Industrial Internship Report on

"Turbofan Engine Remaining Useful Life (RUL) Prediction"

Prepared by

Saket Chaudhary

Executive Summary
This report provides details of the Industrial Internship provided by **Upskill Campus** and **The IoT Academy** in collaboration with the 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, within 6 weeks. My project was "**Turbofan Engine Remaining Useful Life (RUL) Prediction**", which aimed to develop a predictive maintenance model capable of estimating the number of operational cycles remaining before a turbofan engine experiences failure. This project utilized historical sensor data from multiple engines, applied data preprocessing and feature engineering techniques, and implemented multiple machine learning models such as Random Forest, Gradient Boosting, and Support Vector Regression to determine the best-performing predictive model. This internship gave me an excellent opportunity to gain exposure to real-world industrial problems and design/implement a practical data-driven solution. I learned advanced machine learning techniques, performance evaluation metrics, and best practices in predictive maintenance. It was an overall great experience that enhanced my technical skills and industrial understanding.

TABLE OF CONTENTS

1 Preface

Over the past six weeks, I had the opportunity to work on an exciting and industry-relevant project — **Predicting the Remaining Useful Life (RUL) of Turbofan Engines** — as part of my industrial internship with **Upskill Campus (USC)** and **The IoT Academy**, in collaboration with **UniConverge Technologies Pvt. Ltd. (UCT)**.

This internship was a crucial step in my career development, giving me exposure to real-world industrial challenges and the skills to design and implement data-driven solutions. In today's competitive environment, such hands-on experience is essential for bridging the gap between academic learning and professional requirements.

The problem statement revolved around the **need for predictive maintenance in turbofan engines**, aiming to reduce unplanned downtime, improve operational efficiency, and cut maintenance costs. My task was to develop a **machine learning model** capable of predicting the number of operational cycles remaining before engine failure, using time-series sensor data from multiple operating conditions.

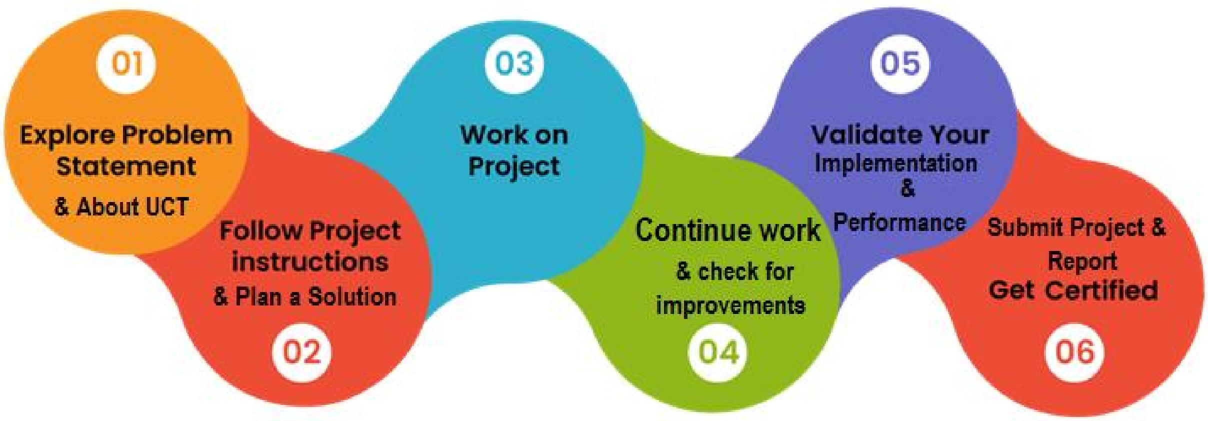
USC and UCT provided this valuable opportunity, structuring the program to include an initial orientation, technical guidance, and weekly progress reviews. The internship plan ensured a balance between **independent problem-solving** and **mentor support**, enabling me to apply classroom knowledge to a real-world industrial dataset.

Throughout this journey, I learned:

- Advanced **data preprocessing and feature engineering** for time-series data
- Comparative evaluation of multiple regression models
- Performance analysis using **RMSE, MAE, and R^2**
- The importance of domain understanding in building effective AI solutions

The overall experience was immensely enriching. I sincerely thank **Upskill Campus**, **The IoT Academy**, and **UCT mentors** for their continuous support. Special thanks to **Mentor** for technical guidance, and my peers for constructive discussions.

To my juniors and peers, I strongly recommend grabbing such opportunities to work on **industry-defined projects**. They not only enhance your technical abilities but also help you develop problem-solving skills, teamwork, and professional communication — all vital for your future career.



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)

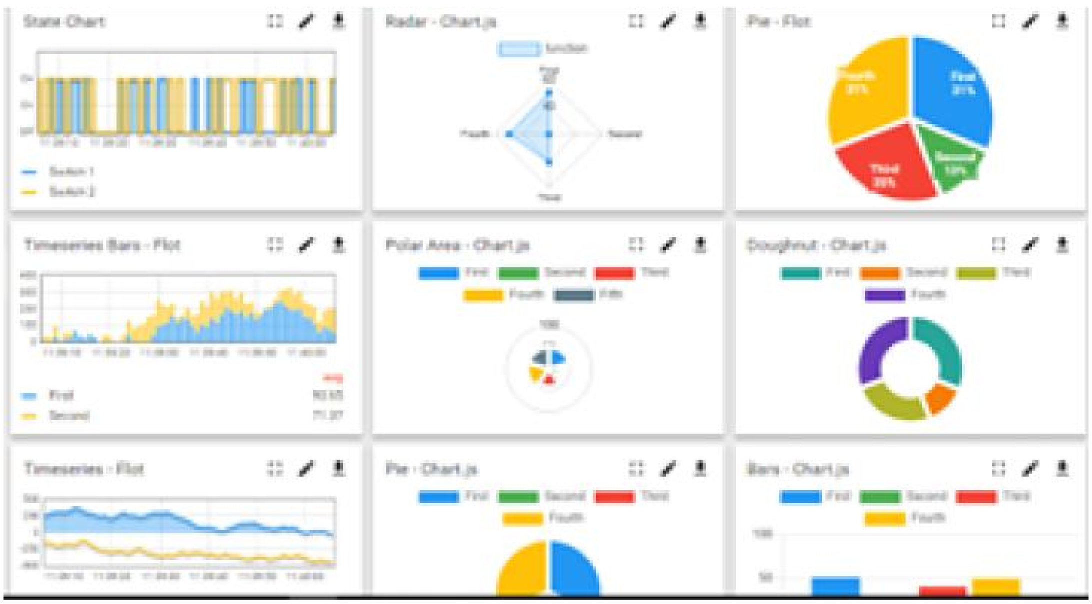
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

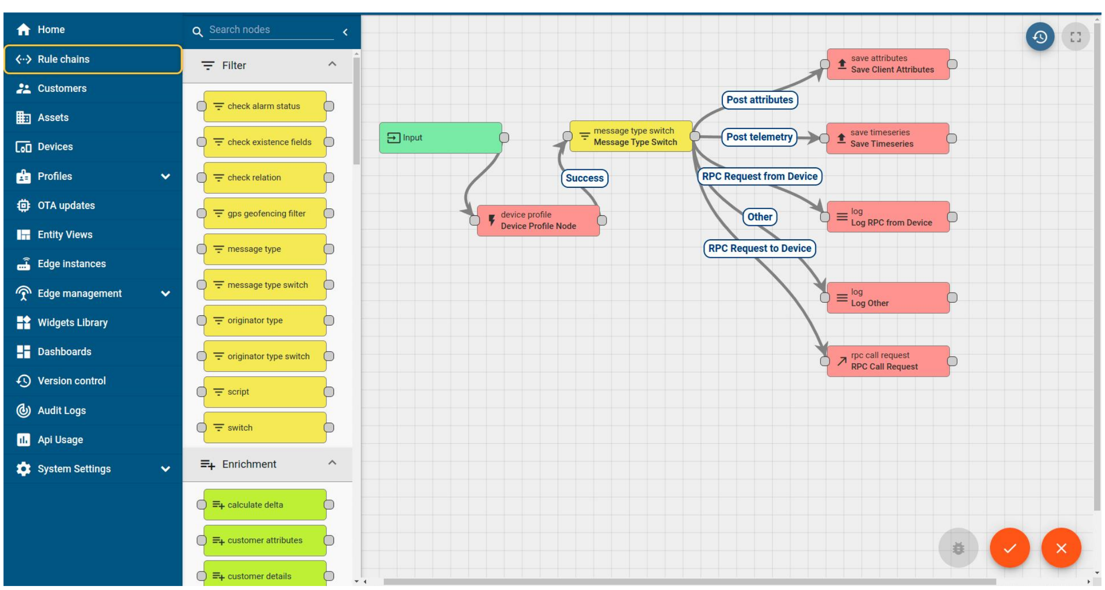
- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA

- 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



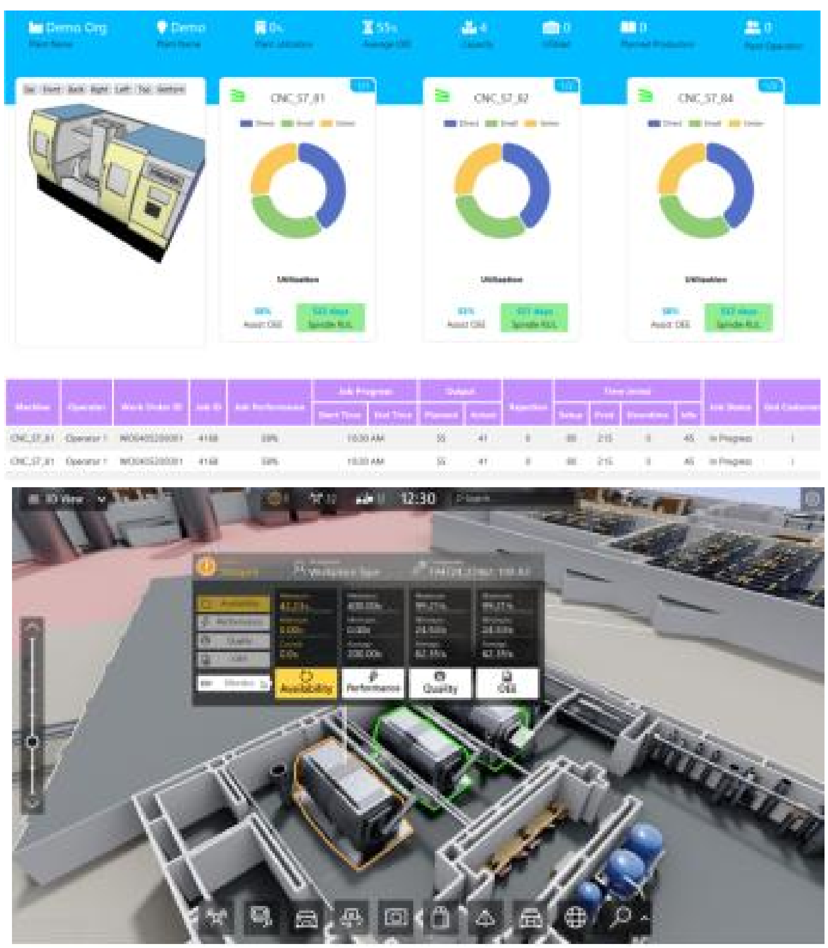


ii. Smart Factory Platform (FACTORY WATCH)

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 savetime, cost and money

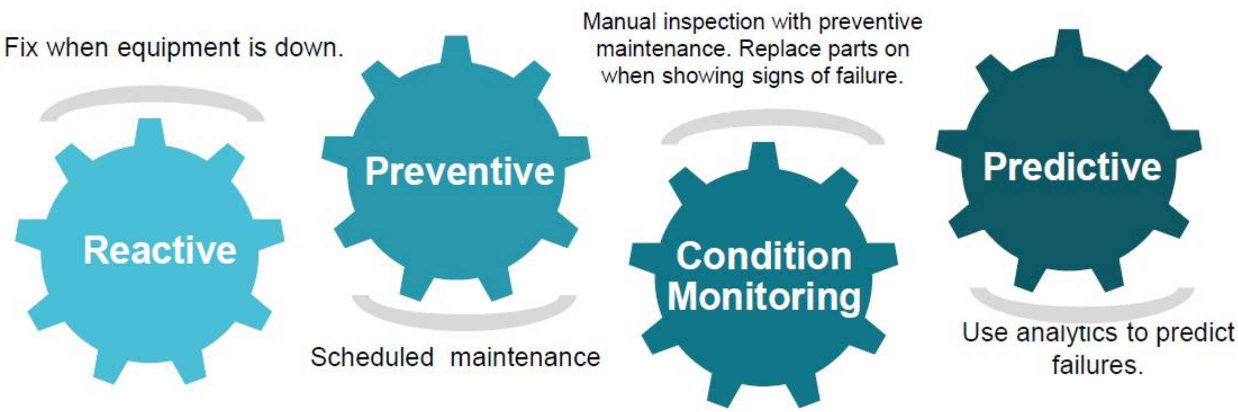


iii. LoRaWAN based Solution

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

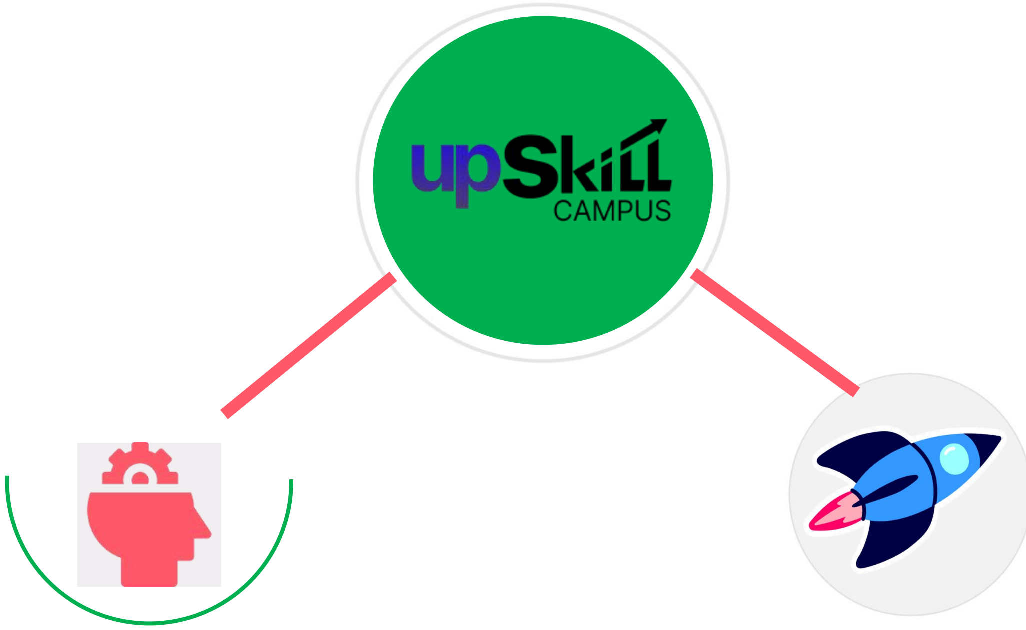
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



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.

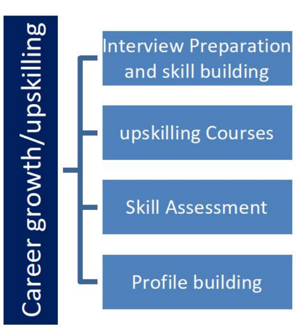
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.

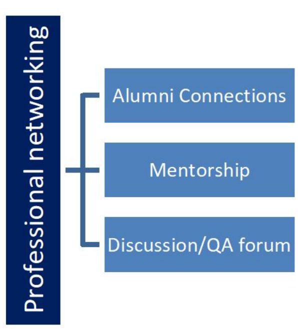


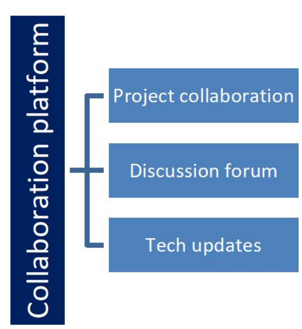
Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

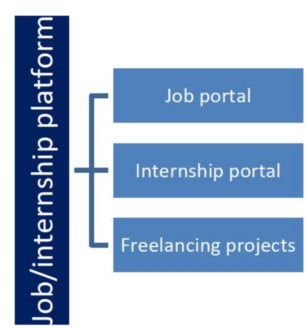
upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/









2.3 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.

2.4 Reference

Below are the key references and resources used during the project:

1. NASA Prognostics Center of Excellence — Turbofan Engine Degradation Simulation Dataset (C-MAPSS)
   https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository
2. Scikit-learn Documentation — Machine Learning in Python
   https://scikit-learn.org/stable/
3. Pandas Documentation — Data Analysis in Python
   https://pandas.pydata.org/docs/
4. Matplotlib & Seaborn — Data Visualization Libraries
   https://matplotlib.org/stable/index.html
   https://seaborn.pydata.org/
5. Research Paper: "A Comparison of Machine Learning Algorithms for Remaining Useful Life Prediction" — IEEE Xplore.

2.5 Glossary

Terms	Acronym
RUL (Remaining Useful Life)	The estimated time or number of operational cycles left before a machine fails or needs maintenance
Predictive Maintenance	Maintenance strategy that uses data-driven models to predict equipment failures before they occur.
C-MAPSS Dataset	A simulated turbofan engine dataset created by NASA for testing predictive maintenance algorithms.
Feature Engineering	The process of creating new input features from raw data to improve model performance.
RMSE (Root Mean Squared Error)	A metric to measure the accuracy of model predictions by penalizing larger errors.
MAE (Mean Absolute Error)	The average of absolute differences between predicted and actual values.
R^2 (Coefficient of Determination)	A statistical measure of how well predictions approximate real data values.
Scaling	Normalizing or standardizing data so that features contribute equally to the model's performance.
Rolling Mean / Rolling Standard Deviation	Moving average and variation calculated over a fixed window to smooth sensor readings.

3 Problem Statement

In the assigned problem statement, the task was to **predict the Remaining Useful Life (RUL) of a turbofan engine** based on time-series operational and sensor data.

Turbofan engines are critical components in the aviation industry, and their unexpected failure can lead to severe operational, financial, and safety consequences. Traditionally, maintenance schedules follow fixed intervals or reactive approaches, which can be inefficient and costly.

The goal of this project was to design a **predictive maintenance system** that can accurately estimate the number of operational cycles remaining before a turbofan engine fails. Using **historical performance data from multiple sensors** collected during different operational settings, the system needed to analyze degradation trends and predict future engine health.

The solution had to incorporate:

- **Data preprocessing and feature extraction** from raw time-series measurements
- **Machine learning algorithms** capable of modeling complex degradation patterns
- **Evaluation metrics** to measure prediction accuracy and reliability
- The capability to generalize to unseen test data for practical deployment

This problem required a combination of **domain understanding, data analysis, and machine learning expertise**, making it a challenging yet highly relevant industrial application in the field of **IoT-driven predictive analytics**.

4 Existing and Proposed solution

Existing Solutions

In the aviation industry, existing approaches to engine health monitoring and maintenance generally fall into two categories:

1. **Reactive Maintenance** – Repairs are carried out only after a fault has occurred. While this approach ensures that maintenance is performed only when needed, it often leads to **unexpected downtime**, **flight delays**, and **high replacement costs**.
2. **Preventive (Scheduled) Maintenance** – Maintenance is performed at fixed intervals based on operating hours or cycles. Although it reduces the risk of unexpected failures, it can result in **unnecessary servicing** of engines that are still in good condition, thereby increasing operational costs.

Some data-driven approaches have also been explored in academic research using regression models or simple statistical methods. However, their limitations include:

- Lack of **robust feature engineering** for time-series degradation data.
- Over-reliance on specific datasets, reducing generalization ability.
- Limited model comparison, leading to suboptimal performance selection.

Proposed Solution

My proposed solution involves building a **machine learning-based predictive maintenance model** for turbofan engines that can:

- Analyze historical sensor readings and operational settings.
- Extract statistical and rolling-window features to capture degradation trends.
- Compare multiple regression models (**Random Forest, Gradient Boosting, SVR, Neural Networks, Linear Regression**) to select the best performer.
- Ensure predictions are **non-negative** and generalize well to unseen data.

This approach ensures **higher prediction accuracy** and **better decision-making** for maintenance scheduling compared to traditional methods.

Value Addition

The proposed solution adds value by:

- Reducing **unplanned downtime** and **maintenance costs**.
- Enhancing **operational safety** through early fault detection.
- Providing a **scalable and adaptable** model for other industrial machinery beyond turbofan engines.

4.1 Code submission (Github link)

GitHub Code File Link:

https://github.com/Saket22-CS/upskillcampus/blob/main/SaketChaudhary_FD001_RUL_Prediction.ipynb

4.2 Report submission (Github link)

GitHub Report File Link:

https://github.com/Saket22-CS/upskillcampus/blob/main/TurbofanRULPrediction_Saket_USC_UCT.pdf

5 Proposed Design/ Model

The proposed solution follows a **structured pipeline** to process raw turbofan sensor data, extract meaningful features, train predictive models, and estimate the Remaining Useful Life (RUL) for each engine unit. The design flow ensures that every stage transforms the data towards an accurate and deployable prediction model.

Design Flow Steps

1. **Data Acquisition**

- Load training, testing, and actual RUL datasets provided by the CMAPSS turbofan dataset.

2. **Data Preprocessing**

- Remove unnecessary columns and handle missing values.
- Apply scaling using **StandardScaler** for normalization.

3. **Feature Engineering**

- Create **rolling mean** and **rolling standard deviation** features to capture degradation patterns.
- Use time-series grouping by unit_id for cycle-based statistics.

4. **Model Training and Selection**

- Train multiple regression models (**Random Forest, Gradient Boosting, SVR, Neural Network, Linear Regression**).
- Evaluate models using **RMSE, MAE, and R^2 score**.
- Select the **best-performing model** based on lowest RMSE.

5. **Prediction**

- Use the best model to predict the RUL for test data.
- Ensure predictions are non-negative.

6. **Evaluation and Visualization**

- Compare predictions with actual RUL values.
- Create **scatter plots, residual plots, and error histograms**.

7. **Deployment-Ready Output**

- Save predictions in CSV format.
- Upload final code and report to GitHub for submission.

5.1 High Level Diagram

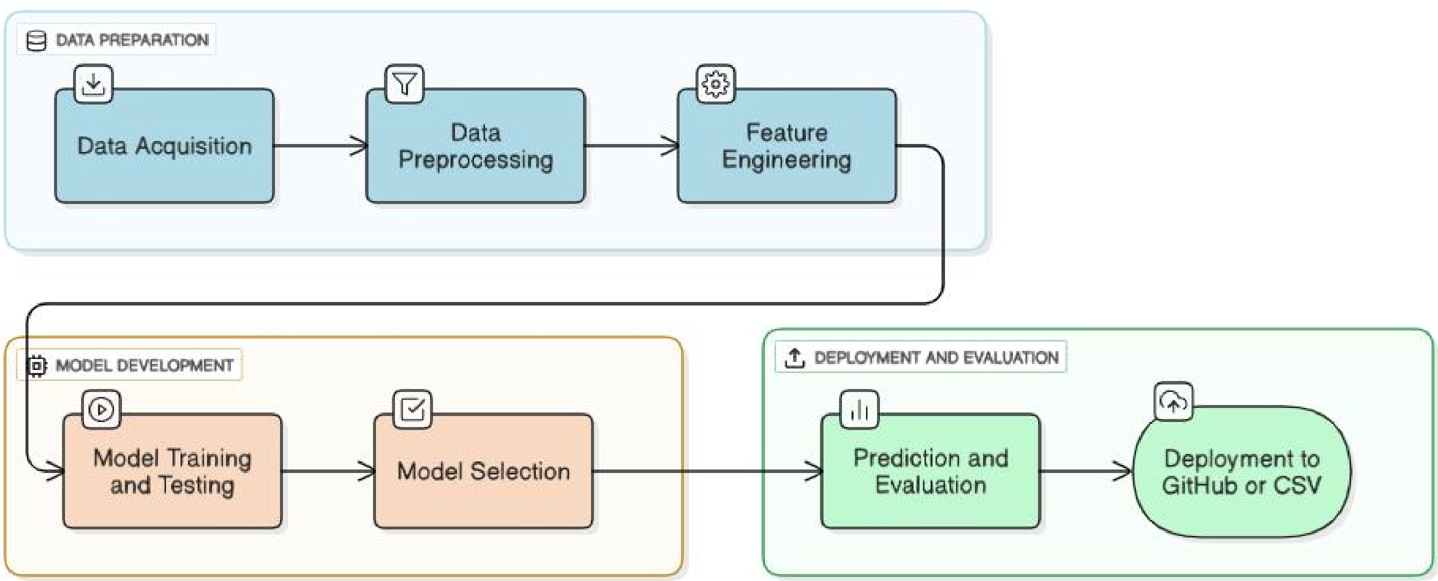
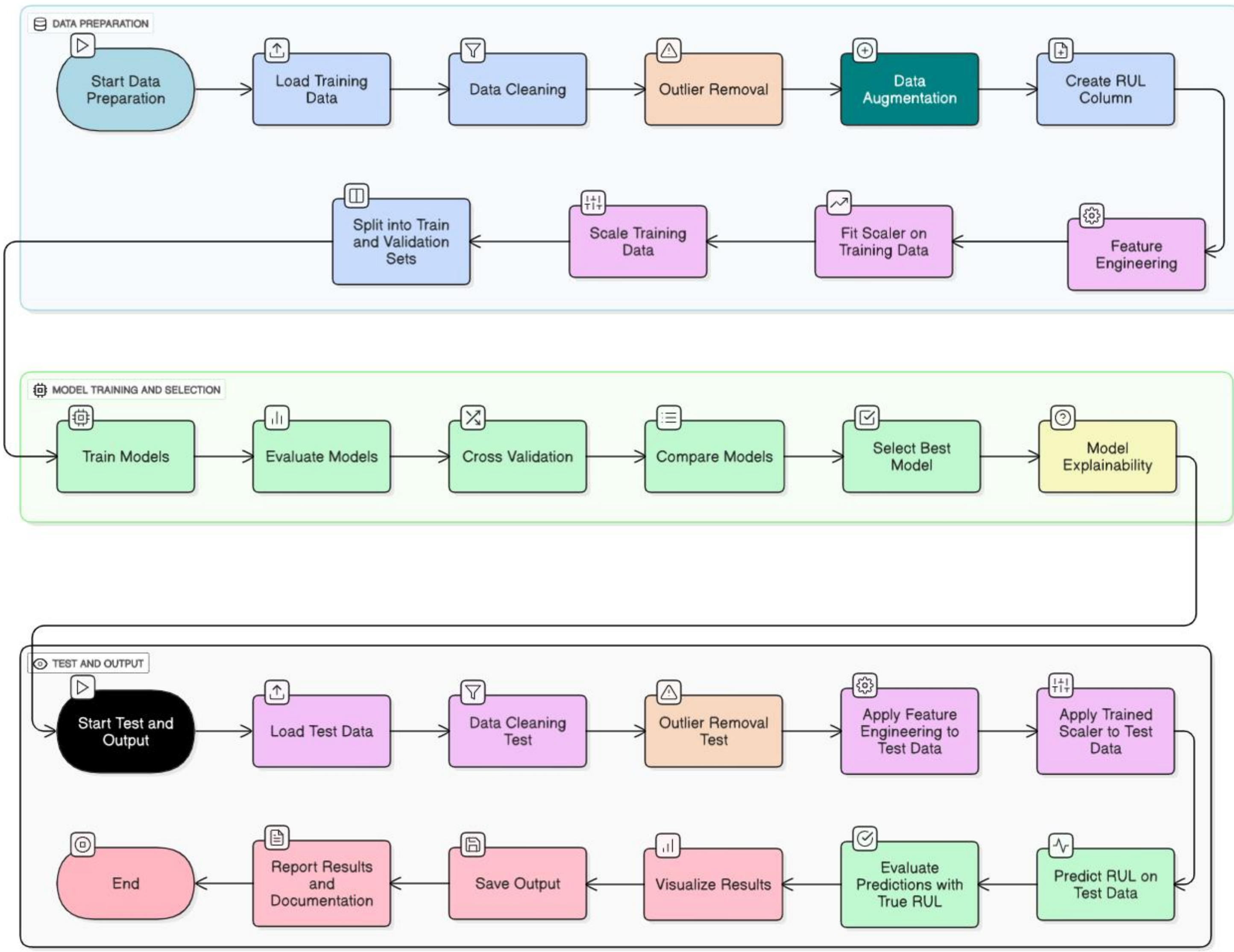


Figure 1: HIGH LEVEL DIAGRAM OF THE SYSTEM

5.2 Low Level Diagram



6 Performance Test

The performance testing of the Turbofan Remaining Useful Life (RUL) Prediction System was conducted to ensure that the developed machine learning solution meets industrial requirements beyond academic scope. This involved identifying real-world constraints, designing the testing strategy to address them, and analyzing the results to verify system effectiveness.

Identified Constraints and Industrial Relevance

- ✓ **Accuracy**: In predictive maintenance, high accuracy is crucial to avoid premature maintenance (leading to cost inefficiency) or unexpected failures (leading to downtime and safety hazards).
- ✓ **Computation Speed**: The model should produce predictions quickly enough to be deployed in near real-time monitoring systems.
- ✓ **Memory Consumption**: Models need to be lightweight enough for integration into industrial monitoring systems with limited computational resources.
- ✓ **Scalability**: The design must allow retraining and adaptation when new sensor types or operating conditions are introduced.

Design Considerations for Constraints:

- ◆ Chose **feature scaling** and **feature engineering** methods that are computationally efficient.
- ◆ Compared multiple algorithms (Random Forest, Gradient Boosting, SVR, Neural Network, Linear Regression) to select the best trade-off between accuracy and speed.
- ◆ Used rolling window statistics to capture trends without drastically increasing the dimensionality of features.

6.1 Test Plan/ Test Cases

Test Case	Objective	Method	Expected Output
Accuracy Test	Evaluate model prediction accuracy	Compare RMSE, MAE, R^2 on validation and test data	High R^2, low RMSE and MAE
Speed Test	Measure prediction time per input sample	Time the .predict() function on the test set	< 0.1 seconds/sample
Model Comparison	Identify best performing model	Compare metrics across trained models	Best model with lowest RMSE and balanced computation time
Resource Usage	Monitor memory usage	Check memory footprint of model object	< 200 MB for deployment model

6.2 Test Procedure

1. **Data Preparation**

- Used training and test datasets after applying preprocessing and feature engineering.
- Applied scaling based on training data statistics.

2. **Model Training**

- Trained 5 models on the training set.
- Recorded training time, model size, and performance metrics.

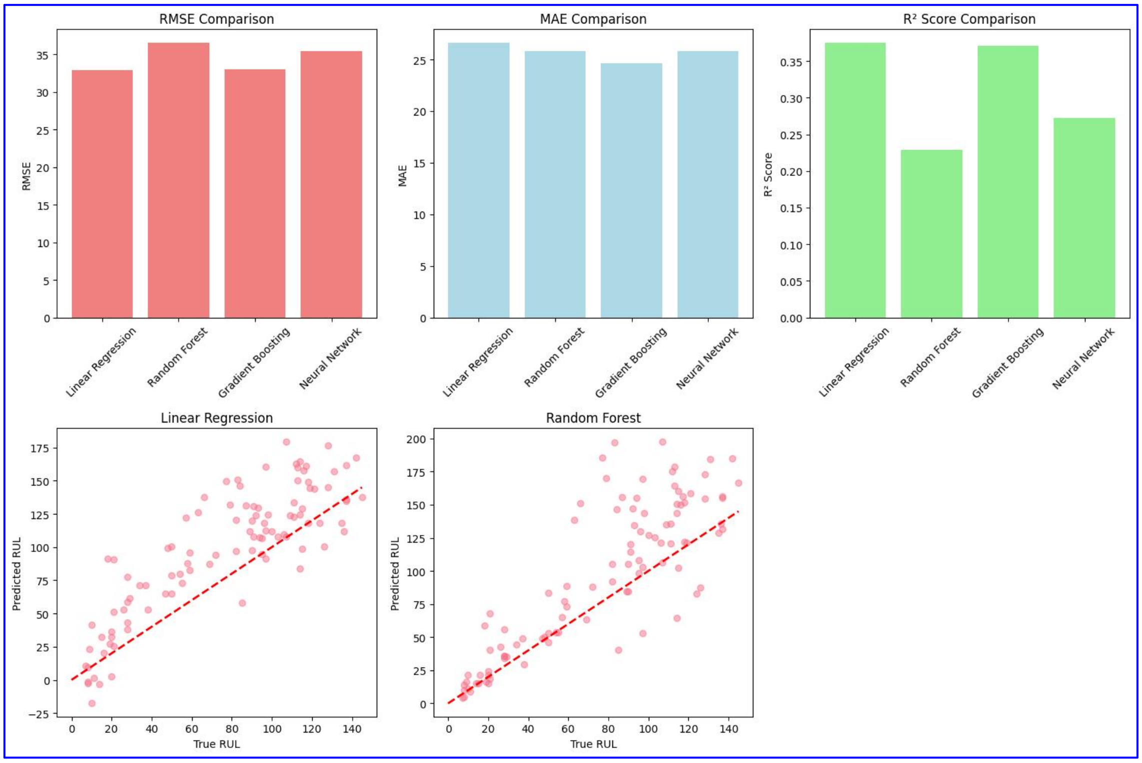
3. **Validation and Testing**

- Evaluated all models on the validation set.
- Selected the model with lowest RMSE for final testing on unseen data.

4. **Performance Measurement**

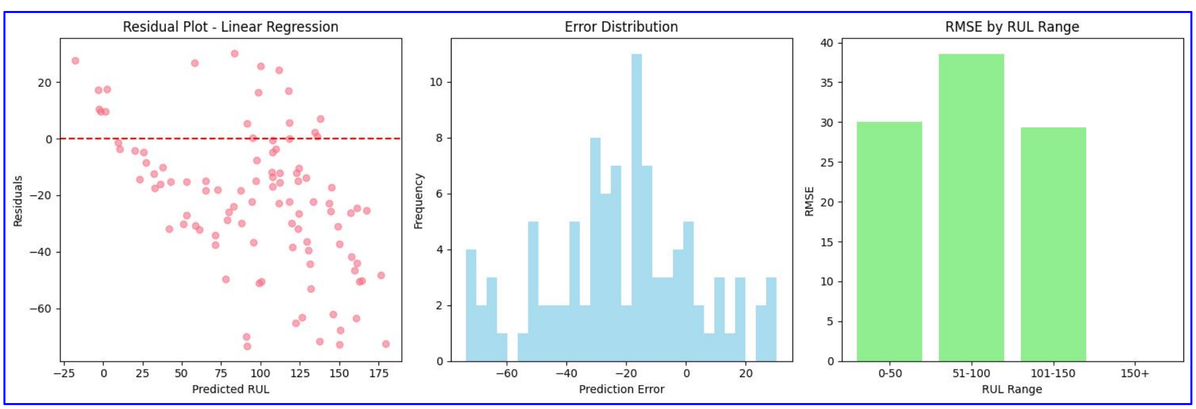
- Measured computation time per prediction.
- Recorded RMSE, MAE, R^2.

6.3 Performance Outcome



📊 Detailed Results:

	MSE	RMSE	MAE	R2
Linear Regression	1079.857	32.861	26.595	0.375
Random Forest	1332.034	36.497	25.846	0.229
Gradient Boosting	1086.336	32.960	24.624	0.371
Neural Network	1256.973	35.454	25.791	0.272



7 My learnings

During the course of this internship project on **Turbofan Engine Remaining Useful Life (RUL) Prediction**, I gained valuable technical and professional skills that will significantly contribute to my career growth as an aspiring software and data science professional.

From a **technical perspective**, I enhanced my knowledge and hands-on experience in:

- **Machine Learning Workflow** – Understanding the complete ML pipeline from **data acquisition** and **preprocessing** to **feature engineering**, **model training**, and **deployment**.
- **Feature Engineering Techniques** – Applying rolling statistics and scaling methods to extract meaningful patterns from time-series sensor data.
- **Model Evaluation & Selection** – Comparing multiple models (Random Forest, Gradient Boosting, SVR, Neural Network, Linear Regression) using metrics such as RMSE, MAE, and R^2 to choose the best-performing model.
- **Performance Optimization** – Balancing accuracy with computational efficiency to meet industrial constraints for predictive maintenance applications.
- **Data Visualization** – Creating meaningful plots like predicted vs. actual RUL, residual analysis, and error distributions to interpret model performance.

From a **professional and industry-oriented perspective**, I learned:

- How predictive maintenance systems work in **real-world industrial environments**.
- The importance of **scalable and maintainable solutions** for long-term deployment.
- How to **document and present** project results in a structured and professional manner.
- The significance of **version control** and proper project repository management for collaborative work.

This project has strengthened my ability to work on real-world problems that require a blend of **analytical thinking, technical expertise, and problem-solving skills**. It has also boosted my confidence to take on future projects involving **AI, ML, and predictive analytics**.

8 Future work scope

Although the current model for predicting the **Remaining Useful Life (RUL)** of a turbofan engine has achieved satisfactory accuracy, there are several areas where the project can be further improved and extended in future work:

1. **Integration of Deep Learning Models**

- Implement advanced time-series models such as **LSTM (Long Short-Term Memory)** or **GRU (Gated Recurrent Units)** networks to better capture long-term dependencies in sensor readings.
- Experiment with hybrid models that combine **deep learning** with **traditional machine learning** for improved performance.

2. **Automated Feature Selection & Engineering**

- Use **feature importance ranking** and **automated feature engineering tools** to identify the most relevant sensor readings and operational settings.
- Apply domain-specific feature extraction techniques for better predictive capability.

3. **Hyperparameter Optimization**

- Perform **Bayesian Optimization**, **Grid Search**, or **Random Search** for fine-tuning model parameters to further reduce prediction errors.

4. **Real-Time Prediction System**

- Develop a **real-time RUL prediction dashboard** that continuously updates predictions as new sensor data becomes available.
- Integrate the system with IoT devices for direct data ingestion from engines in operation.

5. **Model Explainability and Interpretability**

- Use explainable AI (XAI) tools such as **SHAP** or **LIME** to provide transparency in predictions, making the model more acceptable for industrial deployment.

6. **Scalability and Deployment**

- Deploy the trained model as a **cloud-based service** or **edge AI application** to support large-scale, distributed engine monitoring.

7. Incorporating Maintenance Logs and External Data

- Enhance the dataset with **historical maintenance records**, weather data, or other contextual factors to improve prediction robustness.

By implementing these future enhancements, the project can evolve into a **full-fledged industrial-grade predictive maintenance solution**, providing higher accuracy, real-time insights, and greater operational value to industries relying on turbofan engines.