination of approaches to overcome the limitations of existing solutions. It involves using machine learning techniques, such as deep learning models or hybrid models, which integrate multiple data sources and capture both temporal and spatial dependencies in traffic patterns. Some key aspects of the proposed solution include:

Integrating diverse data sources, such as historical traffic data, weather conditions, traffic sensor data, road network information, and events data, to provide a comprehensive view of traffic patterns.

Developing hybrid models that combine the strengths of different approaches, such as time series forecasting, traffic flow modeling, and data-driven machine learning models, to capture both short-term and long-term traffic dynamics.

What value addition are you planning?

The proposed solution aims to provide several value additions:

Improved Accuracy: By leveraging multiple data sources and hybrid models, the proposed solution seeks to enhance prediction accuracy, capturing complex traffic patterns and long-term dependencies more effectively.

Real-time Prediction: The integration of real-time data and adaptive learning enables more timely and up-to-date traffic predictions, facilitating proactive decision-making and efficient traffic management.

Scalability and Adaptability: The solution aims to be scalable and adaptable to different cities and road networks, allowing for customization and transferability across various urban environments.

4.1 Code submission (Github link)

https://github.com/Rohit-Lahori02/Smart_City_Traffic_Prediction/blob/main/Smart_City.ipynb

4.2 Report submission (Github link) : «Link»

5 Proposed Design/ Model

The design flow of traffic prediction using machine learning typically involves several key steps. Here's a proposed design flow for traffic prediction: The design flow of traffic prediction using machine learning typically involves several key steps. Here's a proposed design flow for traffic prediction:

Data Preprocessing: Clean and preprocess the collected data to make it suitable for machine learning models. This step includes handling missing values, data normalization, removing outliers, and aggregating or resampling data to the desired time intervals (e.g., minutes, hours).

Feature Engineering: Extract relevant features from the preprocessed data that can capture traffic patterns and influence traffic conditions. These features may include historical traffic flow, speed, weather attributes (e.g., temperature, precipitation), time of day, day of the week, holidays, road network characteristics, and events information.

Model Selection: Choose appropriate machine learning models for traffic prediction. This can include traditional time series models (e.g., ARIMA, SARIMA), deep learning models (e.g., LSTM, GRU), ensemble models (e.g., Random Forest, Gradient Boosting), or hybrid models that combine different techniques. Consider the characteristics of the data, the complexity of traffic patterns, and the desired prediction horizon.

Model Training: Split the preprocessed data into training and validation sets. Train the selected machine learning model using the training data, optimizing model parameters through techniques like cross-validation or grid search. Monitor the model's performance on the validation set to avoid overfitting

Model Evaluation: Evaluate the trained model's performance using appropriate evaluation metrics for traffic prediction, such as mean absolute error (MAE), root mean squared error (RMSE), or accuracy measures. Assess the model's ability to capture traffic patterns, handle different traffic conditions, and generalize to unseen data.

Model Deployment: Deploy the trained model in a production environment or as part of a traffic management system. This involves integrating the model with data ingestion pipelines, real-time data sources, and appropriate infrastructure to handle predictions at scale. Ensure the model is regularly updated to incorporate new data and maintain accuracy.

Visualization and Communication: Develop visualizations and user interfaces to present the predicted traffic information to stakeholders, such as transportation authorities, traffic operators, and commuters. Provide intuitive visualizations, dashboards, or mobile applications to support decision-making and improve user experience.

5.1 High Level Diagram (if applicable)

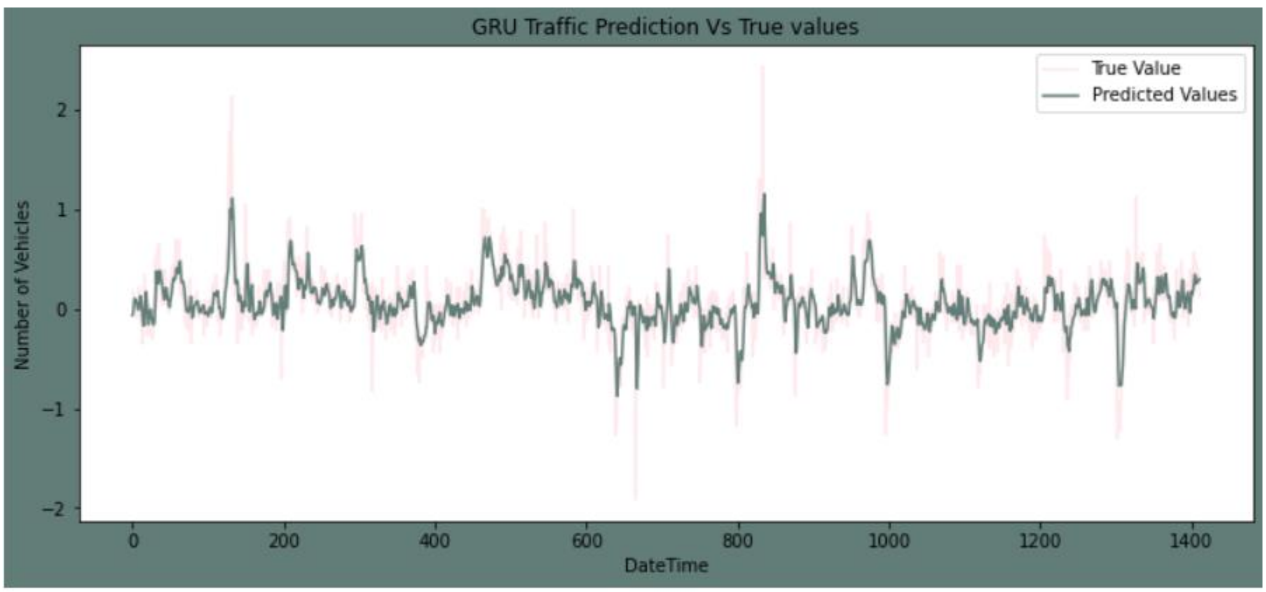
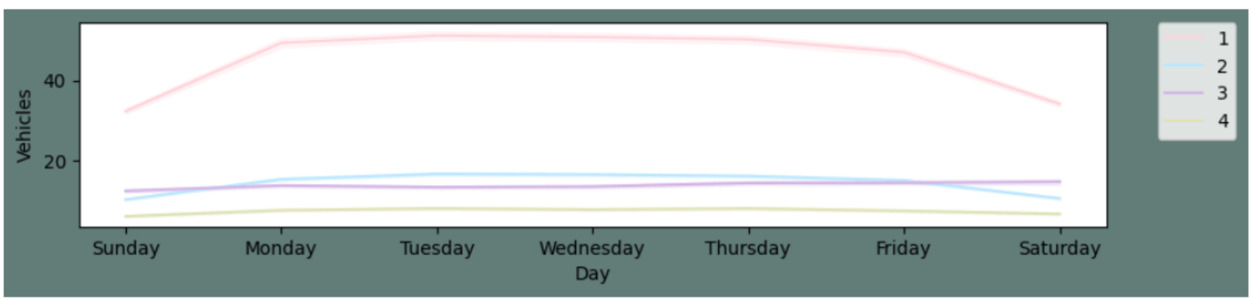
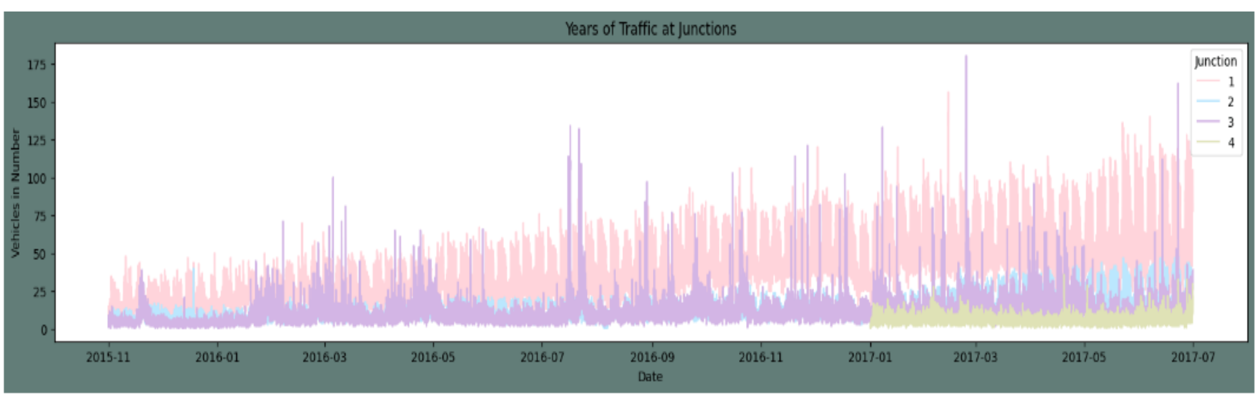


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Interfaces (if applicable)





6 Performance Test

Constraints in design

Dynamic Nature of Traffic: Traffic patterns are inherently dynamic and can change due to various factors such as weather conditions, road construction, accidents, special events, and time of day. Models need to adapt and account for these dynamic changes to provide accurate predictions.

Spatial and Temporal Dependencies: Traffic patterns are influenced by the spatial layout of roads, intersections, and junctions, as well as temporal factors such as rush hours, weekends, and holidays. Capturing the dependencies between different locations and time periods is crucial for accurate traffic predictions.

How were those constraints taken care in your design?

Designing a model that incorporates real-time or near-real-time data feeds can help capture the dynamic nature of traffic. Continuous monitoring of relevant factors, such as weather updates, event schedules, and traffic incident reports, can enable the model to dynamically adjust its predictions and account for changing conditions.

Feature engineering techniques can be employed to encode spatial and temporal dependencies into the model. Spatial features can include road network characteristics, proximity to landmarks, and population density, while temporal features can incorporate time of day, day of the week, and seasonality. Additionally, techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can capture spatial and temporal patterns in the data, respectively.

What were test results around those constraints?

- With the exception of the fourth junction, all junctions have shown a rising yearly tendency. As was already stated above, the fourth junction contains scant data that doesn't go back more than a year.

- We can observe that around June, there is an increase in traffic at the first and second crossroads. This, we assume, may be related to summer vacation and related activities.
- We may observe that there are peaks in the morning and evening and a fall in activity throughout the night for a given day. This is what was predicted.
- Due to fewer automobiles on the roadways on Sundays than on other days of the week, traffic flows more smoothly. The traffic is consistent from Monday through Friday.

<u>Steps taken to handle constraints.</u>

The following differencing procedure should be used to remove seasonality:

- We'll be using the difference in weekly numbers for Junction 1.
- The difference of consecutive days is a preferable option for junction two.
- The difference between the hourly numbers will be used for Junctions 3 and 4.

6.1 Test Procedure

Assigned training and test cases for each junction and then decided to employ a Gated Recurrent Unit for our project (GRU). We are developing a function in this part that the neural network may use to access and fit the data frames for all four junctions. Fit the model on the training data.

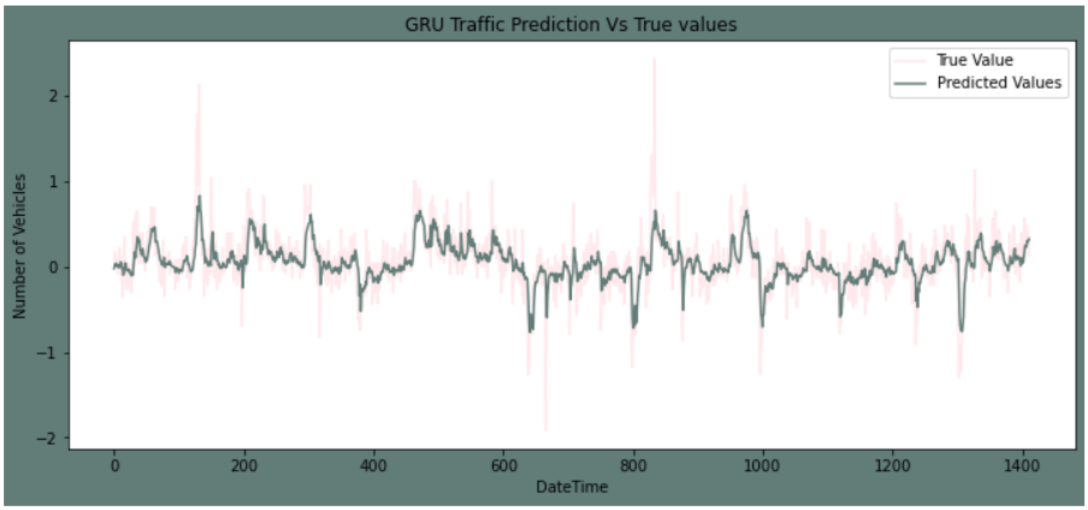
```
In [29]: #Fitting the model
         PredJ1 = GRU_model(X_train_Junction1,y_train_Junction1,X_test_Junction1)

Epoch 1/50
87/87 [==============================] - ETA: 0s - loss: 0.0660WARNING:tensorflow:Early stopping conditioned on metric `val_l
oss` which is not available. Available metrics are: loss
87/87 [==============================] - 25s 212ms/step - loss: 0.0660
Epoch 2/50
87/87 [==============================] - ETA: 0s - loss: 0.0484WARNING:tensorflow:Early stopping conditioned on metric `val_l
oss` which is not available. Available metrics are: loss
87/87 [==============================] - 19s 221ms/step - loss: 0.0484
Epoch 3/50
87/87 [==============================] - ETA: 0s - loss: 0.0479WARNING:tensorflow:Early stopping conditioned on metric `val_l
oss` which is not available. Available metrics are: loss
87/87 [==============================] - 20s 225ms/step - loss: 0.0479
Epoch 4/50
87/87 [==============================] - ETA: 0s - loss: 0.0484WARNING:tensorflow:Early stopping conditioned on metric `val_l
oss` which is not available. Available metrics are: loss
87/87 [==============================] - 20s 229ms/step - loss: 0.0484
Epoch 5/50
87/87 [==============================] - ETA: 0s - loss: 0.0474WARNING:tensorflow:Early stopping conditioned on metric `val_l
oss` which is not available. Available metrics are: loss
```

Testing the model on the test data

```
In [30]: #Results for J1
RMSE_J1=RMSE_Value(y_test_Junction1,PredJ1)
PredictionsPlot(y_test_Junction1,PredJ1,0)
```

```
The root mean squared error is 0.2537800865796726.
```



Followed the same procedure for all junctions.

6.2 Performance Outcome

Out[40]:

	Junction	RMSE
0	Junction1	0.253780
1	Junction2	0.540583
2	Junction3	0.634318
3	Junction4	1.025672

7 My learnings

Traffic prediction projects involve working with real-world data, which often requires extensive cleaning, preprocessing, and feature engineering. Learning these skills equips you with valuable data handling techniques applicable to a wide range of data science projects.

Traffic prediction models require careful selection and engineering of relevant features. Acquiring expertise in feature selection and engineering enables you to identify the most impactful variables and improve model performance across various domains.

Traffic patterns exhibit temporal dependencies, making time series analysis an essential aspect of traffic prediction. Developing skills in time series modeling, forecasting, and handling dynamic data helps you tackle a broader range of time-dependent problems.

Exploring and comparing different machine learning models for traffic prediction allows you to understand their strengths, weaknesses, and applicability to specific scenarios. Evaluating models using appropriate metrics helps you refine your understanding of model performance and make informed decisions.

Working on traffic prediction projects requires a deep understanding of traffic dynamics, road networks, and influencing factors. Building domain knowledge enhances your ability to identify relevant features, interpret results, and propose practical solutions.

Developing traffic prediction models that cater to real-time or low-latency requirements expands your skill set, as it involves optimizing model performance, exploring lightweight models, and addressing computational constraints. These skills are valuable for time-sensitive applications across various industries.

By acquiring these skills and experiences, you enhance your expertise in machine learning, data analysis, and problem-solving within the specific context of traffic prediction. This knowledge can be transferable to other domains facing similar challenges, such as demand forecasting, supply chain management, and predictive maintenance. It also demonstrates your ability to apply advanced analytics techniques to real-world problems, making you an attractive candidate for data science roles and offering opportunities for career growth in areas related to intelligent transportation systems, smart city initiatives, and data-driven decision-making.

8 Future work scope

Some ideas that could be explored in the future for traffic prediction using machine learning can be:

Integrating various data sources, such as traffic sensor data, social media feeds, and weather information, to develop a comprehensive traffic prediction model. This approach could provide a more holistic understanding of traffic patterns and improve prediction accuracy.

Applying reinforcement learning techniques to optimize traffic signal timings and control strategies in real-time. This approach could dynamically adapt traffic signals based on current traffic conditions, reducing congestion and improving traffic flow.

Utilizing graph neural networks (GNNs) to analyze road networks and capture the spatial dependencies between different road segments and junctions. GNNs can help identify critical nodes, congestion propagation patterns, and optimal routes for traffic management.

Developing anomaly detection models to identify and predict traffic incidents such as accidents, road closures, or unusual traffic patterns. This could assist in proactively managing traffic disruptions and providing timely information to drivers and authorities.

Integrating traffic prediction models with existing intelligent transportation systems, such as adaptive traffic signal control systems or dynamic route guidance systems. This collaboration could enhance the efficiency and effectiveness of these systems by incorporating real-time traffic predictions.

Here are some thoughts which I feel can be investigated as well in the future.