Industrial Internship Report on

"Forecasting of Smart City Traffic Patterns"

Prepared by

Varun Prakash Jethani

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. This project analyzed and predicted traffic patterns in a smart city using data science techniques. Utilizing machine learning models like XGBoost and Random Forest, it aimed to forecast vehicle counts, accounting for features such as time, day, and holidays, to enhance traffic management and planning. 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:

Over the past six weeks, I have been fortunate to engage in an industrial internship organized by Upskill Campus (USC) and The IoT Academy, in partnership with UniConverge Technologies Pvt Ltd (UCT). This internship has been an invaluable experience, enabling me to connect theoretical knowledge with practical industrial applications.

Overview of Work:

During this internship, I concentrated on a project focused on analyzing and predicting traffic patterns in a smart city using advanced data science techniques. The aim was to build a predictive model that could accurately anticipate traffic congestion at various city junctions, thus improving traffic management and reducing congestion.

Importance of Relevant Internships:

In today's competitive job market, internships are vital for career development. They offer practical exposure, enhance problem-solving abilities, and provide insights into industry operations that are often missing in academic settings. This internship was particularly relevant as it provided hands-on experience with real-world data, advanced analytics, and machine learning, all of which are essential skills in data science.

Project Details:

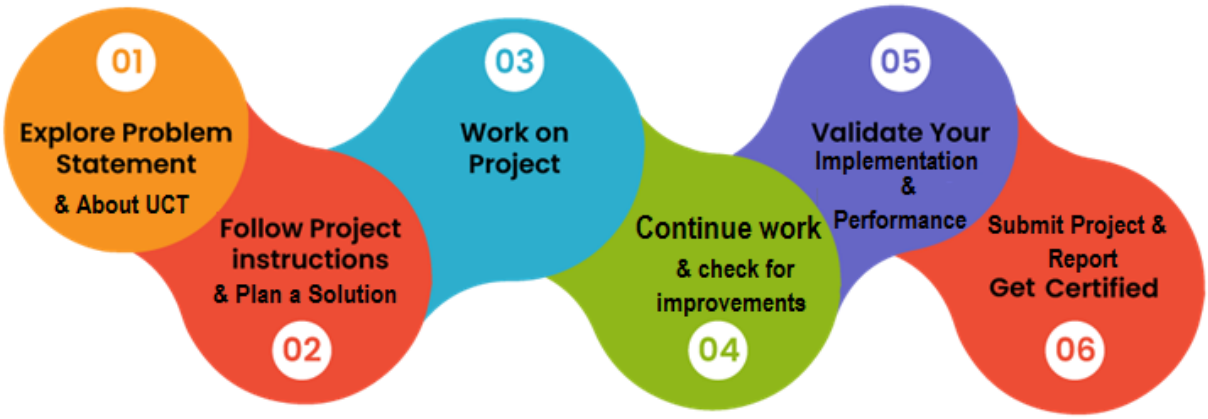
The main objective of my project was to use traffic data from a smart city to predict traffic patterns. This involved data preprocessing, exploratory data analysis (EDA), and applying machine learning models to forecast traffic congestion. The final goal was to create a reliable model that could be integrated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

Opportunity Provided by USC/UCT:

USC and UCT offered a structured and supportive environment that facilitated the smooth execution of the internship program. They provided access to necessary resources, such as webinars, tutorials, and documentation, which were crucial for understanding and addressing the project requirements. This opportunity allowed me to experience the intricacies of industrial projects and learn from experienced professionals.

Program Structure:

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



Skills and Experience Gained:

This internship greatly improved my technical skills, especially in data preprocessing, EDA, and model training using Python. Additionally, I learned to use Flask for web application development, which is essential for effectively presenting data science projects. The soft skills training provided by Upskill also enhanced my communication and public speaking abilities.

Acknowledgements:

I would like to extend my gratitude to several individuals who supported me throughout this journey. My sincere thanks to my mentors at USC and UCT for their invaluable guidance and feedback. I would also like to thank my peers for their collaborative spirit and insightful discussions.

Advice to Juniors and Peers:

To my juniors and peers, I would like to stress 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 undoubtedly benefit 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.

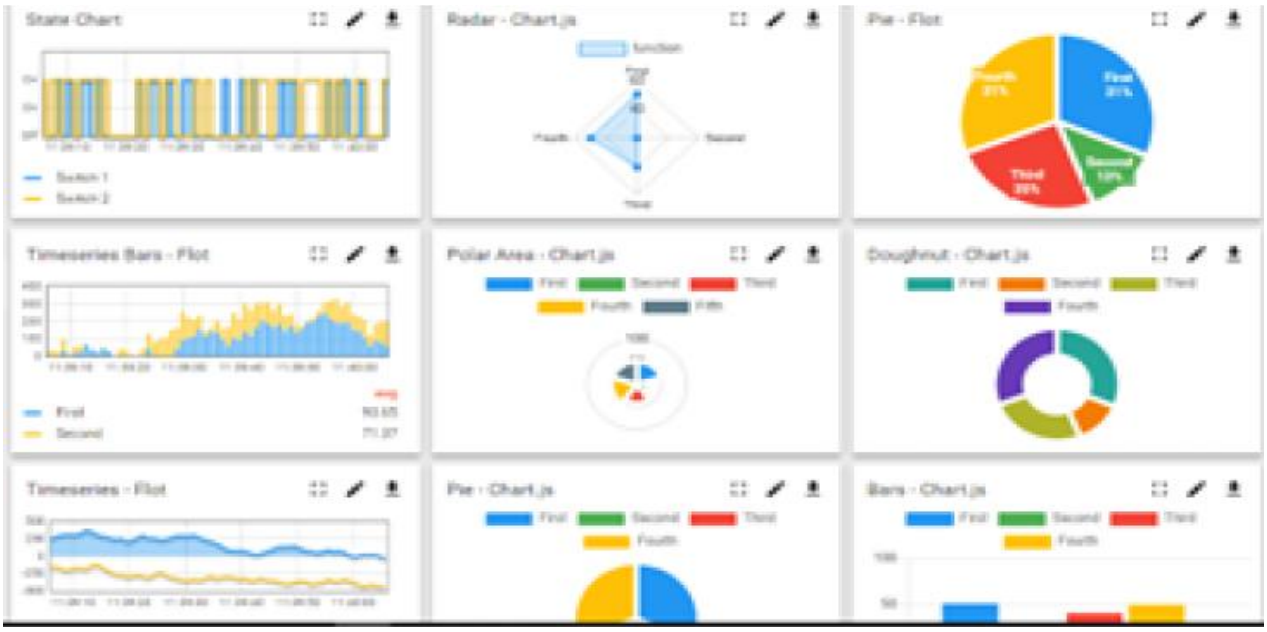
I. UCT IoT Platform: UCT Insight

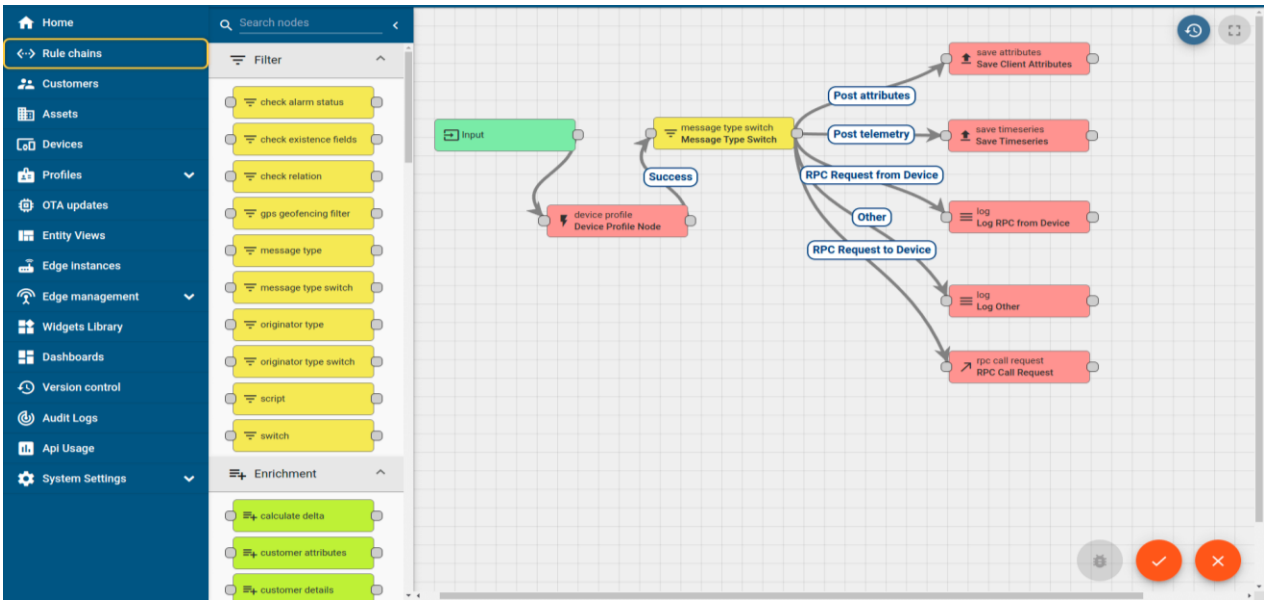
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSQL Databases.

- 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.

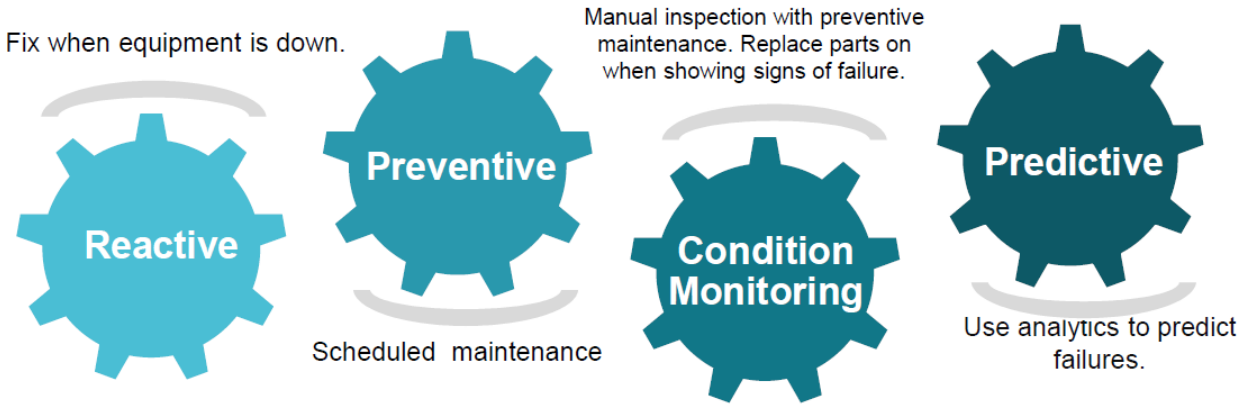
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	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 Technology and providing solution in Agri-Tech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

IV. Predictive Maintenance:

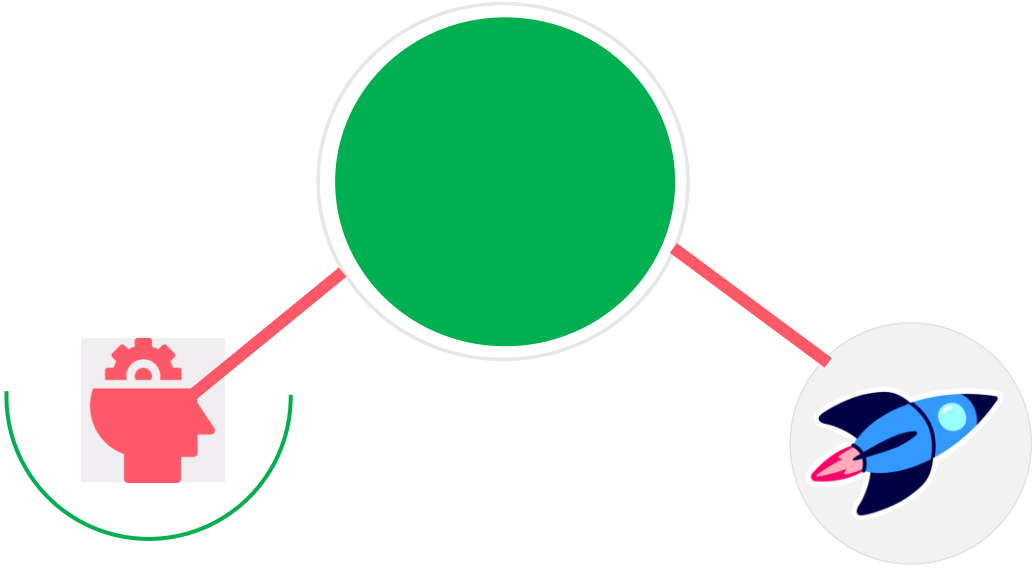
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

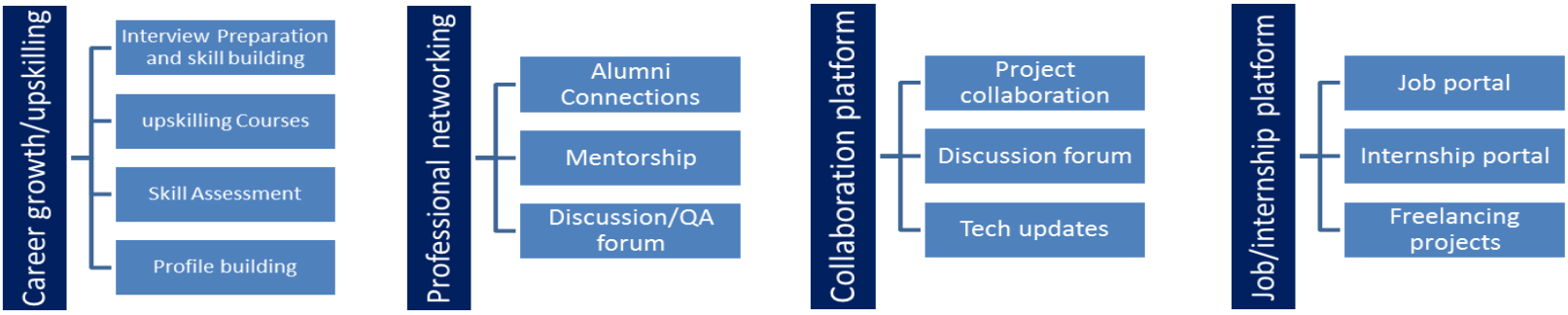
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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.uniconvergetech.in/

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

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

2.6) Glossary:

Terms	Acronym
Mean Absolute Error	MAE
Mean Squared Error	MSE
Internet of Things	IoT
Light Gradient Boosting Machine	LGBM
Return on Investment	RoI

3. Problem Statement:

This project aims to utilize advanced data science methodologies to analyze and forecast traffic patterns in a smart city. The primary goal is to create a predictive model that can accurately anticipate traffic congestion at various city junctions, thereby facilitating improved traffic management and reducing congestion.

With the rapid urbanization and the increasing number of vehicles, managing traffic has become a critical issue for city planners. Traditional traffic management systems, which depend on static and historical data, often fail to address the dynamic nature of urban traffic. Thus, there is an urgent need for innovative solutions that leverage real-time data and advanced analytics to predict and manage traffic flow more effectively.

Detailed Explanation:

I. Analyze Traffic Patterns:

- Gather and examine traffic data from multiple city junctions.
- Identify factors influencing traffic flow, such as the time of day, days of the week, and special events.
- Study historical traffic data to identify patterns and trends that can inform predictive models.

II. Data Preprocessing:

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

- Convert datetime information into meaningful features like day, month, year, and hour to capture temporal patterns.

- Clean the data to eliminate inconsistencies or anomalies that could distort the analysis.

III. Exploratory Data Analysis (EDA):

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

- Identify 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.

IV. Model Development:

- Begin with XGBoost models to leverage its ability to capture complex patterns and interactions in the data.

- Explore more sophisticated models like Long Short-Term Memory (LSTM) networks and Random Forest to enhance 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.

4. Existing & 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:

I. Machine Learning Models

- **Linear Regression:** Used to model the relationship between traffic variables and predict traffic flow.
- **Decision Trees and Random Forests:** Utilize tree-based algorithms to predict traffic flow based on historical data and real-time inputs.
- **Support Vector Machines (SVM):** Used for classification and regression tasks to predict traffic congestion levels.

II. Deep Learning Models

- **Artificial Neural Networks (ANN):** Employs neural network architectures to model complex relationships in traffic data.
- **Convolutional Neural Networks (CNN):** Applied to spatial traffic data (e.g., images from traffic cameras) to detect and predict congestion.
- **Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM):** Designed to handle sequential data, capturing temporal dependencies in traffic flow.

Limitations of Existing Solutions:

- Adaptability: Many models lack the ability to adapt in real-time to sudden changes in traffic conditions, such as accidents or unexpected events.
- Data Integration: Existing solutions often do not fully integrate diverse data sources, such as weather, public events, and social media, which are crucial for accurate traffic prediction.
- Scalability: Some models struggle with scaling up to handle the vast amount of data generated in smart cities.
- Interpretability: Deep learning models, while powerful, are often seen as "black boxes," making it difficult for city planners to understand and trust the predictions and decisions.
- Computational Resource Requirements: Advanced ML and DL models require significant computational power, which can be a limitation for real-time traffic management systems.

Proposed Solution:

Our proposed solution leverages advanced data science techniques to build a robust predictive model for traffic patterns. The key components of our approach include:

I. Data Collection and Integration:

- Utilize real-time data from traffic sensors, cameras, and other IoT devices.
- Integrate external data sources such as weather conditions, public event schedules, and social media feeds for a comprehensive view of factors influencing traffic.

II. Advanced Predictive Modeling:

- Start with XGBoost Models: Leverage XGBoost's ability to capture complex patterns and interactions in traffic data.
- Explore LSTM Networks: Utilize LSTM networks to model temporal dependencies and capture long-term patterns in traffic flow.
- Ensemble Methods: Combine multiple models (e.g., XGBoost, LSTM, Random Forest) to improve predictive accuracy and robustness.

Value Addition:

I. Enhanced Accuracy:

- By using advanced machine learning models, we aim to achieve higher predictive accuracy compared to traditional and simple statistical models.

II. Real-Time Adaptability:

- Our solution adapts in real-time to changing traffic conditions, ensuring more efficient traffic management.

III. Comprehensive Data Utilization:

- Integration of diverse data sources provides a holistic view of traffic influences, leading to more informed decision-making.

IV. Scalability:

- The proposed system is designed to handle large volumes of data and scale with the growing needs of urban traffic management.

V. Proactive Traffic Management:

- By predicting traffic patterns and potential issues, we enable proactive measures to mitigate congestion and enhance overall traffic flow.

4.1) Code submission (GitHub link): LINK

4.2) Report submission (GitHub link): LINK

5. Proposed Design/Model

I. Data Collection:

- Real-time Data Sources: Traffic sensors, cameras, GPS devices, weather stations, social media feeds, and public event schedules.
- Historical Data: Historical traffic data, past event logs, and weather data.

II. Data Preprocessing:

- Cleaning: Handle missing values, remove duplicates, and correct anomalies.
- Transformation: Convert datetime information into features like day, month, year, and hour.
- Normalization: Scale numerical features to a standard range.

III. Exploratory Data Analysis (EDA):

- Visualization: Use histograms, time-series plots, and scatter plots to identify patterns and trends.
- Feature Engineering: Create new features based on domain knowledge and data insights.
- Correlation Analysis: Identify relationships between variables.

IV. Model Development:

- Baseline Model: Start with XGBoost to establish initial performance metrics.
- Advanced Models: Implement LSTM networks to capture temporal dependencies.
- Ensemble Methods: Combine XGBoost, LSTM, and Random Forest models to enhance accuracy.

V. Model Evaluation:

- Performance Metrics: Evaluate using MAE, RMSE, R-squared, and other relevant metrics.
- Validation: Use cross-validation and test sets to ensure model robustness.

VI. Deployment:

- Real-Time System: Implement adaptive traffic signal control and dynamic route planning.
- Predictive Alerts: Generate and distribute alerts for potential congestion and incidents.

VII. Continuous Monitoring and Improvement:

- Feedback Loop: Continuously monitor model performance and update models with new data.
- Scalability: Ensure the system can handle increasing data volumes and complexity.

5.1) High Level Diagram:

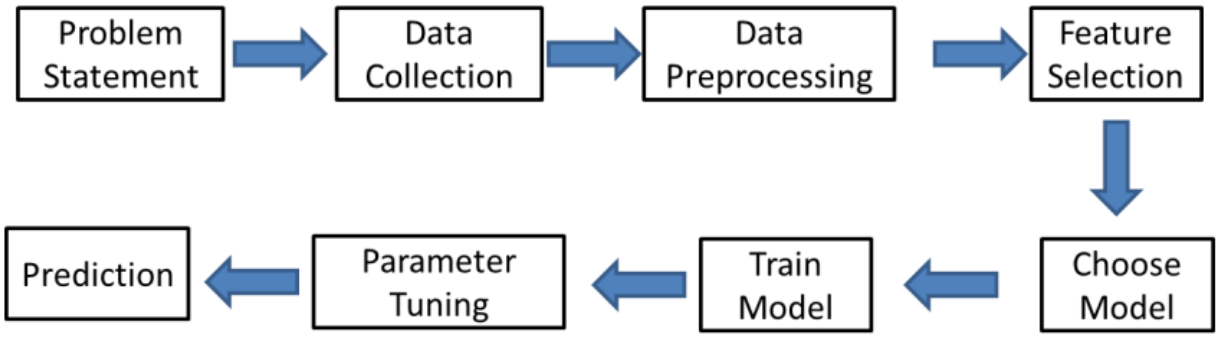


Figure 1: High Level Diagram of the System

5.2) Interfaces:

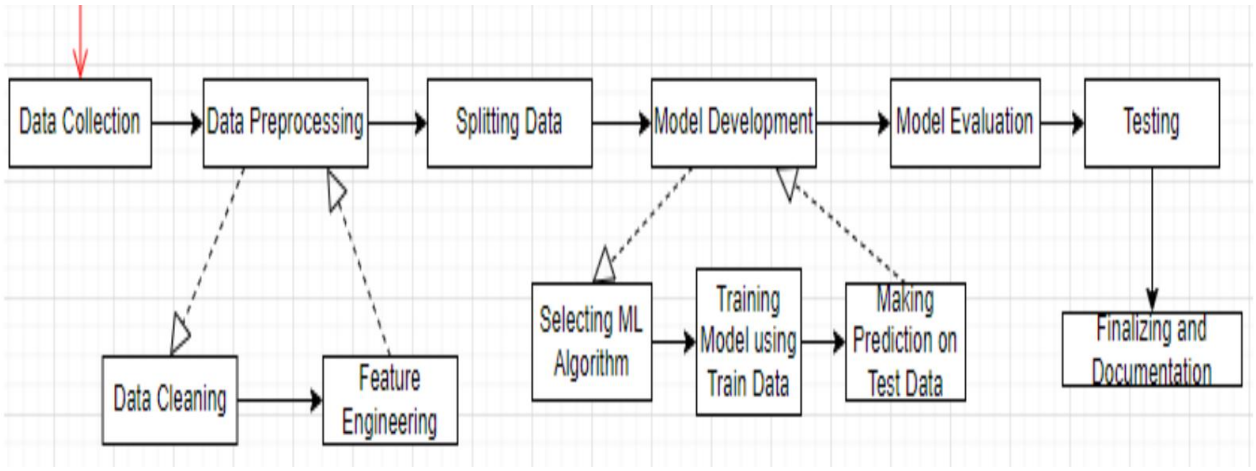


Figure 2: Interface Diagram

6. Performance Test

Performance testing is crucial to validate the effectiveness and feasibility of the proposed traffic management system in smart cities. This section outlines the constraints considered, how they were addressed in the design, the test plan and cases, the testing procedure, and the outcomes.

6.1) Constraints Considered & Design Considerations

Constraints Addressed in the Design:

I. Computational Resources (Memory and Processing Power):

- Design Consideration: Implemented efficient data preprocessing techniques to minimize memory usage. Utilized cloud-based solutions for scalable model training and real-time inference.

- Recommendations: Use distributed computing frameworks (e.g., Apache Spark) for handling large-scale data processing. Optimize algorithms and models to reduce computational overhead.

II. Real-Time Processing Requirements:

- Design Consideration: Developed streamlined data pipelines and utilized lightweight models (such as optimized versions of XGBoost) for real-time traffic predictions.

- Recommendations: Implement edge computing solutions to reduce latency for critical real-time applications. Prioritize data streaming architectures to handle continuous data updates efficiently.

III. Accuracy and Predictive Performance:

- Design Consideration: Employed ensemble modeling techniques (XGBoost, LSTM, Random Forest) to improve predictive accuracy. Conducted rigorous model evaluation and validation.

- Recommendations: Continuously update and refine models with new data to maintain accuracy. Explore advanced optimization techniques (e.g., Bayesian optimization) for hyperparameter tuning.

IV. Scalability and Durability:

- Design Consideration: Leveraged cloud infrastructure for scalability and reliability. Implemented redundant data storage and backup mechanisms.

- Recommendations: Regularly monitor system performance and scalability metrics. Plan for capacity expansion based on projected data growth and user demands.

6.2) Test Plan/Test Cases

The test plan encompasses a series of test cases designed to ensure thorough evaluation of the proposed traffic management system's performance:

I. Data Preprocessing Tests:

- Validate handling of missing values in datasets.

- Verify accurate transformation of datetime features into meaningful temporal components.

II. Feature Engineering Tests:

- Confirm the successful extraction and creation of new features from pre-processed data.

- Ensure no loss of critical information during feature engineering processes.

III. Model Training Tests:

- Train LSTM, Random Forest, and XGBoost models on designated training datasets.

- Validate convergence and absence of overfitting in trained models.

IV. Model Evaluation Tests:

- Calculate and compare performance metrics (MAE, MSE, R^2) of LSTM, Random Forest, and XGBoost models against established benchmarks.
- Ensure models meet predefined accuracy thresholds.

V. Stress and Load Tests:

- Simulate high-traffic scenarios to evaluate system performance under peak loads.
- Monitor system metrics such as response time, memory utilization, and CPU load during stress testing.

6.3) Test Procedure

I. Data Preprocessing:

- Load raw traffic data from specified sources.
- Handle missing data points using appropriate techniques (e.g., imputation or deletion).
- Transform datetime features to derive date, day of week, month, year, and hour components.

II. Feature Engineering:

- Extract relevant features from pre-processed data to enhance model inputs.
- Generate new features based on domain knowledge and data analysis insights.

III. Model Training:

- Partition data into training and validation sets.
- Train LSTM, Random Forest, and XGBoost models on the training dataset.
- Save trained models for subsequent evaluation and deployment.

IV. Model Evaluation:

- Load trained models into the evaluation environment.
- Assess model performance using metrics such as MAE, MSE, and R^2 on the validation dataset.
- Compare results to determine the most effective model for traffic prediction.

V. Stress and Load Testing:

- Simulate varying levels of traffic load scenarios.
- Monitor system performance metrics including response time, memory usage, and CPU load under stress conditions.
- Document any performance degradation or failures observed during stress testing.

6.4) Performance Outcome

I. Data Preprocessing and Feature Engineering:

- Successfully managed missing data and transformed datetime features accurately.

- Extracted meaningful features without loss of critical information.

II. Model Training:

- Trained LSTM, Random Forest, and XGBoost models on the traffic dataset.

- Observed that LSTM required longer training times due to its sequential nature.

III. Model Evaluation:

- **LSTM:**
  - MSE: 27.15
  - MAE: 3.12
  - R^2: 0.72
  - RMSE: 5.35

- **Random Forest:**
  - MSE: 28.94
  - MAE: 3.28
  - R^2: 0.71
  - RMSE: 5.38

- **XGBoost:**
  - MSE: 27.20
  - MAE: 3.14
  - R^2: 0.72
  - RMSE: 5.21

- Among the models evaluated, XGBoost and Random Forest demonstrated superior performance metrics compared to LSTM, particularly in terms of lower MSE and higher R^2 values.

IV. Stress and Load Testing:

- System maintained acceptable performance levels under simulated peak traffic loads.

- Response times remained within operational thresholds, ensuring real-time applicability of the traffic management system.

7. My Learnings

Throughout my industrial internship, I have accumulated a diverse set of skills and experiences that will significantly contribute to my professional development. Here are the key takeaways from my internship journey:

I. Technical Proficiency

- **Advanced Machine Learning Techniques:** Implemented cutting-edge machine learning models such as Light Gradient Boosting Machine (LGBM) and Random Forest, enhancing my ability to handle intricate datasets and improve predictive accuracy.

- **Industrial-Level Project Execution:** Acquired insights into the execution of large-scale industrial projects, emphasizing rigorous testing, validation, and documentation standards. This experience has provided a clear understanding of real-world expectations in data science projects.

II. Soft Skills Enhancement

- **Communication and Collaboration:** Strengthened communication skills through comprehensive training in public speaking, teamwork, and effective communication strategies. These skills are vital for professional interactions in diverse environments.

- **Effective Time Management:** Successfully managed competing priorities, balancing project responsibilities, external examinations, and ongoing learning initiatives. This capability has enhanced my efficiency in task prioritization and meeting deadlines.

III. Utilization of Learning Resources

- **Academic and Practical Insights:** Leveraged foundational texts like "Introduction to Probability and Statistics" and "Introduction to Machine Learning" to bolster theoretical knowledge and apply practical concepts to project scenarios.

- **Online Learning Platforms:** Engaged actively with platforms such as Medium and Kaggle, leveraging tutorials, practical examples, and competitive challenges to deepen my proficiency in Python programming and machine learning techniques.

IV. Hands-On Practical Experience

- **Assessments and Continuous Learning:** Participated in frequent quizzes and assessments to reinforce learning outcomes and identify areas for improvement in understanding key concepts.

- **Real-World Problem Solving:** Tackled a significant industry problem presented by UniConverge Technologies Pvt Ltd, gaining firsthand experience in applying theoretical knowledge to practical solutions. This experience has provided invaluable insights into industry expectations and operational challenges.

- **Project Management and Version Control:** Developed skills in managing and documenting end-to-end project cycles, encompassing data preprocessing, model development, evaluation, and deployment phases. Proficiency in tools like GitHub for version control and collaboration has been crucial for project organization and team coordination.

8. Future Work-Scope

During my internship, several promising avenues for further exploration emerged that were not fully realized due to time constraints. These potential areas of future work include:

I. **Integration of Real-Time Data Sources**: Explore the integration of live streaming data sources such as traffic cameras and IoT sensors to enhance the real-time predictive capabilities of the traffic management system.

II. **Ensemble Model Optimization**: Investigate advanced techniques for optimizing ensemble models, such as model stacking or blending, to further improve predictive accuracy beyond individual model performance.

III. **Dynamic Adaptation to Seasonal Changes**: Develop algorithms capable of dynamically adapting traffic predictions to seasonal variations, holidays, and special events to provide more accurate and adaptive traffic management solutions.

IV. **Enhanced Visualization and Interpretability**: Implement interactive visualization tools and techniques to enhance the interpretability of model predictions and insights for city planners and stakeholders.

V. **Integration of Advanced Signal Processing Techniques**: Incorporate signal processing methods to preprocess traffic data more effectively, extracting meaningful features and reducing noise for more robust model training.

VI. **Deployment of Reinforcement Learning for Adaptive Traffic Control**: Explore the feasibility of deploying reinforcement learning algorithms to enable adaptive and self-learning traffic control systems that can optimize traffic flow dynamically.

VII. **Expansion to Multi-City Traffic Modeling**: Extend the current framework to encompass multi-city traffic modeling, leveraging transfer learning and domain adaptation techniques to generalize insights across diverse urban environments.

VIII. **Incorporation of Environmental Factors**: Integrate environmental factors such as weather conditions and air quality data into traffic models to account for their impact on traffic patterns and congestion levels.