Industrial Internship Report on

"Smart City Traffic Patterns"

Prepared by

Sumit Jana

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 Patterns** and I have to work on the project to convert my current city to 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. I have to work as a data scientist, working to manage the traffic of the city better and to provide input on infrastructure planning for the future. Train data and Test data is provided and I have to work on train data to and build a model that can predict on test data. By this we can forecast the traffic for the city and by knowing the traffic we can be prepared for traffic peaks. If we can predict the traffic condition the we can make alternate ways to reduce the high traffic conditions. And this will help in many ways to make life of every people in this city easier.   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. Also got to know about how to prepare myself for interview and what are things to learn and how to answer questions and what are the things to keep in mind most while answering the questions and also got the exposure to some sample questions what frequently asked in interview.

TABLE OF CONTENTS

1 Preface

1.1 Summary of the whole 6 weeks' work.

During the first week of the internship, I learned about Uniconverge Technologies Private Limited, their domains, products/services, and technologies because at first everyone should know about the company who gave this opportunity to work under them and develop knowledge and skill by working. I choose the 'Smart City Traffic Patterns' project, conducted a Literature Survey, and explored the need for Smart City Traffic Patterns and the problems I can solve.

In the second week, I explored different algorithms, including Random Forest, Decision Tree, and Support Vector Machine, to find a solution for the project. They studied an e-book on Data Science and Machine Learning, learning about their principles and applications.

In week 3, I started implementing the algorithms I had explored in the previous week. I measured the accuracy of each algorithm and visualized the real and predicted traffic patterns at various junctions. Decision Tree yielded the highest accuracy. I also studied an e-book on Probability and Statistics and learned about the differences between AI and Data Science.

In the fourth week, I focused on finding a solution using the Decision Tree algorithm and successfully implemented it for the test dataset. I studied topics such as Introduction to Machine Learning, Linear Functions, and Optimization techniques.

During the fifth week, I thoroughly reviewed their project code and explored other algorithms to potentially improve the model's performance. I didn't find any significant errors but had time constraints preventing further exploration. Also gone through some frequently asked interview questions on data science and got some very useful tips on how to prepare for an interview and how to answer questions during interview.

Overall, I had a productive and educational experience during their internship, learning about various algorithms, data science, machine learning, and how to apply them to address real-world challenges like Smart City Traffic Patterns.

1.2 Need of this Internship in career development.

The internship on "Smart city traffic patterns detection using Data science and machine learning" holds significant importance in my career development as a final year student of Computer Science Engineering with specialization in AI and ML. Here are several reasons why this internship is valuable for my career growth:

1. Practical Application of Knowledge: The internship provides the opportunity to apply the theoretical concepts and skills I've learned throughout your academic journey in a real-world

scenario. Practical experience is highly valued by employers and can enhance my understanding of AI and ML concepts.

2. Relevant Industry Experience: Working on a project related to Smart City Traffic Patterns exposes me to a relevant and cutting-edge domain. This experience aligns with the current trends and demands in the technology industry, making me a competitive candidate for future job opportunities in similar domains.

3. Skill Development: The internship allows me to enhance my technical skills in data science, machine learning, and programming. I have also gained experience in data analysis, visualization, and the implementation of algorithms, which are crucial skills sought after by employers in the AI and ML field.

4. Building a Portfolio: Completing an internship in this specialized domain enables me to add a meaningful project to your portfolio. A well-documented project showcasing my work on Smart City Traffic Patterns detection can impress potential employers and demonstrate my proficiency in AI and ML.

5. Networking Opportunities: During my internship, I have collaborated with experienced professionals, mentors, and fellow interns. Building a network within the industry can open doors to future job prospects and career growth.

6. Problem-Solving Abilities: Working on a complex project like Smart City Traffic Patterns detection challenges my problem-solving abilities. I have learned to analyze issues, identify potential solutions, and make informed decisions, all of which are crucial skills for career development.

7. Exposure to Real Data: Handling real-world datasets during the internship equips me with the skills to manage and analyze large-scale data, which is an essential aspect of many AI and ML applications in the industry.

8. Understanding Project Lifecycle: From project initiation to completion, I have gained insights into the entire project lifecycle. Understanding how projects are planned, executed, and delivered prepares me for working on similar projects in the future.

9. Increased Confidence: Successfully completing an internship on a complex topic like Smart City Traffic Patterns detection have boosted my confidence in tackling real-world challenges and reinforce my belief in your skills and abilities.

10. Fostering Innovation: Engaging in research and development during the internship encourages innovation and creative thinking. These qualities are highly valued in the AI and ML industry, where new solutions are constantly sought to improve various applications.

In conclusion, this internship experience has significantly contributed to my career development by providing hands-on experience, skill enhancement, exposure to relevant industry domains, and the opportunity to create a strong professional network. The knowledge and expertise gained from this internship can serve as a solid foundation for pursuing a successful career in the field of AI and ML.

1.3 Brief about Your project/problem statement.

My project was Smart City Traffic Patterns and I have to work to work on the project to convert my current city to 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. I have to work as a data scientist, working to manage the traffic of the city better and to provide input on infrastructure planning for the future.

Train data and Test data is provided and I have to work on train data to and build a model that can predict on test data. By this we can forecast the traffic for the city and by knowing the traffic we can be prepared for traffic peaks. If we can predict the traffic condition the we can make alternate ways to reduce the high traffic conditions. And this will help in many ways to make life of every people in this city easier

1.4 Opportunity given by USC/UCT.

As a final year student of computer science engineering with a specialization in AI and ML, I am extremely grateful to USC/UCT also IOT Academy for providing me with the opportunity to undertake the internship on "Smart city traffic patterns detection using Data science and machine learning". USC/UCT and IOT Academy has played a vital role in shaping my career and contributing to my professional growth in several ways:

1. Industry Exposure: USC/UCT is a reputed company known for its expertise in the field of data science and machine learning, IoT etc. Being associated with such a company has exposed me to real-world industry practices, cutting-edge technologies, and advanced methodologies used in solving complex problems.

2. Supportive Environment: Throughout my internship journey, USC/UCT has fostered a supportive and inclusive environment. They have provided me with the necessary resources, guidance, and mentorship, enabling me to make the most of my learning experience and challenging me to perform at my best.

3. Diverse Projects: USC/UCT's diverse range of projects allows interns to work on varied topics and domains. The opportunity to choose the "Smart city traffic patterns detection" project aligned perfectly with my interests in AI and ML and offered a challenging and rewarding experience.

4. Guidance: As this was a self-driven internship but then also, they have guided me throughout the internship period which helped me to complete the project work . Everything what I have to do was written on the project document.

5. Real-World Impact: One of the most rewarding aspects of interning with USC/UCT is knowing that my work contributes to solving real-world challenges. The "Smart City Traffic Patterns detection" project has the potential to impact the lives of many, making it a meaningful and purposeful endeavor.

6. Continuous Learning: USC/UCT places a strong emphasis on continuous learning and encourages interns to explore new technologies, attend workshops, and engage in skill-building activities. This culture of learning has inspired me to stay curious and updated with the latest advancements in the AI and ML domain.

7. Recognition of Efforts: USC/UCT acknowledges and appreciates the efforts put forth by interns. Recognition for the work done boosts motivation and confidence, fostering a positive and productive internship experience.

In conclusion, my internship experience with USC/UCT has been invaluable in terms of knowledge, skill development, and industry exposure. The opportunity to work on the "Smart City Traffic Patterns detection" project has been a stepping stone in my career journey, and I am thankful to USC/UCT for providing me with the platform to grow and excel in the field of AI and ML.

1.5 How Program was planned

I planned the 6-week execution program flow of this internship according to the plan provided by USC/UCT. This helped me to complete my project work on time. Program flow is given below:
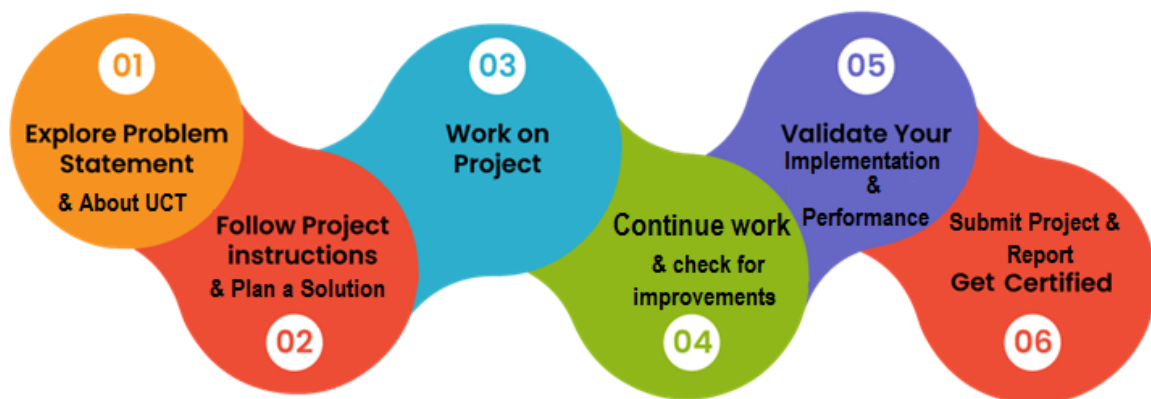
1.6 My Learnings and overall experience.

During my internship on "Smart City Traffic Patterns Detection using Data Science and Machine Learning" with USC/UCT, I have had an enriching and transformative experience. The internship has not only allowed me to apply my theoretical knowledge but has also equipped me with practical skills that will undoubtedly shape my future career in AI and ML. Here are some of the key learnings and insights from my internship:

1. Application of Data Science and ML Concepts: I gained hands-on experience in applying data science and machine learning concepts to solve real-world challenges. From data preprocessing to

model implementation, I learned how to tackle complex problems systematically and derive meaningful insights from data.

2. Exploration of Advanced Algorithms: The internship exposed me to a variety of machine learning algorithms like Random Forest, Decision Tree, and Support Vector Machine. Understanding the strengths and weaknesses of each algorithm helped me make informed decisions in selecting the most suitable approach for the "Smart City Traffic Patterns Detection" project.

3. Data Analysis and Visualization: I learned the importance of data analysis and visualization in deriving meaningful patterns and trends from data. Visualizing real and predicted traffic patterns at various junctions helped me communicate insights effectively and aided in making informed decisions.

4. Project Management and Planning: Working on a project with multiple phases taught me the significance of project management and planning. Breaking down the project into manageable tasks and setting realistic milestones ensured smooth progress and timely completion.

5. Critical Thinking and Problem-Solving: The project involved several challenges that required critical thinking and problem-solving skills. I learned to approach problems analytically, experiment with different solutions, and adapt my strategies when necessary.

6. Collaboration and Teamwork: During the internship, I had the opportunity to collaborate with experienced professionals and fellow interns. Working in a team environment taught me the value of effective communication, sharing ideas, and collective problem-solving.

7. Continuous Learning and Adaptability: The field of AI and ML is ever-evolving, and this internship emphasized the importance of continuous learning and adaptability. I developed a mindset of staying updated with the latest advancements in the industry to remain relevant and innovative.

8. Real-World Impact: Knowing that the project's outcome could have a positive impact on smart city traffic patterns was motivating and fulfilling. Understanding how data-driven solutions can be applied to solve real-world problems enhanced my passion for using AI and ML for societal benefits.

Overall, my internship on "Smart City Traffic Patterns Detection using Data Science and Machine Learning" has been a transformative experience that has deepened my understanding of AI and ML concepts. It has provided me with a solid foundation for my future career aspirations. I am thankful to USC/UCT for the support, guidance, and exposure to real-world challenges that have been instrumental in shaping my professional growth. This internship has not only given me valuable technical skills but has also instilled in me a sense of purpose and commitment to using AI and ML for creating a smarter and more sustainable future.

Though I have done this project all alone but without help of mentors and some intern mates who helped me to complete my project and other works on time. Thankyou once to The IOT Academy, Upskill Campus and Uniconverge Technologies and everyone who helped me directly and indirectly.

1.7 My message to juniors and peers.

To my juniors and peers,

As my internship on "Smart City Traffic Patterns Detection using Data Science and Machine Learning" concludes, I want to share some advice. Embrace opportunities, be curious, and keep learning in the evolving AI and ML field. Face challenges with courage, collaborate, and celebrate achievements. Learn from failures, stay humble, and always seek growth. Find your passion and pursue projects aligned with it. Support each other in this competitive field and believe in your abilities. Remember, this is just the beginning of a lifelong journey of knowledge and growth. Embrace every experience, learn from it, and shape the best version of yourself. Wishing you success, fulfillment, and joy in your academic pursuits and future careers. Keep striving for excellence, and I am excited to witness your great accomplishments.

Best wishes,

Sumit Jana

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
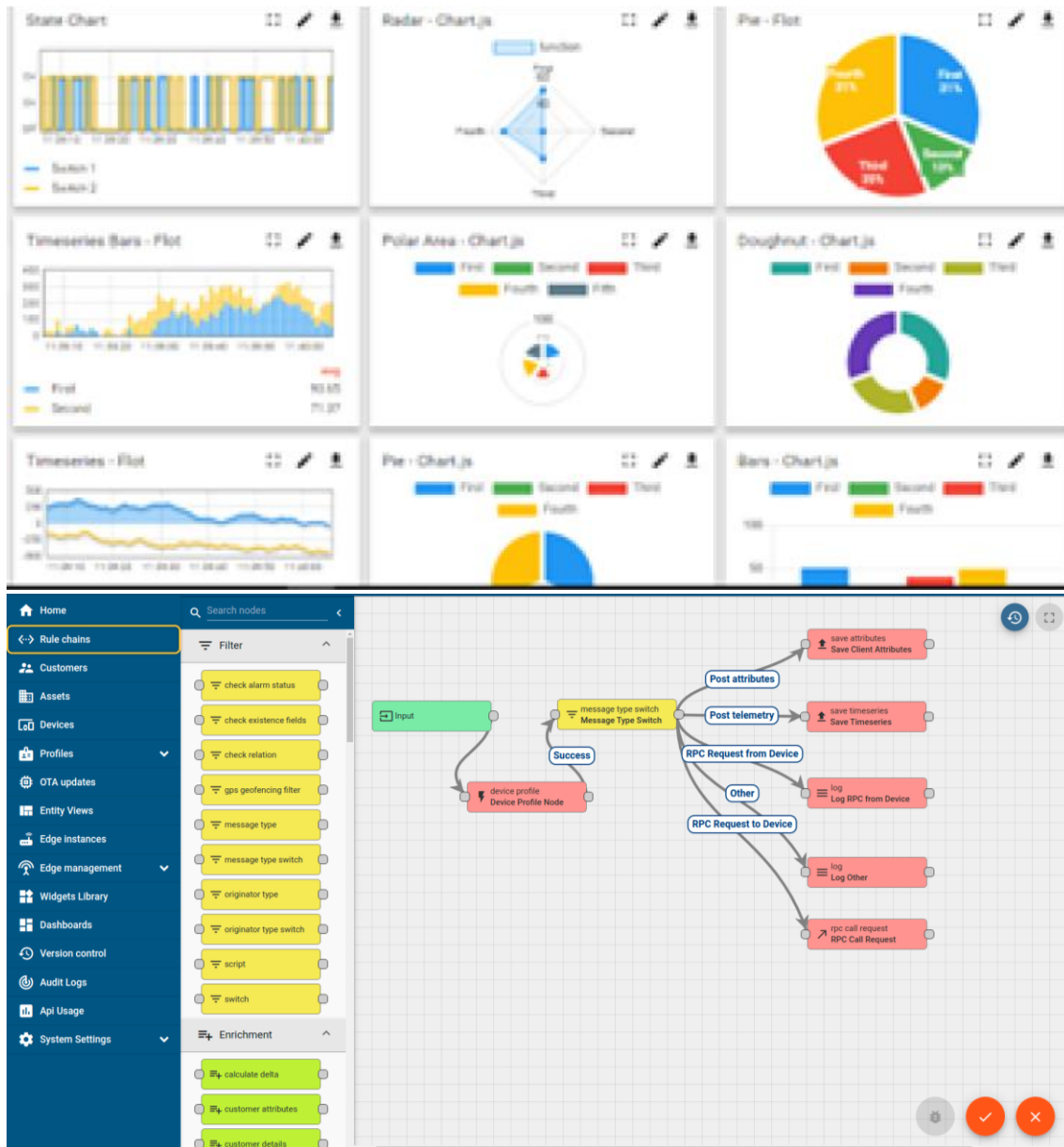
ii. Smart Factory Platform ( )

Factory watch 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.
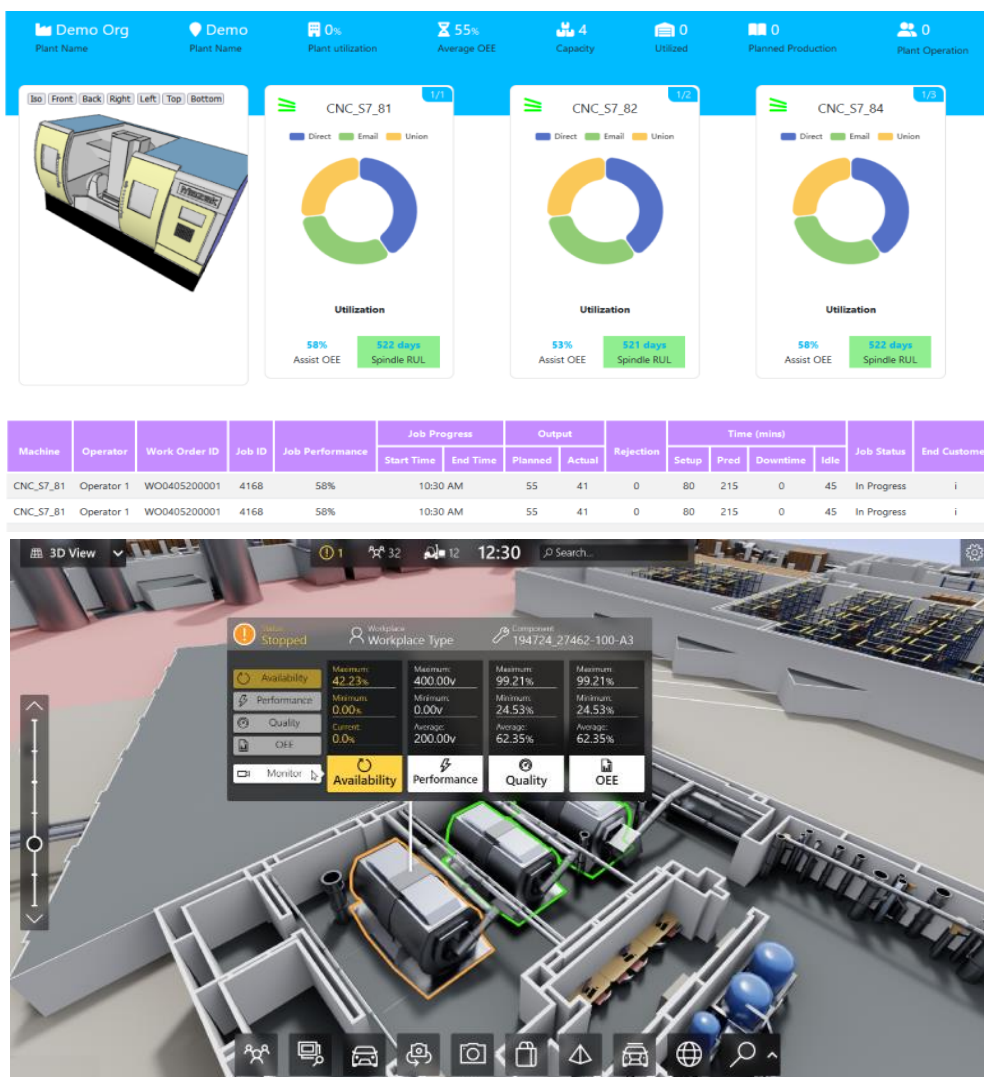
iii. **LoRaWAN** based Solution

UCT is one of the early adopters of LoRAWAN technology 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.
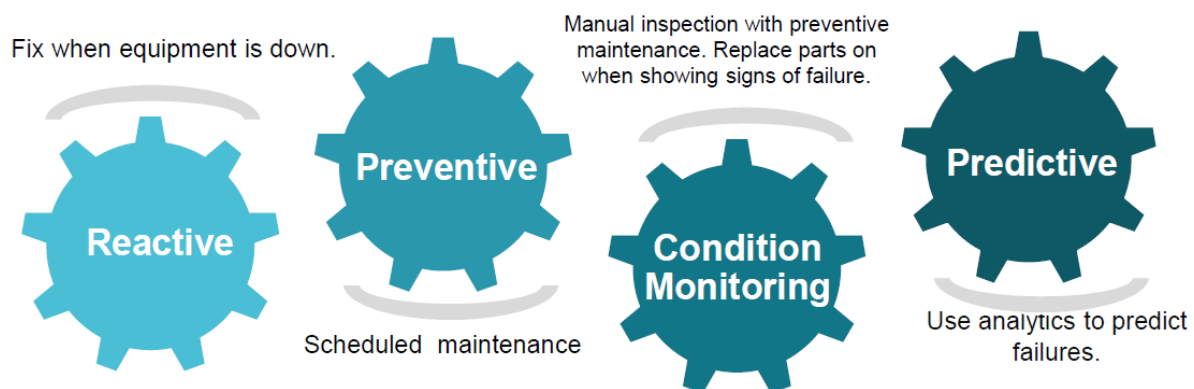
2.2 About upskill Campus (USC)

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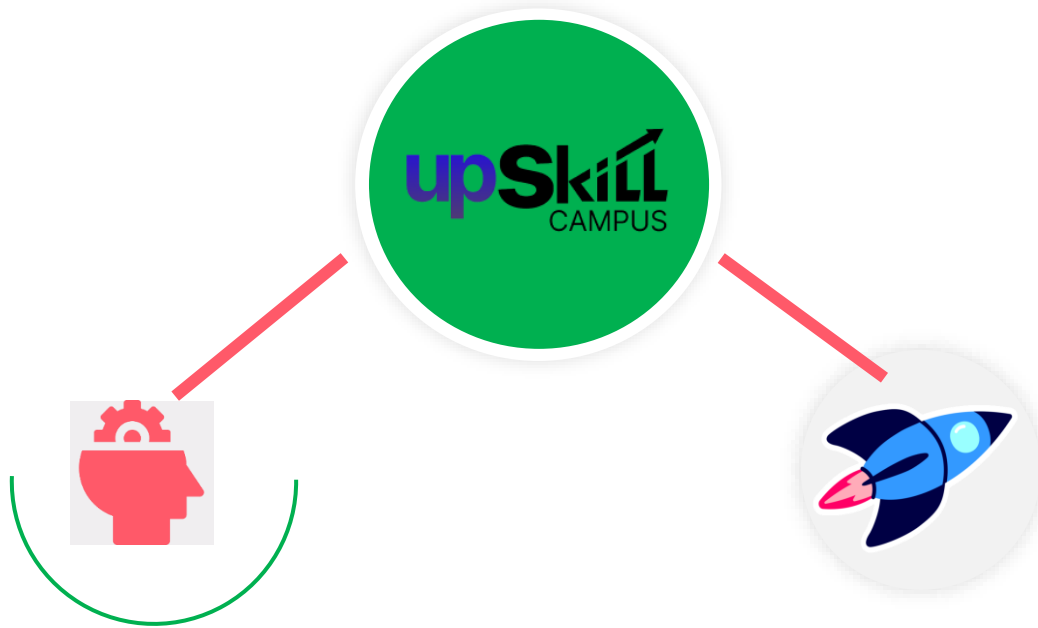
Seeing need of upskilling in self-paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career

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

https://www.upskillcampus.com/

Career growth/upskilling	Professional networking	Collaboration platform	Job/internship platform
Interview Preparation and skill building	Alumni Connections	Project collaboration	Job portal
upskilling Courses	Mentorship	Discussion forum	Internship portal
Skill Assessment	Discussion/QA forum	Tech updates	Freelancing projects
Profile building			

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.researchgate.net/publication/363677388_Smart_City_Traffic_Patterns_Prediction_Using_Machine_Learning

2. https://www.geeksforgeeks.org/decision-tree/

3. https://www.javatpoint.com/machine-learning-random-forest-algorithm

4. https://www.geeksforgeeks.org/support-vector-machine-algorithm/

5. https://stackoverflow.com/

6. https://towardsdatascience.com/decision-tree-in-machine-learning-e380942a4c96

7. https://towardsdatascience.com/what-when-how-extratrees-classifier-c939f905851c

8. https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-interview-questions

9. https://chat.openai.com/

2.6 Glossary

Terms	Acronym
AI & ML	Artificial Intelligence and Machine Learning
Data Science	An interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data.
Smart City	A city that uses digital technologies and data to enhance the quality and performance of urban services, reduce costs and resource consumption, and improve the overall well-being of its citizens.
Traffic Patterns	The regular or predictable flow of vehicular and pedestrian movement in a city or specific areas, which can be analyzed to optimize traffic management and infrastructure planning.
Decision Tree	A predictive modeling technique that maps decisions and their possible consequences in the form of a tree-like structure
Random Forest	An ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting in predictions.
Support Vector Machines(SVM)	A supervised machine learning algorithm used for classification and regression tasks by finding an optimal hyperplane that best separates data points of different classes.
Portfolio	A collection of projects, works, or achievements that showcase an individual's skills, experiences, and expertise.
Data Analysis	The process of inspecting, cleaning, transforming, and modeling data to discover meaningful information and make informed decisions.
Data Preprocessing	The process of cleaning, transforming, and preparing raw data before feeding it into a machine learning model for analysis.
Optimization Techniques	Algorithms and methods used to find the best possible solution to a problem within defined constraints.
ExtraTreeClassifier	ExtraTreesClassifier is an ensemble learning method fundamentally based on decision trees. ExtraTreesClassifier, like Random Forest, randomizes certain decisions and subsets of data to minimize over-learning from the data and overfitting.
USC/ UCT	UpSkill Campus/ Uniconverge Technologies

3   Problem Statement

You are working with the government to transform your city 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.

Explanation of problem statement

The problem statement involves transforming the city into a smart city with a focus on improving the efficiency of services for its citizens. As a data scientist, the specific challenge assigned is to address the issue of traffic congestion within the city. The government aims to implement a robust traffic management system that can effectively handle traffic peaks and variations in traffic patterns during different occasions, such as holidays and special events.

The key objectives are:

1. Traffic Analysis: Conduct a comprehensive analysis of traffic patterns at the four junctions of the city. This includes understanding the flow of vehicles, identifying peak hours, and detecting variations in traffic on holidays and other occasions compared to regular working days.

2. Forecasting: Develop accurate traffic forecasting models that can predict the expected traffic volume at each junction for different days of the week and special occasions. This will help in better resource allocation and planning for managing traffic effectively.

3. Infrastructure Planning: Provide valuable insights to aid in infrastructure planning. This involves suggesting appropriate improvements, such as road expansions, traffic signal optimizations, and the possible implementation of smart technologies to alleviate traffic congestion.

4. Smart Traffic Management: Propose the implementation of an intelligent traffic management system that can dynamically adjust traffic signals, reroute traffic, and provide real-time updates to drivers to optimize traffic flow during peak periods.

5. Data-Driven Decision Making: Utilize data-driven approaches to make informed decisions on traffic management strategies. Analyze data from various sources, such as traffic cameras, GPS devices, and mobile apps, to gain valuable insights into traffic patterns and identify potential areas of improvement.

By addressing these objectives, the aim is to create a traffic management system that enhances the overall mobility and quality of life for the citizens, reduces commute times, minimizes congestion, and contributes to the city's transformation into a smart and intelligent urban center.

4 Existing and Proposed solution

Existing Solution:

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. The paper used five machine learning models: Bagging (BAG), K-Nearest Neighbors (KNN), Multivariate Adaptive Regression Spline (MARS), Bayesian Generalized Linear Model (BGLM), and Generalized Linear Model (GLM) to predict traffic pattern in a smart city. The dataset consists of 48,120 rows and 4 columns from which the weekday, year, month, date, and time were extracted. Analysis results show that increase in the number of junctions of the city can alleviate problem being faced on the road by commuters. The Root Mean Square Error (RMSE) of BAG, KNN, MARS, BGLM, GLM are 13.09, 9.23, 23.34, 8.7, and 8.6 respectively. Experimental results demonstrated that GLM attained minimal prediction error compared to other machine learning models such as BAG, KNN, MARS, and BGLM used in this study.

Limitations: They explored BAG, KNN, MARS, BGLM and GLM algorithms and calculated root means squared errors.

But there are some other algorithms which can be used to predict solution

GLM is attaining minimal prediction error but there may be some algorithm who can have lesser prediction error.

So, I decided to explore other than these algorithms

Proposed Solution: As they have explored BAG, KNN, MARS, BGLM and GLM machine learning models so I decided to explore Random Forest, Decision Tree and Support Vector Machines. The data set I have used is same as them which consisting two datasets one for training and another for testing of predicting. Train dataset consists of 48120 rows and 4 columns, and column names are DateTime, Junction, Vehicle and ID and from DateTime weekday, year, month, date, and time were extracted. Test dataset consists 11808 rows and 3 columns which are DateTime, Junction and ID. And I have used train data to find number of vehicles in test dataset. Root Mean Square Error and Accuracy on prediction of Random Forest, Decision Tree and Support Vector Machine are 7.86 & 20.5%, 0.00 & 100%, 11.45 & 8.3%.

Experimental results demonstrated that Decision Tree(DT) attained minimal prediction error compared to other machine learning models used in this study.

What value addition are you planning?

I have planned to use Decision Tree to solve the predict number of vehicles on test dataset.

Then planned to find the best variable among them and build the model

Then I will use test dataset to predict number of vehicles and find peeks of traffic.

4.1   Code submission (GitHub link) :
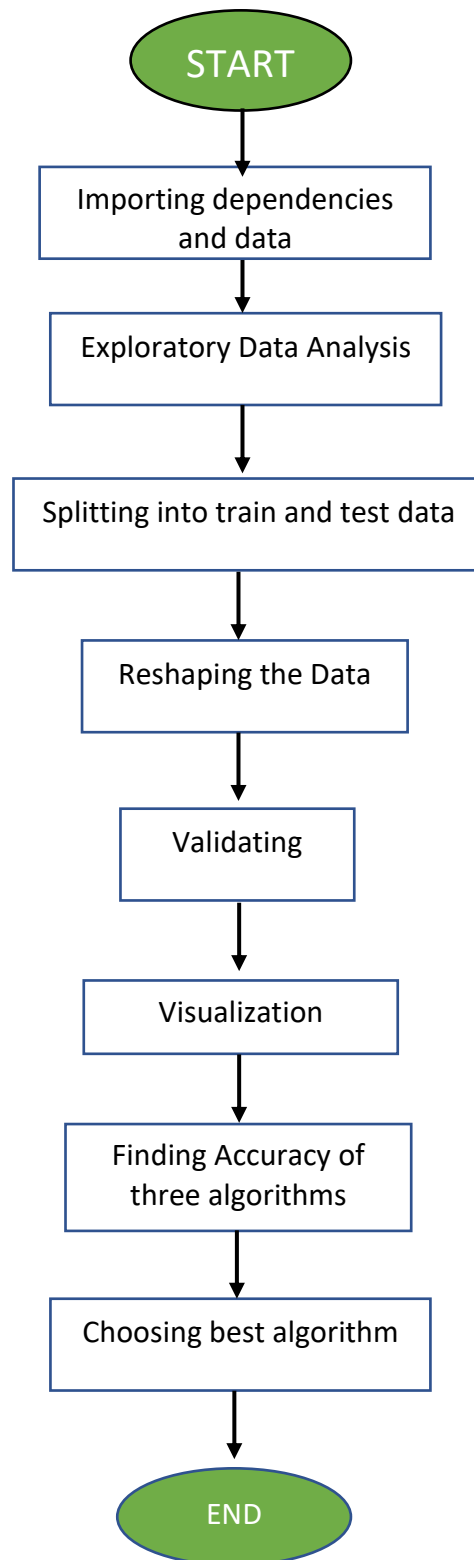
https://github.com/SumitJana404/UpSkill-Campus.git

4.2   Report submission (GitHub link) :

https://github.com/SumitJana404/UpSkill-Campus.git

Both links are same and of the link to the repository where codes and report are uploaded.

5   Proposed Design/ Model

Flow Diagram of Selection of algorithm:

```
                      ┌─────────────┐
                      │    START    │
                      └─────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │ Importing dependencies │
                  │        and data        │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │ Exploratory Data Analysis │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │ Splitting into train and test data │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │   Reshaping the Data   │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │      Validating        │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │     Visualization      │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │   Finding Accuracy of  │
                  │    three algorithms    │
                  └────────────────────────┘
                             │
                             ▼
                  ┌────────────────────────┐
                  │  Choosing best algorithm │
                  └────────────────────────┘
                             │
                             ▼
                      ┌─────────────┐
                      │     END     │
                      └─────────────┘
```

Flow diagram of designing solution:
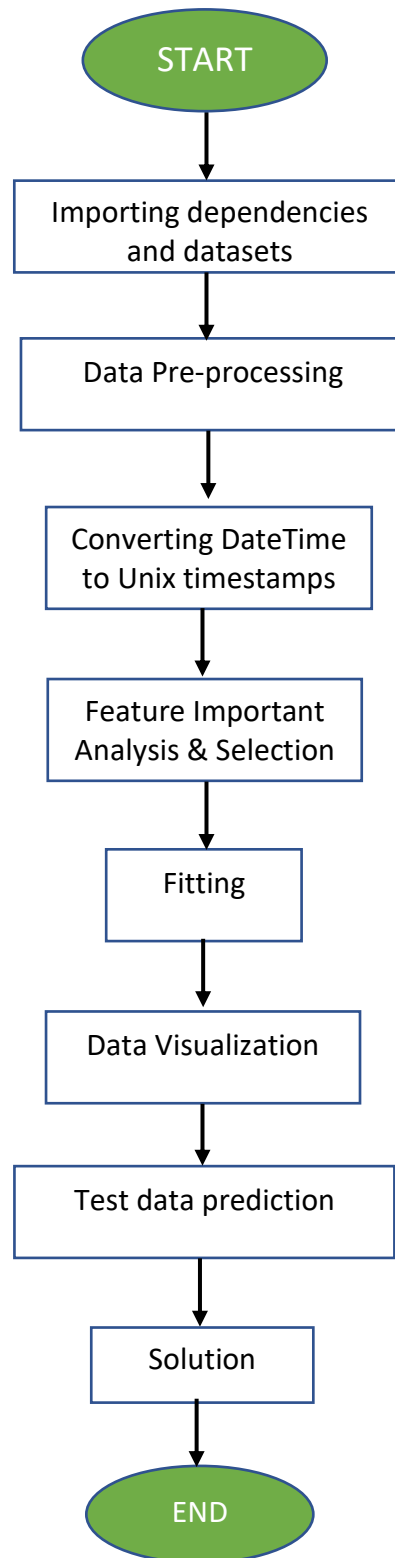
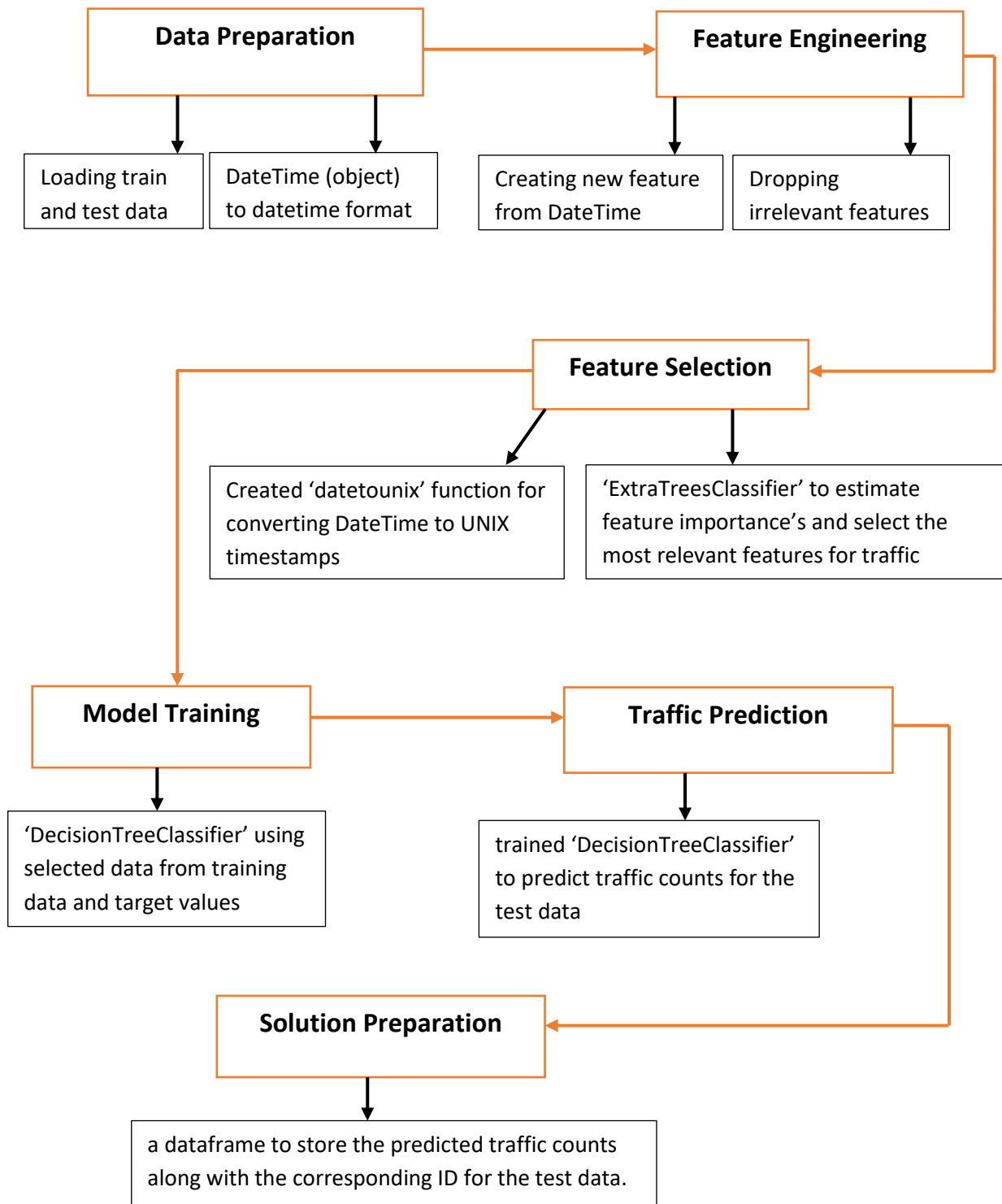## 5.1 High Level Diagram

Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

6 Performance Test

As I have not tested them, but to reach the final solution I have tested and chooses the best to predict more accurate result. For example, at first while implementing three algorithms I have measured Root Mean Square Error and Accuracy then choose Decision Tree because it was giving maximum accuracy and less error on prediction.
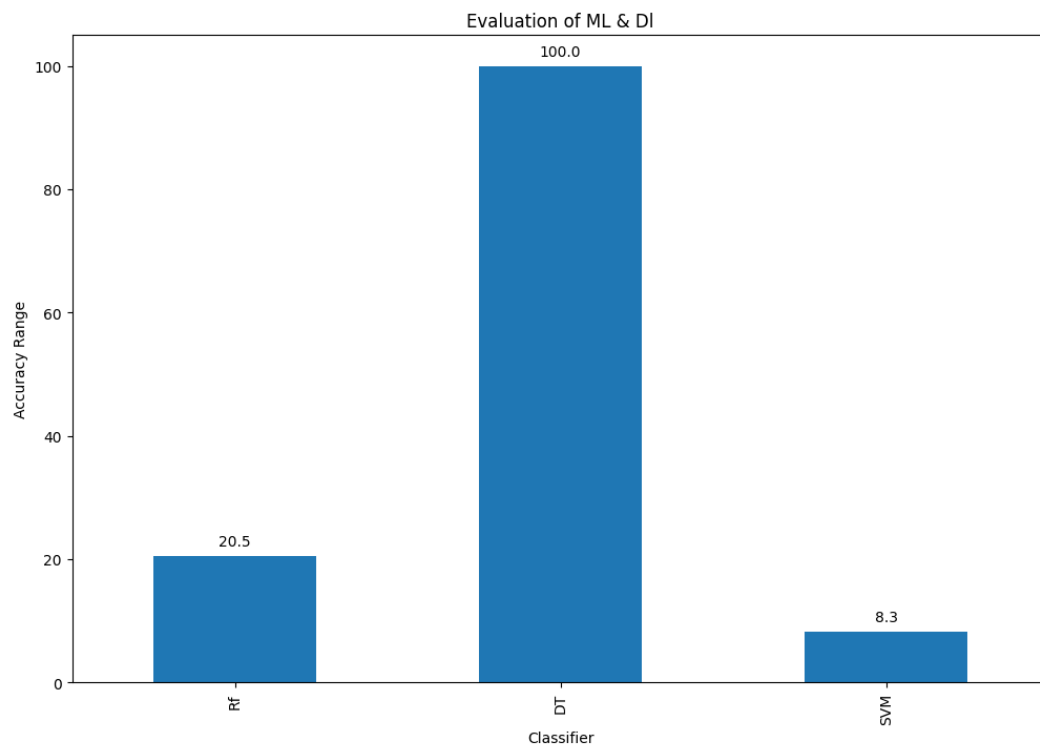
After that while implementing Decision Tree I have extracted many features from 'DateTime' feature and used 'ExtraTreeClassifier' for selection of most important and dropped irrelevant features which will help to predict more accurately.
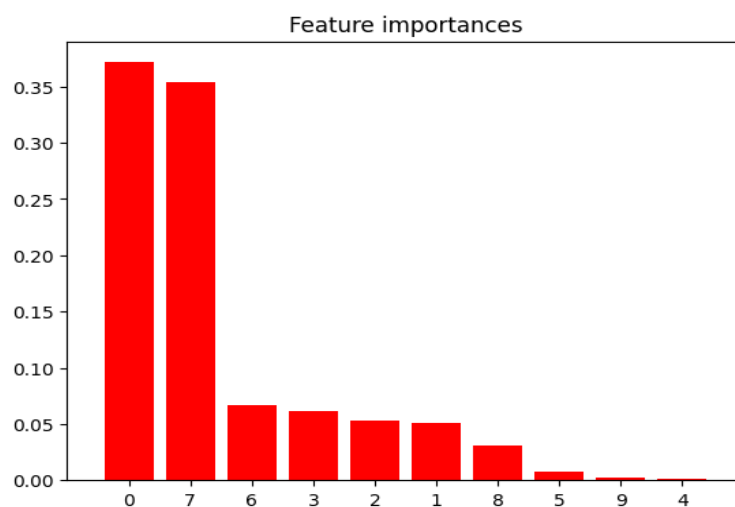
After that I have fitted the model. In this way I have prepared the solution.

But still, I am mentioning how identified constraints can impact my design, and what are recommendations to handle them.

My Device Specification

I3, 11th Generation

8GB RAM, 256 GB SSD

Platform : Google Colab

Memory Usage of training dataset : 1.5 MB

Memory Usage of test dataset : 277 KB

Memory Usage during training : 10709 MB

Memory Usage during prediction : 10685 MB

Time Taken to execute the code: 24 Sec.

Identified Constraints:

1. Memory Usage: The code uses various libraries and data structures, which can lead to high memory consumption, especially during model training and prediction. The memory usage is measured using the 'memory_profiler' library. Depending on the dataset size and available system memory, this could be a potential constraint, especially when dealing with large datasets. It may lead to slower execution or even cause the program to crash if the memory limit is exceeded.

2. Training Time: Decision trees and ensemble methods like 'ExtraTreesClassifier' can be computationally expensive, especially with large datasets or complex feature engineering. The 'time' library is used to measure the training time, which can impact the overall efficiency of the predictive model.

3. Feature Importance: The code also measures feature importances using 'ExtraTreesClassifier'. However, it is essential to ensure that the selected features are truly relevant and contribute significantly to the predictive power of the model. Incorrect feature selection or ignoring important features can lead to suboptimal performance.

4. Data Preprocessing: The code converts the 'DateTime' column into various time-related features, which can be beneficial for modeling. However, the data preprocessing steps need to be carefully examined to ensure that they are correctly handled and do not introduce any bias or errors in the dataset.

Recommendations:

1. Memory Optimization: To handle memory constraints, consider using more memory-efficient data structures or reducing the dataset size, if possible. Additionally, try to free up memory by removing unnecessary variables from memory after their use. Efficiently handling data and minimizing unnecessary copies can help optimize memory usage.

2. Feature Selection: Before training the model, perform a thorough feature selection process to identify the most relevant features. Techniques like recursive feature elimination, feature importance ranking, or univariate feature selection can be employed. This will not only improve the model's performance but also reduce computation time and memory consumption.

3. Use Efficient Algorithms: Decision trees and ensembles can be resource-intensive. Consider exploring other machine learning algorithms that are computationally less expensive, such as linear models or light gradient boosting machines (LightGBM/XGBoost). These algorithms can provide comparable performance while being faster and less memory-demanding.

4. Data Sampling: If the dataset is large and memory constraints are severe, consider using data sampling techniques like random sampling or stratified sampling. This will create a smaller representative subset of the data, which can be used for model training and preliminary analysis.

5. Model Evaluation: Assess the model's performance thoroughly using appropriate evaluation metrics like accuracy, precision, recall, F1-score, or mean absolute error (MAE) for regression tasks. Cross-validation can help provide a more robust estimate of the model's generalization performance.

6. Feature Engineering: Continuously explore and experiment with different feature engineering techniques to enhance the predictive power of the model. Consider domain knowledge and incorporate additional relevant features to capture complex patterns in the data.

7. Scaling: If applicable, apply feature scaling (e.g., StandardScaler) to ensure that features are on a similar scale. This can improve the convergence and stability of some machine learning algorithms.

8. Monitoring Resource Usage: Keep track of the system's resource usage during model training and prediction. Monitoring CPU and memory usage can help identify potential bottlenecks and prevent system instability.

9. Cloud Computing: For large-scale datasets and resource-intensive tasks, consider leveraging cloud computing platforms that offer scalable resources. Cloud-based solutions can handle complex computations and alleviate memory constraints.

By addressing these constraints and implementing the recommendations, you can build a more efficient and accurate traffic prediction model, enhancing its usability and performance in real-world scenarios.

7 My learnings

Internship Learning Report

During my internship, I had the opportunity to explore various topics related to data science, machine learning, artificial intelligence, and data engineering. I gained knowledge from different sources, such as Wikipedia, YouTube, e-books, webinars, and the internship content provided. Here's a comprehensive report on my learnings during the internship:

1. Introduction to Data Science*:

I learned about the significance of data science in solving complex problems by combining mathematics, statistics, programming, and domain knowledge. Exploratory Data Analysis (EDA) and data cleaning techniques were highlighted as critical steps in the data science process.

2. Introduction to Machine Learning:

Machine learning algorithms and their applications, including supervised and unsupervised learning techniques, were introduced. I gained an understanding of model evaluation and validation, big data, and distributed computing technologies like Hadoop and Spark.

3. Electric Vehicles (EV):

I attended a webinar on EVs, where I learned about their architecture, components, technologies used, status in India, and challenges in the EV market.

4. Introducing Data Science (e-book):

This comprehensive guidebook covered various concepts in data science. Key takeaways included data science definition, exploratory data analysis, data cleaning, and pre-processing, as well as machine learning algorithms and model evaluation techniques.

5. The Unmatched Impact of Big Data on Business:

I understood how big data influences businesses, leading to data-driven decision-making, predictive analytics, and better customer insights across various industries like retail, e-commerce, and manufacturing.

6. Pros and Cons of Being a Data Analyst vs. Data Scientist:

I gained insights into the roles and responsibilities, skills, and tools used by data analysts and data scientists. This understanding helped me compare the two career paths and assess their respective pros and cons.

7. Lessons Learned from "An Introduction to Probability and Statistics":

I learned about the importance of probability theory, descriptive statistics, statistical inference, regression analysis, and other statistical concepts for data analysis, model building, and prediction.

8. AI vs. Data Science: Mapping Your Career Path:

This topic explored the relationship between AI and data science, their applications, and career opportunities. I understood the in-demand job roles and salary considerations for both fields.

9. Top 5 Data Scientist Skills for 2023:

I learned about the essential technical skills, data science certifications, career opportunities, and continuous learning required to excel as a data scientist in 2023.

10. Linear Algebra and Linear Functions in Data Science:

I gained insights into the fundamental role of linear algebra and linear functions in data representation, analysis, and prediction, including their applications in machine learning and neural networks.

11. Optimization Techniques in Machine Learning:

I understood the central role of optimization in machine learning and the use of various techniques like gradient descent for model parameter tuning.

12. Preparing for Data Science Interviews:

I learned about the skills, interview process, and common topics in data science interviews, and some very important and frequently asked questions which will help me in my future job search.

13. Application of Data Science and ML Concepts: I gained hands-on experience in applying data science and machine learning concepts to solve real-world challenges. From data preprocessing to model implementation, I learned how to tackle complex problems systematically and derive meaningful insights from data.

14. Exploration of Advanced Algorithms: The internship exposed me to a variety of machine learning algorithms like Random Forest, Decision Tree, and Support Vector Machine. Understanding the strengths and weaknesses of each algorithm helped me make informed decisions in selecting the most suitable approach for the "Smart City Traffic Patterns Detection" project.

15. Data Analysis and Visualization: I learned the importance of data analysis and visualization in deriving meaningful patterns and trends from data. Visualizing real and predicted traffic patterns at various junctions helped me communicate insights effectively and aided in making informed decisions.

16. Project Management and Planning: Working on a project with multiple phases taught me the significance of project management and planning. Breaking down the project into manageable tasks and setting realistic milestones ensured smooth progress and timely completion.

17. Critical Thinking and Problem-Solving: The project involved several challenges that required critical thinking and problem-solving skills. I learned to approach problems analytically, experiment with different solutions, and adapt my strategies when necessary.

18. Collaboration and Teamwork: During the internship, I had the opportunity to collaborate with experienced professionals and fellow interns. Working in a team environment taught me the value of effective communication, sharing ideas, and collective problem-solving.

19. Continuous Learning and Adaptability: The field of AI and ML is ever-evolving, and this internship emphasized the importance of continuous learning and adaptability. I developed a mindset of staying updated with the latest advancements in the industry to remain relevant and innovative.

20. Git and GitHub:

I attended a session on Git and GitHub, where I learned the basics of version control, creating repositories, uploading files, and using common Git commands.

Overall, this internship has been an enriching experience, helping me develop a solid foundation in data science, machine learning, and related fields. I am now better equipped to pursue a career in these domains and make data-driven decisions in solving real-world challenges. I am grateful for the diverse sources of knowledge provided during the internship, and I look forward to applying these skills in future projects and opportunities.

8 Future work scope

Smart city traffic pattern detection is essential for optimizing traffic management, resource utilization, environmental impact, safety, emergency response, and overall urban planning. It directly contributes to a more sustainable and efficient urban environment, ensuring the well-being and convenience of city dwellers. So, I worked on this project.
While the project was successful in building a predictive model using Decision Trees, there are several areas that could be explored in the future to enhance the project's capabilities and provide more comprehensive insights:

Other Ensemble Models: Investigate the use of ensemble techniques, such as Gradient Boosting, to further improve prediction accuracy by combining multiple decision trees.

Time Series Analysis: Explore time series forecasting models, like ARIMA or Prophet, to handle traffic patterns' temporal nature more effectively and capture seasonality and trends.

Data Augmentation: Consider augmenting the existing dataset with external data sources, such as social events or road construction information, to capture more factors influencing traffic patterns.

Feature Selection Techniques: Apply feature selection methods, such as Recursive Feature Elimination or LASSO regression, to identify the most relevant features for traffic prediction, reducing model complexity and enhancing interpretability.

Model Hyperparameter Tuning: Perform an extensive hyperparameter tuning process to optimize the decision tree model and prevent overfitting.

Predictive Analytics Dashboard: Develop an interactive dashboard that visualizes traffic predictions, allowing city authorities to make data-driven decisions for efficient traffic management.

Conclusion: The Smart City Traffic Patterns Prediction project successfully implemented a decision tree model to forecast traffic conditions in a smart city environment. Future work scope includes exploring some other ensemble models and conducting a more extensive algorithm comparison. By addressing these aspects, the project can offer more accurate predictions and valuable insights, ultimately contributing to the effective management and planning of traffic in smart cities.