Industrial Internship Report on

”Forecasting of Smart city traffic patterns”

Prepared by

Anandita Sharma

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 traffic prediction and management. Among the most difficult problems affecting cities is designing and developing an effective transportation system. As the population of the city grows, so do the city's governmental and non-governmental transport networks, and even minor faults and accidents can have a negative impact on a system that is already stretched to its limits. Traffic jam and delays on transportation systems can cost a city loss in terms economy and have a negative impact on inhabitants' standard of living as they spend a lot of time traveling from one location to another. The goal of a smart city is to deliver good services to its citizens by using modern technology and data analytics on data collected by sensors. Technology driven infrastructure, community development programs, smart transportation systems, the use of technology to minimize crimes and burglary, giving safety to residents, and other factors all contribute to a city's smartness. 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

Week 1: I carefully reviewed some relevant publications throughout the week, I discovered that the existing literature does not provide any work that uses BAG, GLM, BGLM, MARS, and KNN techniques to solve the problem of traffic congestion prediction in smart city. Furthermore, the prediction accuracy of some of the existing work is relatively low while in some cases the authors did not use performance metrics to evaluate the performance of their proposed system. Also, some authors did not compare their work with high performing machine learning or deep learning models. Considering these shortcomings, the novelty of this work centers on the use of machine learning for traffic pattern prediction.

The summary of my contribution is as follows: (i) This paper proposed five (5) machine learning techniques for traffic pattern prediction in smart city. (ii) (ii) The Bagging (BAG), K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM) models for solving the problem of traffic congestion prediction in smart city was presented in this paper.

Week 2: In the 1st week we had explored the problem statement and came across various machine learning algorithms that can be implemented on the data collected to predict traffic.
This week, as we moved forward in the project, we realized that the data collected in its raw form would be useless. The machine learning algorithm need to be implemented on data that has been processed, segregated, and classified. Then the dataset needs to be separated into training sets and validation sets. The training sets will then be used to create the learned model by applying appropriate machine learning model.

Week 3: Throughout the course of last week, we worked on how to plan a solution and how to prepare the data for creating the training sets and validation set for the machine learning algorithms to be implemented on and to create a learned model from it. This week we further explored the problem statement and used the machine learning algorithms we had established in the first week, understood their working and used them with the real time data that was provided.

Week 4: Last week we had worked on various machine learning algorithms, understood their working, and used them with the real time data that was provided. This week we worked on the methodology of the proposed traffic prediction system.

Proposed traffic prediction system:
The proposed Smart city traffic pattern was borne out of the problem associated with urban city. Smart cities use big data for decision-making and a revolution in the Internet of a Thing (IoT) allows such data to be collection and transmission of such data. Transportation is a fundamental element influencing metropolitan areas and a crucial use case for smart cities. Smart transportation technologies, such as GPS Bus, traffic cameras, taxis, and geospatial technology, which is frequently employed in a traffic control centre to monitor and coordinate a wide network of sensors, are discovering creative methods to reduce traffic congestion and urban mobility. Smart city traffic pattern incorporates all the big data obtained from GPS Bus, traffic camera, GPS Taxi, and geospatial technology. The traffic pattern system consists of data visualization, data cleaning to remove unwanted and other irrelevant values.
Week 5: Last week we had worked on identifying the trends and pattern in data and plotting their relationships on a graph to understand them better. This week we evaluated the performance and

implementation of the project throughout the week which can be represented in the form of graph to better understand them. Last week we identified the traffic pattern in data through various constraints such as days, various durations during the day (Morning, Noon, Afternoon, Evening) and applied the various machine learning algorithms to obtain a learned model. This week we further elaborate on the that for implementation and performance.

Traffic congestion varies within the day event may cause surges in traffic at unexpected times. In Figs. 8, 9 and 10, we considered Morning, Afternoon, Evening, and Noon. The study revealed that in 2015 there is traffic surge in the Morning followed by Afternoon as shown in Fig. 8. In 2017 the traffic congestion increased exponentially which results in traffic surge in Afternoon than in the Noon as shown in Fig. 9. Depicted in Fig. 10 is the vehicular movement in 2017 which shows that Afternoon has the highest number of vehicles plying the road followed by Noon.

Need of relevant internship in career development: Just having a good degree is no longer enough to secure that all-important graduate job offer in today's world. Pertinent work experience is now just as valuable as your degree and exam results when it comes to building a successful career. As a result, internships have become an essential way to help candidates make themselves stand out.

Internships provide exposure to the real world

Unfortunately, in today's job market, passing exams with high scores and getting a degree does not offer the much-needed work experience, you will need to succeed in a workspace.

By partaking in an internship, you will be able to gain real-life exposure, grow your knowledge and determine if you are in the right career field. Internships not only provide you with the first-hand experience in the real working world but also enable you to understand the career trajectory for your desired job title. You can learn how to apply the knowledge you have acquired during an internship to your future workplaces.

Internships allow you to learn more about yourself

You may start as an intern in a specific field. But the more you explore it, the more you will discover about it. Every industry has its pros and cons. Working in a real-world environment will help you understand the depths of your field and will help you determine whether it is a career you wish to pursue in the future.

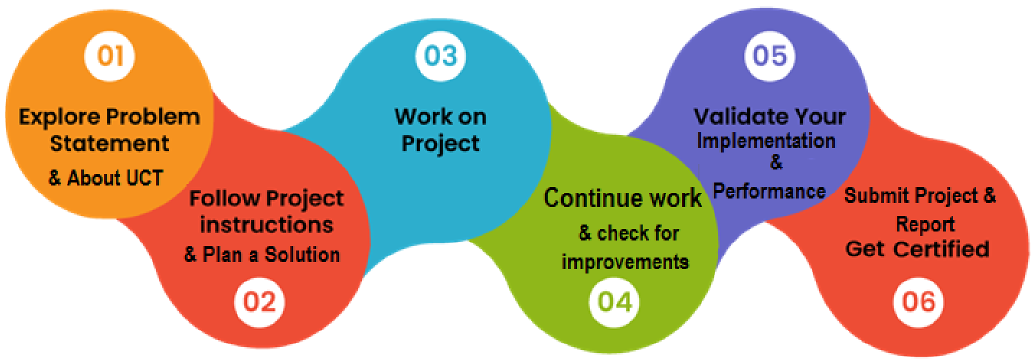
An internship will help you learn about your capabilities and ultimately encourage you to have a greater understanding of your strengths and weaknesses.

Problem Statement: 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.

USC/UCT gave me the opportunity to work on a 6-week internship where I got to work on the project of Smart City Traffic forecasting under the guidance of mentors and experts. It gave me an excellent opportunity to get to know the real-world application of machine learning algorithms and to work with real time data and the importance of segregating, classifying, and organizing data. Each project that we received had scope varying from current scenario to industrial demand, thus gave me a lot of exposure towards the industrial and IT world

The program was planned according to the given below diagram:



I learned a lot while working on this project, learned various new machine learning algorithms and went through step-by-step procedure of creating a learned model from the training set in combination with the machine learning algorithm. Data analysis and data science were a huge part of this project. This project helped me gain a lot of exposure towards big data and the importance of understanding the trends and patterns in data. Overall, it was an excellent learning experience which is guaranteed to help me in the future. Thankful to all the mentors who guided us throughout the internship by helping us with our doubts and errors, and provided us with helpful guidance.

To all my peers, as we prepare to say goodbye, I would like to thank you for all your hard work. I hope you found your time with us enjoyable and educational. You can count on us for a positive reference as you go forward in life. Best of luck to you, and thanks again.

To my juniors , study hard and best of luck for your internship

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.

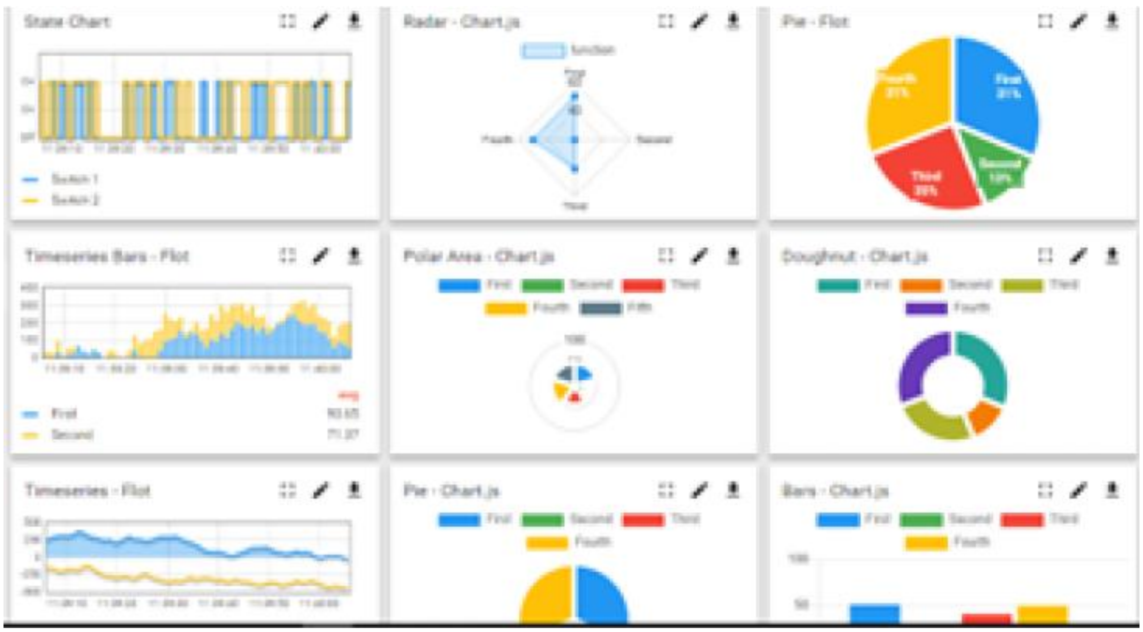
For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

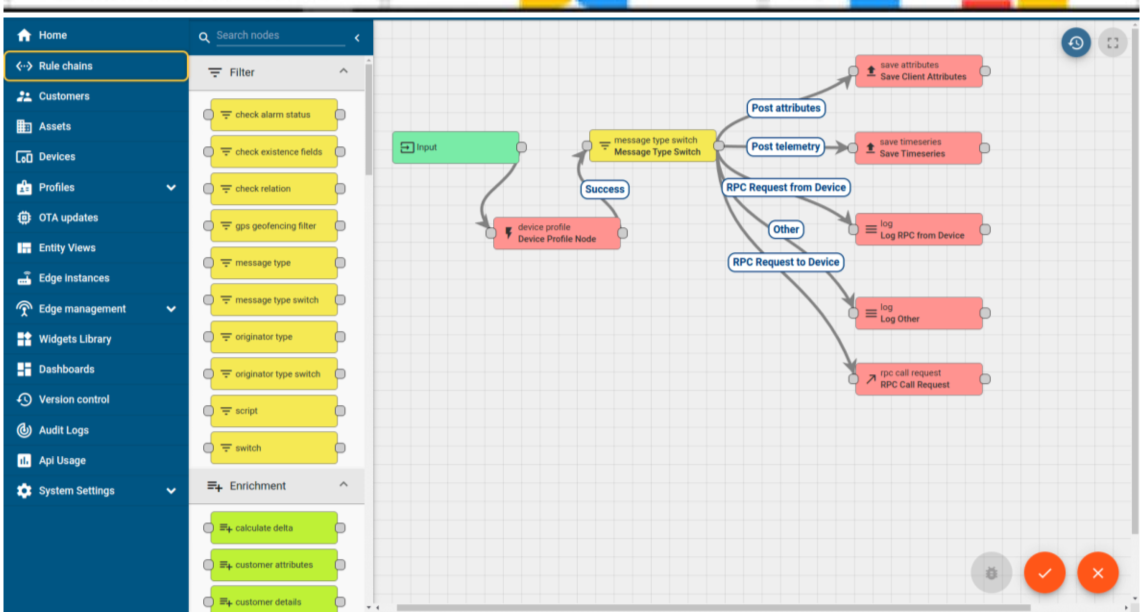
i. UCT IoT Platform (uct Insight)

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





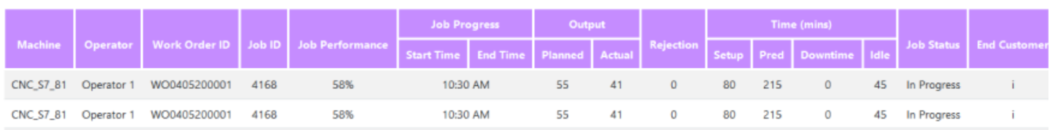
ii. Smart Factory Platform (FACTORY WATCH)

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.



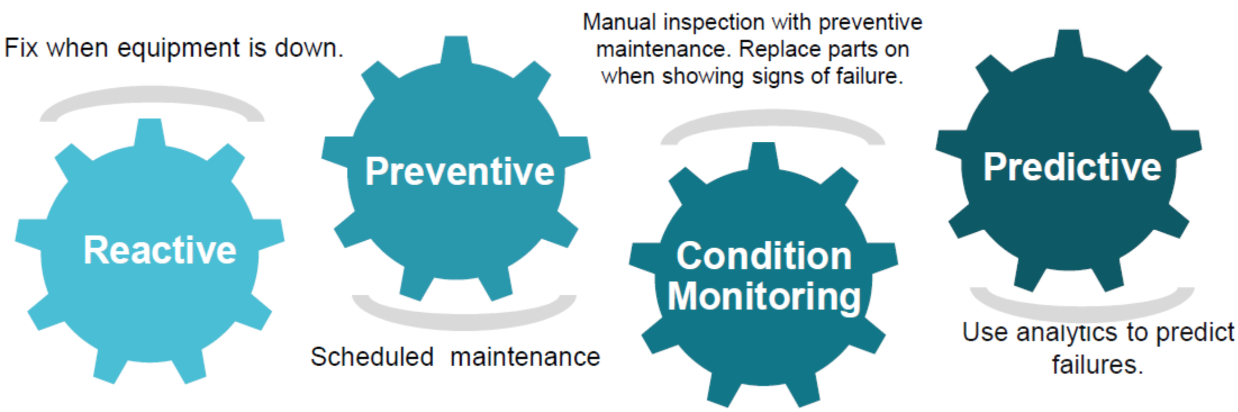
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

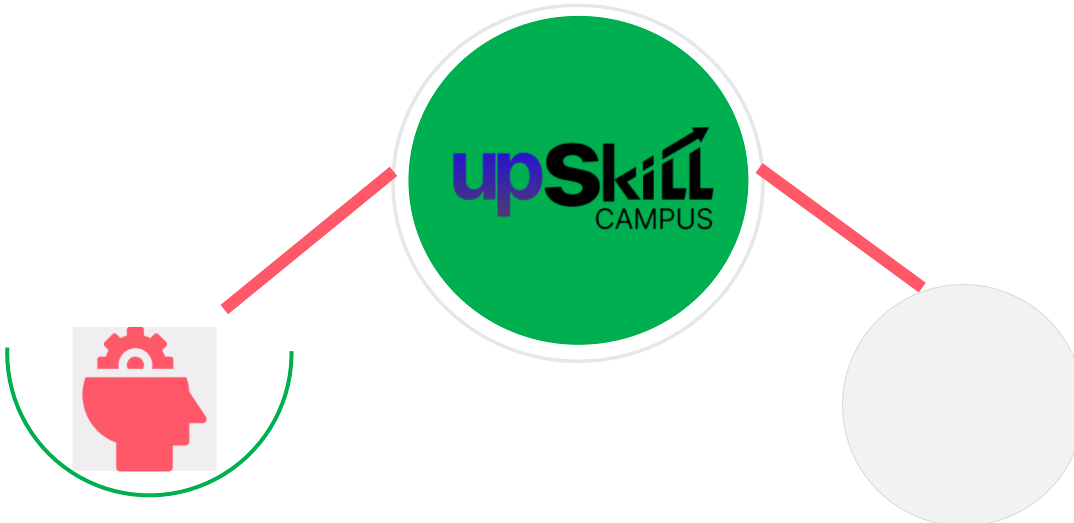
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

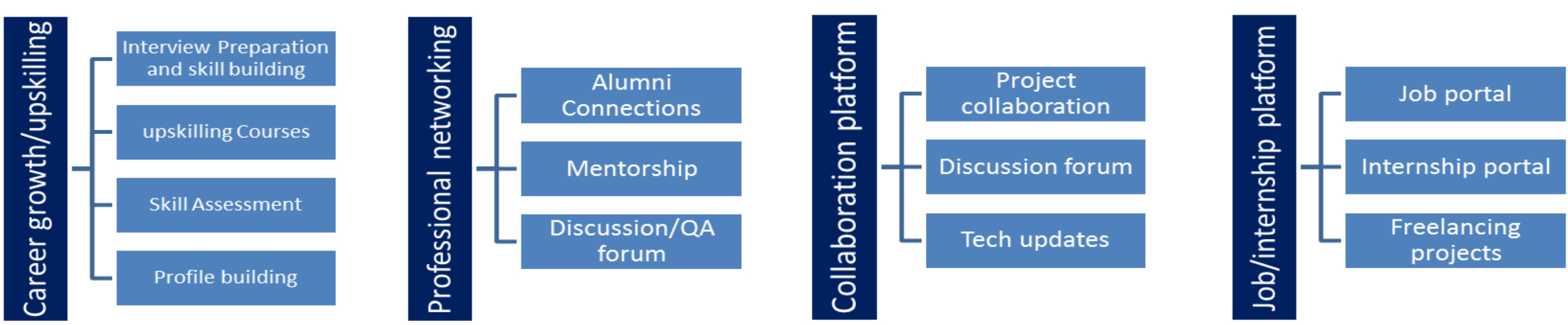
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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 References

[1]Adeniyi, D. A., Wei, Z., & Yongquan, Y. (2016). Automated web usage data mining and recommendation system using K-nearest neighbor (KNN) classification method. Applied Computing and Informatics, 12(1), 90–108.

An, C., & Wu, C. (2020). Traffic big data assisted V2X communications toward smart transportation. Wireless Networks, 26(3), 1601–1610.

Bhattacharya, S., Somayaji, S. R. K., Gadekallu, T. R., Alazab, M., Maddikunta, P. K. R. (2020) A review on deep learning for future smart cities. Internet Technology Letters, e187.

[2] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00542-7

[3] https://www.javatpoint.com/traffic-prediction-using-machine-learning

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.

Explanation:

Traffic affects every citizen's life in many ways by how long it takes for him or her to travel from home to office, the air condition he or she inhales, the strain generated by traffic jams, sleep, and workouts induced by time spent in traffic. Since motorists cannot see the entire traffic system, the urban traffic system must be anticipated in order to sensitize residents about their mobility choices and the subsequent impact on the environment, as well as to implement smart transport system.

Among the most difficult problems affecting cities is designing and developing an effective transportation system. As the population of the city grows, so do the city's governmental and non-governmental transport networks, and even minor faults and accidents can have a negative impact on a system that is already stretched to its limits. Traffic jam and delays on transportation systems can cost a city loss in terms economy and have a negative impact on inhabitants' standard of living as they spend a lot of time traveling from one location to another. Traffic gridlock has several negative consequences, including lost time, environmental degradation, and safety hazards, as well as a negative influence on economic growth and a deterioration of individuals' relationship with their municipal authorities. The idea rt cities became a reality thanks to advancement in Information Technology (IT). The use of Advanced Traveler (ATS) and Advanced Traffic Management Systems (ATMS) to effectively managed, control and manage traffic flows is a key element of the smart or intelligent cities of the future. The ATMS/ATS aims to improve the overall performance of the traffic system, for example, to reduce emissions, noise, and travel times. Various types of transport models are routinely employed in the estimation of traffic status.

4 Existing and Proposed solution

There were various machine learning (ML) models used for traffic prediction, such as time series models, recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) networks, and convolutional neural networks (CNNs). Here is a summary of their characteristics and limitations:

Time Series Models:

Summary: Time series models use historical traffic data to forecast future traffic patterns based on seasonality, trends, and other temporal patterns.

Limitations: These models may struggle with capturing complex nonlinear relationships and long-term dependencies in traffic data. They might not consider other relevant factors like weather, events, and road infrastructure changes.

Recurrent Neural Networks (RNNs):

Summary: RNNs are designed to handle sequential data and can capture temporal dependencies in traffic patterns.

Limitations: Traditional RNNs suffer from the vanishing gradient problem, making it difficult to model long-term dependencies. As a result, they may struggle with long-range traffic predictions.

Long Short-Term Memory (LSTM) Networks:

Summary: LSTMs are a type of RNN that address the vanishing gradient problem and can capture long-term dependencies better.

Limitations: While LSTMs improved over traditional RNNs, they may still struggle with capturing complex spatial relationships in traffic data.

Convolutional Neural Networks (CNNs):

Summary: CNNs are primarily used in image recognition but have been adapted for traffic prediction by treating traffic data as 2D images (e.g., spatial traffic density maps).

Limitations: Using CNNs for traffic prediction may not fully capture the temporal aspects of traffic data, potentially leading to less accurate predictions.

Hybrid Models:

Summary: Hybrid models combine multiple ML techniques to leverage the strengths of different approaches for better traffic prediction performance.

Limitations: Developing and tuning hybrid models can be complex and computationally expensive, requiring substantial expertise and resources.

My proposed solution: The models cannot cover all aspects of the real system and the models must be supplemented with observed traffic status data, for example, traffic numbers and speed/travel measures, in order to have a proper depiction of reality. Many Internet of Things (IoT) sensors are put across so many sites in a smart city to gather data on traffic, drainage, passenger movement, and so on, and the revelations derived from these data are used to better manage resources, assets, and the like. Many researchers have employed Machine Learning substantially on data collected by IoT sensors in a smart city. The growth of the Internet of Things, as well as big data analytics and machine learning, has made the concept of smart city a possibility. The goal of a smart city is to deliver good services to its citizens by using modern technology and data analytics on data collected by sensors.

After carefully reviewing some relevant publications, it was discovered that the existing literature does not provide any work that uses BAG, GLM, BGLM, MARS, and KNN techniques to solve the problem of traffic congestion prediction in smart city. Furthermore, the prediction accuracy of some of the existing work is relatively low while in some cases the authors did not use performance metrics to evaluate the performance of their proposed system. Also, some authors did not compare their work with high performing machine learning or deep learning models.

- **(i)** My project proposed five (5) machine learning techniques for traffic pattern prediction in smart city.
- **(ii)** The Bagging (BAG), K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM) models for solving the problem of traffic congestion prediction in smart city was presented in this paper.
- **(iii)** Minimal Root Mean Square Error (RMSE), Mean Average Square Error (MASE), and Mean Square Error (MASE) for GLM model was attained. The proposed models were evaluated using different performance metrics. (iv) A recent review of state-of-the-art proposals for traffic patterns prediction in smart city is presented.

4.1 Code submission (Github link):

https://github.com/Anandita1233/SmartCity_Anandita_USC_UCT

4.2 **Report submission (Github link):**

https://github.com/Anandita1233/SmartCity_Anandita_USC_UCT

5 Proposed Design/ Model

Throughout the course of the entire internship, I worked on how to plan a solution and how to prepare the data for creating the training sets and validation set for the machine learning algorithms to be implemented on and to create a learned model from it.

The machine learning algorithms are:

1. **Bagging (BAG):**
   In bagging model only one prediction may be made from the decisions of different learners. In the case of classification, voting is simply the sum of those decisions. Individual models are created by a variety of methods of bagging and boosting. Models in bagging are given the same weights, whereas more successful models in boosting are given more weighting since an executive may base alternative outcomes on a variety of expert information based on their prior estimations. Individual decision trees are brought together by requiring them to vote on each test. If one class receives more votes than other classes because there are more votes, projections based on a larger number of votes are more reliable.
2. **K-nearest Neighbour:**
   A lazy approach is the K-Nearest Neighbour (KNN) classification, who merely keeps trainings, because there is no obvious process of training. It learns via analogy, meaning that a given test tuple is compared with similar training tuples. These tuples should be the closest to the unidentified tuple. A Euclidean distance measures the closest neighbour and the unknown tuple is chosen from its closest neighbours as the most common class.
   The basic KNN algorithm is:
   1. The value of K used in the KNN algorithm in this work is 3. This enables us to achieve desired accuracy.
   2. Calculate the distance of the test sample with all the samples of the training dataset. 3. Sort the distance and define the closest neighbours by minimum K-th distance.
   4. Assemble the closest neighbourhood categories. 5. Make use of simple majority of the closest neighbourhood category as a new data item prediction value.

3. Multivariate Adaptive Regression Spline (MARS):

Multivariate Adaptive Regression Spline (MAR) is a versatile regression method and non-parametric approach that incorporates piecewise linear regression function referred as basic function (bf). In order to estimate performance of MAR, it uses basic functions (bf) for capturing the hidden non-linear relations between independent input variables. Bf is therefore the main component in the generation of a MAR model.

4.Bayesian Generalized Linear Model (BGLM):

The Bayes Generalized Linear Model (BGL) assumes preliminary or prior distribution based on preliminary or prior information and subsequent distribution is achieved through the integration of sample information with such prior information. In general, information collected from the post or posterior distribution is closer to true information, since it brings together sample data and expert views.

The dataset on which the ML algorithms are implemented is:

Description of the dataset:

Terms	Meaning	Data
DateTime	A digital record of occurrence of a particular event	Date and time
Vehicles	Different types of transportation plying the road such as car, lorry, or cart	Number of vehicles at different intervals
Junction	A point where two or more things are joined	1–4 junction
ID	ID of each vehicle DateTime	

Comprehensive dataset of smart city traffic pattern:

Date	Vehicles	Junction	Weekday	Year	Month	Day	Time
01–11–15	15	1	1	2015	11	1	00
01–11–15	13	1	1	2015	11	1	01
01–11–15	10	1	1	2015	11	1	02
01–11–15	7	1	1	2015	11	1	03
01–11–15	9	1	1	2015	11	1	04
01–11–15	6	1	1	2015	11	1	05
01–11–15	9	1	1	2015	11	1	06
01–11–15	8	1	1	2015	11	1	07
01–11–15	11	1	1	2015	11	1	08

The proposed Smart city traffic pattern was borne out of the problem associated with urban city. Smart cities use big data for decision-making and a revolution in the Internet of a Thing (IoT) allows such data to be collection and transmission of such data. Transportation is a fundamental element influencing metropolitan areas and a crucial use case for smart cities. Smart transportation technologies, such as GPS Bus, traffic cameras, taxis, and geospatial technology, which is frequently employed in a traffic control centre to monitor and coordinate a wide network of sensors, are discovering creative methods to reduce traffic congestion and urban mobility. Smart city traffic pattern incorporates all the big data obtained from GPS Bus, traffic camera, GPS Taxi, and geospatial technology. The traffic pattern system consists of data visualization, data cleaning to remove unwanted and other irrelevant values.

5.1 High Level Diagram

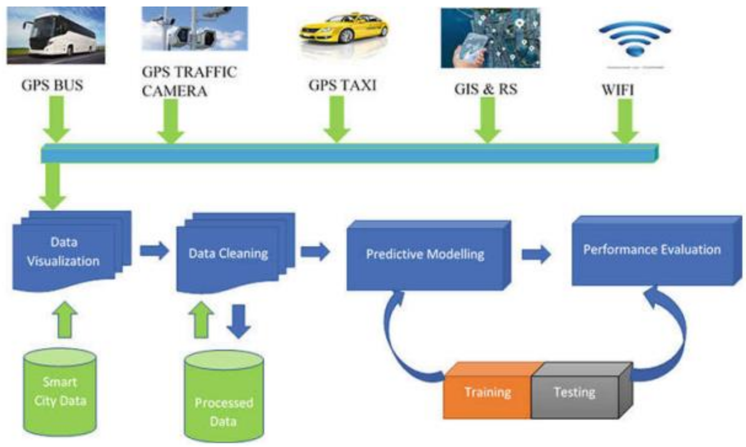


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM(Block diagram of smart city traffic pattern)

Results and Discussions:

The conditions of road traffic must be collected in order to manage and control traffic flows. The traffic condition can be characterized on a given section of the road with speed, flow, and density. In this study, traffic management is of paramount importance in a smart city.
Figure 2 shows the exponential increases in the number of vehicles between 2015 and 2017. This shows that there will be increase in the traffic congestion during this period. Figure 3 displays the increase in the number of junctions between 2015 and 2016. There is exponentially increase in the number of junctions every year. By comparing Figs. 2 and 3, we observed that as the number of vehicles increases, government also increases the number of junctions which leads to the reduction of traffic between 2016 and 2017. Congestion in traffic is characterized in transportation by slower speeds, longer travel durations, and an increase in vehicle queuing. Figure 4 shows the longer trip times experience on the road and vehicular queues.

In 2016–2017, the number of vehicles increases to 54,000 against 5700 obtained in 2015 which leads to increase in vehicular movement in (Morning, Afternoon, Evening, and Noon) time but as the number of junctions increases in the city, it resulted into decrease in vehicular queuing in 2017 to 48,000 as also shown in Figs. 5 and 6. The study reveals that increase in the number of junctions on a high way may lead to reduction in vehicular movement on the road.

Depicted in Fig. 7 are the output results of Actual (Vehicles) that are plying the roads in the city and the predicted values of the number of vehicles in 2017 using BAG, KNN, MARS, BGLM, and GLM models.

Figure 7 shows the machine learning algorithms utilized in this project. Figure 6 shows the actual value of the vehicles and the predicted value of the machine learning such as BAG, KNN, MARS, BGLM, and GLM.

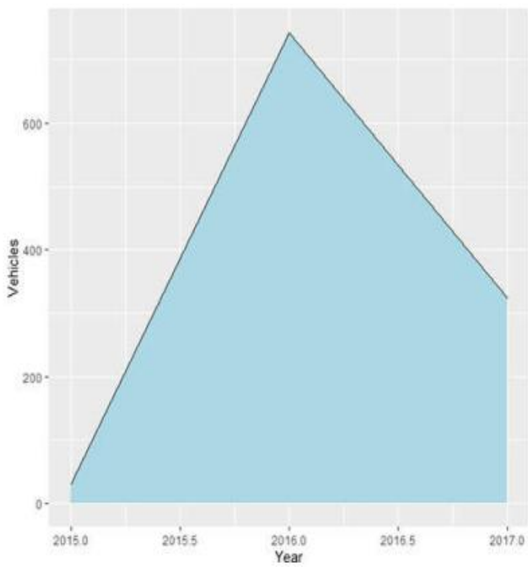


Figure 2: No. of vehicles v/s year

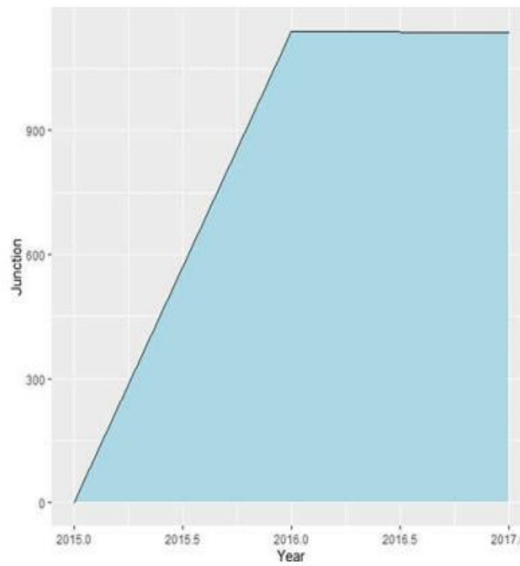


Figure 3: No. of junction’s v/s year

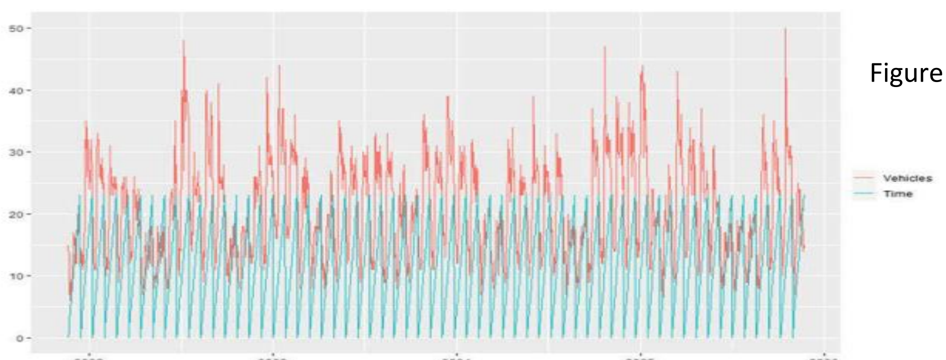


Figure 4: Vehicles v/s time in 2015

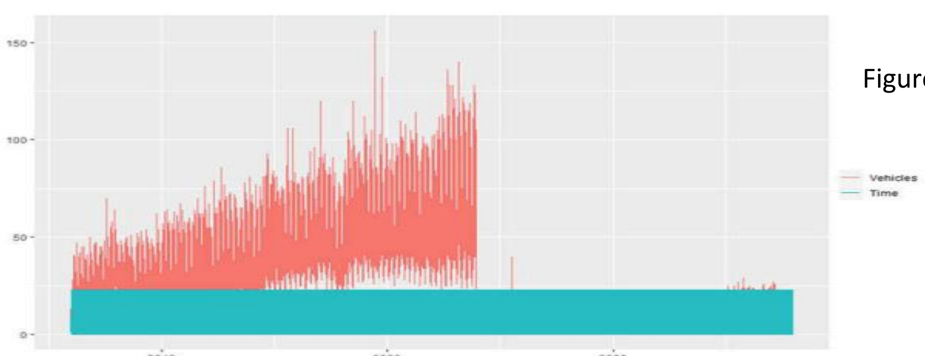


Figure 5: Vehicles v/s time in 2016

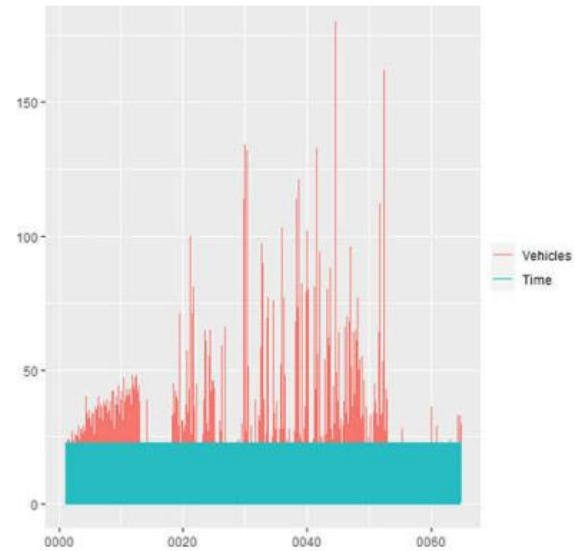


Figure 6: Vehicles v/s time in 2017

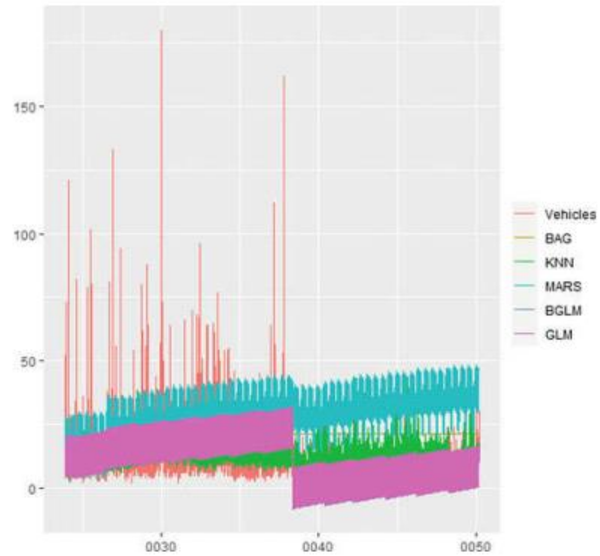


Figure 7: Output results of actual (Vehicles) and predicted (BAG, KNN, MARS, BGLM, and GLM)

6 Performance Test

Validating the Implementation and Performance of the Project:

Performance Evaluation:

The accuracy test of the smart city dataset is evaluated using:

1) Mean Absolute Error (MAE)

   Consider a set of the actual closing price Cp and the predicted values C cp . MAE are given as follows:

$$\frac{1}{n}\sum_{n=1}^{n}\left|C_p - \widehat{C_p}\right|$$

2) Root Mean Square Error (RMSE)

$$\sqrt{\frac{1}{n}\sum_{n=1}^{n}\left(C_p - \widehat{C_p}\right)^2}$$

3) Mean Square Error (MSE)

   MSE is given as

$$\frac{1}{n}\sum_{n=1}^{n}\left(C_p - \widehat{C_p}\right)^2$$

4) Mean Absolute Scaled Error (MASE)

   MASE is given as

$$\frac{1}{n}\sum_{n=1}^{n}\frac{\left|C_p - \widehat{C_p}\right|}{\frac{1}{n-m}\sum_{n=m+1}^{n}\left|C_p - \widehat{C_p}\right|}$$

   where m is the seasonal period of the closing price and n is the trading days.

Last week we had worked on identifying the trends and pattern in data and plotting their relationships on a graph to understand them better. This week we evaluated the performance and implementation of the project throughout the week which can be represented in the form of graph to better understand them. Last week we identified the traffic pattern in data through various constraints such as days, various durations during the day (Morning, Noon, Afternoon, Evening) and applied the various machine learning algorithms to obtain a learned model. This week we further elaborate on the that for implementation and performance.

Traffic congestion varies within the day event may cause surges in traffic at unexpected times. In Figs. 8, 9 and 10, we considered Morning, Afternoon, Evening, and Noon. The study revealed that in 2015 there is traffic surge in the Morning followed by Afternoon as shown in Fig. 8. In 2017 the traffic congestion increased exponentially which results in traffic surge in Afternoon than in the Noon as shown in Fig. 9. Depicted in Fig. 10 is the vehicular movement in 2017 which shows that Afternoon has the highest number of vehicles plying the road followed by Noon.

The performance of each of the machine learning is shown in Table 1 using Root Mean Square Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Scaled Error (MASE). GLM performs better than three the remaining four algorithms such as BAG, KNN, MARS, and BGLM having the least accuracy of 8.6802, 75.3468, 1.0498 using RMSE, MSE, MASE. From Table 3, we can deduce that GLM is a better predictor compared to the other models used in this work. Therefore, it is a promising algorithm for predicting vehicular movements.

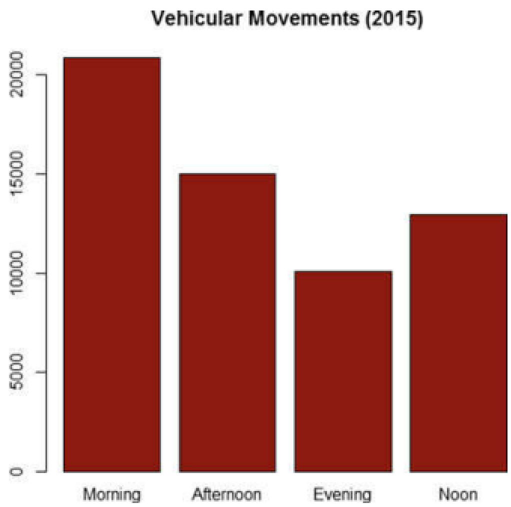


Fig-1 Vehicular movements in 2015 during the day (Morning, Afternoon, Evening, Noon)

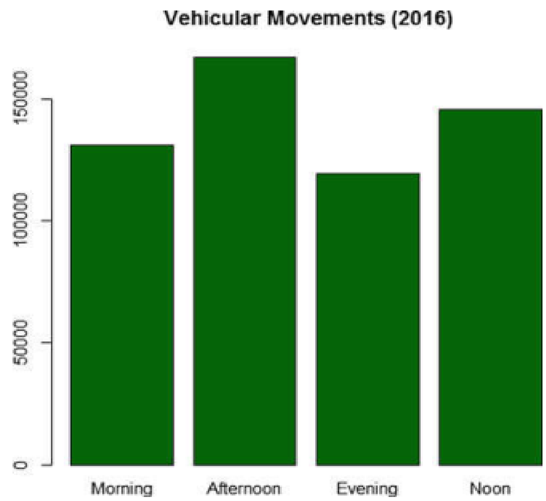


Fig-2 Vehicular movements in 2016 during the day (Morning, Afternoon, Evening, Noon)

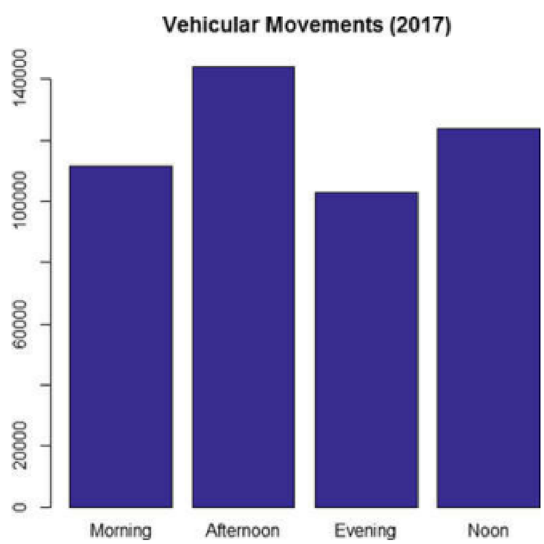


Fig-3 Vehicular movements in 2017 during the day (Morning, Afternoon, Evening, Noon)

MODEL	RMSE	MSE	MASE
BAG	13.09740	171.5419	1.958938
KNN	9.236216	85.30769	1.093255
MARS	23.34668	545.0675	3.775596
BGLM	8.703736	75.75501	1.054976
GLM	8.680255	75.34683	1.049848

Table-1 Performance metrics of smart city traffic patterns

Smart city building, a new form of information technology-based urban development plan, is dedicated to changing the traditional urban organising and management paradigm to enable effective data and resource distribution and address urbanisation issues. This project investigated the smart city's traffic congestion. This research fills this vacuum in the literature by examining whether intelligent city-buildings, such as road junctions, may reduce traffic in cities using data from the Kaggle database, which has 48,120 entries from 2015 to 2017. The construction of smart cities, such as road intersections, has been shown to significantly reduce traffic congestion and boost urban efficiency. The mechanism check demonstrates that the growth of information technology and smart cities is a solution to address traffic problems while also reducing the congestion of urban highways by fostering urban overall creativity. This project report has important practical ramifications that will help to advance the growth of smart cities and relieve traffic congestion. The dataset's inability to provide information on the number of buses, taxis, and traffic cameras at each intersection prevents us from being able to calculate the pace at which vehicles are backed up at each road junction, which is one of the work's limitations.

7 My learnings

It was wonderful to learn and engage in a project to anticipate traffic in smart cities. These initiatives offered insightful information about how to improve urban planning and optimize transportation networks.

1) Working with real-world traffic data taught me the value of high-quality data collection and preparation. Before creating my prediction models, I came across problems with missing data, outliers, and noise that needed to be fixed.
2) Engineering features that work well is essential for precise traffic forecast. I gained knowledge on how to extract pertinent elements from my dataset, such as past traffic trends, weather patterns, special events, road closures, and holidays.
3) Evaluation Metrics: I employed a variety of evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to evaluate the effectiveness of your prediction models.
4) Temporal Patterns: Throughout the course of my study, I've discovered recurrent temporal patterns in traffic data, such as daily rush hours, weekly changes, and seasonal trends. Having this information was crucial for increasing forecast accuracy.
5) External Factors: I could never forget about how the weather, public events, road upkeep, and holidays may affect traffic patterns. Including these elements in my models helps produce forecasts that are more precise.

Working on a complex project like traffic prediction allowed me to develop and refine a wide range of skills, including data preprocessing, feature engineering, model selection, evaluation, and interpretation. These skills are highly valuable in the field of data science.

8 Future work scope

The future scope of smart city traffic prediction is promising and encompasses various advancements and applications. Here are several key areas in traffic prediction in smart cities that I could not incorporate into the project:

Data Integration and IoT: Smart cities will continue to gather data from various sources, including sensors, cameras, GPS devices, social media, and more. Integrating and analyzing this diverse data through the Internet of Things (IoT) will enable more accurate and real-time traffic predictions.

Real-Time Updates: With the proliferation of mobile devices and applications, real-time traffic updates can be delivered directly to users' smartphones, helping them make informed decisions about route planning and travel time estimation.

Traffic Flow Optimization: Traffic prediction can be integrated with smart traffic management systems to optimize traffic flow dynamically. This could involve adjusting traffic signals, rerouting vehicles, and managing congestion based on real-time predictions.

Multi-Modal Transportation: Smart city traffic prediction systems will increasingly consider multi-modal transportation options, including buses, trains, bicycles, and pedestrian pathways. This holistic approach will help citizens choose the most efficient mode of transportation.

Environmental Impact: Future systems might consider the environmental impact of traffic congestion and offer suggestions for reducing emissions and energy consumption through optimized traffic management.

Public Participation and Feedback: Citizen engagement can provide valuable data for traffic prediction. Gathering feedback from residents about their travel plans, preferences, and experiences can help refine prediction models.