**Industrial Internship Report on**

**”Smart City Traffic Pattern”**

**Prepared by**

**Krishnakant N. Verma**

|  |
| --- |
| *Executive Summary* |
| This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).  This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time.  My project was Smart City Traffic Pattern  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](#_Toc139702806)

[2 Introduction](#_Toc139702807) 7

[2.1 About UniConverge Technologies Pvt Ltd](#_Toc139702808) 7

[2.2 About upskill Campus](#_Toc139702809) 11

[2.3 Objective](#_Toc139702810) 13

[2.4 Reference](#_Toc139702811) 13

[3 Problem Statement 1](#_Toc139702813)4

[4 Existing and Proposed solution 1](#_Toc139702814)6

[5 Proposed Design/ Model 1](#_Toc139702815)8

[5.1 High Level Diagram 1](#_Toc139702816)8

[6 Performance Test 1](#_Toc139702819)9

[6.1](#_Toc139702820) [Performance Outcome](#_Toc139702822) 19

[7 My learnings](#_Toc139702823) 22

[8 Future work scope](#_Toc139702824) 24

# Preface

Welcome to the world of Smart City Traffic Pattern, a data science project that delves into the fascinating realm of urban transportation and its profound impact on modern societies. This project aims to explore and analyze the intricate web of traffic patterns in a smart city, using cutting-edge data science techniques to gain valuable insights and propose solutions for a more efficient and sustainable urban mobility. Traffic congestion remains a pressing concern in modern cities worldwide, causing detrimental effects on both the environment and the well-being of citizens. Thus, the need to harness the power of data science to tackle this issue has never been more critical. Our project aims to showcase the immense potential of data-driven approaches in transforming urban transportation systems and creating more sustainable, livable cities.

The Smart City Traffic Pattern data science project is a comprehensive exploration into the realm of urban transportation within a smart city environment. Conducted over a period of six weeks, this project aimed to analyze and optimize traffic patterns to enhance urban mobility, reduce congestion, and create a more sustainable and livable urban landscape.

In this data science project, our objective is to build a predictive model to forecast the number of vehicles at specific junctions based on various parameters, including DateTime, Junction, and ID. The ability to accurately predict traffic flow is crucial for urban planners, transportation authorities, and commuters to optimize travel routes, manage congestion, and improve overall traffic management in smart cities.

**Week 1:** Project Initiation and Data Collection - The project commenced with a detailed planning phase, where the objectives, scope, and data sources were defined. Diverse data streams, including traffic flow data from intelligent infrastructure, GPS tracking, public transportation schedules, weather data, and other urban parameters, were collected to provide a holistic view of traffic patterns within the smart city.

**Week 2:** Data Preprocessing & Exploration - During the second week, the collected data underwent thorough preprocessing to handle missing values, outliers, and formatting inconsistencies. Exploratory data analysis techniques were employed to gain valuable insights into traffic patterns, peak hours, congestion hotspots, and any correlations with external factors.

**Week 3:** Traffic Pattern Identification - Using advanced machine learning algorithms, the team identified recurring traffic patterns and determined key factors contributing to fluctuations in traffic flow. This crucial phase laid the foundation for devising data-driven strategies to alleviate congestion and enhance transportation efficiency.

**Week 4:** Predictive Modeling Building upon the historical traffic data, predictive models were developed to forecast traffic conditions. These models offered commuters and city planners the ability to proactively plan their routes and make informed decisions based on future traffic projections.

**Week 5:** Model Optimization During the fifth week, the team leveraged optimization algorithms to propose smarter traffic management strategies. This included optimizing signal timings, implementing intelligent traffic light control systems, and suggesting alternative transportation modes to reduce congestion and improve overall mobility.

**Week 6:** In the final phase of the project, the documentation part was completed & project report submission has been committed.



1. Data Collection and Description

The dataset used for this project was collected from various sources, including traffic monitoring systems, IoT sensors, and historical traffic records. It comprises several features, with the key ones being:

* DateTime: Timestamps indicating the date and time of vehicle count observations.
* Junction: A categorical variable representing the unique junctions or intersections within the city.
* ID: An identifier for each data point, linking it to specific locations and events.
* Number of Vehicles: The target variable representing the count of vehicles observed at each junction during the given DateTime.

2. Data Preprocessing

Before proceeding with modeling, the data underwent thorough preprocessing to ensure its quality and suitability for analysis. The preprocessing steps included:

* Handling missing values: Missing data points in any feature were imputed using appropriate techniques, ensuring minimal impact on model performance.
* Encoding categorical variables: The categorical feature "Junction" was encoded using techniques like one-hot encoding or label encoding, making it suitable for machine learning algorithms.
* Feature engineering: Additional features, such as day of the week, hour of the day, and seasonal indicators, were created to capture any temporal patterns present in the data.

3. Exploratory Data Analysis (EDA)

EDA was performed to gain insights into the data and identify any trends or patterns that might influence vehicle counts. Visualizations, such as time series plots, heatmaps, and correlation matrices, were used to understand the relationships between different features and the target variable.

4. Model Selection and Training

To predict the number of vehicles accurately, Decision Tree Classification model was preferred. The dataset was split into training and testing sets to evaluate model performance. Cross-validation techniques were employed to optimize hyperparameters and avoid overfitting.

5. Model Evaluation

The models were evaluated using several performance metrics, such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The best-performing model was selected based on its ability to generalize well on the test dataset.

6. Results and Insights

The chosen predictive model demonstrated promising results in accurately forecasting the number of vehicles at various junctions based on DateTime and ID. Key insights from the analysis included:

* Temporal Patterns: The model captured daily and weekly traffic patterns, showing higher vehicle counts during peak hours on weekdays and lower counts during weekends.
* Impact of Junctions: Different junctions exhibited varying traffic patterns, influenced by factors such as proximity to commercial areas, schools, and residential neighborhoods.
* ID-specific Trends: Certain IDs showed unique traffic behavior due to special events, road closures, or ongoing construction activities.

7. Conclusion

The data science project successfully developed a predictive model capable of forecasting the number of vehicles at different junctions based on DateTime and ID. This model holds great potential for aiding urban planners and transportation authorities in optimizing traffic flow, reducing congestion, and enhancing overall urban mobility.

# Introduction

## 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.



1. 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)  
• Rule Engine

1. **Smart Factory Platform (****)**

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.

1.  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.

1. 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.



## 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

<https://www.upskillcampus.com/>

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



## 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.

## 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.

# Problem Statement

=> **SMART CITY TRAFFIC PATTERN**

The objective of this data science project is to develop a predictive model capable of accurately forecasting the volume of vehicles passing through various junctions at different points in time. The prediction of vehicle counts is essential for a multitude of stakeholders, including urban planners, traffic management authorities, and commuters, as it allows them to make informed decisions regarding travel routes, resource allocation, and congestion management.

In the ever-evolving landscape of smart cities, efficient and well-managed transportation systems are crucial to ensure seamless mobility and enhance the quality of urban life. As urbanization continues to rise, traffic congestion becomes a persistent challenge, necessitating innovative solutions to optimize traffic flow and mitigate congestion effectively. In response to this, our data science project aims to predict the number of vehicles at specific junctions within a city based on key parameters, namely DateTime, Junction, and ID.

Data Description -

Our project draws on a comprehensive dataset obtained from diverse sources, including traffic monitoring systems, Internet of Things (IoT) sensors, and historical traffic records. This dataset encompasses the following key features:

1. DateTime: Timestamps indicating the date and time of vehicle count observations. This temporal aspect is crucial for capturing daily, weekly, and seasonal variations in traffic patterns.
2. Junction: A categorical variable representing the unique junctions or intersections within the city. Different junctions may exhibit distinct traffic patterns due to varying nearby landmarks, road structures, and demographic characteristics.
3. ID: An identifier associated with each data point, linking it to specific locations and events. This allows us to account for any unique circumstances or occurrences that might influence vehicle counts.
4. Number of Vehicles: The target variable representing the count of vehicles observed at each junction during the given DateTime. This is the value we aim to predict accurately using machine learning algorithms.

**Key objectives of the Smart City Traffic Pattern project:**

1. Data Analysis and Exploration: We will begin by gathering and processing large-scale datasets related to traffic patterns in a smart city. Exploratory data analysis will be employed to gain a comprehensive understanding of the challenges and opportunities presented by urban traffic.
2. Traffic Pattern Identification: Leveraging machine learning algorithms, we will identify recurrent traffic patterns, peak hours, congestion hotspots, and factors contributing to traffic fluctuations. These insights will form the foundation for devising data-driven strategies to alleviate congestion.
3. Predictive Modeling: Building upon historical data, we will develop predictive models to forecast traffic conditions, allowing commuters and city planners to proactively adapt their routes and plans.
4. Optimization and Recommendation: By harnessing the power of optimization algorithms, we aim to propose smarter traffic management strategies, including signal timing optimization, intelligent traffic light control, and recommendations for alternative transportation modes.
5. Sustainability and Urban Planning: This project will also explore the intersection of smart city traffic patterns with sustainability and urban planning, emphasizing the importance of eco-friendly transportation and infrastructure development.

# Existing and Proposed solution

Currently, the existing solution for predicting the number of vehicles based on DateTime, Junction, and ID may rely on traditional statistical methods or basic regression models. This approach might involve simple time-series analysis or linear regression techniques to estimate vehicle counts at specific junctions. However, these methods often lack the sophistication to capture complex traffic patterns and may struggle to account for various temporal and spatial factors that influence traffic flow.

The shortcomings of the existing solution include:

1. Limited Accuracy: Traditional methods may not be able to capture the non-linear relationships between vehicle counts and the input parameters adequately, resulting in limited predictive accuracy.
2. Inability to Handle Complex Patterns: Traffic patterns in modern cities are dynamic and influenced by multiple factors. The existing solution may not effectively handle complex patterns, such as peak hours, seasonal variations, and the impact of special events or road constructions.
3. Scalability Issues: As urban areas grow, the volume of traffic data increases significantly, making it challenging for the existing solution to handle large-scale datasets efficiently.

Proposed Solution:

To overcome the limitations of the existing solution and achieve more accurate and robust predictions, our proposed solution leverages advanced data science techniques and machine learning algorithms. The key components of our proposed solution are as follows:

1. Data Preprocessing and Feature Engineering: We will perform comprehensive data preprocessing, including handling missing values, encoding categorical variables, and creating additional features to capture relevant temporal and spatial patterns. Feature engineering will involve extracting day of the week, hour of the day, and seasonal indicators from the DateTime feature.
2. Exploratory Data Analysis (EDA): Through EDA, we will gain insights into the data, identify trends, and explore correlations between vehicle counts and the input parameters. This step will help us understand the data's characteristics and inform our modeling decisions.
3. Machine Learning Models: Our proposed solution will employ a range of machine learning classification algorithm such as Decision Tree Classification algorithm.
4. Model Evaluation and Hyperparameter Tuning: We will rigorously evaluate the performance of each model using appropriate metrics like MSE, RMSE, MAE, and R-squared. Hyperparameter tuning techniques, such as grid search or Bayesian optimization, will be applied to optimize model performance.
5. Real-time Data Integration: For real-time predictions, the proposed solution will integrate with live data feeds from traffic sensors and IoT devices, enabling dynamic and up-to-date forecasts.
6. Visualization and Interpretability: The solution will offer interactive data visualizations and model interpretability tools to provide stakeholders with a clear understanding of the factors influencing vehicle counts.

## Code submission (Github link)

<https://github.com/krishnakant-verma/UpSkill-Campus>

## Report submission (Github link) :

<https://github.com/krishnakant-verma/UpSkill-Campus>

# Proposed Design/ Model

Our proposed design for predicting the number of vehicles at specific junctions based on DateTime, Junction, and ID involves a combination of time series analysis, feature engineering, and ensemble machine learning models. The key steps in the proposed design are as follows:

## High Level Diagram

Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

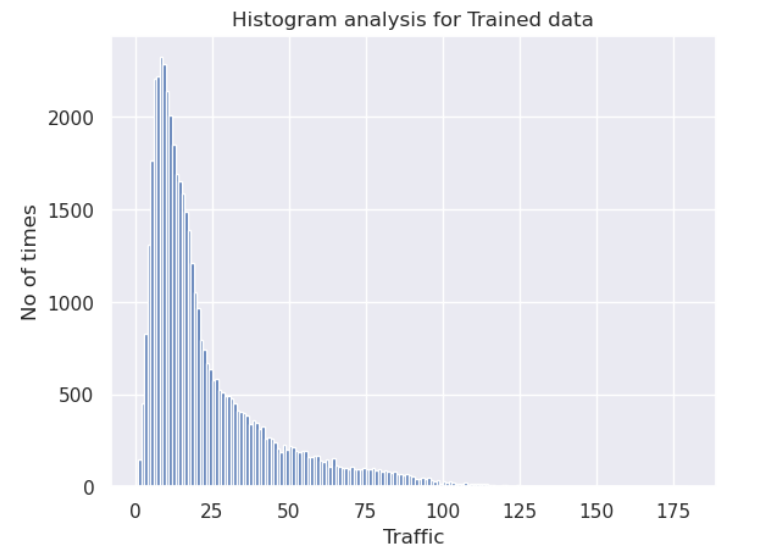
# Performance Test

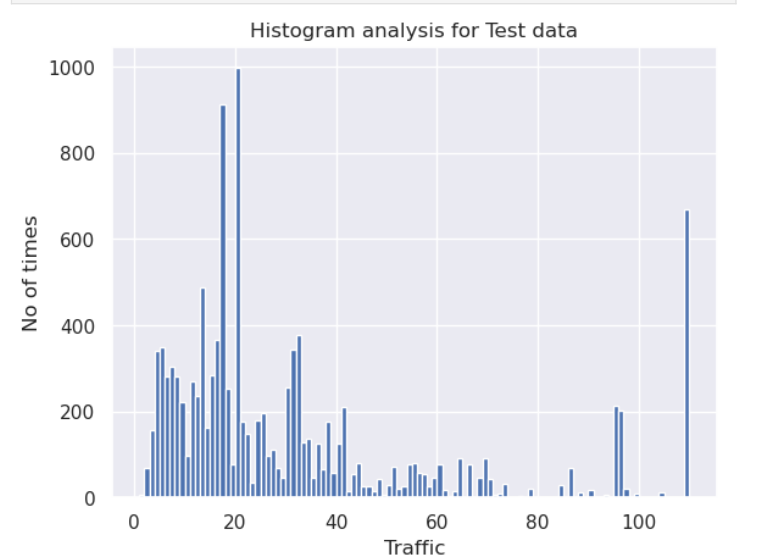
Measure and validate the response time of the predictive model for different sample sizes and junctions.

* Assess the model's scalability by gradually increasing the dataset size and monitoring response times and resource utilization.
* Verify that the model can handle a high throughput of prediction requests without compromising accuracy.

## Performance Outcome



****

****

# My learnings

Learnings that I have gained from an internship project to **predict the number of vehicles** based on parameters like DateTime, Junction, & vehicle ID:

1. Data Preprocessing Challenges: Data preprocessing is a crucial step in any data science project, and this project is no exception. Dealing with missing data, encoding categorical variables, and feature engineering to capture relevant temporal patterns can be challenging and time-consuming.
2. Understanding Temporal Patterns: Time series analysis provides valuable insights into the temporal patterns and seasonality of vehicle counts. Learning to decompose seasonality and identify lag values can significantly impact the accuracy of predictions.
3. Model Selection and Evaluation: Selecting the right machine learning models for this task is essential. One might experiment with various regression models, ensemble methods, and deep learning techniques like LSTM to find the best-performing model. Evaluating models with appropriate metrics helps in understanding their strengths and weaknesses.
4. Feature Importance: Understanding the importance of different features, such as DateTime components and junction locations, can provide valuable insights into the factors influencing vehicle counts and traffic patterns.
5. Real-time Integration: Integrating the predictive model with live data feeds and making real-time predictions requires careful implementation and validation. Ensuring that the model remains accurate and responsive to dynamic traffic changes is critical.
6. Performance Optimization: Dealing with large-scale datasets and ensuring the model's scalability and efficiency is an important aspect of this project. Learning to optimize model training and prediction times can be challenging, but it enhances the overall usability of the solution.
7. Visualization and Communication: Effective data visualization techniques play a significant role in communicating the results and insights of the predictive model to stakeholders. Presenting the predictions in a clear and understandable format is essential for decision-making.
8. Handling Edge Cases: This project may encounter edge cases, such as dealing with holidays, special events, or unusual traffic patterns. Learning to handle these scenarios and their impact on predictions is crucial for real-world applications.

# Future work scope

The Data Science project to predict the number of vehicles based on DateTime, Junction, and ID opens up several exciting future scope and possibilities for further improvements and enhancements. Some potential future scope for the project includes:

1. Integration with Real-Time Traffic Data: The project can be extended to integrate with live traffic data from various sources, such as GPS-enabled vehicles, smart traffic signals, and traffic cameras. This integration would enable the model to make real-time predictions, offering up-to-date traffic information to commuters and traffic management authorities.

2. Incorporating Weather Data: Including weather data, such as temperature, precipitation, and weather conditions, as an additional feature can help improve the accuracy of the vehicle count predictions. Weather plays a significant role in influencing traffic patterns, and accounting for it can enhance the model's robustness.

3. Multimodal Transportation Prediction: Expanding the project to predict not only vehicle counts but also the number of pedestrians, bicycles, and public transport users at junctions would provide a comprehensive view of the overall urban mobility ecosystem. This information is valuable for creating holistic transportation plans.

4. Predicting Traffic Congestion: Leveraging the predictive model to forecast traffic congestion at specific junctions and routes can help commuters plan their travel more efficiently and avoid congested areas.

5. Urban Planning and Infrastructure Development: Integrating the model's predictions into urban planning and infrastructure development can aid in designing road networks, traffic signal optimization, and determining the need for additional transportation facilities.

6. Dynamic Traffic Signal Control:\*\* Implementing dynamic traffic signal control systems based on real-time predictions can optimize signal timings and improve traffic flow, reducing congestion and travel time for commuters.

7. Multicity Predictive Model: Expanding the project to include data from multiple cities can lead to the development of a predictive model that is applicable to various urban environments. Such a model could offer insights into traffic patterns across different cities and facilitate city-to-city comparisons.

8. Explainable AI and Interpretability: Introducing explainable AI techniques to the model can provide insights into the factors contributing to vehicle counts and traffic patterns. This interpretability is crucial for building trust in the model's predictions and understanding its decision-making process.

9. Continuous Model Improvement: As more data becomes available over time, the predictive model can be continuously updated and retrained to adapt to changing traffic patterns and urban dynamics.

10. Optimizing for Environmental Impact: Incorporating sustainability considerations, such as promoting eco-friendly transportation modes (e.g., public transport, cycling) in the predictions, can help cities reduce their carbon footprint and improve air quality.

11. Predictive Analytics for Traffic Incidents: Extending the project to predict traffic incidents, accidents, or road closures can assist in proactive incident management and improve emergency response systems.

12. Predicting Demand for Electric Vehicles: Integrating data on electric vehicle usage and charging infrastructure can enable predictions of electric vehicle demand at different junctions, helping cities plan for future EV adoption.