**Industrial Internship Report on**

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

**Prepared by**

**Vaibhav Vinod Dhande**

|  |
| --- |
| *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 Objective:  The goal of this project is to develop machine learning models that predict the remaining useful life (RUL) of aircraft turbofan engines.  RUL represents the number of operational cycles an engine has left before it requires maintenance.  Dataset:  The dataset contains information on 249 engines (identified by engine\_no).  Each engine is monitored over time (time\_in\_cycles).  For each cycle, operational settings and sensor measurements are recorded.  The dataset is divided into training and testing subsets.  Engines start with varying degrees of initial wear and manufacturing variation.  The data is also contaminated with sensor noise.  Engines operate normally initially and develop faults during the series.  Task:  In the training set, faults grow until system failure.  In the test set, time series end before system failure.  The objective is to forecast the number of remaining operational cycles before failure in the test set.  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 5](#_Toc139702806)

[2 Introduction 6](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 6](#_Toc139702808)

[2.2 About upskill Campus 11](#_Toc139702809)

[2.3 Objective 13](#_Toc139702810)

[2.4 Reference 13](#_Toc139702811)

[2.5 Glossary 13](#_Toc139702812)

[3 Problem Statement 15](#_Toc139702813)

[4 Existing and Proposed solution 17](#_Toc139702814)

[5 Proposed Design/ Model 19](#_Toc139702815)

[3](#_Toc139702815)[6 Performance Test 21](#_Toc139702819)

[7 My learnings 23](#_Toc139702823)

[8 Future work scope 25](#_Toc139702824)

# 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 (with names), who have helped you directly or indirectly.

Your message to your juniors and peers.

# Introduction

## 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.



1. 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

 

1. **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.

 

1.  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.

1. 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.



## 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

<https://www.upskillcampus.com/>

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



## 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.

## 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.

## Reference

[1] Edu net Training program

[2] upskill campus

[3] Teachers

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Problem Statement

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

**Experimental Scenario**

Data sets consists of multiple multivariate time series. Each data set is further divided into training and test subsets. Each time series is from a different engine – i.e., the data can be considered to be from a fleet of engines of the same type. Each engine starts with different degrees of initial wear and manufacturing variation which is unknown to the user. This wear and variation is considered normal, i.e., it is not considered a fault condition. There are three operational settings that have a substantial effect on engine performance. These settings are also included in the data. The data is contaminated with sensor noise.

The engine is operating normally at the start of each time series, and develops a fault at some point during the series. In the training set, the fault grows in magnitude until system failure. In the test set, the time series ends some time prior to system failure. The objective of the competition is to predict the number of remaining operational cycles before failure in the test set, i.e., the number of operational cycles after the last cycle that the engine will continue to operate. Also provided a vector of true Remaining Useful Life (RUL) values for the test data.

The data are provided as a zip-compressed text file with 26 columns of numbers, separated by spaces. Each row is a snapshot of data taken during a single operational cycle, each column is a different variable.

The columns correspond to:

1) unit number

2) time, in cycles

3) operational setting 1

4) operational setting 2

5) operational setting 3

6) sensor measurement 1

7) sensor measurement 2

...

26) sensor measurement 26

Data Set: FD001

Train trajectories: 100

Test trajectories: 100

Conditions: ONE (Sea Level)

Fault Modes: ONE (HPC Degradation)

Data Set: FD002

Train trajectories: 260

Test trajectories: 259

Conditions: SIX

Fault Modes: ONE (HPC Degradation)

Data Set: FD003

Train trajectories: 100

Test trajectories: 100

Conditions: ONE (Sea Level)

Fault Modes: TWO (HPC Degradation, Fan Degradation)

Data Set: FD004

Train trajectories: 248

Test trajectories: 249

Conditions: SIX

Fault Modes: TWO (HPC Degradation, Fan Degradation)

# Existing and Proposed solution

Existing solutions and research related to turbofan engine failure prediction are:

NASA Turbofan Failure Prediction (GitHub Project):

This data analytics and machine learning project investigates the relationship between sensor readings and the onset of failure (in terms of remaining engine cycles) for simulated turbofan engine data from a NASA research project.

The project objectives include:

Analyzing relationships between engine behavior and remaining useful lifetime (RUL).

Developing predictive models for RUL.

The dataset was explored using Jupyter Notebooks, and various machine learning models were developed, including linear regression, decision tree regressors, random forest regressors, and neural network regressors (using TensorFlow and Keras).

The project repository can be found on GitHub: NASA Turbofan Failure Prediction.

Prognostic Health Management for Turbofan Engines (Stanford University):

Researchers explore deep learning solutions to model the spatio-temporal relationships exhibited by NASA turbofans.

They propose a novel CNN-LSTM architecture and explore the power of LSTMs to model sequential data for predicting the Remaining Useful Life (RUL) of a system.

Remaining Useful Life Prediction of an Aircraft Turbofan Engine:

This research uses machine learning to provide a prediction framework for an aircraft’s RUL based on the entire life cycle data and deterioration parameter data.

A Deep Layer Recurrent Neural Network (DL-RNN) model is presented for lifetime assessment.

Fault Prognosis of Turbofan Engines:

This study significantly expands upon past efforts by broadening the research scope to encompass eventual failure prediction.

Accurate prediction and isolation of the reason for failure have important implications for maintenance

## Code submission (Github link)

https://github.com/VaibhavDhande/upskillCampus/blob/main/project%206.ipynb

## Report submission (Github link) :

# Proposed Design/ Model

Given more details about design flow of your solution. This is applicable for all domains. DS/ML Students can cover it after they have their algorithm implementation. There is always a start, intermediate stages and then final outcome.

Data Acquisition:

The model will require historical data on turbofan engine performance. This data can come from various sensors that monitor engine health, including:

Vibration sensors

Temperature sensors

Pressure sensors

Fuel flow sensors

Speed sensors

Each data point should be associated with the remaining cycles before failure (RUL) for that specific engine. This "labeled data" is crucial for training the model.

Data Preprocessing:

The raw sensor data might need cleaning and pre-processing steps like:

Handling missing values

Identifying and removing outliers

Feature scaling (ensuring all features have similar scales)

Model Selection and Training:

Several machine learning models are suitable for RUL prediction

Overall Design Modal:

The project follows a supervised learning approach:

Data Acquisition and Preprocessing

Model Selection and Training

Prediction and Evaluation

# Performance Test

Evaluating the performance of a turbofan engine RUL prediction model is crucial to assess its effectiveness and identify areas for improvement. Here are some common performance tests used for this purpose:

**1. Error Metrics:**

* **Root Mean Squared Error (RMSE):** This metric measures the average difference between the predicted RUL and the actual RUL. Lower RMSE indicates better prediction accuracy.
* **Mean Absolute Error (MAE):** Similar to RMSE, MAE calculates the average absolute difference between predicted and actual RUL, providing another measure of prediction error.

**2. Distribution Metrics:**

* **Scatter Plots:** Visualizing the distribution of predicted vs. actual RUL can reveal patterns and biases in the model's predictions. Ideally, the data points should cluster around a diagonal line representing perfect prediction.
* **Quantile Plots:** These plots compare the distribution of predicted RUL with the actual RUL distribution. They help identify potential biases towards overestimating or underestimating RUL at different quantiles.

**3. Statistical Significance Tests:**

* **t-test:** This test compares the means of the predicted and actual RUL to assess if there's a statistically significant difference. A non-significant difference indicates the model's predictions are on par with the actual values on average.

**4. Prognostic Metrics:**

* **Survival Plots:** These plots depict the probability of an engine surviving for a certain number of cycles beyond a specific point in time. They allow for evaluating the model's ability to predict the likelihood of failure over time.
* **Concordance Index (C-Index):** This metric measures how well the model ranks engines based on their remaining useful life. A higher C-index indicates better agreement between the model's predictions and the actual engine failures.

**Performing these tests:**

* Split the available data into training and testing sets. The model is trained on the training data and evaluated on the unseen testing data to avoid overfitting.
* Calculate the chosen metrics on the testing data to assess the model's generalizability to new engine data.
* Compare the model's performance with other existing RUL prediction models to understand its strengths and weaknesses relative to the state-of-the-art.

By employing these performance tests, you can gain valuable insights into the effectiveness of your turbofan engine RUL prediction model. This allows for data-driven decisions on model refinement, exploration of alternative algorithms, or potentially incorporating additional features to enhance prediction accuracy.

# My learnings

Here are some key learnings you can take away from a project predicting turbofan engine RUL using machine learning:

**Data is Crucial:**

* The model's accuracy hinges on the quality and quantity of data used for training.
* Diverse data encompassing various operating conditions, engine types, and failure modes is essential for generalizability.

**Feature Engineering Matters:**

* Beyond raw sensor readings, extracting meaningful features from the data can significantly improve prediction accuracy.
* Techniques like signal processing and domain knowledge can help identify degradation patterns.

**Model Selection is Key:**

* Different machine learning models have strengths and weaknesses.
* Understanding the data characteristics and task (regression for RUL prediction) helps choose the most suitable model (e.g., RNNs for sequential data).

**It's an Iterative Process:**

* The initial model is unlikely to be perfect.
* Evaluating performance metrics, refining data or feature engineering, and potentially trying different models are essential for continuous improvement.

**Domain Expertise is Valuable:**

* Collaboration between data scientists and engineers with turbofan engine knowledge can significantly enhance the project.
* Understanding the physics of engine degradation can guide feature selection and model interpretation.

**Real-World Benefits:**

* An accurate RUL prediction model can lead to significant benefits:
  + Improved safety by enabling preventive maintenance and avoiding unexpected failures.
  + Optimized maintenance schedules by focusing resources on engines nearing failure.
  + Cost savings by reducing unnecessary maintenance and downtime.

**Limitations to Consider:**

* Machine learning models are not perfect and can be susceptible to errors.
* Unexpected failure modes not present in the training data might not be predicted well.
* The model's reliability depends on the quality and completeness of the data used.

Overall, this project highlights the power of machine learning in predictive maintenance for critical systems like turbofan engines. By understanding the learnings and potential limitations, you can leverage this technology to improve safety, optimize maintenance strategies, and gain valuable insights into engine health.

# Future work scope

The project focusing on turbofan engine RUL prediction using machine learning has a promising future scope with several exciting possibilities for advancement:

**1. Increased Accuracy and Generalizability:**

* **Incorporate More Data Sources:** Integrate data from additional sensors, flight logs, and maintenance records to capture a more comprehensive picture of engine health.
* **Synthetic Data Generation:** Utilize techniques like Generative Adversarial Networks (GANs) to create realistic synthetic engine data with diverse failure modes, expanding the training dataset and improving model generalizability to unseen scenarios.
* **Explainable AI (XAI):** Develop models that provide explanations for their predictions. This can build trust in the model's outputs and allow for targeted maintenance actions based on the identified degradation patterns.

**2. Multi-Modal Learning:**

* Explore combining sensor data with physics-based models of engine degradation. This can leverage the strengths of both data-driven and physics-driven approaches for more robust and interpretable predictions.

**3. Real-time Monitoring and Integration:**

* Deploy the model for real-time monitoring of engine health during operation. This allows for continuous assessment of RUL and immediate notifications for potential issues, enabling proactive maintenance strategies.
* Integrate the model with existing aircraft health monitoring systems for a holistic view of the aircraft's health and potentially predict failures in other critical components.

**4. Advanced Anomaly Detection:**

* Develop algorithms to detect anomalies in sensor data that deviate from normal operation patterns. This can provide early warnings of potential failures even if they are not yet reflected in the predicted RUL.

**5. Engine Health Optimization:**

* Utilize the RUL predictions to optimize engine operating conditions and maintenance schedules. This could involve adjusting flight profiles or implementing preventative maintenance actions to maximize engine lifespan and efficiency.

**6. Cloud-based Deployment and Collaboration:**

* Leverage cloud platforms to deploy the RUL prediction model, facilitating access for airlines and maintenance providers. This allows for centralized data storage, model updates, and collaboration on improving prediction accuracy across a vast fleet of engines.

By exploring these future directions, the turbofan engine RUL prediction project can evolve into a powerful tool for ensuring aviation safety, optimizing maintenance practices, and extending engine lifespan.