Industrial Internship Report on

Smart Traffic Prediction

Prepared by

Jeet Patel

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 on the topic of Smart Traffic Prediction. 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

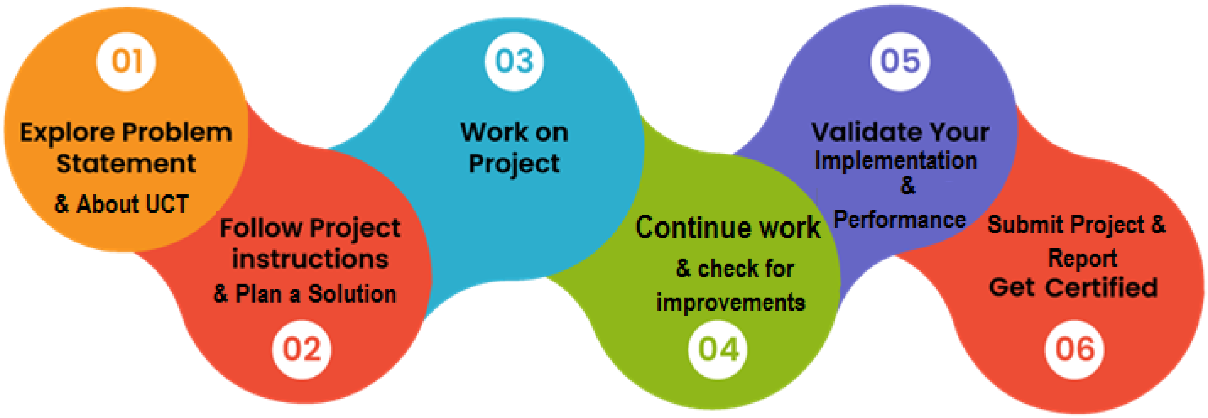
1 Preface

Throughout the duration of this project, a meticulously crafted Profile Report of the dataset was diligently generated, harnessing the powerful capabilities of Pandas Profiling. The central focus of this endeavor revolved around the complex task of predicting intricate traffic patterns within the context of a Smart City. This considerable challenge was met head-on by skillfully employing an array of data visualization techniques, with a particular emphasis on the judicious use of Time-Series Plots. These plots provided a wealth of invaluable insights into the temporal trends and intricate patterns embedded within the data.

The project unfurled seamlessly across a series of pivotal phases, each carrying its own weight of significance. The outset involved foundational steps of Data Loading and Preprocessing, which provided the bedrock for subsequent stages. Following this, rigorous Feature Engineering techniques were applied, aimed at extracting the most salient facets of the data to empower the ensuing forecasting model. Subsequent to this, a visualization phase ensued, featuring a diverse array of graphical representations that further illuminated the underlying dynamics of the data. This phase not only facilitated an intuitive grasp of the traffic patterns but also contributed to a deeper understanding of the data's inherent complexities.

Feature Transformation methods were adroitly employed to distill the features into a more representative form, laying the groundwork for the subsequent Data Preparation phase. A well-considered Train-Test Split protocol was meticulously executed to ensure the resilience and robustness of the subsequent model. The predictive mantle was then assumed by the esteemed Random Forest Regressor, chosen for its aptitude in handling intricate relationships present within the data. Following the deployment of the model, a comprehensive Model Evaluation strategy was implemented, encompassing a wide spectrum of metrics to holistically assess its efficacy and performance.

In summary, this project provided a holistic encapsulation of the multifaceted nature of forecasting Smart City Traffic Patterns. It traversed through a comprehensive array of indispensable phases and techniques, culminating in the development of a robust, insightful model. This model stands as a testament to the harmonious fusion of domain expertise and analytical acumen, showcasing the successful navigation of complexities inherent to real-world data analysis and prediction.



Engaging in this project has been a profoundly enriching learning experience, one that has illuminated the intricate pathways of data analysis and forecasting within the context of a Smart City. As I delved into the project's various phases, the process of meticulously generating a Profile Report using Pandas Profiling highlighted the significance of comprehensive data exploration as a foundation for insightful insights. The focal point of predicting complex traffic patterns within a Smart City context not only underscored the challenges but also showcased the potential for leveraging data to drive informed decisions.

Through the guidance of IoT Academy blogs and resources, I was able to grasp the nuances of the Internet of Things (IoT) and its application in urban environments. These resources served as beacons of knowledge, shedding light on the intricate interplay between data and the urban landscape, further enriching my understanding of Smart City dynamics. The deployment of various data visualization techniques, particularly Time-Series Plots, not only provided a practical avenue for conveying trends but also deepened my appreciation for visual representation as a powerful means of conveying complex information.

Each phase, from Data Loading and Preprocessing to Feature Engineering and beyond, showcased the intricacies involved in transforming raw data into actionable insights. The insights gleaned from the visualization phase, inspired by IoT Academy's insights, augmented my ability to comprehend the underlying data dynamics and make informed decisions. The employment of Feature Transformation techniques, as illuminated by the academy's resources, was instrumental in refining the data for modeling, emphasizing the art of distilling complexity into a more comprehensible form.

The pivotal role of the Train-Test Split protocol, guided by the principles outlined in the academy materials, underscored the significance of robust model evaluation and preparation. The choice of the Random Forest Regressor, informed by insights from IoT Academy's blogs, showcased how domain expertise can influence model selection for intricate datasets. The implementation of a comprehensive Model Evaluation strategy, fortified by the academy's guidance, illuminated the multifaceted nature of assessing model performance.

In summation, this project has been an immersive journey into the world of data-driven insights and predictive analytics within the realm of Smart Cities. The convergence of project experiences with the insights from IoT Academy's blogs has magnified my appreciation for the fusion of domain knowledge and analytical techniques. It has not only deepened my understanding of data analysis but also illuminated the transformative potential of IoT within urban landscapes, offering a well-rounded learning experience that bridges theory and practical application.

Thanks to Kaushlendra Singh Sisodia Sir’s, Apurv Sir’s and Tushar Mehra Sir’s, who have helped me a lot while implementing the above project and their valuable guidance as and when needed.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

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

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

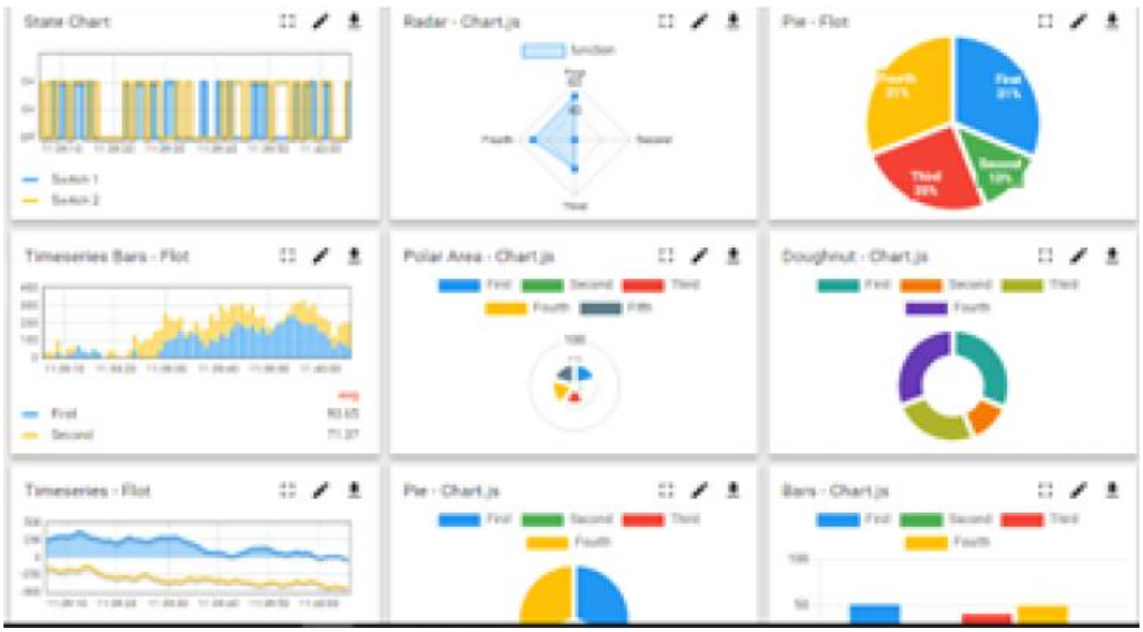
i. UCT IoT Platform ( )

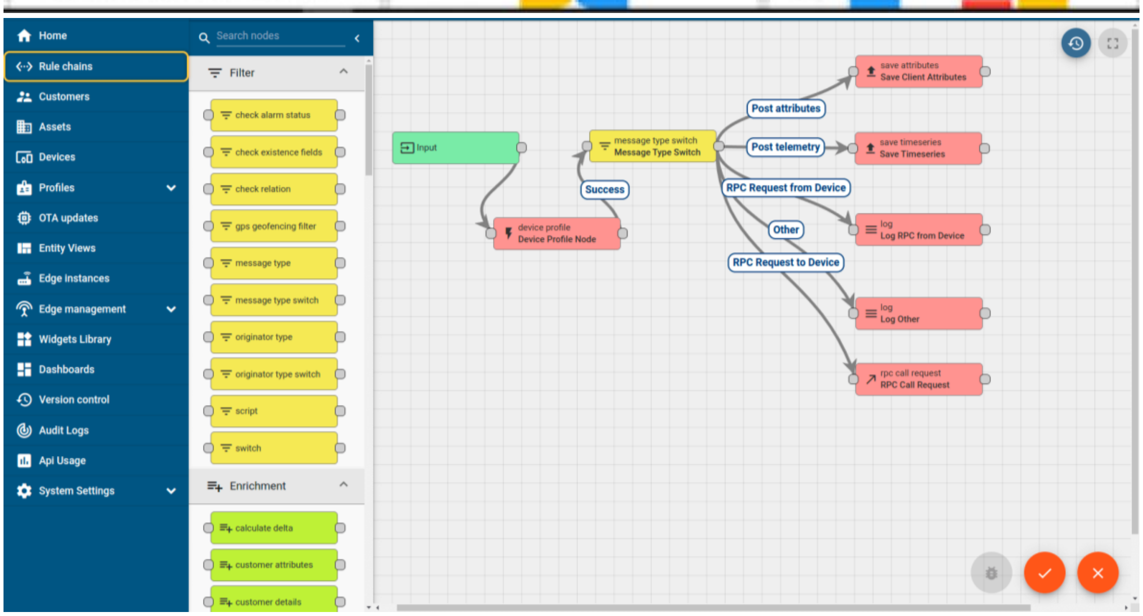
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

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

It has features to

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

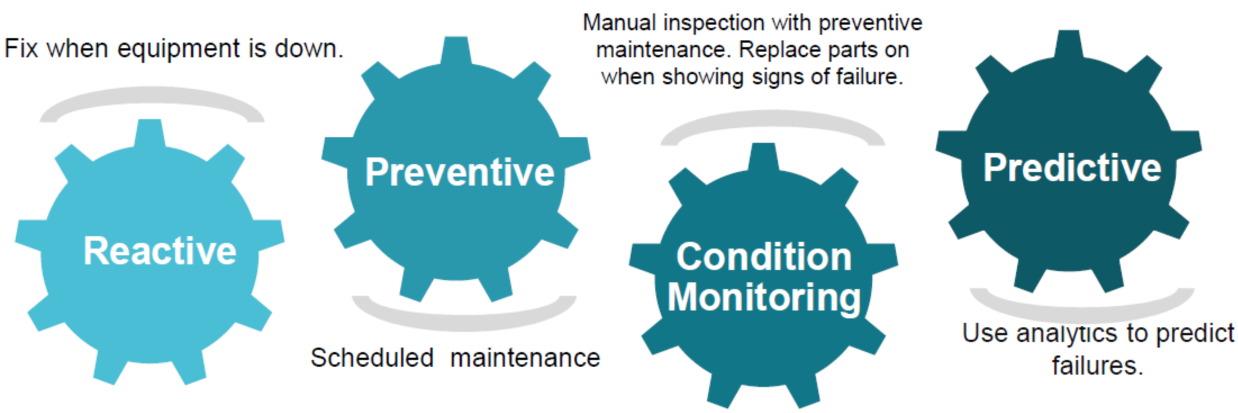
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

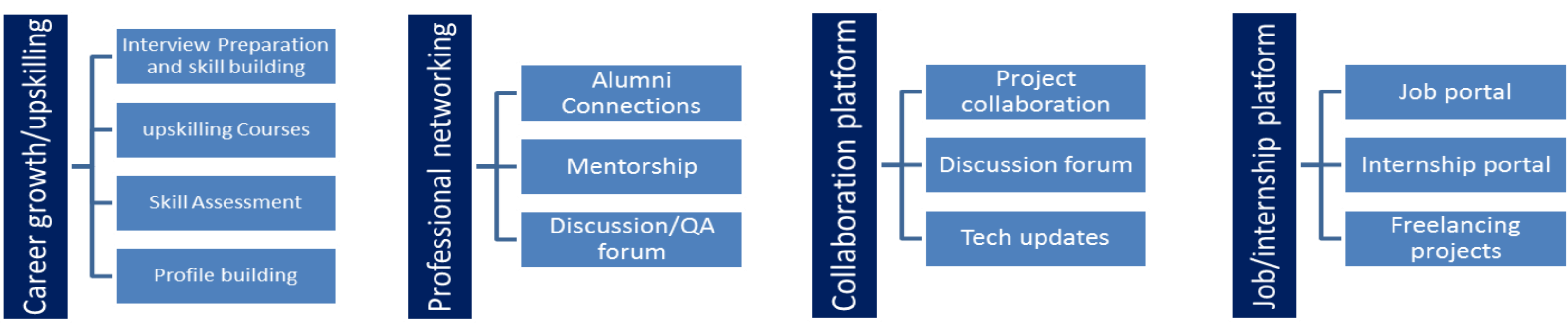
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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]

[2]

[3]

2.6 Glossary

Terms	Acronym

3 Problem Statement

Forecasting of Smart City Traffic Patterns

We are working with the government to transform various cities into a smart city. The vision is to convert it into a digital and intelligent city to improve the efficiency of services for the citizens. One of the problems faced by the government is traffic. You are a data scientist working to manage the traffic of the city better and to provide input on infrastructure planning for the future.

The government wants to implement a robust traffic system for the city by being prepared for traffic peaks. They want to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays, as well as on various other occasions during the year, differ from normal working days. This is important to take into account for your forecasting.

Dataset Link -
https://drive.google.com/file/d/1y61cDyuO9Zrp1fSchWcAmCxk0B6SMx7X/view?usp=sharing

4 Existing and Proposed solution

Existing Solution –

The existing solution was founded upon the utilization of the Decision Tree Classifier, accompanied by a limited incorporation of Data Visualizations. In contrast, the approach I undertook aimed to enhance the analytical depth of the project by adopting a multifaceted strategy. In line with this, a comprehensive Pandas Profiling report was meticulously generated, offering a comprehensive overview of the dataset's characteristics. This approach, inspired by the desire for a holistic understanding of the data, laid a robust foundation for subsequent analytical endeavors.

Building upon this groundwork, I opted for the Random Forest Regressor, a more intricate algorithm, to undertake the predictive task. This choice was influenced by the insights gained from prior research and the recognition of the algorithm's ability to handle complex relationships present within the data. Additionally, I embraced an expansive array of Data Visualizations to unravel intricate patterns and trends embedded within the dataset. By deploying these visual aids, I endeavored to provide a more intuitive and comprehensive interpretation of the data's nuances, facilitating a deeper grasp of its underlying dynamics.

In summary, my approach diverged from the established solution by integrating a Pandas Profiling report to comprehensively profile the dataset and by employing the Random Forest Regressor for predictive modeling. Moreover, the extensive utilization of Data Visualizations underscored my commitment to unraveling the intricacies of the data. This methodology not only enhanced the analytical robustness of the project but also paved the way for a more comprehensive understanding of the dataset's intricacies and the potential to extract valuable insights.

Proposed Solution –

1. Data Loading and Preprocessing: The code starts by importing required libraries, including data manipulation (NumPy, Pandas), visualization (Matplotlib, Seaborn), time handling (datetime), and machine learning tools (scikit-learn). It reads two datasets, dfTrain and dfTest, containing traffic-related data. The datetime columns are converted to datetime objects for both train and test datasets using pd.to_datetime.
2. Feature Engineering: New features are created from the datetime information for both train and test datasets. These features include 'Weekday', 'Year', 'Month', 'Day', 'Time', 'Week', and 'Quarter'. Time-series plots are generated using Seaborn to visualize the relationship between datetime and vehicle count for different junctions.
3. Visualization: A time-series plot illustrates the variation of vehicle counts over time, with different colors representing different junctions. Another visualization shows the distribution of vehicle counts across different years for each junction.

4. Feature Transformation: A function named datetounix1 is defined to convert datetime objects to Unix timestamps (seconds since epoch). This function is applied to both train and test dataframes.
5. Data Preparation: Data is prepared for modeling by storing predictor features in the 'X' array and the target variable 'Vehicles' in the 'y' array. One-hot encoding is applied using Pandas' get_dummies function to convert categorical variables into binary columns.
6. Train-Test Split: The dataset is split into training and testing sets using train_test_split, with a test size of 33% and a specified random seed.
7. Random Forest Regressor: A Random Forest Regressor model is created with 100 estimators (trees) and a random seed of 42. The model is trained on the training data using the fit method.
8. Model Evaluation: Predictions are made on the test set using the trained model. Evaluation metrics are calculated, including Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2) score. The evaluation metrics are printed to assess the model's performance.

4.1 Code submission (Github link) –

Link – https://github.com/jjeetpatell/UpSkillCampus/blob/main/SmartTrafficPrediction.ipynb

4.2 Report submission (Github link) –

Link –
https://github.com/jjeetpatell/UpSkillCampus/blob/main/SmartTrafficPrediction_Jeet_USC_UCT.pdf

5 Proposed Design/ Model

Various Phases include –

1. Data Loading and Preprocessing
2. Feature Engineering
3. Visualization
4. Feature Transformation
5. Data Preparation
6. Train Test Split
7. Random Forest Regressor
8. Model Evaluation

6 Performance Test

The performance assessment and model evaluation phase was instrumental in gauging the efficacy of the developed model. Key evaluation metrics were computed, shedding light on its predictive accuracy and overall performance. The mean squared error (MSE), a measure of the average squared differences between actual and predicted values, provided insights into the model's precision in capturing the data's variance. The mean absolute error (MAE), representing the average absolute differences between actual and predicted values, quantified the model's absolute predictive accuracy. Moreover, the R-squared (R2) score, a valuable coefficient indicating the proportion of variance captured by the model, offered a comprehensive view of its explanatory power. By meticulously calculating and subsequently printing these evaluation metrics, the assessment process not only quantified the model's performance but also empowered a clear and concise interpretation of its predictive capabilities.

7 My learnings

Engaging in this project has been an invaluable journey of continuous learning and growth. From its inception to the final stages, every step has enriched my understanding of data analysis, predictive modeling, and their practical applications. Working on profiling the dataset through Pandas Profiling has underscored the significance of comprehensive data exploration as a precursor to meaningful insights. The adoption of the Random Forest Regressor as the predictive model has provided me with a deeper grasp of algorithm selection and its impact on complex data relationships.

Furthermore, the emphasis on Data Visualizations has been a revelation in enhancing data interpretation. Through the visualization techniques employed, I've come to appreciate their role in conveying intricate patterns and trends to both technical and non-technical stakeholders. The process of evaluating the model using metrics like mean squared error, mean absolute error, and R2 score has elucidated the importance of quantifying model performance in a clear and interpretable manner.

Throughout this endeavor, I've also recognized the transformative power of domain knowledge and its integration with analytical techniques. The insights gained from IoT Academy's resources have illuminated the real-world relevance of Smart City dynamics and how IoT technologies are shaping urban environments. This project has solidified my belief in the iterative nature of learning, where each challenge and triumph contributes to a deeper understanding of data science and its potential to drive informed decision-making.

In retrospect, this project has not only honed my technical skills but has also fostered a deeper appreciation for the holistic and interdisciplinary nature of data-driven problem-solving. As I look ahead, I am excited to apply these learnings to future projects and endeavors, armed with a heightened sense of curiosity and a stronger foundation in data analysis and predictive modeling.

8 Future work scope

The project's completion marks a significant milestone, yet it also opens the door to potential future work that could further enrich the insights and capabilities of the developed model. Time constraints during this phase restricted the exploration of certain avenues, but these areas present promising directions for future investigation.

The exploration of ensemble methods, beyond the Random Forest Regressor, could be pursued to ascertain whether combining the strengths of multiple algorithms leads to improved predictions. The integration of deep learning approaches, like neural networks, might also offer novel insights into the data's complexities, although this was beyond the scope of the current project.

Another avenue worth exploring is the application of advanced optimization techniques to fine-tune hyperparameters. This process could potentially lead to enhanced model performance and robustness. Finally, a more comprehensive analysis of temporal patterns using advanced time-series techniques, such as ARIMA or LSTM networks, could provide a deeper understanding of cyclical trends and further improve forecasting accuracy.

In conclusion, while this project has achieved its objectives within the given timeframe, its culmination marks the beginning of a range of potential future enhancements. By delving into unexplored directions such as external data integration, advanced algorithmic approaches, and hyperparameter optimization, the model's capabilities could be extended to new horizons, delivering even more accurate and insightful predictions for Smart City traffic patterns.