Industrial Internship Report on

1. Prediction of Agriculture Crop Production in India
2. Forecasting of Smart City Traffic Patterns

Prepared by

Ariz Ejaz Khan

Executive Summary

This report documents my industrial internship with **upskill Campus** and **The IoT Academy** in collaboration with **UniConverge Technologies Pvt Ltd (UCT)**.
The internship involved solving real-world problems using **Data Science and Machine Learning**, specifically focusing on two projects:

1. **Prediction of Agriculture Crop Production in India**
2. **Forecasting of Smart City Traffic Patterns**

Over the course of 6 weeks, I gained practical exposure to handling complex datasets, applying machine learning models, and interpreting results.

- **Prediction of Agriculture Crop Production in India** – aimed at building a system that can analyze historical agricultural datasets, handle inconsistencies in multiple data sources, and accurately forecast crop yields. This project is important for food security and policy-making, as reliable predictions can assist in planning and resource allocation.
- **Forecasting of Smart City Traffic Patterns** – aimed at predicting the number of vehicles at city junctions using time-series and categorical features. This work contributes to urban development by supporting better traffic management, reducing congestion, and improving mobility in smart cities.

This internship provided hands-on experience with real industrial challenges, significantly enhanced my technical expertise, and gave me confidence in applying data science for impactful problem solving.

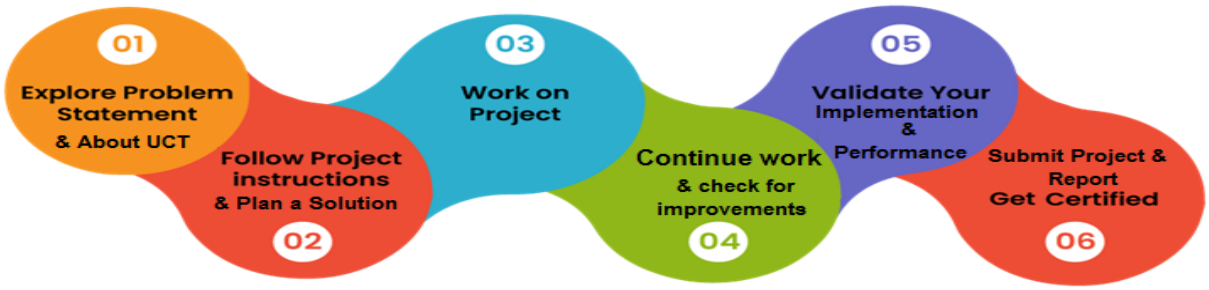
TABLE OF CONTENTS

1 Preface

This internship has been a valuable journey in connecting academic learning with real-world applications. I worked on two projects: **Agriculture Crop Production Prediction** and **Smart City Traffic Forecasting**, which gave me the opportunity to apply data science and machine learning techniques to practical problems.

Through these projects, I gained hands-on experience in **data preprocessing, exploratory analysis, feature engineering, and model building**. I also explored ensemble methods like **RandomForest** and boosting algorithms such as **LightGBM**, which proved effective for prediction tasks.

I am grateful to **upskill Campus**, **The IoT Academy**, and **UniConverge Technologies Pvt Ltd (UCT)** for providing this opportunity and for their continuous support throughout the program.



This internship has enhanced my technical expertise, problem-solving skills, and confidence in handling industrial datasets. I believe the skills and experiences gained will play a crucial role in shaping my career in data science.

Finally, I would like to extend my thanks to my family, friends, and colleagues who supported me throughout this journey. To my juniors and peers, I encourage you to take such opportunities seriously, as internships like this are invaluable stepping stones to building a strong professional foundation.

To my peers and juniors, I encourage you to take internships like this seriously and view them as opportunities to grow beyond academics. Stay curious, be open to learning new tools and techniques, and don’t hesitate to ask for guidance. Consistency and hands-on practice will not only build your technical skills but also prepare you for real-world challenges.

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.

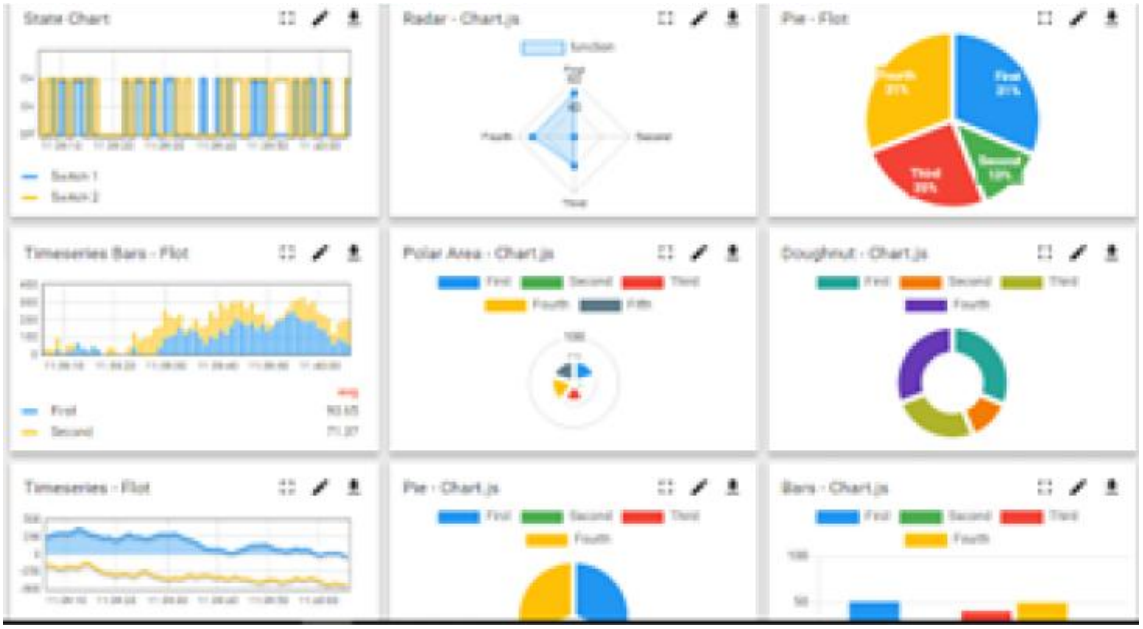
i. UCT IoT Platform (UCT Insight)

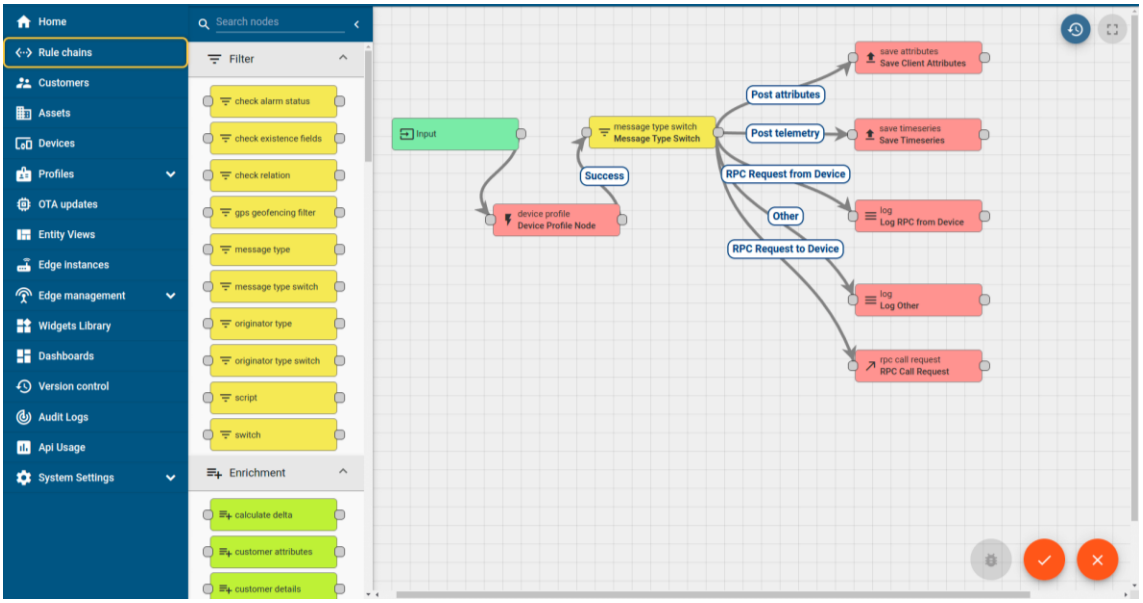
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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.

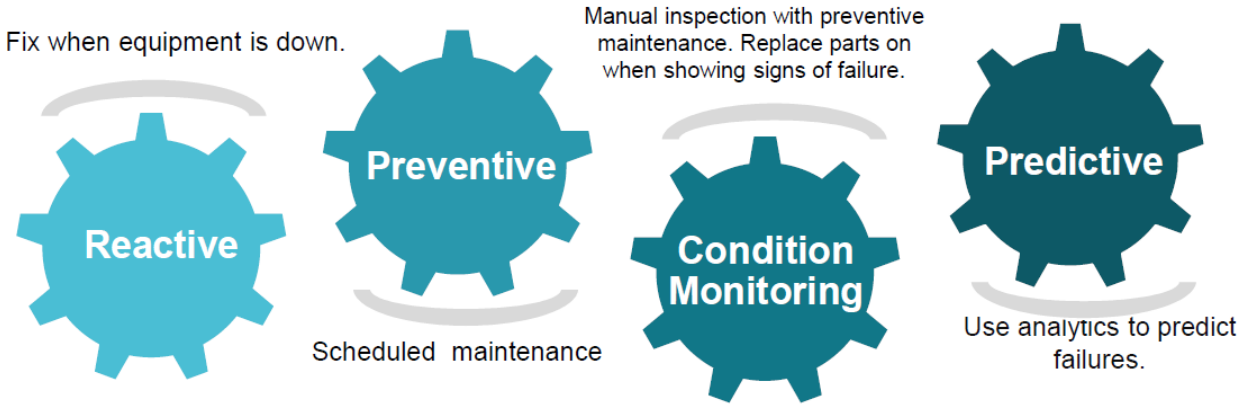
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1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

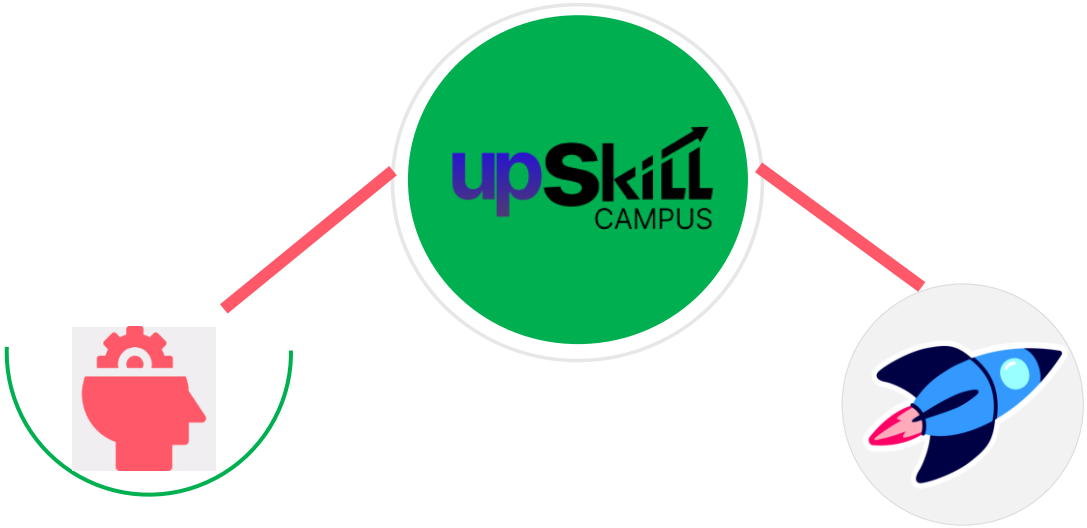
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

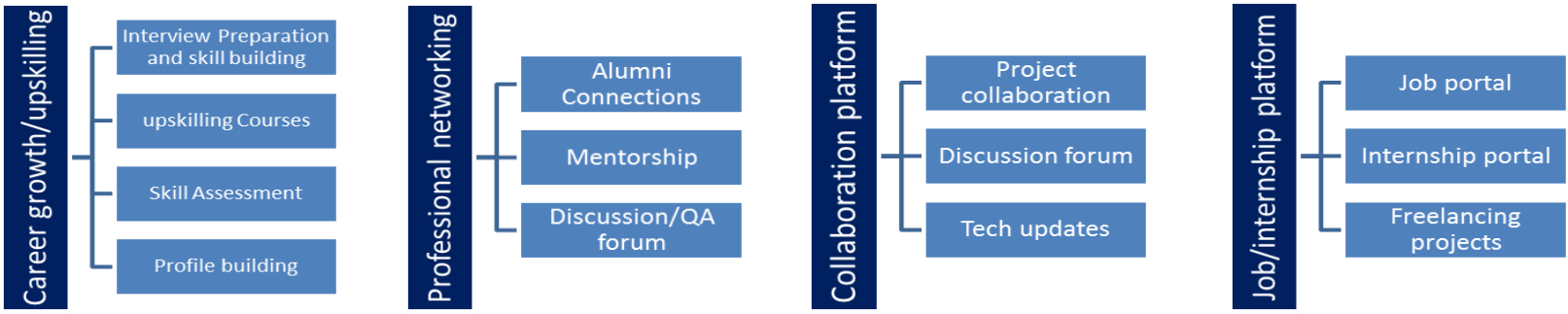
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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 Reference

[1] Scikit-learn Documentation- https://scikit-learn.org/stable/

[2] Kaggle Datasets and Notebooks- https://www.kaggle.com/

[3] LightGBM Documentation- https://lightgbm.readthedocs.io/en/stable/

[4] Pandas Documentation – https://pandas.pydata.org/docs/

[5] NumPy Documentation – https://numpy.org/doc/

[6] Seaborn Documentation – https://seaborn.pydata.org/

[7] Matplotlib Documentation – https://matplotlib.org/stable/contents.html

[8] UCT & Upskill Campus Internship Materials and Project Guidelines

2.6 Glossary

Terms	Acronym
Exploratory Data Analysis	**EDA**
Root Mean Squared Error	**RMSE**
Light Gradient Boosting Machine	**LightGBM**
Random Forest	**RF**
Machine Learning	**ML**
Artificial Intelligence	**AI**
Internet of Things	**Iot**

3 Problem Statement

- **Prediction of Agriculture Crop Production in India**
  Accurate forecasting of crop production is crucial for food security, efficient resource allocation, and agricultural planning. The challenge lies in the variability of agricultural data across crops, seasons, and regions, which makes prediction complex. The problem statement was to build a system that can analyze agricultural datasets and provide reliable forecasts of crop production.

- **Forecasting of Smart City Traffic Patterns**
  Urban cities face increasing challenges of traffic congestion, leading to delays, higher fuel consumption, and reduced quality of life. To support smart city initiatives, the task was to develop a system that can forecast vehicle counts at city junctions based on historical traffic data, enabling better planning and congestion management.

4 Existing and Proposed solution

- **Existing Solutions**
    - Agriculture crop forecasting often relies on **statistical methods** (e.g., linear regression, ARIMA), but these methods fail to capture complex, non-linear interactions.
    - Traffic forecasting is often attempted with **basic time-series models**, which do not scale well to large datasets or categorical junction-level data.
- **Proposed Solution**
    - For **Agriculture**, I proposed using **RandomForest Regression** because it can handle mixed data types, nonlinear relationships, and high-cardinality features.
    - For **Traffic**, I proposed using **LightGBM** (a boosting algorithm) and **RandomForest** for comparative analysis, as both can capture complex dependencies and handle time-based features effectively.

4.1 Code submission (Github link)

https://github.com/Ariz253/upskillcampus/blob/main/CropPrediction.ipynb.ipynb

https://github.com/Ariz253/upskillcampus/blob/main/TrafficPrediction.ipynb.ipynb

4.2 Report submission (GitHub link):

https://github.com/Ariz253/upskillcampus/blob/main/CropTrafficPrediction_Ariz_Ejaz_Khan_USC_UCT.pdf

5 Proposed Design/ Model

5.1 High Level Diagram (if applicable)

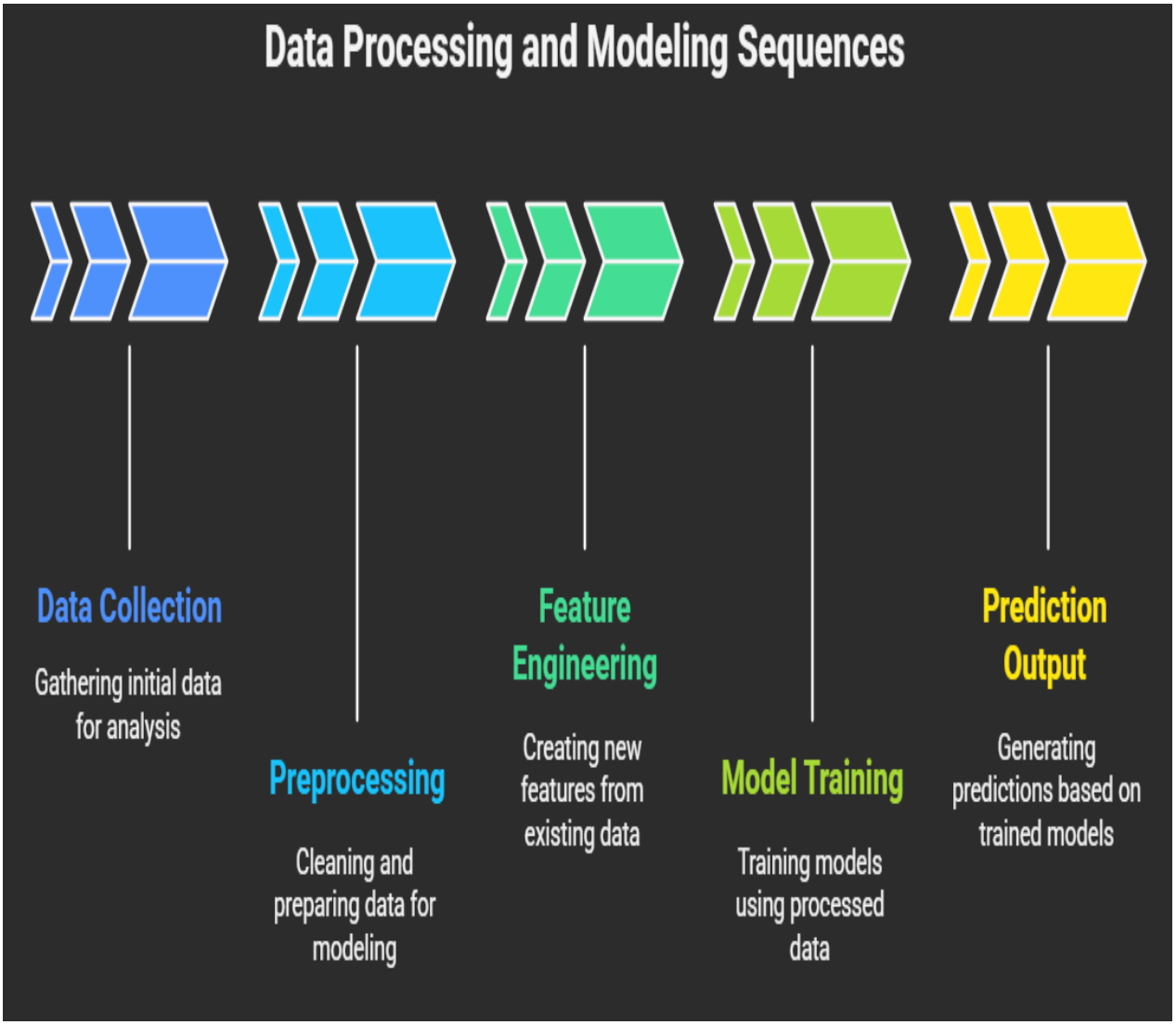
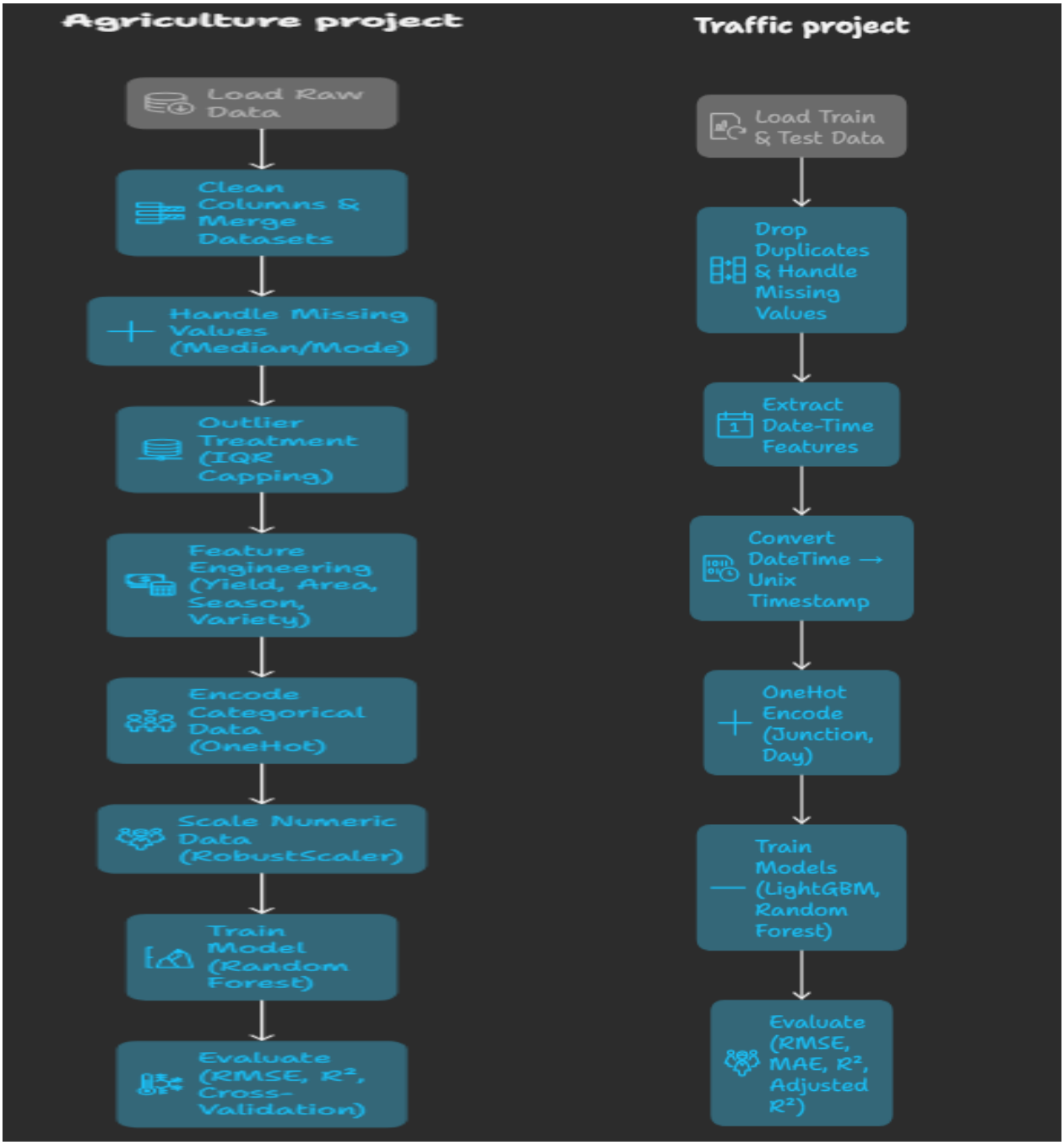


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram (if applicable)



5.3 Interfaces (if applicable)

For both projects, the primary interfaces were between the **datasets, preprocessing pipeline, models, and outputs**.

- **User–System Interface**:
    - The system was implemented in Python (Jupyter Notebooks).
    - The user interacts with the system through code execution, dataset upload, and visualization outputs.
- **Data Interfaces**:
    - **Agriculture Project**: CSV files containing production, yield, cost, and crop information served as inputs. The system produced cleaned datasets and forecasted production values as outputs.
    - **Traffic Project**: CSV files containing traffic counts at junctions served as inputs. The system output predicted vehicle counts for future timestamps.
- **Model Interfaces**:
    - For agriculture, the RandomForest model served as the predictive interface, accepting preprocessed crop features and returning estimated production values.
    - For traffic, LightGBM and RandomForest models served as interfaces, mapping historical traffic patterns to future forecasts.
- **Visualization Interfaces**:
    - Graphs, plots, and performance metrics provided feedback to the user about model accuracy and data patterns.

6 Performance Test

6.1 Test Plan/ Test Cases

For both projects, the test plan was designed to evaluate model performance on unseen data.

- **Agriculture Project**: The dataset was split into **training and testing sets (80:20)**. Test cases focused on predicting crop production values and comparing them to actual outcomes. Cross-validation (5-fold) was also used to validate stability.
- **Traffic Project**: The dataset was split into **training and testing sets (67:33)**. Test cases included forecasting vehicle counts for different junctions and verifying accuracy across metrics such as RMSE, MAE, and R^2.

6.2 Test Procedure

• Preprocessed and cleaned datasets were fed into the models.

• For the **Agriculture Project**, RandomForest Regressor was trained on the training set, and predictions were generated for the test set. Performance was evaluated using RMSE and R^2.

• For the **Traffic Project**, two models (LightGBM and RandomForest) were trained. Predictions for vehicle counts were generated and compared with actual test values. Evaluation metrics included RMSE, MAE, R^2, and Adjusted R^2.

• Cross-validation was applied to ensure the models generalized well and to reduce overfitting risk.

6.3 Performance Outcome

- **Agriculture Project**:
  - RandomForest achieved consistent results with low RMSE and high R^2.
  - Cross-validation confirmed that the model performed reliably across multiple folds, showing strong generalization.
- **Traffic Project**:

- LightGBM achieved lower RMSE and MAE compared to RandomForest, indicating higher accuracy in forecasting traffic.
- RandomForest performed slightly worse but provided interpretable feature importance.
- Adjusted R^2 confirmed that the models captured a significant portion of the variance in traffic data.

Overall, the performance outcomes demonstrated that the proposed models were effective in addressing the problem statements and can serve as a baseline for further improvements.

7 My learnings

This internship provided me with a valuable opportunity to enhance both my technical expertise and professional skills.

From a **technical perspective**, I gained hands-on experience in:

- Working with **large, real-world datasets** that required extensive cleaning and preprocessing.
- Applying **Exploratory Data Analysis (EDA)** to identify patterns, outliers, and key insights.
- Performing **feature engineering** to transform raw data into meaningful predictors.
- Implementing and evaluating machine learning models such as **Random Forest** and **LightGBM**.
- Using performance metrics like **RMSE, MAE, R^2, and Adjusted R^2** to assess model effectiveness.
- Comparing different models to select the most suitable one for a given problem.

From a **professional development perspective**, I learned to:

- Approach real-world problems systematically, breaking them down into smaller, solvable tasks.
- Manage time effectively to complete weekly goals within deadlines.
- Document progress clearly through weekly reports, ensuring consistent communication of results.
- Adapt to challenges such as inconsistent data formats, high computational costs, and the need for efficient solutions.

Overall, this internship not only strengthened my foundations in **data science and machine learning** but also improved my ability to work independently, think critically, and deliver solutions aligned with industry needs.

8 Future work scope

Although the internship projects were successfully completed within the given timeline, there are several areas where the work can be extended and improved in the future:

- **Advanced Modeling Approaches**: For the agriculture project, experimenting with algorithms such as **XGBoost, CatBoost, or Neural Networks** could further improve prediction accuracy. For the traffic project, deep learning models like **LSTM or GRU** could better capture time-series dependencies.
- **Inclusion of External Data**: Both projects can be enhanced by integrating additional features. For agriculture, data on **weather, soil quality, and rainfall** could provide richer insights. For traffic, **weather conditions, holidays, or special events** could improve forecasting accuracy.
- **Real-Time Implementation**: Extending the traffic forecasting project into a **real-time prediction system** could support smart city operations by dynamically adjusting traffic lights or routing. Similarly, agriculture predictions could be connected to **dashboards for policymakers**.
- **Hyperparameter Tuning & Optimization**: Applying automated tuning techniques such as **Grid Search, Random Search, or Bayesian Optimization** could yield further improvements in model performance.
- **Deployment & Visualization**: Packaging the trained models into **interactive dashboards or APIs** would make the results more accessible and actionable for stakeholders.