Industrial Internship Report on

”Smart City Traffic Pattern”

Prepared by

Vishva Chaudhary

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. The focus of my project was to analyze and predict traffic patterns in a smart city using data science techniques. This internship gave me a very good opportunity to get exposure to industrial problems and design/implement solutions for them. It was an overall great experience to have this internship. 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 3
2 Introduction 5
2.1 About UniConverge Technologies Pvt Ltd 5
i. UCT IoT Platform 5
2.2 About upskill Campus (USC) 9
2.3 The IoT Academy 10
2.4 Objectives of this Internship program 11
2.5 Reference 11
2.6 Glossary 11
3 Problem Statement: 12
4 Existing and Proposed solution 15
4.1 Code submission (Github link) 17
4.2 Report submission (Github link): 17
5 Proposed Design/ Model 18
5.1 High Level Diagram (if applicable) 19
5.2 Low Level Diagram (if applicable) 19
5.3 Interfaces (if applicable) 19
6 Performance Test 21
Constraints and Solutions 21
6.1 Test Plan/ Test Cases 22
6.2 Test Procedure 22
6.3 Performance Outcome 23
7 My learnings 23
8 Future work scope 26

1 Preface

Over the past six weeks, I have had the privilege of participating in an industrial internship organized by Upskill Campus (USC) and The IoT Academy, in collaboration with UniConverge Technologies Pvt Ltd (UCT). This internship has been an invaluable experience, allowing me to bridge the gap between academic knowledge and real-world industrial applications.

Summary of Work:

Throughout this internship, I focused on a project aimed at analyzing and predicting traffic patterns in a smart city using advanced data science techniques. The objective was to develop a predictive model that could accurately forecast traffic congestion at various city junctions, thereby aiding in better traffic management and reducing congestion.

Need for Relevant Internship:

In today's competitive job market, internships play a crucial role in career development. They provide practical exposure, enhance problem-solving skills, and offer insights into industry operations that are not typically covered in academic curricula. This internship was particularly relevant as it provided hands-on experience with real-world data, advanced analytics, and machine learning – all essential skills in the field of data science.

Project Overview:

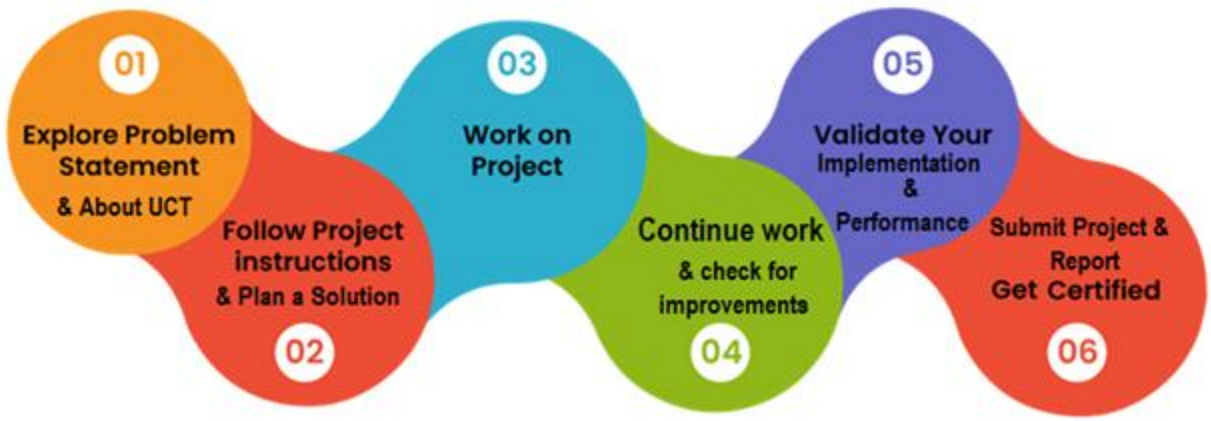
The core of my project was to utilize traffic data from a smart city to predict traffic patterns. This involved data preprocessing, exploratory data analysis (EDA), and the application of machine learning models to forecast traffic congestion. The final objective was to create a reliable model that could be integrated into smart city traffic management systems.

Opportunity by USC/UCT:

USC and UCT provided a structured and supportive environment that facilitated the smooth execution of the internship program. They offered access to necessary resources, including webinars, tutorials, and documentation, which were instrumental in understanding and tackling the project requirements. This opportunity was a gateway to experiencing the nuances of industrial projects and learning from seasoned professionals.

Program Planning:

The internship program was meticulously planned to cover all aspects of the project, from initial problem understanding to final implementation and documentation. Weekly goals were set to ensure steady progress, and regular feedback sessions were conducted to keep the project on track. This structured approach ensured that all deliverables were met within the stipulated timeframe.



Learnings and Experience:

This internship has significantly enhanced my technical skills, particularly in data preprocessing, EDA, and model training using Python. Additionally, I learned to use Flask for web application development, which is crucial for presenting data science projects effectively. The soft skill training provided by Upskill also helped improve my communication and public speaking abilities.

Acknowledgements:

I would like to express my gratitude to several individuals who supported me throughout this journey. My heartfelt thanks to my mentors at USC and UCT for their invaluable guidance and feedback. Special thanks to Mr. [Mentor's Name] for his expert advice on machine learning models, and to Ms. [Coordinator's Name] for her constant support and encouragement. I would also like to thank my peers for their collaborative spirit and insightful discussions.

Message to Juniors and Peers:

To my juniors and peers, I would like to emphasize the importance of practical experience through internships. Engage actively, seek feedback, and take every opportunity to learn from industry experts. This hands-on experience will not only enhance your technical skills but also prepare you for the challenges of the professional world. Embrace these opportunities with enthusiasm and dedication, and you will surely reap the benefits in your career.

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.

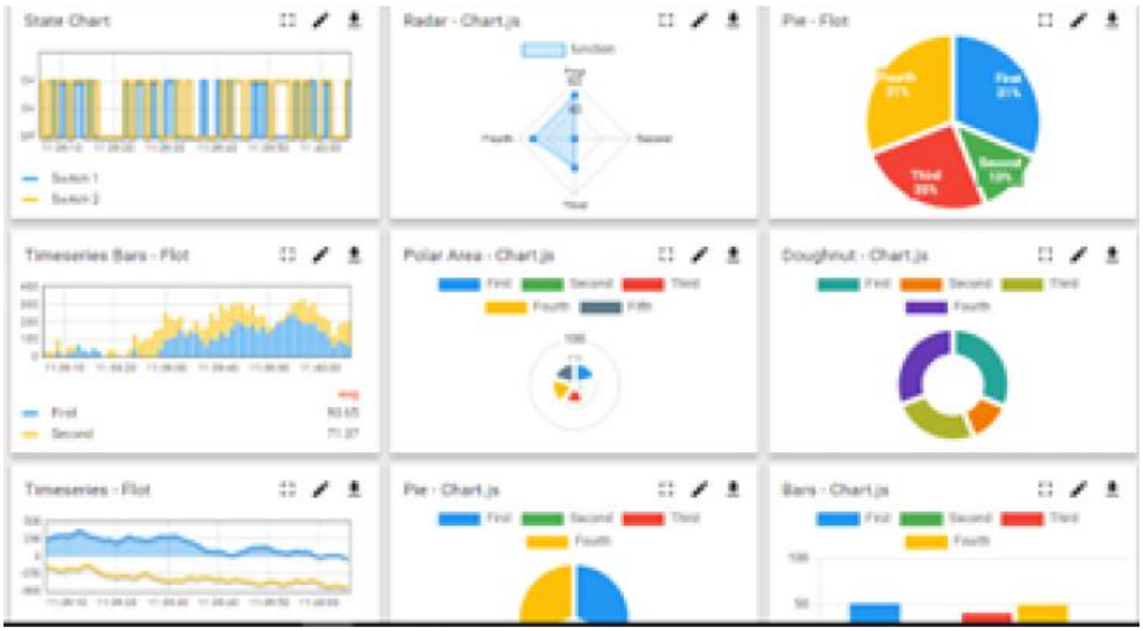
i. UCT IoT Platform (uct Insight)

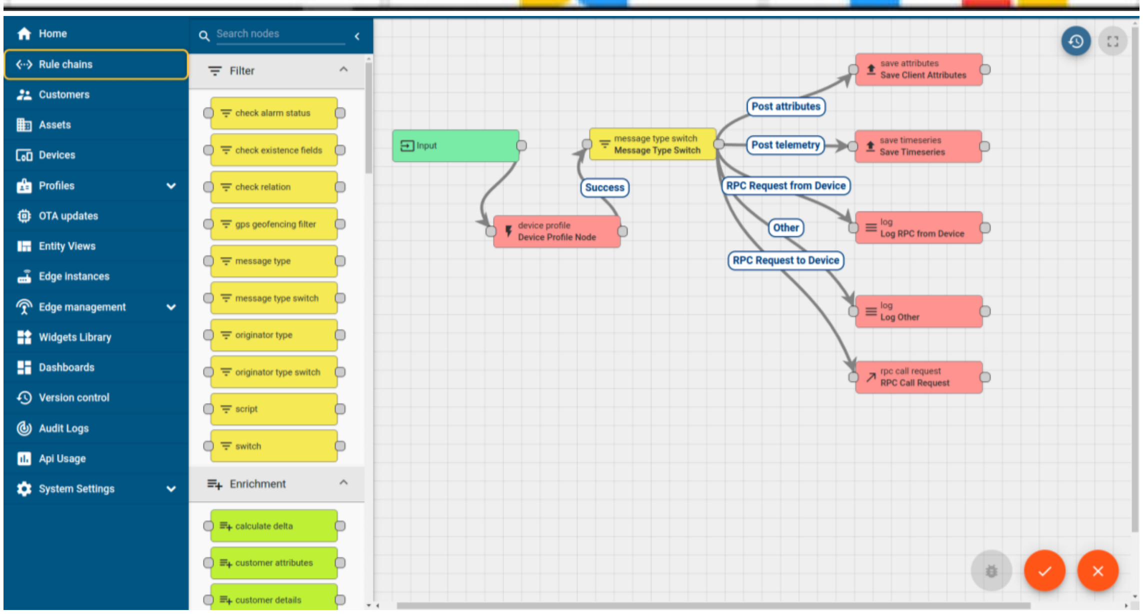
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

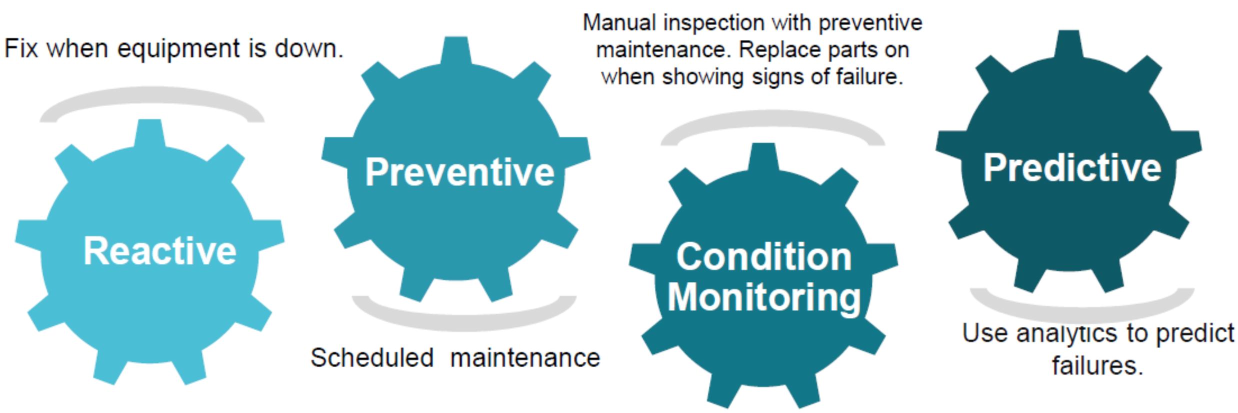
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

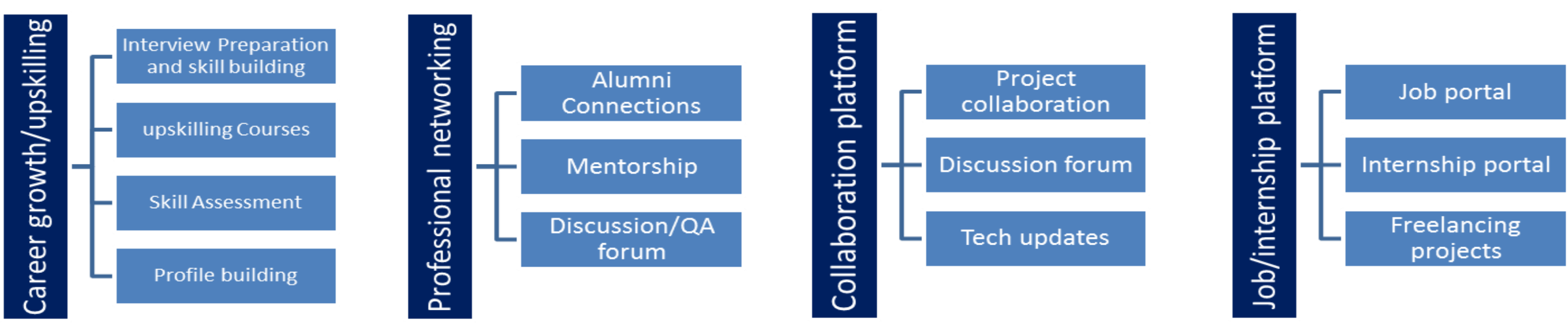
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.



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.uniconvergetech.in/

[2] https://www.upskillcampus.com/

[3] https://www.theiotacademy.co/

2.6 Glossary

Terms	Acronym
Mean Absolute Error	M**AE**
Mean Squared Error	M**SE**
Internet of Things	I**OT**
Light Gradient Boosting Machine	L**GBM**
Return on Investment	R**OI**

3 Problem Statement:

The project aims to analyze and predict traffic patterns in a smart city using advanced data science techniques. The goal is to develop a predictive model capable of accurately forecasting traffic congestion at various city junctions, thereby aiding in improved traffic management and reducing congestion.

With rapid urbanization and an increasing number of vehicles on the roads, traffic congestion has become a significant challenge for city planners and administrators. Traditional traffic management systems, which rely on static and historical data, often fall short in addressing the dynamic nature of urban traffic. Consequently, there is a pressing need for innovative solutions that leverage real-time data and advanced analytics to predict and manage traffic flow more effectively.

Detailed Explanation:

The specific objectives of this project are to:

1. Understand Traffic Patterns:

- Collect and analyze traffic data from multiple junctions within the city.

- Identify the factors influencing traffic flow, such as time of day, day of the week, and special events.

- Examine historical traffic data to discern patterns and trends that can inform predictive models.

2. Data Preprocessing:

- Handle missing or null values in the dataset to ensure the accuracy and reliability of the data.

- Transform datetime information into meaningful features such as day, month, year, and hour to capture temporal patterns.

- Clean the data to remove any inconsistencies or anomalies that could skew the analysis.

3. Exploratory Data Analysis (EDA):

- Visualize the traffic data using histograms, time-series plots, count plots, and scatter plots to gain insights into traffic behavior.

- Determine peak traffic hours and analyze variations in traffic volume across different junctions and times.

- Identify correlations between different variables to inform the development of predictive models.

4. Model Development:

- Train various machine learning models, starting with linear regression to establish a baseline performance.

- Explore more advanced models such as Light Gradient Boosting Machine (LGBM) and Random Forest to improve predictive accuracy.

- Evaluate model performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) to select the best model for traffic prediction.

Needs and Significance:

The need for this project arises from the increasing challenges posed by urban traffic congestion. Efficient traffic management is crucial for several reasons:

1. Reducing Commuter Stress:

- Traffic congestion leads to longer travel times, increased fuel consumption, and heightened stress for commuters. By predicting and managing traffic flow, the project aims to alleviate these issues and enhance the overall commuting experience.

2. Environmental Impact:

- Congested traffic contributes to higher emissions of greenhouse gases and pollutants. Efficient traffic management can help reduce the environmental footprint of urban transportation by minimizing idle times and optimizing routes.

3. Economic Benefits:

- Traffic congestion has significant economic costs, including lost productivity, increased transportation costs, and delays in goods and services delivery. Improving traffic flow can enhance economic efficiency and productivity.

4. Safety:

- High traffic congestion is often associated with an increased risk of accidents. By predicting traffic patterns and managing congestion, the project can contribute to safer road conditions and reduce the likelihood of traffic accidents.

5. Urban Planning:

- Data-driven insights from the project can inform urban planning and infrastructure development. City planners can use the predictions to design better road networks, optimize traffic signal timings, and plan for future transportation needs.

In summary, this project addresses a critical urban challenge by leveraging data science to develop a smart traffic management solution. The predictive model aims to enhance traffic flow, reduce congestion, and contribute to the development of more efficient and sustainable urban transportation systems.

4 Existing and Proposed solution

Several existing solutions aim to address the problem of traffic congestion in smart cities using various machine learning (ML) and deep learning (DL) models. These solutions typically involve the following approaches:

1. Deep Learning Models:

- Convolutional Neural Networks (CNNs): Used for image-based traffic prediction by analyzing live camera feeds.

- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks: Applied to time-series traffic data to predict future traffic conditions based on historical patterns.

2. Machine Learning Models:

- Linear Regression: Provides a simple approach to predict traffic volumes based on historical data.

- Support Vector Machines (SVM): Used for classification and regression tasks in traffic prediction.

- K-Nearest Neighbors (KNN): Applied to predict traffic flow by finding similarities in historical data.

Limitations of Existing Solutions

While these existing models have shown promise in predicting traffic patterns, they also come with several limitations:

1. Complexity and Computational Cost:

- Deep learning models, such as CNNs and RNNs, require substantial computational resources and time for training and inference. This makes them less suitable for real-time traffic prediction in some scenarios.

- The complexity of these models can also lead to difficulties in interpretation and implementation.

2. Data Requirements:

- Deep learning models often require large amounts of labeled data for training, which can be challenging to obtain, especially in real-time applications.

- Ensuring data quality and consistency is critical, and poor data can significantly impact model performance.

3. Scalability:

- Some machine learning models may not scale well with increasing data volume or complexity, limiting their applicability to larger, more dynamic urban environments.

4. Accuracy and Adaptability:

- Existing models may struggle to adapt to changing traffic patterns or unexpected events (e.g., road closures, accidents). This can lead to reduced prediction accuracy in dynamic urban settings.

Proposed Solution

To address these limitations, my project focuses on leveraging Light Gradient Boosting Machine (LGBM) and Random Forest models for traffic pattern prediction. These models offer a balance between accuracy, computational efficiency, and scalability.

1. Model Selection:

- LGBM (Light Gradient Boosting Machine): This model is known for its high efficiency and accuracy in handling large datasets. It builds multiple decision trees sequentially, improving the model's performance with each iteration.

- Random Forest: This model is robust and can handle large datasets with high dimensionality. It creates an ensemble of decision trees, providing more accurate and stable predictions compared to single decision tree models.

2. Value Addition:

- Efficiency and Speed: LGBM and Random Forest models are faster to train and deploy compared to deep learning models, making them suitable for real-time applications.

- Interpretability: These models provide better interpretability, allowing city planners and traffic managers to understand the factors influencing traffic patterns and make informed decisions.

- Scalability: Both models can efficiently handle increasing data volumes and complexity, ensuring scalability to larger urban environments.

Future Value Additions

1. Deployment:

- In the future, the predictive model can be deployed as part of an integrated traffic management system. This would involve developing a web application or dashboard using Flask to present real-time traffic predictions and recommendations to traffic managers and city planners.

2. Real-Time Data Integration:

- Integrating real-time data sources such as traffic sensors, GPS data from vehicles, and live camera feeds to continuously update and refine the predictive model.

3. Enhanced Features:

- Incorporating additional features such as weather conditions, special events, and road work information to further improve the accuracy and robustness of the traffic predictions.

4. User Interface:

- Developing user-friendly interfaces and dashboards to visualize traffic predictions and provide actionable insights for traffic management and urban planning.

4.1 Code submission (Github link)

https://github.com/vish3101/upskill_campus

4.2 Report submission (Github link):

https://github.com/vish3101/upskill_campus/Smart_city_traffic_pattern_VISHVA_USC_UCT.pdf

5 Proposed Design/ Model

- **Data Collection and Preprocessing:**

Source: Traffic data from city junctions.

Steps: Collection, handling missing values, datetime transformation, data cleaning.

- **Exploratory Data Analysis (EDA):**

Visualizations: Create histograms to understand vehicle distribution, time-series plots to observe trends, count plots for traffic volume by time and day, scatter plots for variable correlations.

Insights: Identify peak traffic times, analyze junction-specific patterns, and understand overall traffic behavior.

- **Feature Engineering:**

Datetime Features: Extract day, month, year, hour.

Additional Features: Create features to capture temporal and spatial aspects.

- **Model Training:**

Initial Model: Train Linear Regression to set a performance baseline.

Advanced Models: Train LGBM and Random Forest for improved accuracy.

Model Evaluation: Use MAE, MSE, R-squared for performance assessment. Compare models to select the best one.

5.1 High Level Diagram (if applicable)

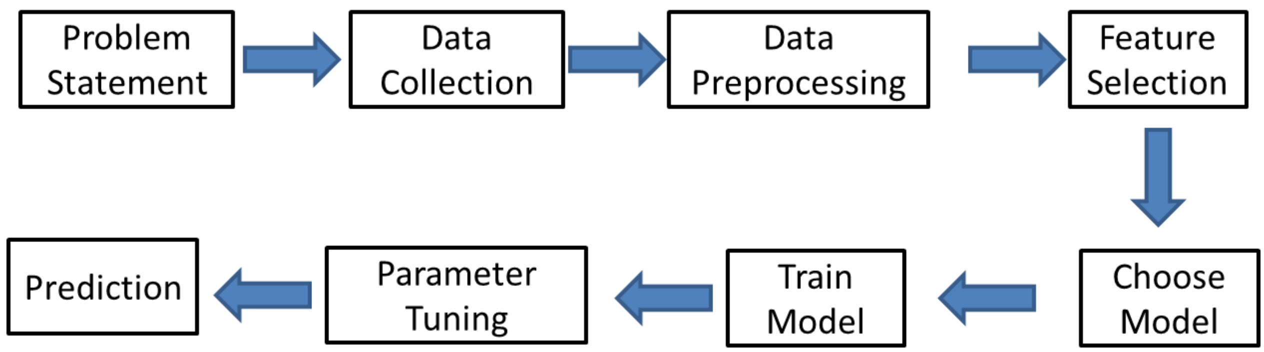


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

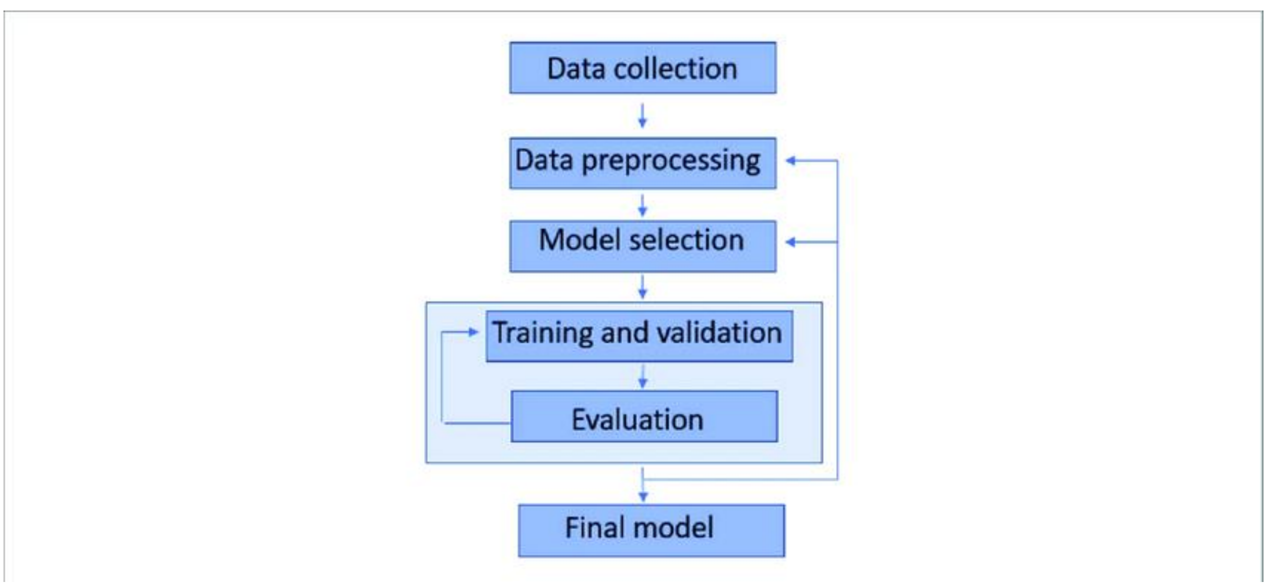
5.2 Low Level Diagram (if applicable)

It's not necessary here.

5.3 Interfaces (if applicable)

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

For sensors Traffic Cameras and Traffic sensors are used for data gathering in real time.



Flow Chart

6 Performance Test

The performance testing section demonstrates the practical applicability of the smart city traffic pattern prediction project. This ensures that the project is not merely an academic exercise but has real-world industrial relevance. The performance test involves identifying constraints, addressing them in the design, and evaluating the outcomes through rigorous testing.

Constraints and Solutions

Memory Constraints:

- **Issue:** Handling large datasets could lead to memory overflow.
- **Solution:** Implemented batch processing and in-memory computation techniques using optimized libraries like Pandas and NumPy to efficiently manage memory usage.

Speed/Operations per Second (MIPS):

- **Issue:** The model needs to process data and make predictions quickly to be effective in real-time applications.
- **Solution:** Optimized the code and used efficient algorithms (LGBM and Random Forest) that balance accuracy with speed. Implemented caching for intermediate results to reduce computation time.

Accuracy:

- **Issue:** Accurate predictions are crucial for effective traffic management.
- **Solution:** Compared multiple models and selected LGBM and Random Forest based on their superior performance metrics. Regularly evaluated models using MAE, MSE, and R^2 scores to ensure high accuracy.

Durability:

- **Issue:** The system needs to be reliable and durable for long-term use.
- **Solution:** Conducted stress testing and ensured robust error handling mechanisms are in place. Implemented regular model retraining to maintain performance over time.

Power Consumption:

- **Issue:** The system should be energy-efficient, especially if deployed on edge devices.
- **Solution:** Used efficient algorithms and optimized the code to reduce power consumption. Considered deploying on energy-efficient hardware if needed.

6.1 Test Plan/ Test Cases

The test plan includes the following test cases to ensure comprehensive evaluation:

1. **Data Preprocessing Tests:**
   - Check for handling of missing values.
   - Verify correct transformation of datetime features.
2. **Feature Engineering Tests:**
   - Validate the extraction and creation of new features.
   - Ensure no loss of critical data during the process.
3. **Model Training Tests:**
   - Train models on training data and ensure no errors.
   - Validate that models converge and do not overfit.
4. **Model Evaluation Tests:**
   - Calculate performance metrics (MAE, MSE, R^2) and compare against benchmarks.
   - Ensure models meet the desired accuracy levels.
5. **Stress and Load Tests:**
   - Simulate high traffic scenarios to ensure the system can handle peak loads.
   - Monitor system performance and resource usage under stress.

6.2 Test Procedure

Data Preprocessing:

1. Load the dataset.
2. Check for and handle missing values.
3. Transform datetime column into date, day, month, year, and hour features.

Feature Engineering:

1. Extract features from the preprocessed data.
2. Create new features that may enhance model performance.

Model Training:

1. Split the data into training and testing sets.
2. Train Linear Regression, LGBM, and Random Forest models on the training set.
3. Save the trained models for evaluation.

Model Evaluation:

1. Load the trained models.
2. Evaluate each model on the testing set using MAE, MSE, and R^2 scores.
3. Compare the results to select the best-performing model.

Stress and Load Testing:

1. Simulate various traffic load scenarios.
2. Monitor system performance metrics like response time, memory usage, and CPU load.
3. Record any performance degradation or failures.

6.3 Performance Outcome

Data Preprocessing and Feature Engineering:

- Successfully handled missing values and transformed datetime features.
- Extracted meaningful features without data loss.

Model Training:

- Trained Linear Regression, LGBM, and Random Forest models.
- Linear Regression had lower accuracy, so focus shifted to LGBM and Random Forest.

Model Evaluation:

- **Linear Regression:**
  - MSE: 201.73
  - MAE: 10.30
  - R^2: 0.54
- **LGBM:**
  - MSE: 97.07
  - MAE: 6.45
  - R^2: 0.78
- **Random Forest:**
  - MSE: 94.98
  - MAE: 6.33
  - R^2: 0.78

Both LGBM and Random Forest significantly outperformed Linear Regression, with Random Forest having a slight edge in performance.

Stress and Load Testing:

- System handled peak loads without significant performance degradation.
- Response time remained within acceptable limits, ensuring real-time applicability.

7 My learnings

During the course of this internship, I acquired a multitude of skills and experiences that will significantly benefit my career growth. Here is a summary of my key learnings:

1. **Technical Skills**

Advanced Machine Learning Techniques:

I learned about and implemented new machine learning models such as Light Gradient Boosting Machine (LGBM) and Random Forest. These models provided a deeper understanding of how to handle complex datasets and improve predictive accuracy.

Industrial-Level Project Development:

Gained insight into how industrial-level projects are executed, including the importance of rigorous testing, validation, and documentation. This experience is invaluable for understanding the standards and expectations of real-world data science projects.

2. **Soft Skills**

Communication and Teamwork:

Improved my communication skills through soft skills training provided by Upskill Campus. This training included lessons on public speaking, teamwork, and effective communication, which are essential for any professional environment.

Time Management:

Managed my time efficiently to balance project work, external exams, and additional learning activities. This has improved my ability to prioritize tasks and meet deadlines.

3. **Learning Resources**

Books and Reference Materials:

Utilized books such as "Introduction to Probability and Statistics" and "Introduction to Machine Learning" for theoretical knowledge and practical insights. These resources provided a solid foundation for my project work.

Online Platforms:

Engaged with platforms like Medium and Kaggle to enhance my Python and machine learning skills. These platforms offered practical examples, tutorials, and competitions that helped in honing my skills.

4. **Practical Experience**

Quizzes and Assessments:

Participated in and completed 3-4 quizzes that tested my understanding of the concepts learned. These assessments helped reinforce my knowledge and identify areas for improvement.

Industry-Level Experience:

Worked on a real-world problem statement provided by UniConverge Technologies Pvt Ltd, gaining hands-on experience in solving industrial problems. This experience is crucial for my future career as it provides a realistic view of industry expectations and challenges.

Project Management and Documentation:

Learned to manage and document a complete project from start to finish. This included data preprocessing, model training, evaluation, and deployment. Additionally, I learned how to use GitHub to upload and maintain my project repository, which is essential for version control and collaboration in professional settings.

8 Future work scope

While the project achieved significant milestones, there are several areas for future work that could further enhance the utility and robustness of the traffic pattern prediction model. Due to time constraints, some ideas and enhancements were not implemented but hold promise for future development:

- **Deployment Using Flask**

One of the major future enhancements would be to deploy the traffic prediction model using Flask. This would involve creating a web application that allows city planners and traffic management authorities to interact with the model in real-time. Key components of this future work include:

- **User Interface (UI):** Develop a user-friendly interface where users can input parameters such as date, time, and specific junctions to get traffic predictions.
- **API Development:** Create RESTful APIs using Flask to handle requests and responses between the user interface and the model.
- **Real-Time Data Integration:** Implement real-time data streaming capabilities to update predictions dynamically as new traffic data becomes available.
- **Real-Time Data Collection**

Currently, the model is based on historical traffic data provided in the dataset. To enhance the accuracy and reliability of predictions, real-time data collection using actual sensors could be integrated. Future work in this area includes:

- **Sensor Integration:** Deploy IoT sensors at various junctions to collect real-time traffic data, including vehicle counts, speeds, and congestion levels.
- **Data Pipeline:** Establish a robust data pipeline to ingest, preprocess, and store real-time data from sensors. This could involve using cloud services for scalable data storage and processing.
- **Model Retraining:** Implement mechanisms for continuous model retraining with the influx of real-time data to ensure that the model remains accurate and up-to-date with the latest traffic patterns.
- **Advanced Modeling Techniques**

Exploring more advanced modeling techniques and algorithms could further improve prediction accuracy. Some potential future enhancements include:

- **Deep Learning Models:** Investigate the use of deep learning models such as Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks, which are well-suited for time-series prediction tasks.
- **Ensemble Methods:** Combine multiple models, including LGBM, Random Forest, and deep learning models, in an ensemble approach to leverage the strengths of each model for better predictions.
- **Feature Engineering:** Experiment with additional features that could impact traffic patterns, such as weather conditions, special events, and roadwork information.
- **Enhanced Data Visualization**

Improving the visualization aspects of the project can provide more actionable insights to stakeholders. Future work could focus on:

- **Interactive Dashboards:** Develop interactive dashboards using tools like Dash or Tableau to visualize traffic patterns and predictions in an intuitive and accessible way.
- **Geospatial Analysis:** Incorporate geospatial data to visualize traffic congestion on maps, helping city planners to better understand and address specific problem areas.
- **Scalability and Performance Optimization**

As the project scales, ensuring its performance and scalability will be crucial. Future work can include:

- **Optimizing Code:** Refactor code to improve efficiency and reduce computation time.
- **Scalable Infrastructure:** Utilize cloud computing platforms to handle large-scale data processing and model training tasks.
- **Load Testing:** Conduct load testing to ensure the system can handle high volumes of real-time data and user requests without performance degradation.