**Industrial Internship Report on**

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

**Prepared by**

**[Romel.TK]**

|  |
| --- |
| *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 predicting the number of remaining operational cycles (RUL) before failure for turbofan engine.  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 3](#_Toc139702806)

[2 Introduction 4](#_Toc139702807)

[2.1 About UniConverge Technologies Pvt Ltd 4](#_Toc139702808)

[2.2 About upskill Campus 8](#_Toc139702809)

[2.3 Objective 10](#_Toc139702810)

[2.4 Reference 10](#_Toc139702811)

[2.5 Glossary 10](#_Toc139702812)

[3 Problem Statement 11](#_Toc139702813)

[4 Existing and Proposed solution 12](#_Toc139702814)

[5 Proposed Design/ Model 13](#_Toc139702815)

[6 Performance Test 14](#_Toc139702819)

[6.1 Test Procedure 14](#_Toc139702821)

[6.2 Performance Outcome 15](#_Toc139702822)

[7 My learnings 17](#_Toc139702823)

[8 Future work scope 18](#_Toc139702824)

# Preface

First and foremost let me express my gratitude to Upskill campus and UniConverge Technologies for providing me with this wonderful opportunity. I am currently pursuing my master’s in Big Data Analytics , and therefore an internship with industry standards was imperative to me. An internship would provide me the necessary exposure of how the real industry works and also give me a feel of the corporate world.

The duration of the internship was 6 weeks. I had to meet the requirements of each week. This imbibed in me the work ethic as well as the necessities of meeting deadlines. The project that I undertook was predicting the number of remaining operational cycles (RUL) before failure for turbofan engine.

This comes under predictive maintenance. Predictive maintenance is a proactive maintenance strategy that uses data and advanced analytics techniques to predict when equipment or machinery is likely to fail and schedule maintenance accordingly. It aims to minimize unexpected breakdowns, reduce downtime, and optimize maintenance activities by identifying potential issues before they cause major problems or failures. Predictive maintenance is particularly relevant and beneficial for turbofan engines used in aircraft. Hence by predicting the remaining useful life of a turbofan engine we could have numerous benefits.



This project gave me an understanding of what predictive maintenance is and also the need of such maintenance in industrial premises. I also got to know about a parameter called RUL, which applies in general to all machineries. Once again grateful to all the volunteers who guided me in successfully completing the project and thereby the internship.

# 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 technology 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.

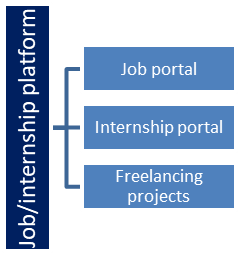
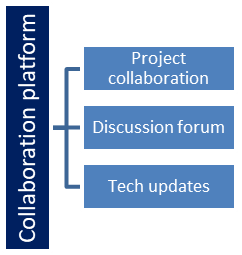
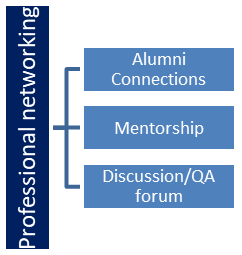
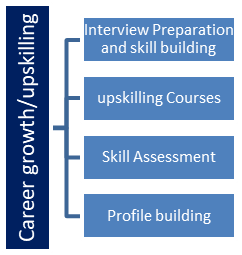
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



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] Damage Propagation Modeling for Aircraft Engine Run-to-Failure Simulation.

Abhinav Saxena, Member IEEE, Kai Goebel, Don Simon, Member, IEEE, Neil Eklund, Member IEEE

[2]

## Glossary

|  |  |
| --- | --- |
| Terms | Acronym |
| Remaining useful lifetime | RUL |
| K-nearest neighbours | KNN |
| Long-short term memory | LSTM |
|  |  |
|  |  |

# Problem Statement

Predicting the number of remaining operational cycles before failure for a turbofan engine is a typical application of predictive maintenance. The problem statement revolves around estimating the remaining useful life (RUL) of the engine, which refers to the number of operational cycles or flight hours the engine has left before it is likely to experience a failure or reach a predefined failure threshold. This prediction is crucial for optimizing maintenance strategies, ensuring safety, and minimizing downtime and operational disruptions.

The primary goal of predicting the number of remaining operational cycles before failure for turbofan engines is to enable proactive maintenance planning. By predicting the RUL accurately, maintenance teams can schedule maintenance activities in advance, ensuring that necessary inspections, repairs, or part replacements are performed at the right time. This approach helps to prevent unexpected failures, reduce downtime, optimize maintenance costs, and enhance the overall safety and reliability of aircraft operations.

# Existing and Proposed solution

Existing works on this project include the use of numerous regression algorithms like linear regression,decision tree regresssor etc . Due to the presence of lot of noise and unimportant features in the data , accuracy of these models are not satisfactory. My project aims to achieve better accuracy using ensemble methods. Performing k-fold cross validation for a more impartial performance. I have used ensemble models like gradient boosting regressor and Random forest regressor to enhance the performance.

## Code submission (Github link)

<https://github.com/Romeltk/Upskill-Campus/blob/598f3aa0e16aa4a5c7e4e8f8176d2d31511987cc/Turbofan%20Engine.ipynb>

## Report submission (Github link) : first make placeholder, copy the link.

# Proposed Design/ Model

1. **Data Collection:** To predict the remaining operational cycles before failure, relevant data from the turbofan engine needs to be collected. This includes data on engine parameters, performance metrics, sensor readings, and maintenance history. The data may be obtained from various sources, such as on board sensors, engine monitoring systems, flight data recorders, maintenance logs, and historical databases.
2. **Data Preparation:** The collected data needs to be preprocessed and cleaned to ensure its quality and reliability. This step involves handling missing values, removing outliers, and formatting the data in a suitable way for analysis.
3. **Feature Engineering:** Feature engineering involves selecting or creating relevant features from the collected data that can help in predicting the remaining operational cycles. These features may include engine operating parameters, temperature, pressure, vibration, fuel consumption, and any other data that could be indicative of engine health. This included Computing the remaining useful life values and incorporating the RUL\_FD001.txt to the test set .
4. **Model Development:** Machine learning algorithms or statistical techniques are applied to develop a predictive model. The model is trained on historical data where the remaining operational cycles are known, and the features are used to learn the patterns and relationships between engine condition and RUL.
5. **Model Validation and Evaluation:** The developed model is validated using a separate set of data that was not used during training. The performance of the model is evaluated using appropriate metrics, such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE), to assess how well it predicts the remaining operational cycles.
6. **RUL Prediction:** Once the model is trained and validated, it can be used to predict the remaining operational cycles for new instances of turbofan engines. When the engine is in operation, its real-time or periodic data can be fed into the model to estimate its current RUL and determine when maintenance should be scheduled before the expected failure.

# Performance Test

* Boxplots revealed that sensors 1, 5, 6, 10, 16, 18 and 19 have constant values hence they are not good for predicting.so had to drop these features on the test as well as train sets. Also found that operational setting 3 also has only 2 unique values, hence it is almost constant. Therefore had to drop it as well. Also by employing the heatmap and checking the correlation of all the variables, I found that sm\_9 and sm\_14 are highly correlated hence had to drop any one of them. Also we could drop the unit\_id, operational setting1 and 2 since they have very low correlation with the target variable RUL.Since all the values of the variables are of different measurement units , we have to normalize the train and test sets. These were the initial constraints that had to be taken care of before implementing the predictive model.

## Test Procedure

* Algorithms Employed :

* Linear Regression
* Decision Tree
* KNN
* Random Forest
* Gradient Boost
* Tools used for Validation:
* Mean squared error
* Root mean squared error
* R2 score
* Mean absolute percentage error
* Cross validation

## Performance Outcome

* Linear Regression

mean squared error : 3025.700820868507

Root mean squared error : 55.00637072983917

R2 score : 0.13020927529142412

mean absolute percentage error : 0.324808216218696

LinearRegression() has an accuracy score of 64.0

* Decision Tree

mean squared error : 4700.484612447499

Root mean squared error : 68.56008031243472

R2 score : -0.3512366752667442

DecisionTreeRegressor() has an accuracy score of 38.0

* Random Forest

mean squared error : 2979.5128416113025

Root mean squared error : 54.58491404785119

R2 score : 0.14348681934795016

RandomForestRegressor() has an accuracy score of 69.0

* KNN

mean squared error : 3464.5363940435277

Root mean squared error : 58.860312554755666

R2 score : 0.004058299429163581

KNeighborsRegressor() has an accuracy score of 63.0

* Gradient Boost

training RMSE: 31.68462295682572

testing RMSE: 64.59557313996085

Although there is a difference between the performances on training and test sets, the model is good

GradientBoostingRegressor(max\_features='sqrt', n\_estimators=1000,

random\_state=42) has an accuracy score of 69.0

Performing k-fold cross validation for a more impartial performance

average RMSE: 38.57581340862455

The average RMSE (38.57) is better than the RMSE (64.59) obtained in the testing RMSE.

Hence by dividing the dataset into training and testing more times we can achieve better performance for the model.

* Evaluating the feature importances

These are scores that represent how each feature variable contributes to the model.

* Time\_cycle is the feature that contributes most to the model
* Performance of the model can be further enhanced if the least important oper\_set1 and oper\_set2 are avoided.

# My learnings

* This project has offered me a holistic growth in understanding how IIOT industry works and also predictive maintenance in particular.
* This journey has provided me an exposure to what predictive maintenance is and how it can be applied in industries of various sort.
* I got to know about what is Remaining Useful Life of a machinery.
* I was not familiar with a lot of machine learning models especially ensemble models
* Boosting is technique that combines weak learners and convert them to strong learners.
* Out of all the features present in the dataset, hardly 13 features were important, in fact dropping a few unimportant ones will enhance the performance of the model.
* I also realized that this particular project need not be limited to the models I have employed. Deep learning models would give better results.
* Applying dimensionality reduction techniques will definitely improve the accuracy of the model.
* Neural networks can provide a better accuracy.

# Future work scope

* Dimensionality of the dataset could be reduced using dimensionality reduction techniques like principal component analysis. This will improve the accuracy.
* Support vector regression is another regression model that might give better results.
* Neural networks like LSTM could also be implemented for better performance.
* Due to the time constraint, here I have utilized only the fleet of engines with

Conditions: ONE (Sea Level)

Fault Modes: ONE (HPC Degradation)

The same implementation can be extended to the rest of the engines.