Industrial Internship Report on

” Smart City Traffic Forecasting”

Prepared by

Aryan Dhull

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 Forecasting 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

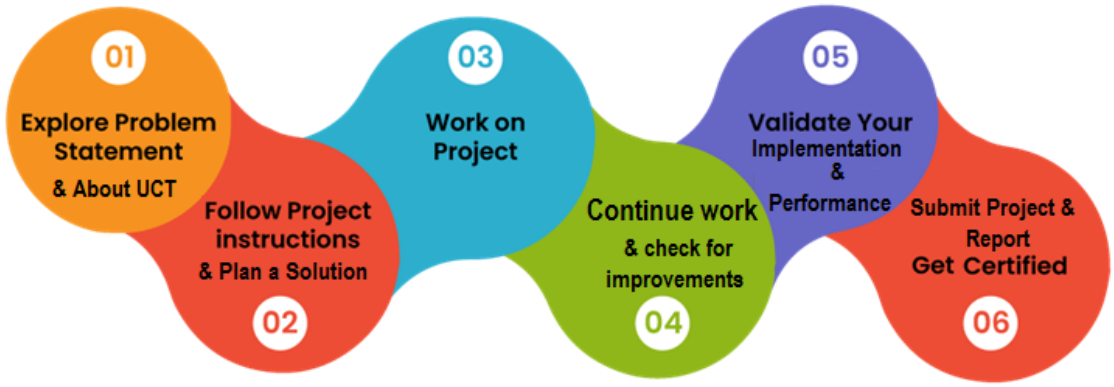
The project aims to predict future traffic patterns based on historical data. It begins by reading the training dataset and performing feature engineering, extracting relevant information such as date, time, day of the week, month, and year from the DateTime column. The project then explores the data through various visualizations and statistical analyses to gain insights into the traffic patterns.

Next, the Project uses python libraries to splits the data into training and testing sets and trains four different machine learning models: Linear Regression, Decision Tree Regression, Random Forest Regression, and Neural Network Regression. The performance of each model is evaluated using mean absolute error (MAE) and root mean squared error (RMSE) metrics.

The project also reads a separate dataset containing future dates and applies the trained models to forecast the traffic patterns for these future dates. It visualizes the predicted traffic patterns for each junction and provides average traffic by day of the week, by month, and by junction.

In summary, the project provides an initial approach to predict future traffic patterns using machine learning models. It performs data preprocessing, model training, and evaluation, and provides visualizations for analyzing the data and model predictions

How Program was planned



Thanks to Kaushlendra Sir for providing me the opportunity to work on this project and gain valuable experience in the field of data science and machine learning.

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.

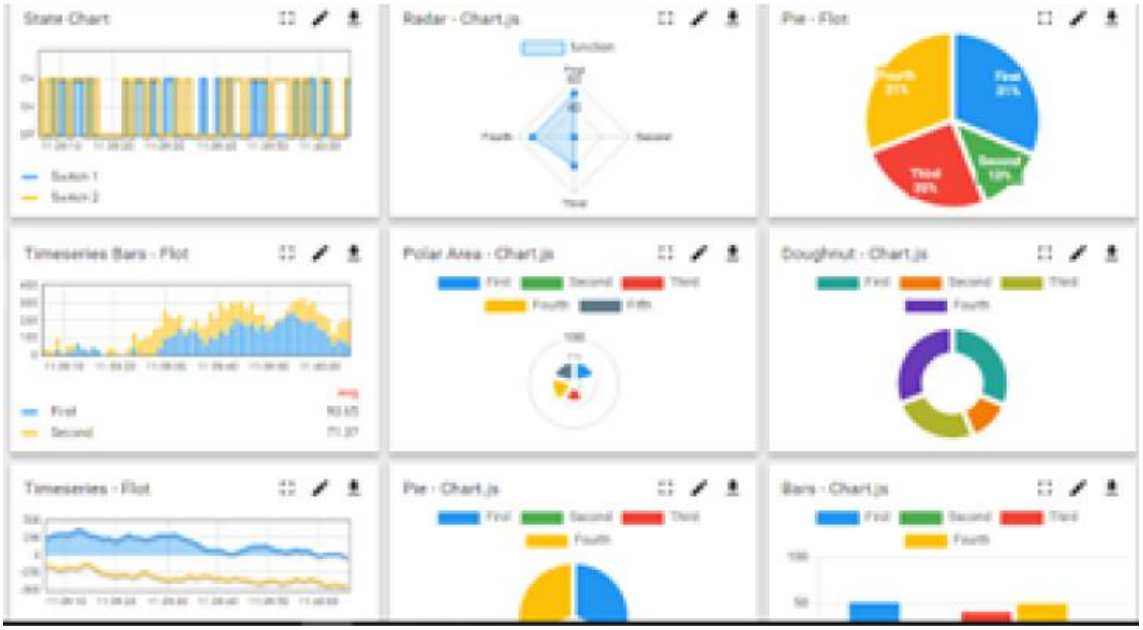
It has features to

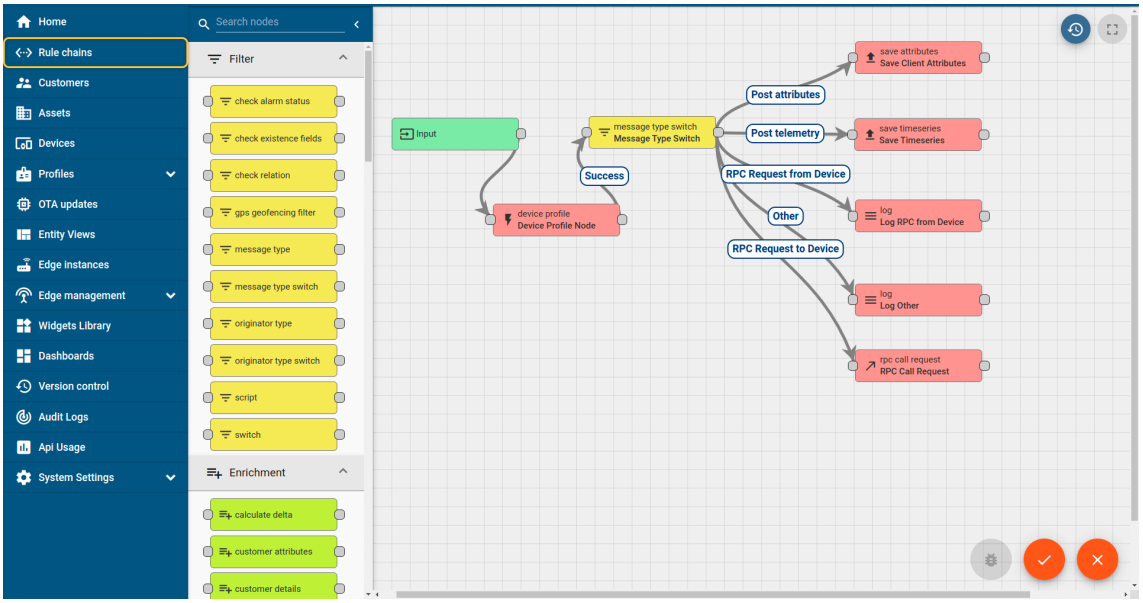
• Build Your own dashboard

• 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.

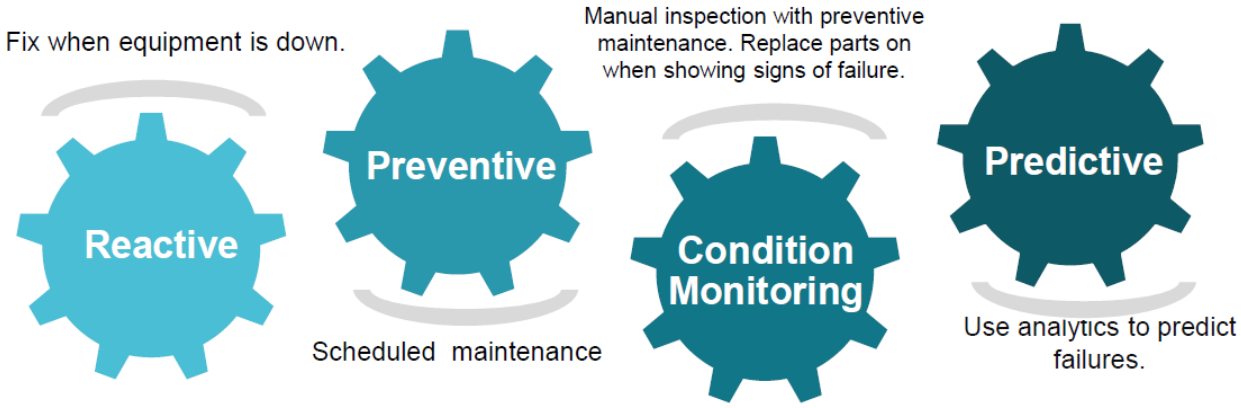
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	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 technology and providing solution in Aggrotech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

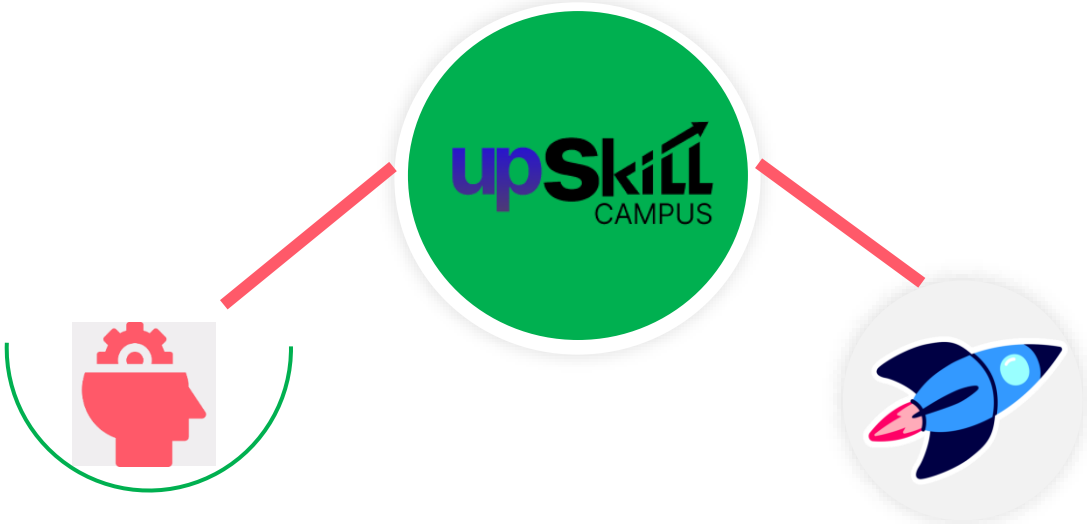
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

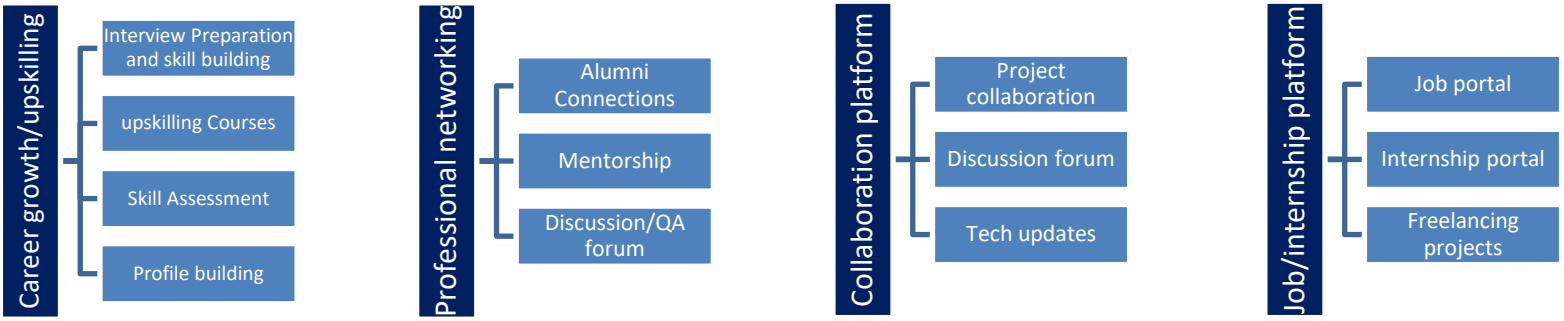
upskill Campus along with The IoT Academy and in association with UniConverge technologies has facilitated the smooth execution of the complete internship process.

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



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] Linear Regression in Machine learning - GeeksforGeeks

[2] Decision Tree - GeeksforGeeks

[3] Machine Learning Random Forest Algorithm - Javatpoint

2.6 Glossary

Terms	Acronym
M**AE**	M**ean Absolute Error**
RMSE	Root Mean Squared Error
EDA	Exploratory Data Analysis
CSV	Comma Separated Values

3 Problem Statement

The government is working to transform various cities into smart cities. A smart city is a city that uses information and communication technologies (ICT) to improve the quality of life for its citizens. This includes things like improving traffic management, providing better public transportation, and making it easier for citizens to access government services.

One of the problems that cities face is traffic congestion. Traffic congestion can lead to delays, pollution, and accidents. The government wants to implement a robust traffic system for the city that will help to reduce traffic congestion.

You are a data scientist who is working to help the government manage traffic better. You will need to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays and other special occasions will differ from traffic patterns on normal working days. This is important to take into account when you are forecasting traffic patterns.

For example, on holidays, there may be more traffic as people travel to visit family and friends. On special occasions, such as sporting events or concerts, there may be even more traffic. You will need to factor these factors into your forecasting model so that you can provide accurate predictions of traffic patterns.

4 Existing and Proposed solution

By understanding the traffic patterns of the city, one can help the government to implement a robust traffic system that will reduce traffic congestion and improve the quality of life for citizens.

Here are some specific things that one can do to help the government manage traffic better:

- Analyse traffic data to identify patterns and trends.
- Develop models to forecast traffic patterns.
- Recommend changes to traffic signals and other infrastructure.
- Work with the government to implement these changes.

By taking these steps, one can help the government to make the city a more liveable place.

4.1 Code submission (GitHub link)

Aryan-Dhull/Smart_CIty_Traffic_Forecasting (github.com)

5 Proposed Design/ Model

The proposed design/model in the provided code is to predict future traffic patterns using different machine learning models. The code implements the following models for prediction:

Linear Regression: It is a simple linear model that assumes a linear relationship between the input features and the target variable (traffic volume). The model is trained on the historical data and then used to predict future traffic patterns.

Decision Tree Regression: This model builds a decision tree based on the input features and their corresponding target values. It predicts the target variable by traversing the tree based on the feature values. The decision tree is trained using the historical data and utilized for future traffic predictions.

Random Forest Regression: Random Forest is an ensemble model that combines multiple decision trees to make predictions. It creates a set of decision trees with random subsets of features and averages their predictions to obtain the final result. The random forest model is trained on the historical data and used to forecast future traffic patterns.

Neural Network Regression: The neural network model consists of multiple layers of interconnected nodes (neurons) that perform computations to predict the target variable. The model utilizes the input features (hour, month, and year) to learn complex patterns and relationships in the data. It is trained on the historical data and employed for predicting future traffic patterns.

The proposed design/model involves feature engineering steps to extract relevant information from the DateTime column and derive additional features such as hour, month, and year. The data is split into training and testing sets, and the models are trained on the training data. The performance of each model is evaluated using mean absolute error (MAE) and root mean squared error (RMSE) metrics. Finally, the trained models are used to forecast future traffic patterns for different junctions and visualize the results.

6 Performance Test

The code provided includes various machine learning models for predicting future traffic patterns based on historical data. The performance of the models is evaluated using mean absolute error (MAE) and root mean squared error (RMSE) metrics. The performance of each model is measured on both the training and testing sets. Additionally, the code includes visualizations to compare the predicted traffic patterns with the actual data.

6.1 Test Plan/ Test Cases

- **Test Case 1: Data Loading and Feature Engineering:**

  Verify that the dataset is successfully loaded from the CSV file.

  Check if the feature engineering steps (e.g., converting DateTime columns, extracting date-related features) are correctly implemented.

- **Test Case 2: Exploratory Data Analysis (EDA):**

  Verify the correctness of the computed metrics, such as daily traffic, average traffic by day of the week, and traffic volume distribution.

  Check if the line plot of traffic volume over time for each junction is correctly displayed.

  Validate the accuracy of the correlation matrix calculation.

- **Test Case 3: T-Test:**

  Verify the correctness of the T-Test calculation and ensure the obtained p-value and t-statistic are within the expected range.

- **Test Case 4: Traffic Analysis:**

  Check if the calculated metrics for junction analysis, holiday analysis, seasonal analysis, and time-of-day analysis are accurate.

  Validate the accuracy of the correlation matrix heatmap.

- **Test Case 5: Model Training and Testing**

  Verify that the dataset is split into training and testing sets with the correct proportions.

  Check if the models (Linear Regression, Decision Tree, Random Forest, Neural Network) are successfully trained.

  Validate the performance metrics (MAE and RMSE) obtained for each model on both training and testing sets.

  Verify the correctness of the traffic forecast plots for different junctions and overall traffic.

- **Test Case 6: Performance Visualization**

  Check if the visualizations of average traffic by day of the week, by month, and by junction are correctly displayed.

6.2 Test Procedure

To execute the test plan, follow these steps:

- Step 1: Load the code and required libraries.
- Step 2: Verify the data loading and feature engineering steps.
- Step 3: Inspect the generated visualizations and metrics for exploratory data analysis.
- Step 4: Check the accuracy of the T-Test calculations.
- Step 5: Verify the correctness of the traffic analysis metrics and correlation analysis.
- Step 6: Validate the model training and testing steps, along with the performance metrics.
- Step 7: Check the accuracy of the traffic forecast plots.
- Step 8: Verify the correctness of the performance visualizations.

6.3 Performance Outcome

The performance outcome is measured using mean absolute error (MAE) and root mean squared error (RMSE) metrics. The code provides the training and testing MAE and RMSE values for each model (Linear Regression, Decision Tree, Random Forest, Neural Network). These metrics quantify the accuracy of the models in predicting the traffic volume. Additionally, the code includes various visualizations to compare the predicted traffic patterns with the actual data.

7 My learnings

By analyzing this code, you can learn the following:

- How to load and preprocess data using pandas library.
- How to perform feature engineering to extract relevant information from datetime columns.
- How to perform exploratory data analysis (EDA) using various visualizations and statistical tests.
- How to train and evaluate machine learning models for regression tasks using scikit-learn library.
- How to interpret and analyze the performance of different models using evaluation metrics.
- How to visualize and interpret the traffic patterns based on different factors such as time, day of the week, and junction.

8 Future work scope

Based on the provided code, some potential future work scope includes:

- Further fine-tuning and optimization of the existing machine learning models to improve their performance.
- Exploring other regression models and algorithms to compare their performance with the existing models.
- Incorporating additional features or external factors that might influence traffic patterns (e.g., weather data, holidays, special events).
- Conducting more in-depth statistical analysis and feature engineering to identify significant patterns and correlations in the data.
- Implementing a comprehensive test suite to ensure the correctness and robustness of the code across different scenarios and datasets.
- Deploying the trained models into a production environment or integrating them into a larger traffic management system for real-time predictions and decision-making.