Industrial Internship Report on

" Smart City Traffic Patterns"

Prepared by

Jagrut Padariya

Sahaj Parmar

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 aims to transform cities into smart, digitally intelligent urban environments by optimizing traffic management. By forecasting traffic patterns at key junctions, especially during holidays and special occasions, the government can enhance infrastructure planning. This report highlights the progress made in understanding the role of a data engineer and the practical application of data engineering skills to support this endeavor. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

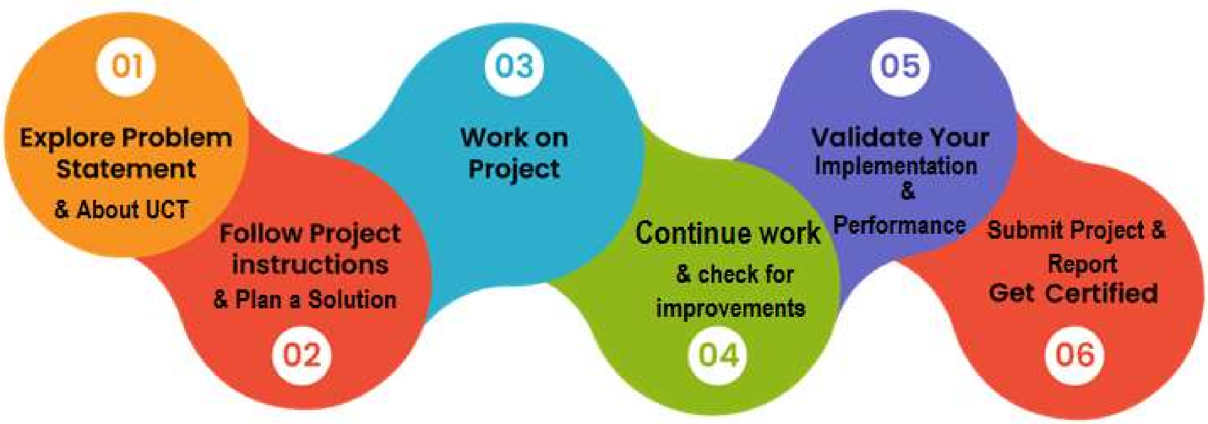
Summary of the whole 6 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

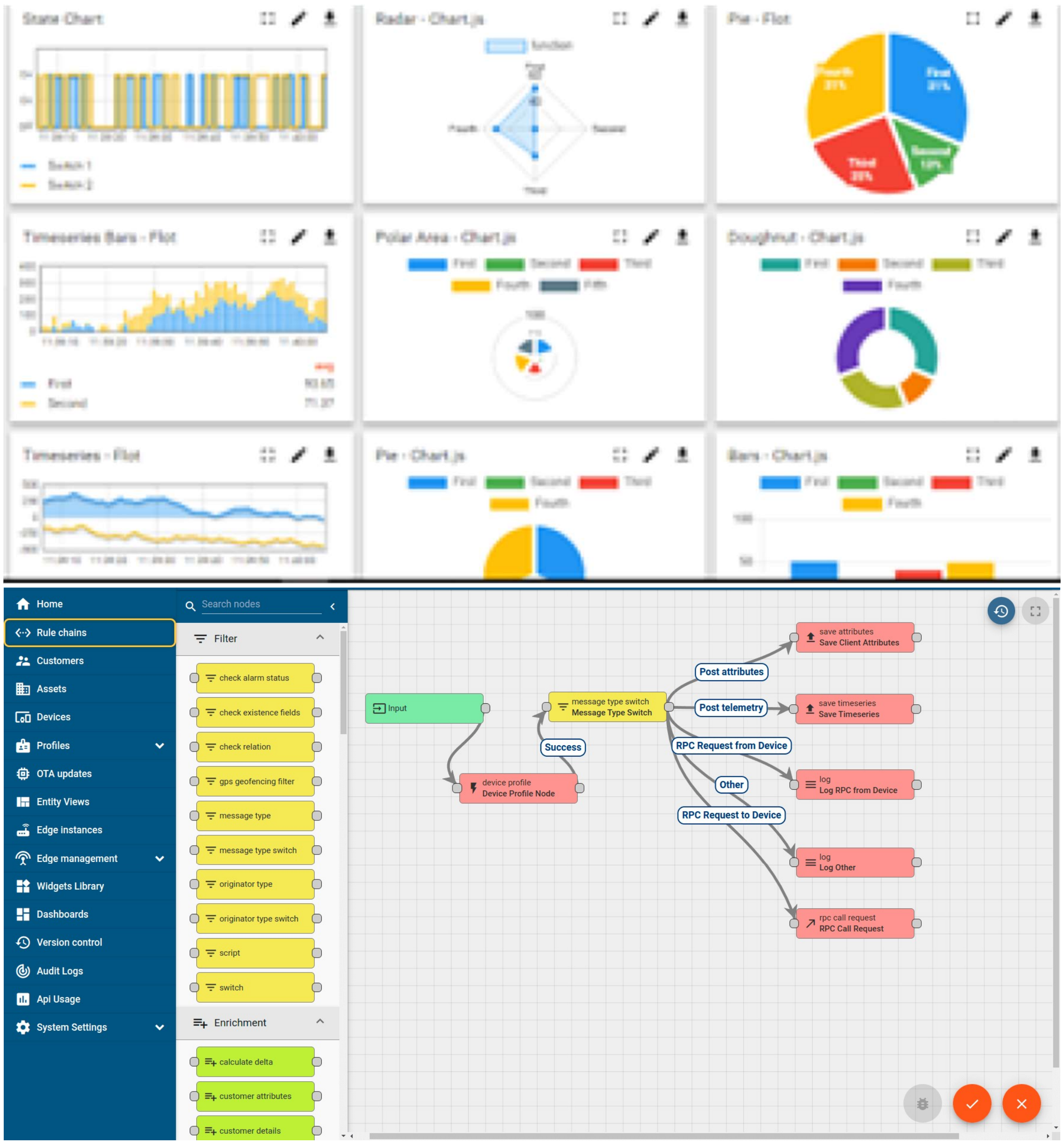
i. UCT IoT Platform (uct Insight)

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

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to

- Build Your own dashboard
- Analytics and Reporting
- Alert and Notification
- Integration with third party application(Power BI, SAP, ERP)
- Rule Engine



ii. Smart Factory Platform (FACTORY WATCH)

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

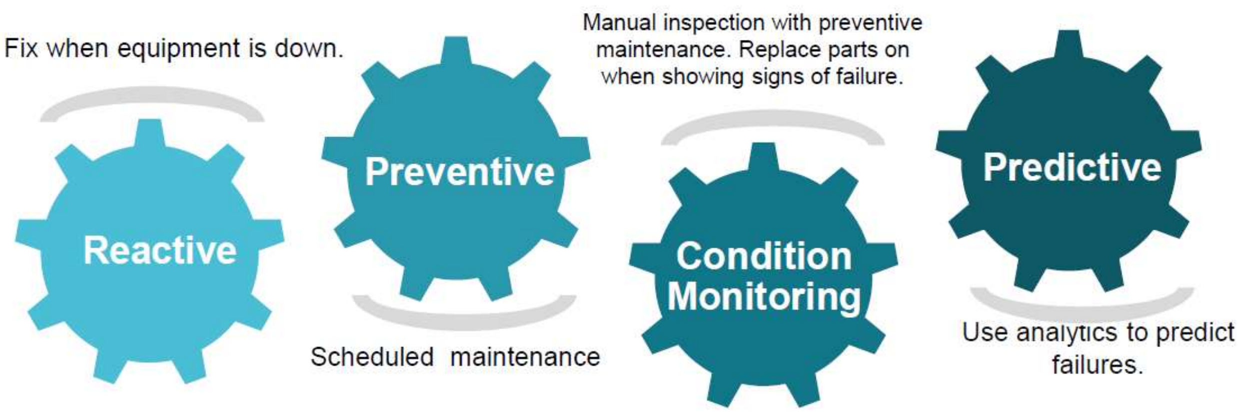
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

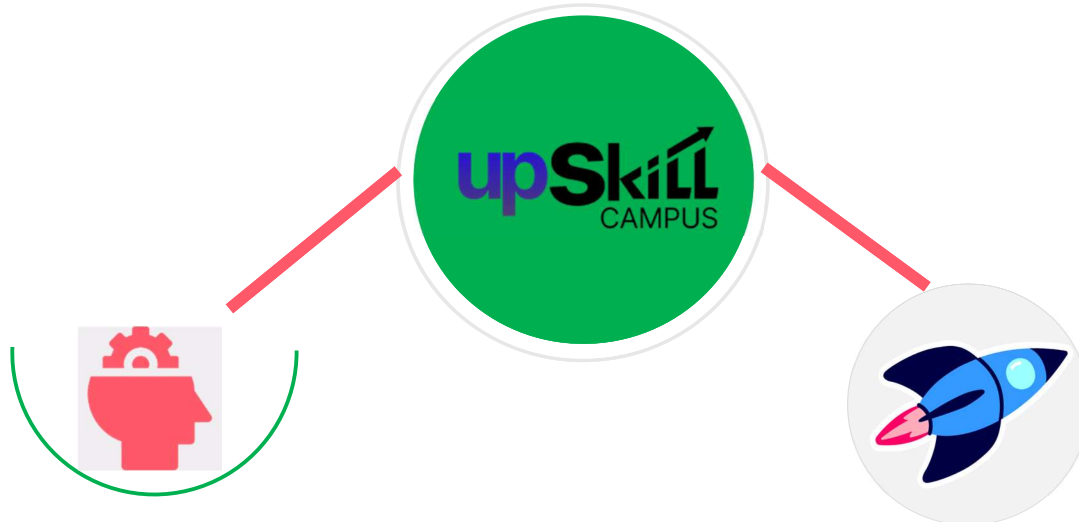
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

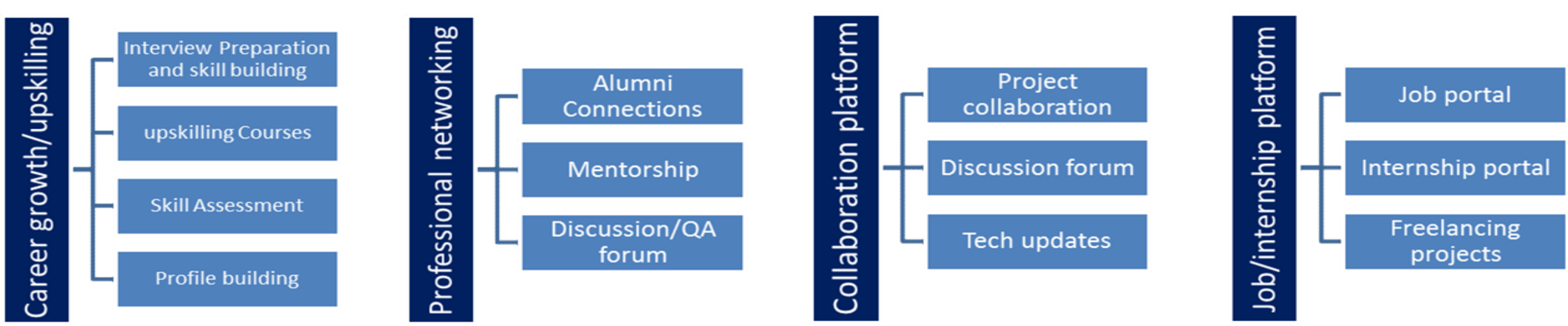
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

☛ get practical experience of working in the industry.

☛ to solve real world problems.

☛ to have improved job prospects.

☛ to have Improved understanding of our field and its applications.

☛ to have Personal growth like better communication and problem solving.

2.5 Reference

[1] https://www.kaggle.com/utathya/smart-city-traffic-patterns

[2] https://www.diva-portal.org/smash/get/diva2:975199/FULLTEXT01.pdf

[3] https://www.mdpi.com/2071-1050/15/7/5949

2.6 Glossary

Terms	Acronym
Traffic Pattern	A "traffic pattern" refers to the typical or recurring way in which vehicles and pedestrians move and interact on roads, streets, or intersections within a specific area over a period of time.
Machine Learning	Machine Learning is a subset of artificial intelligence (AI) that involves the development of algorithms and models that enable computers to learn and make predictions or decisions based on data without explicit programming.
Data Science	Data science is an interdisciplinary field that uses scientific methods, algorithms, processes, and systems to extract knowledge and insights from structured and unstructured data.
Data Analysis	Data analysis is the process of inspecting, cleaning, transforming, and interpreting data with the goal of discovering useful information, drawing conclusions, and supporting decision-making.

3 Problem Statement

Objective: The objective of this project is to transform cities into smart, digitally intelligent urban environments by optimizing traffic management. The government aims to implement a robust traffic system to prepare for traffic peaks and to understand traffic patterns at key junctions, especially during holidays and special occasions.

Challenges: The challenges include handling missing data, dealing with high-dimensional datasets from multiple sensors, incorporating dynamic factors such as holidays and events into traffic forecasting, and ensuring infrastructure planning aligns with traffic insights.

Solution: The solution involves data analysis, preprocessing, and the development of predictive models to forecast traffic patterns accurately. Additionally, infrastructure planning recommendations will be provided based on the model's insights.

Benefit: The project's success will lead to reduced traffic congestion, optimized resource allocation, improved citizen experiences, and efficient traffic management in smart cities.

Scope: The scope encompasses data collection, preprocessing, model development, evaluation, and infrastructure recommendations.

Stakeholders: The primary stakeholders include government authorities responsible for urban planning, citizens, and data scientists working on traffic optimization.

4 Existing and Proposed solution

1. **Traffic Signal Optimization:**
   - **Solution:** Using advanced algorithms to optimize traffic signal timings based on real-time traffic data.
   - **Limitations:** Often works well for fixed schedules but struggles to adapt to unpredictable traffic patterns. Real-time data accuracy is crucial, and implementation can be costly.

2. **Machine Learning for Prediction:**
   - **Solution:** Applying machine learning models to historical traffic data for prediction.
   - **Limitations:** Requires large amounts of historical data for accurate predictions. May struggle with sudden, unexpected events.

3. **Smart Traffic Cameras and Sensors:**
   - **Solution:** Installing cameras and sensors at junctions to monitor traffic in real-time.
   - **Limitations:** Costly infrastructure setup. Limited coverage and data quality issues.

4. **Public Transportation Improvement:**
   - **Solution:** Expanding and improving public transportation systems to reduce the number of private vehicles on the road.
   - **Limitations:** Requires significant investment and time. May not be feasible in all cities.

5. **Data Collection and Integration:**
   - Gather data from various sources, including traffic cameras, sensors, GPS devices, and historical records.
   - Integrate data into a centralized data repository or data lake for analysis.

6. **Data Preprocessing:**
   - Clean and preprocess the data to handle missing values, outliers, and inconsistencies.
   - Transform the data into a format suitable for analysis, including time series data for traffic patterns.
7. **Data-Driven Traffic Insights:**
   - By leveraging advanced data analysis and machine learning techniques, the proposed solution aims to provide accurate and actionable insights into traffic patterns. This includes the ability to forecast traffic conditions, identify peak times, and understand how holidays and special events impact traffic flow. These insights are crucial for informed decision-making in traffic management.

8. **Infrastructure Planning:**
    - The solution goes beyond traffic forecasting to provide recommendations for infrastructure planning. It will suggest measures such as road expansions, public transportation improvements, and traffic signal optimizations. These recommendations are data-driven and tailored to the specific needs of the smart city, ensuring smarter and more sustainable urban development.
9. **Real-Time Monitoring and Adaptation:**
    - Implementing a real-time monitoring system allows for continuous data collection and analysis. This capability enables immediate responses to changing traffic conditions, allowing traffic management strategies to adapt dynamically, reducing congestion, and improving safety.
10. **Optimized Resource Allocation:**
    - The project's recommendations will assist government authorities in optimizing resource allocation for traffic management. This means more efficient deployment of traffic personnel, better timing of traffic signals, and improved coordination of road maintenance activities, all contributing to reduced congestion and enhanced traffic flow.

4.1 Code submission (Github link)

https://github.com/JagrutPadariya/upskillcampus.git

4.2 Report submission (Github link) :

upskillcampus/Smartcitytrafficpatterns_Jagrut_USC_UCT.pdf at main · JagrutPadariya/upskillcampus (github.com)

5 Proposed Design/ Model

Random forest Model

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees.

- **Project code for implementation algorithm of Random Forest**

Create a Random Forest regressor

```
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)

# Train the model

rf_regressor.fit(X_train, y_train)

# Make predictions on the test set

y_pred = rf_regressor.predict(X_test)

#for i in range(15880):

#    print(y_pred[i],y_test.iloc[i])
```

Evaluating the model

```
mse = mean_squared_error(y_test, y_pred)

mae = mean_absolute_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

# Printing the evaluation metrics

print("Mean Squared Error:", mse)

print("Mean Absolute Error:", mae)

print("R2 Score:", r2)
```

6 Performance Test

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

Here we need to first find the constraints.

How those constraints were taken care in your design?

```
Mean Squared Error
```

```
Mean Absolute Error
```

```
R2 Score
```

What were test results around those constraints?

```
Mean Squared Error: 15.803522852644837
Mean Absolute Error: 2.5130220403022667
R2 Score: 0.9641402212958443
```

6.1 Test Plan/ Test Cases

- Miscellany: Each tree has a unique attribute, variety and features concerning other trees. Not all trees are the same.
- Immune to the curse of dimensionality: Since a tree is a conceptual idea, it requires no features to be considered. Hence, the feature space is reduced.
- Parallelization: We can fully use the CPU to build random forests since each tree is created autonomously from different data and features.
- Train-Test split: In a Random Forest, we don't have to differentiate the data for train and test because the decision tree never sees 30% of the data.
- Stability: The final result is based on Bagging, meaning the result is based on majority voting or average.

6.2 Test Procedure

Step 1: Select random samples from a given data or training set.

Step 2: This algorithm will construct a decision tree for every training data.

Step 3: Voting will take place by averaging the decision tree.

Step 4: Finally, select the most voted prediction result as the final prediction result.

6.3 Performance Outcome

```
R2 Score: 0.9641402212958443

Model Accuracy : 96%
```

7 My learnings

Through my involvement in the "Smart City Traffic Patterns" project, I have acquired valuable skills, knowledge, and experiences that are instrumental in advancing my career as a data scientist and in contributing to the broader field of urban data analytics. Here are the key takeaways:

1. Data Engineering and Preprocessing Skills: I have gained proficiency in data engineering, including data collection, integration, and preprocessing. These skills are essential in handling real-world data, ensuring data quality, and preparing data for analysis.

2. Machine Learning and Predictive Modeling: Building predictive models for traffic forecasting has deepened my understanding of machine learning algorithms and their practical applications. This expertise is transferable to various domains requiring predictive analytics.

3. Feature Engineering: I have learned the importance of feature engineering in extracting meaningful insights from data. Identifying relevant features and engineering them effectively enhances the accuracy of predictive models.

4. Domain Knowledge: The project has provided me with domain-specific knowledge in traffic management and urban planning. Understanding the intricacies of traffic patterns, holidays, and special events is valuable for addressing complex urban challenges.

5. Real-World Application: Working on a tangible, real-world project has allowed me to apply theoretical knowledge to practical scenarios. This experience is invaluable in bridging the gap between theory and real-world problem-solving.

6. Communication and Collaboration: Regular reporting and communication with government stakeholders have improved my ability to convey technical findings to non-technical audiences. Collaboration with multidisciplinary teams has honed my teamwork and project management skills.

7. Problem-Solving and Critical Thinking: The project has presented complex challenges that require creative problem-solving and critical thinking. These skills are transferable to various data science and analytical roles.

8. Long-Term Impact: Contributing to the transformation of cities into smart, efficient, and citizen-friendly environments is not only professionally rewarding but also aligns with my passion for making a positive impact on society.

8 Future work scope

The future work scope for the "Smart City Traffic Patterns" project encompasses several key areas to further enhance traffic management and infrastructure planning in smart cities:

Advanced Machine Learning Models: Continue research and development of advanced machine learning models for traffic forecasting. Explore deep learning techniques, reinforcement learning, and hybrid models that can adapt to dynamic traffic patterns and external factors.

Enhanced Data Sources: Expand the sources of data for a more comprehensive understanding of traffic patterns. This may include integrating data from IoT devices, social media, and additional sensors to capture a wider range of traffic-related information.

Real-Time Predictions: Implement real-time traffic predictions with low-latency capabilities. Develop a system that can provide up-to-the-minute traffic forecasts, enabling immediate response to changing conditions.

Dynamic Traffic Management: Incorporate machine learning models into traffic management systems to enable dynamic adjustments of traffic signals, lane configurations, and variable speed limits based on real-time predictions.

Intelligent Transportation Systems (ITS): Explore the integration of intelligent transportation systems, such as connected vehicles and smart traffic signals, to further improve traffic flow and safety.

Public Transportation Integration: Enhance public transportation systems by integrating real-time traffic data to optimize routes, schedules, and capacity planning.

Citizen Engagement: Develop mobile applications and communication channels that allow citizens to access real-time traffic information and contribute to data collection, such as reporting road incidents or accidents.

Sustainability Focus: Incorporate sustainability goals into traffic management, encouraging the use of public transportation, carpooling, and eco-friendly modes of transport to reduce congestion and emissions.

City Expansion: As the smart city expands, adapt the traffic management system to cover new areas and junctions, considering the evolving needs of a growing urban population.

Continuous Monitoring and Evaluation: Implement continuous monitoring of the system's performance, regularly assess the accuracy of predictions, and gather feedback from stakeholders for ongoing improvements.

Infrastructure Upgrades: Work closely with government authorities to execute infrastructure recommendations based on data-driven insights. Prioritize and schedule road expansions, maintenance, and other improvements accordingly.

Data Security and Privacy: Strengthen data security and privacy measures to ensure the responsible collection and use of citizen data. Comply with data protection regulations and ensure data anonymity where necessary.

Community Outreach: Engage with the local community to educate citizens about the benefits of smart traffic management, encourage participation, and address concerns.

Scalability and Resilience: Ensure that the system is designed to scale with the city's growth and is resilient to technical failures or cyber threats.

Research and Innovation: Stay updated on the latest advancements in data science, machine learning, and urban planning to continually innovate and adapt to emerging challenges.