Industrial Internship Report on

”Predict the number of remaining operational cycles before failure for Turbofan engine”

Prepared by

[Prathibha.Golusula]

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” Predict the number of remaining operational cycles before failure for Turbofan engine” The objective of this project is to develop a machine learning model to predict the remaining useful life of aircraft turbofan engines. The Remaining Useful Life (RUL) is the amount of cycles an engine has left before it needs maintenance.This project aims to develop an advanced predictive maintenance system for turbofan engines used in aircraft. The primary objective is to predict the number of remaining operational cycles before failure, also known as the Remaining Useful Life (RUL), for each turbofan engine based on historical data and real-time engine parameters. 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

Maintenance of equipment is a critical activity for any business involving machines. Predictive maintenance is the method of scheduling maintenance based on the prediction about the failure time of any equipment. The prediction can be done by analyzing the data measurements from the equipment. Machine learning is a technology by which the outcomes can be predicted based on a model prepared by training it on past input data and its output behavior. The

model developed can be used to predict machine failure before it actually happens. In this paper, a comparative study of existing set of machine learning algorithm to predict the Remaining Lifetime of aircraft's turbo fan engine is done. The machine learning model were constructed based on the datasets from turbo fan engine data from the Prognostics Data Repository of NASA. Using a training set, a model was constructed and was verified with a test data set. The results obtained were compared with the actualresults to calculate the accuracy and the algorithm that results in maximum accuracy is identified.

TABLE OF CONTENTS

1 Preface

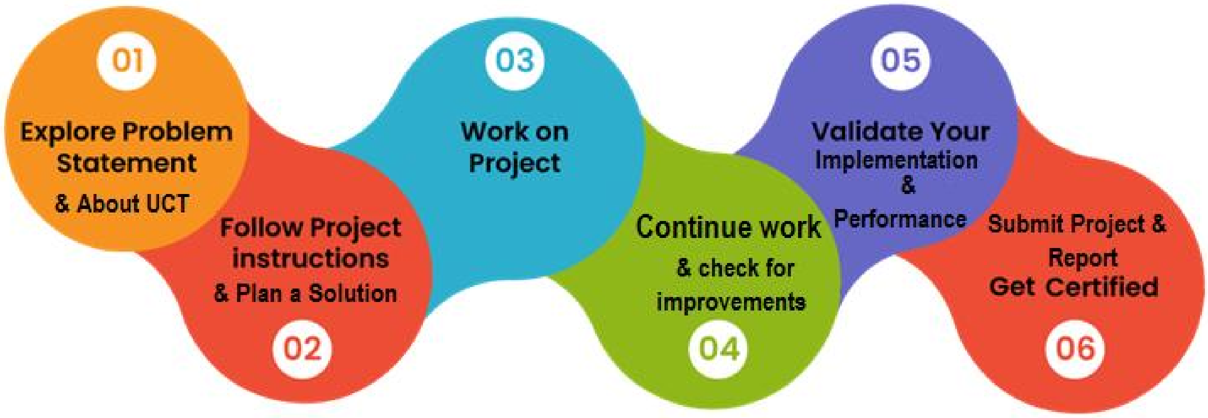
Summary of the whole 6 weeks' work.

About need of relevant Internship in career development.

Brief about Your project/problem statement.

Opportunity given by USC/UCT.

How Program was planned



Your Learnings and overall experience.

Thank to all ,Kaushlendra Singh sir and Nitin Tyagi sir,Apurv sir who have helped me a lot even though it is virtual session , am enjoyed a lot while learning new things and enhance my knowledge

The internship is offered by **Upskill campus** in partnership with its industrial partner **UniConverge Technologies** (UCT). UCT offers a wide variety of services and solutions across the world in IOT, Wireless Communication, Industry 4.0 & Predictive Maintenance. For developing its products and solutions it is leveraging various Cutting Edge Technologies e.g., Internet of Things (IoT), Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end etc.

Orientation was held on the 1st day of internship where all the information related to internship program was provided. During this week of internship, the initial phase laid strong foundation about the fundamental concepts on core principles of data science and machine learning. Familiarized with algorithms and techniques used for data processing. The week is marked by theoretical understanding and setting up for the project work and excited to continue building upon this progress and contributing to the project's success.

The prediction of remaining lifetime of a turbofan engine is a critical task in the maintenance and operation of aircraft. Machine learning algorithms can be used to analyze sensor data and other factors to predict the remaining useful lifetime of a turbofan engine .The process typically involves several steps, such as data collection, feature extraction, model training, and prediction .Data collection involves gathering sensor data from the engine during its operation. This data can include measurements such as temperature, pressure, vibration, and other parameters that can affect the engine's performance. Prediction involves using the trained model to predict the remaining useful lifetime of the engine based on the current sensor data. This can be done using a variety of techniques such as survival analysis, time-series analysis, or other methods that account for the degradation of the engine over time.

The prediction of RUL using machine learning algorithm can help improve the efficiency and safety of aircraft operation by enabling proactive maintenance and reducing the risk of unexpected failures.

To estimate remaining useful life of an turbofan engine.

To measure how many life cycles remains
before engine failure. To save time and energy
by avoiding unnecessary activities.

To Predict if it can enable proactive maintenance by identifying potential issues before they become critical.

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.

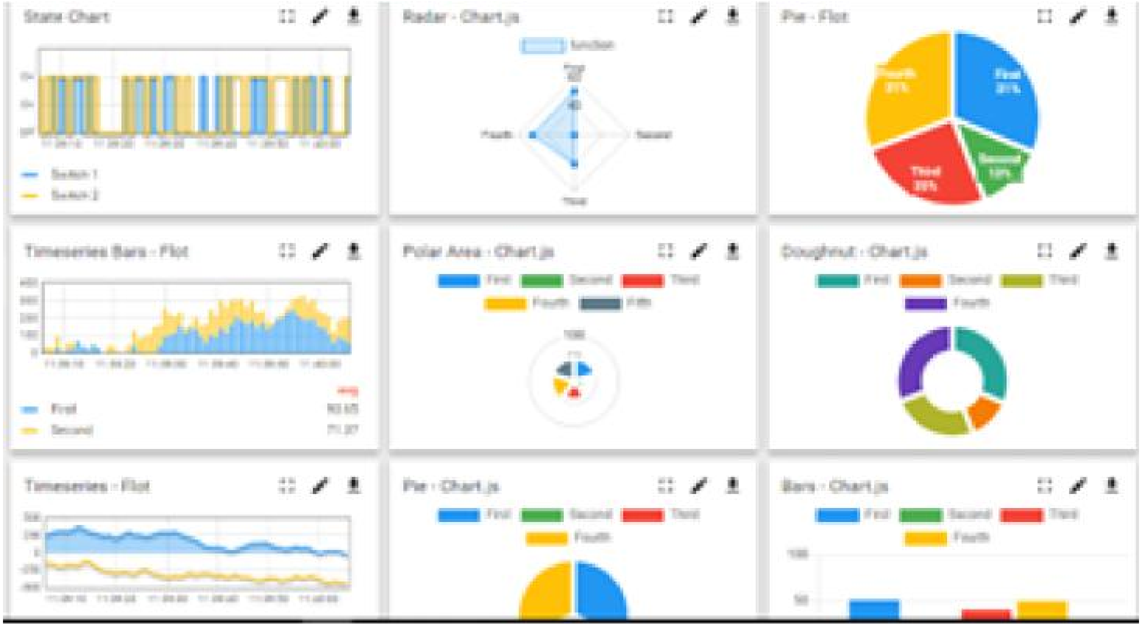
i. UCT IoT Platform (uct Insight)

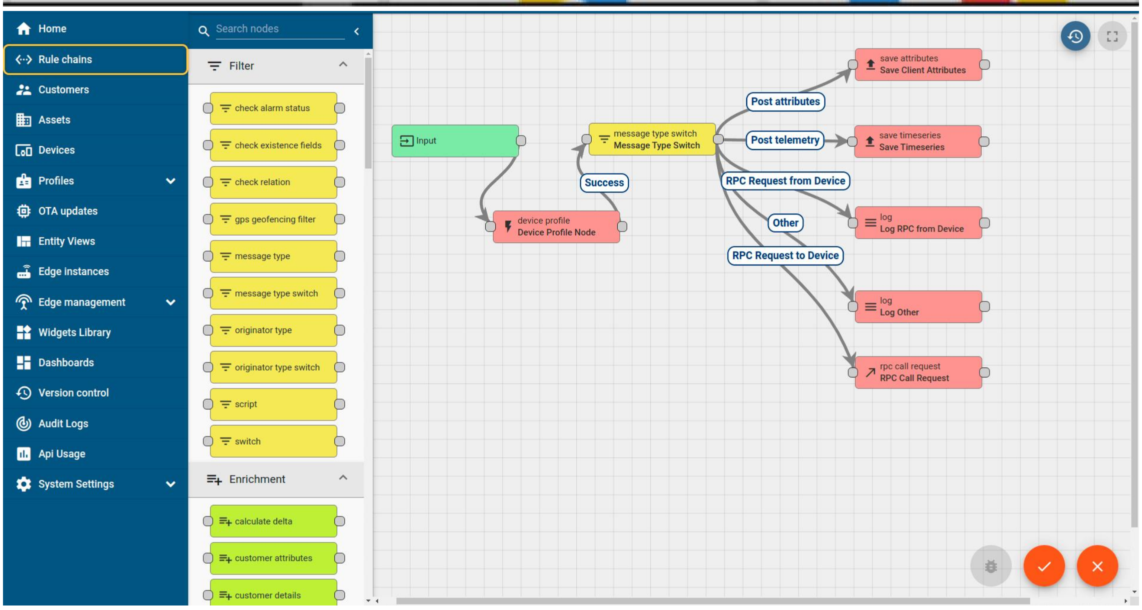
UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable “insight” for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- 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)

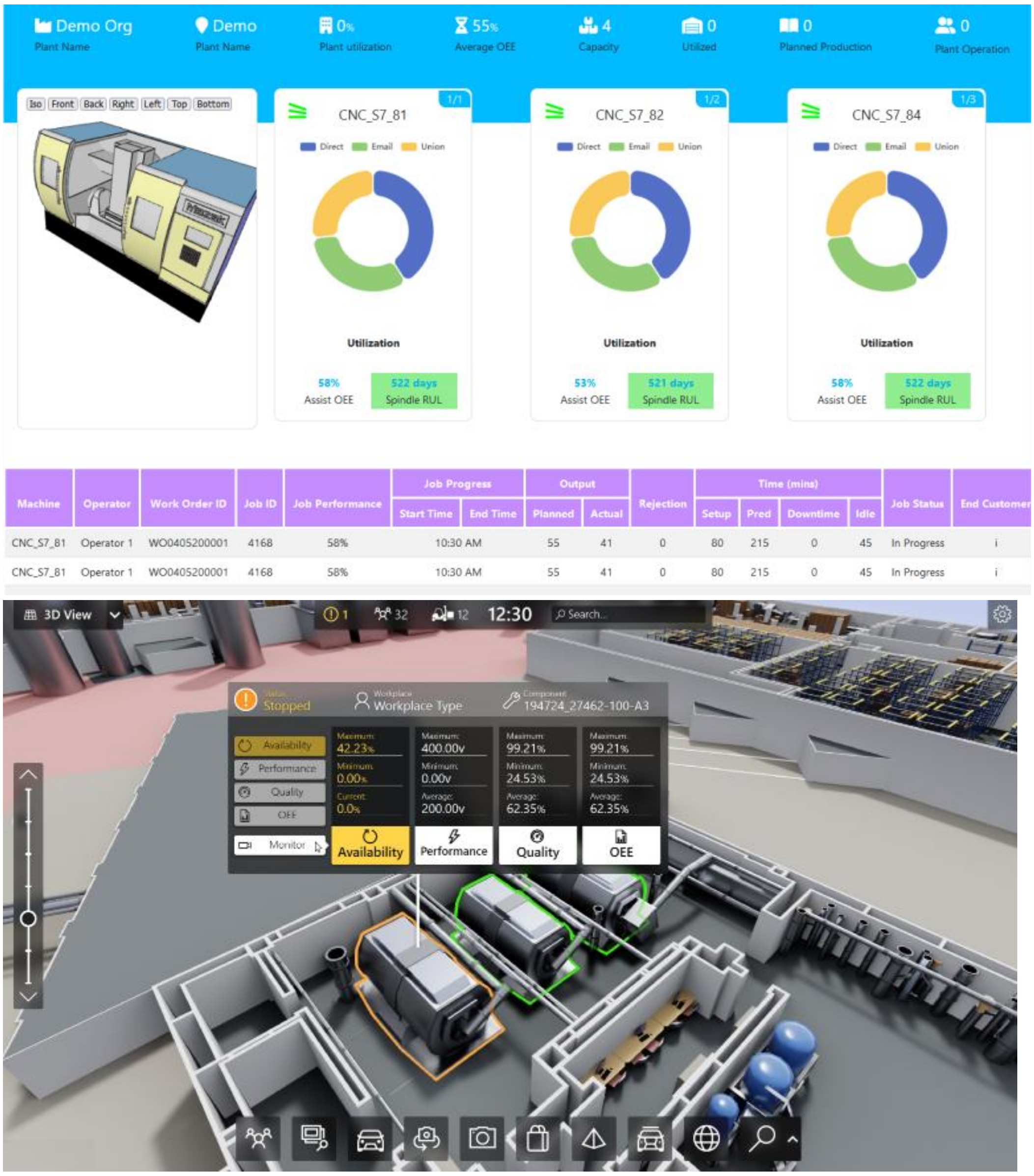
Factory watch is a platform for smart factory needs.

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.

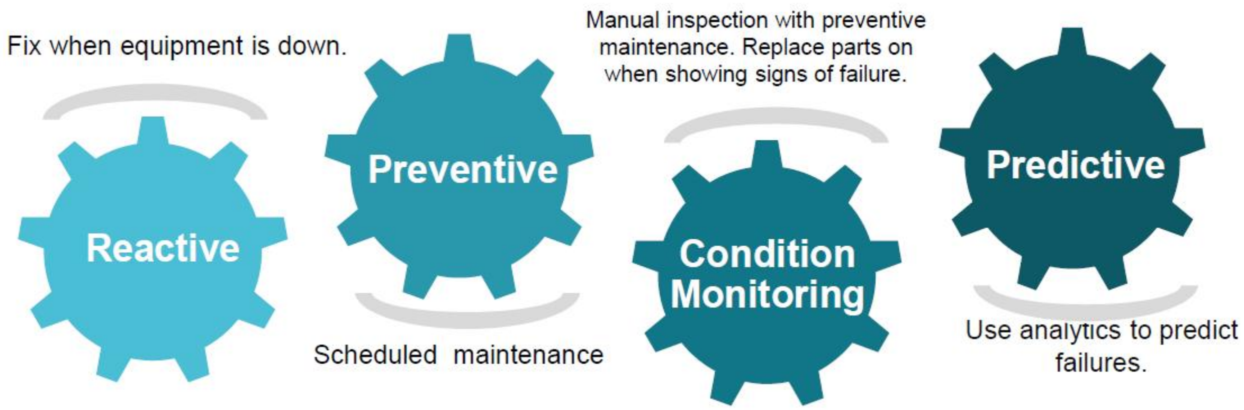


iii. LoRaWAN based Solution

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

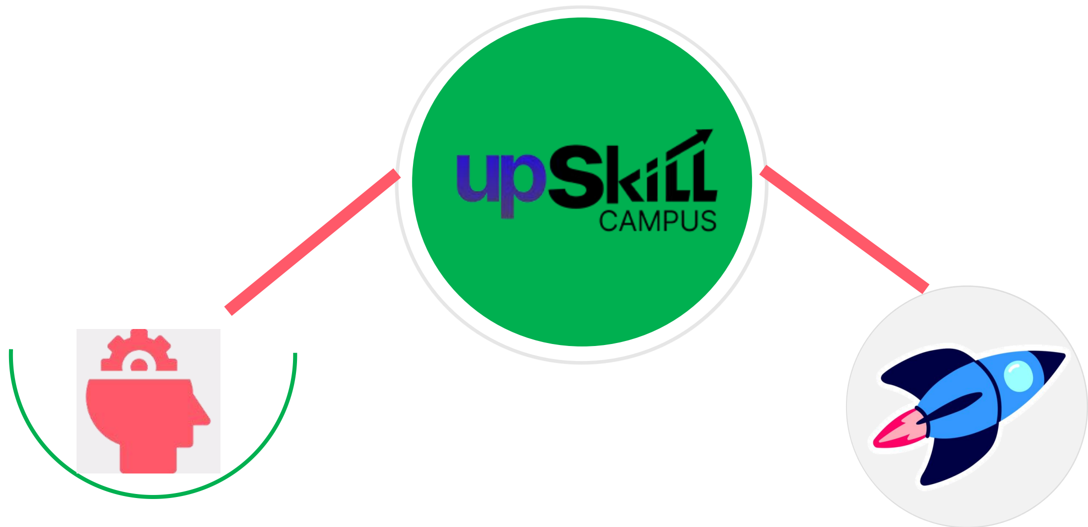
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

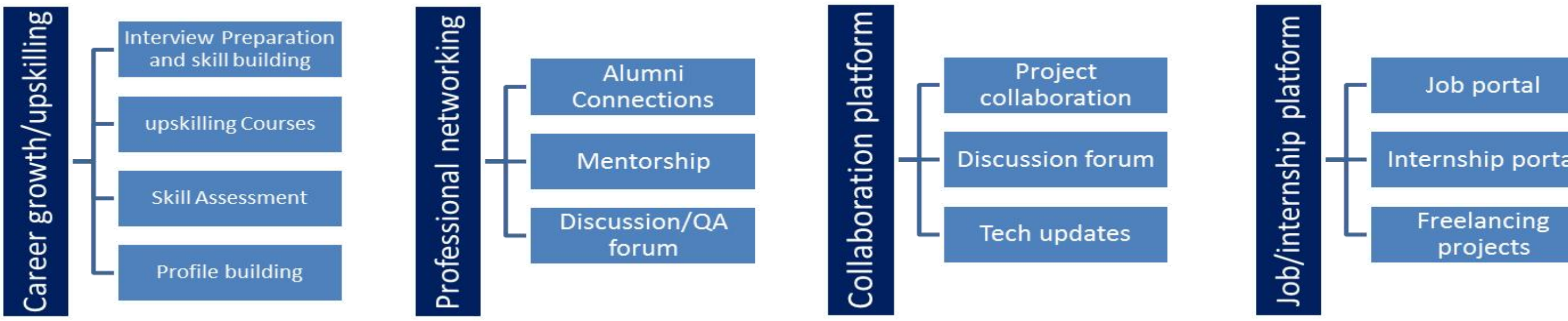
upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

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

upSkill Campus aiming to upskill 1 million



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] [1] W. F. Godoy, I. d. S. Nunes , A. Goedtel, R. H. Cunha Palacios and T. D. Lopes, "Application of intelligent

tools to detect and classify broken rotor bars in three-phase induction motors fed by an inverter," IET Electric

Power Applications, vol. 10, no. 5, pp. 430-439, 2016.

[2] M. Canizo, E. Onievay, A. Condez, S. Charramendietax and S. Trujillo, "Real-time predictive maintenance

for wind turbines using big data frameworks," in IEEE International Conference on Prognostics and Health

Management (ICPHM), 2017.

[3] C. Yang, Q. Chen, Y. Yang and N. Jiang, "Developing predictive models for time to failure estimation," in

Proceedings of the 2016 IEEE 20th International Conference on Computer Supported Cooperation Work in

Design, 2016.

[4] J. J. A. Costello, G. M. West and S. D. J. McArthur, "Machine learning model for eventbased prognostics in

gas circulator condition monitoring," IEEE Transactions on Reliability, vol. DOI 10.1109/TR.2017.2727489,

2017.

[5] G. Wang, T. Xu, H. Wang and Y. Zou, "AdaBoost and least square based failure prediction of railway

turnouts," in 9th International Symposium on Computational Intelligence and Design, 2016.

[6] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, A. Beghi and G. A. Susto, "Machine learning for

predictive maintenance: A multiple classifier approach," IEEE Transactions on Industrial Informatics, vol. 11, no. 3, pp. 812-820, 2015.

[7] Z. Li and Q. He, "Prediction of railcar remaining useful life by multiple data source fusion," IEEE

Transactions on Intelligent Transportaion Systems, vol. 16, no. 4, pp.22265-22355, 2015.

[8] C. Jennings, D. Wu and J. Terpenny, "Forecasting obsolescence risk and product life cycle with machine

learning," IEEE TRansactions on Components Packaging and Manufacturing Technology, vol. 6, no. 9, pp.

1428-1439, 2016.

[9] T. Pitakrat, A. v. H. Hoorn and L. Grunske, "A comparison of machine learning algorithms for proactive hard

disk drive failure detection," Proceedings of the 4th international ACM Sigsoft symposium on Architecting

critical systems, pp. 1-10, 2013.

[10] W. Lin, Z. Wu, L. Lin, A. Wen and J. Li, "An ensemble random forest algorithm for insurance big data

[2]

Terms	Acronym

3 Problem Statement

While predicting RUL, the main objective is to reduce the error between the actual RUL and the predicted

RUL. For each dataset, the test results were compared with the actual values of the RUL available in the data

set. The mean absolute error were plotted and it is observed that the best results were obtained by long – short

term memory.

When predicting the remaining lifetime of a turbofan engine using a machine learning algorithm, the

performance analysis typically involves evaluating the accuracy, precision, recall, and other relevant metrics of

the predictive model. The specific metrics used can vary depending on the approach and problem formulation.

Evaluate the trained model on the testing dataset to assess its performance. Calculate the relevant regression

metrics mentioned above to gauge the accuracy and precision of the predictions. Additionally, you can plot the

predicted remaining lifetime against the actual remaining lifetime to visually inspect the model's performance.

According to Figure 2, It indicates the proportion of the variance in the target variable. To create an Rsquared

graph, you would typically plot the predicted remaining lifetime values against the actual remaining

lifetime values for the test dataset. The x-axis would represent the actual remaining lifetime values, while the yaxis

would represent the predicted remaining lifetime values. Each data point on the graph would correspond to

an individual instance from the test dataset.

According to Figure 3, It measures the average absolute difference between the predicted and actual values

of the target variable (remaining lifetime). The lower the MAE, the better the model's predictions align with the

actual values.

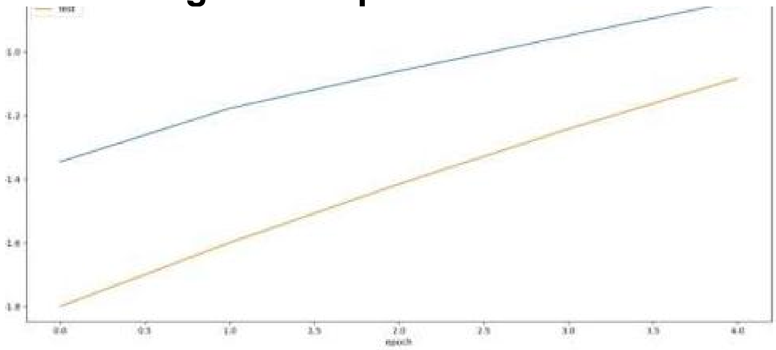
According to Figure 4, A model loss graph for predicting the remaining useful lifetime of a turbofan engine

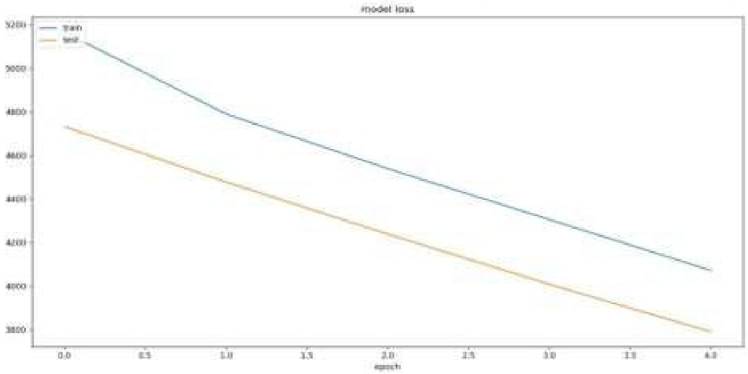
using a machine learning algorithm typically shows the training and validation loss values over different epochs

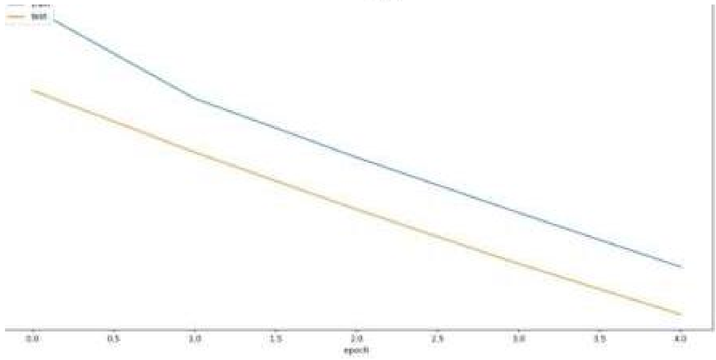
or iterations during the training process. Loss refers to the error or discrepancy between the predicted remaining

lifetime and the actual remaining lifetime.

4 Existing and Proposed solution





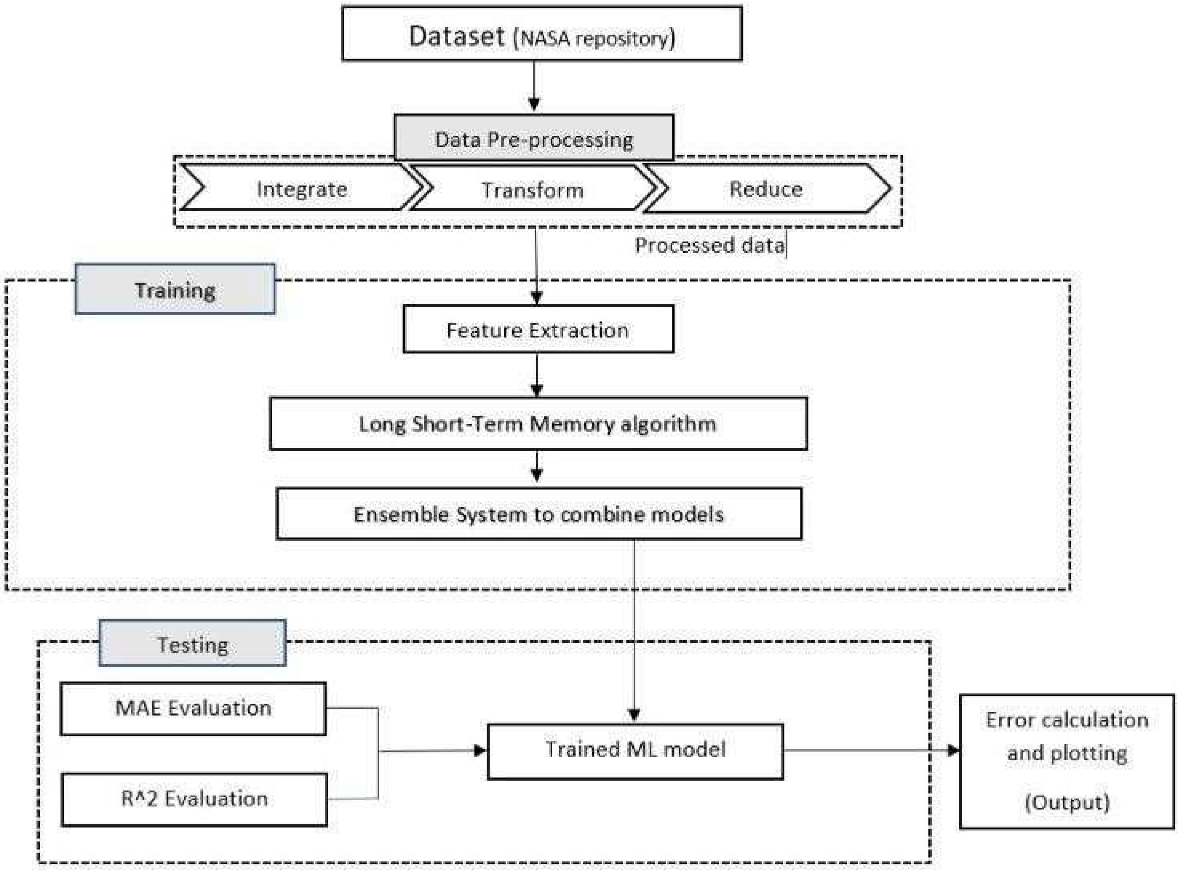


4.1 Code submission (Github link):https://github.com/prathibhagolusula/upskillcampus/blob/main/prediction of failure turbofan engine.ML&DS

4.2 Report submission (Github link) : https://github.com/prathibhagolusula/upskillcampus/blob/main/Prediction of failure turbo fan-USC-UCT.pdf

4.3

5 Proposed Design/ Model



The system architecture defines the overall structure of the system, including its components, their interactions, and the data flow between them. In this system, we have identified the following components:

Data Preprocessing: Preprocessing data is a common first step in the deep learning workflow to prepare raw data in a format that the network can accept. For example, you can resize image input to match the size of an image input layer. You can also preprocess data to enhance desired features or reduce artifacts that can bias the neural network.

Data Splitting: This model is commonly used In deep learning to split data into a train and test set. The training data set is used to train and develop models. Training sets are commonly used to estimate different parameters or to compare different model performance. The testing data set is used after the training is done. The training and test data are compared to check that the final model works correctly.

Model Training : In this models 80% are trained by using large sets of labeled data

and neural network architectures that learn features directly from the data withoutThe system architecture defines the overall structure of the system, including its components, their

interactions, and the data flow between them. In this system, we have identified the following components:

Data Preprocessing: Preprocessing data is a common first step in the deep learning workflow to prepare raw data in a format that the network can accept. For example, you can resize image input to match the size of an image input layer. You can also preprocess data to enhance desired features or reduce artifacts that can bias the neural network.

Data Splitting: This model is commonly used In deep learning to split data into a train and test set. The training data set is used to train and develop models. Training sets are commonly used to estimate different parameters or to compare different model performance. The testing data set is used after the training is done. The training and test data are compared to check that the final model works correctly.

Model Training : In this models 80% are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without

6 Performance Test

This is very important part and defines why this work is meant of Real industries, instead of being just academic project.

Here we need to first find the constraints.

How those constraints were taken care in your design?

What were test results around those constraints?

Constraints can be e.g. memory, MIPS (speed, operations per second), accuracy, durability, power consumption etc.

In case you could not test them, but still you should mention how identified constraints can impact your design, and what are recommendations to handle th

Performance Outcome

In this section, we describe about the existing work. Our Paper deals with the problem associated with the existing system and also gives user a clear knowledge on how to deal with the existing problems and how to provide solution to the existing problems.The author in the paper[1], This paper presents a comprehensive study on the application of intelligent tools for detecting and classifying broken rotor bars in three-phase induction motors that are fed by an inverter. The detection and classification of broken rotor bars are crucial for ensuring the reliable and efficient operation of induction motors. The proposed work focuses on leveraging advanced techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and fuzzy logic systems to develop a robust and accurate system for fault detection and classification. [2]. This paper proposes a real-time predictive maintenance framework for wind turbines leveraging big data frameworks. Wind turbines are complex systems prone to various faults and failures that can lead to costly downtime and maintenance. The proposed work aims to enhance the maintenance strategies by employing big data analytics techniques to collect, process, and analyze large volumes of data generated by wind turbines. By utilizing real-time data monitoring, predictive analytics, and machine learning algorithms, the proposed framework enables early detection of potential failures and allows for proactive maintenance, leading to improved turbine performance and reduced operational costs. [3]. This paper proposes a comprehensive framework for developing predictive models to estimate the time to failure of critical systems. Predicting the remaining useful life of assets is crucial for effective maintenance planning, reducing downtime, and optimizing resource allocation. The proposed work focuses on leveraging machine learning and statistical techniques to develop accurate and reliable models that can estimate the time to failure of assets based on historical data. The developed models enable proactive maintenance strategies, enhancing operational efficiency and reducing maintenance costs.

[4]. This paper proposes a machine learning-based approach for event-based prognostics in gas circulator condition monitoring. Gas circulators play a critical role in various industrial applications, and accurate prognosis of their remaining useful life (RUL) is essential for optimal maintenance planning and preventing unexpected failures. The proposed work focuses on developing a machine learning model that can predict the occurrence of critical events and estimate the remaining life of gas circulators based on real-time sensor data. The model enables proactive maintenance strategies, improving operational efficiency and reducing downtime. The uniqueness of [4 Review existing research on prognostics and condition monitoring in gas circulators.

Discuss the utilization of machine learning algorithms, event detection techniques, and feature engineering approaches in prognostics. [5] This paper proposes a novel approach for failure prediction of railway turnouts using the AdaBoost algorithm and least square-based techniques. Railway turnouts, or switches, are critical components in railway infrastructure, and timely detection of potential failures is crucial for ensuring safe and reliable operations. The proposed work aims to develop a predictive model that can effectively identify the probability of failure in railway turnouts based on historical data. By combining the ensemble learning capability of AdaBoost and the robustness of least square regression, the proposed approach enhances the accuracy and reliability of failure prediction, enabling proactive maintenance strategies and minimizing disruptions in railway operations.

[6] This paper proposes a novel approach for predictive maintenance using machine learning, specifically employing a multiple classifier approach. Predictive maintenance plays a crucial role in optimizing asset performance, minimizing downtime, and reducing maintenance costs. The proposed work focuses on developing a robust and accurate predictive maintenance model by leveraging the ensemble learning capabilities of multiple classifiers. The combination of different classifiers enhances prediction accuracy, provides robustness against variability in data, and enables more reliable maintenance decision-making. [7] This paper proposes a novel approach for predicting the remaining useful life (RUL) of railcars by fusing data from multiple sources. Accurate prediction of the RUL is crucial for optimizing railcar maintenance, reducing downtime, and ensuring safe and efficient operations. The proposed

work focuses on integrating diverse data sources, including sensor data, historical maintenance records, operational parameters, and external factors, to develop a comprehensive predictive model. By leveraging the fusion of multiple data sources, the proposed approach enhances prediction accuracy, robustness, and enables effective maintenance planning for railcar assets. [8] This paper proposes a novel approach for forecasting obsolescence risk and predicting the product life cycle using machine learning techniques. Obsolescence risk refers to the likelihood of a product becoming outdated or obsolete due to technological advancements, changing market dynamics, or other factors. Accurate forecasting of obsolescence risk and product life cycle is essential foreffective inventory management, product development, and strategic decision-making. The proposed work aims to leverage machine learning algorithms to develop predictive models that can estimate obsolescence risk and forecast the product life cycle, enabling proactive measures to mitigate risks and maximize the value of products.[9] This paper presents a comprehensive comparison of various machine learning algorithms for proactive hard disk drive (HDD) failure detection. Early detection of HDD failures is critical for preventing data loss, minimizing downtime, and optimizing maintenance activities. The proposed work aims to evaluate the performance of different machine learning algorithms in accurately predicting HDD failures and identifying potential warning signs. By comparing the algorithms' predictive capabilities, this study aims to provide insights into the most effective approach for proactive HDD failure detection, aiding in the development of reliable and efficient maintenance strategies.

[10]This paper proposes an ensemble random forest algorithm for the analysis of insurance big data. The insurance industry generates massive amounts of data, including customer information, claims history, policy details, and risk factors. Extracting valuable insights from this data is crucial for decision-making, risk assessment, and developing effective insurance strategies. The proposed work aims to leverage the power of ensemble learning and the random forest algorithm to address the challenges posed by insurance big data. By combining multiple decision trees and aggregating their predictions, the ensemble random forest algorithm enhances predictive accuracy, robustness, and interpretability for insurance data analysis.

In this proof of concept the goal is to maintain aircraft gas turbine engines in time before they fail as failures of these engines are very expensive for the operating company. To achieve this we will predict the remaining useful life (RUL) for the engines and switch them into maintenance a few cycles before we think a failure will

happen.

This task is aggravated by the fact that this use case is from the perspective of the manufacturer of the engines who are selling them and has no direct access to the engines operating data as the operating companies consider this data as confidential. The manufacturer still wants to offer the described failure early warning system.

There is some data available to the manufacturer from internal turbofan engines that will be used for training an initial machine learning model. All engines on the market will be expanded by a software component reading in the sensor measurements of the engine and predicting the RUL using this model and reacting on a low RUL with performing a maintenance. During a maintenance the theoretical moment of failure will be estimated by the maintenance staff, in this proof of concept the engine will be set in maintenance mode and the emulation data will continue to figure out the moment of failure. Complete data series up to a failure will then be used the regularly re-train the model to improve prediction quality over time.

7 My learnings

- Gained detail information about Data science and machine learning.
- Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more.
- Collecting the datasets, way of implementing the project. Knowledge of data preprocessing and data analysis in machine learning.
- Python concepts, installation and usage of libraries and improved the skills with quiz activity.
- Resume and Interview preparation-based Webinar enhanced our presentation skills.
- Got to know about the company (UCT).
- Basic knowledge about Data Science and Machine Learning.
- Various technologies used in this domain.
- Applications of Data Science and Machine Learning.
- Collecting the datasets.
- Way of implementing the project.

8 Future work scope

The Remaining Useful Lifetime prediction has been carried out so as to plan the maintenance requirements of

the turbo fan engine. By doing predictive maintenance, failures can be predicted and maintenance can be

scheduled in advance. This reduces the cost and effort for doing maintenance. By predicting the RUL, ML

algorithms can assist in scheduling maintenance activities proactively. This can minimize downtime and reduce

the overall maintenance costs associated with engine replacements or repairs. It can analyze various sensor data

and operational parameters to optimize the performance of turbofan engines. This estimation helps in planning

maintenance activities and avoiding unexpected failures.

In future, the algorithm can be tested for more real time data and always be one step ahead in predicting the

maintenance requriments.