Predictive Maintenance

Objective:

The main objective of these project is to perform extensive Exploratory Data Analysis on the Aerospace domain Dataset and build appropriate machine learning that will predict RUL(Remaining time).

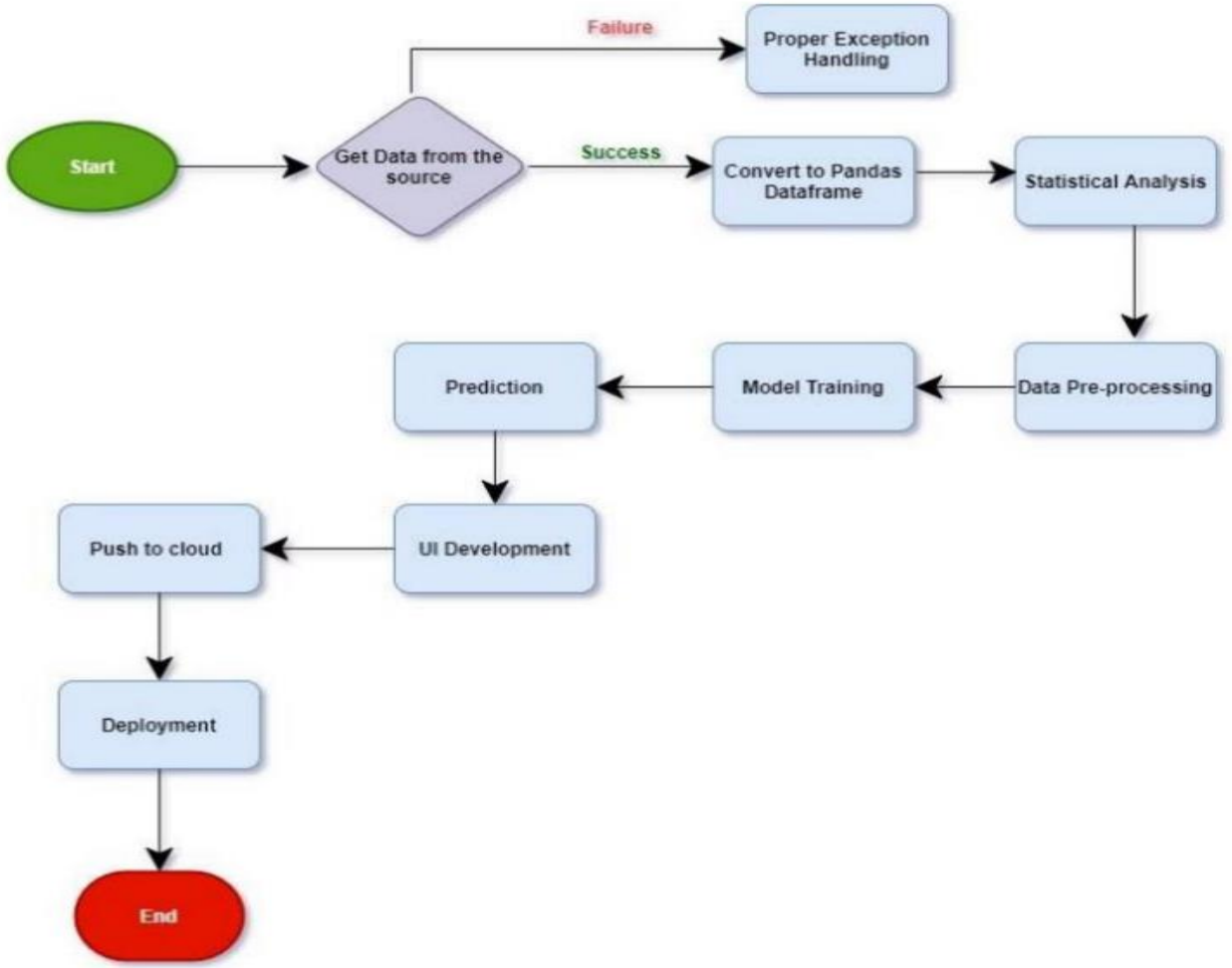
Benefits:

1. Using Organization data into real world Business –case.
2. Predicting remaining useful life.
3. Helps increase profits to organization.
4. Optimum services provided by app.

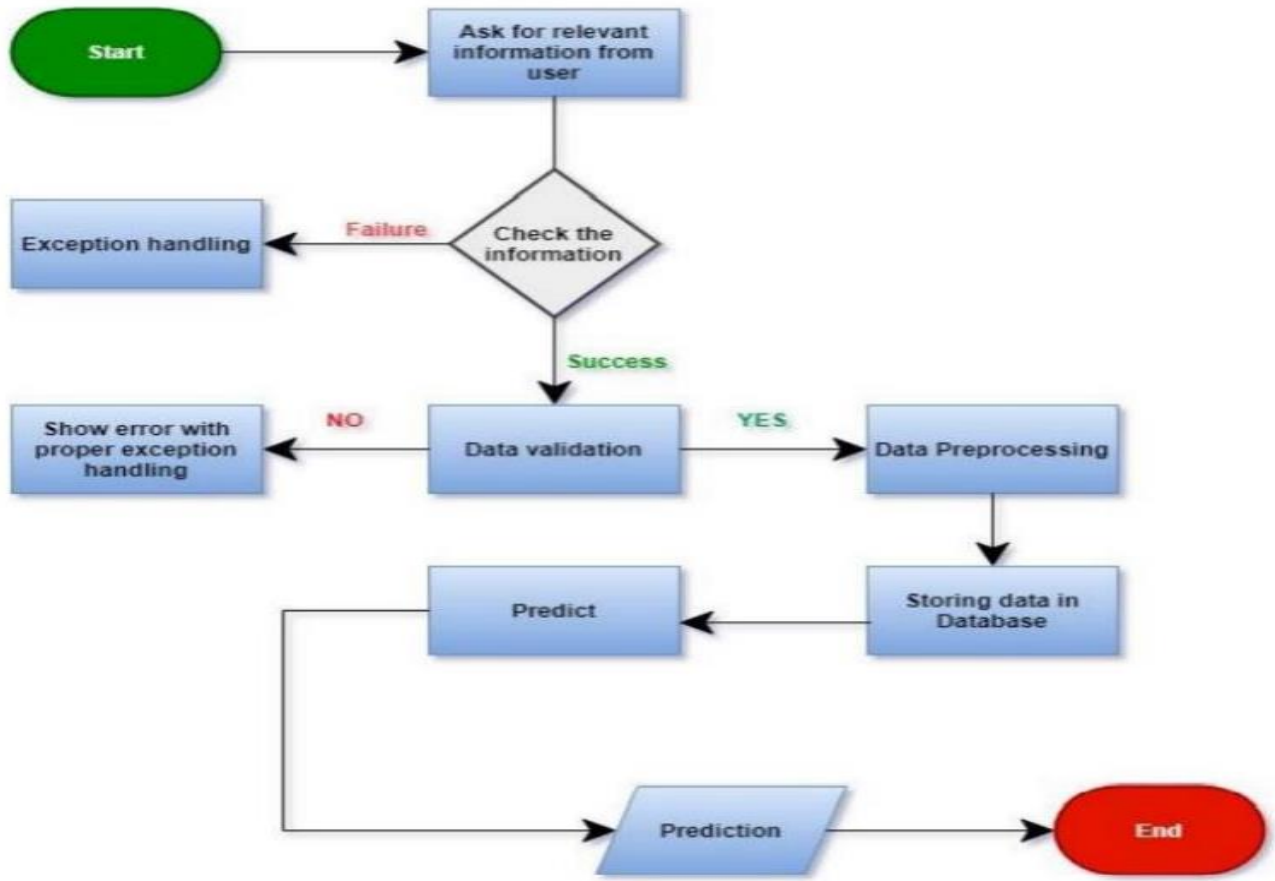
Data:

source_file_is
https://www.kaggle.com/datasets/behrad3d/nasa-cmaps
shape of the data 20,630*23
20630 rows
23 columns

Architecture
Machine Learning Model



6. User I/O Workflow



Model Training:

* Data exported from csv:
Loading csv data using python pandas and extracting all data into frame and calculating Rul.

Data Preprocessing:

*Performing EDA to get insight of data like identifying distribution, outliers.
*Check the NULL values in the columns. If present remove them.
*Perform Feature selection and extract all the necessary features from data.

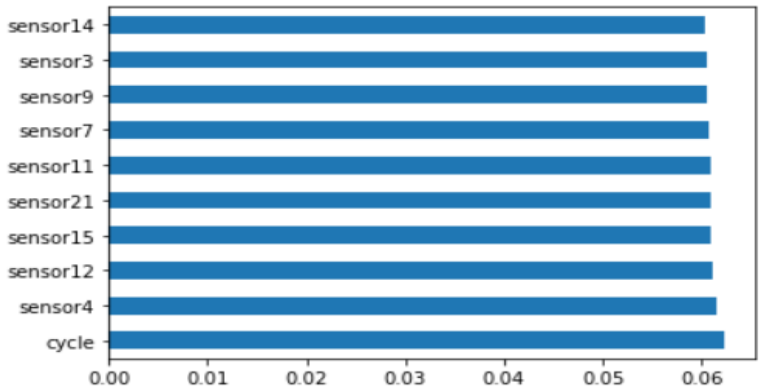
Feature Selection:

```
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
```

```
X = data.drop('RUL', axis=1)
y = data.RUL
```

```
from sklearn.ensemble import ExtraTreesClassifier
import matplotlib.pyplot as plt
model = ExtraTreesClassifier()
model.fit(X,y)
print(model.feature_importances_) #use inbuilt class feature_importances of tree based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```

```
[0.05694353 0.06224131 0.05528592 0.         0.         0.06032026
 0.06043793 0.06157118 0.         0.00269742 0.060603   0.05657214
 0.06052612 0.         0.06083329 0.061048   0.05646868 0.06033904
 0.06091653 0.         0.04259616 0.         0.         0.05968605
 0.06091343]
```



Train and Test Split:

* Train data 75% of whole data.
* Test data 25% of whole data.
* Data is randomly spitted.
* There is only train and test data.

Model Selection:

```
from sklearn.ensemble import AdaBoostClassifier
```

```
ada = AdaBoostClassifier()
ada.fit(X_train, y_train)
```

```
AdaBoostClassifier()
```

```
y_pred = ada.predict(X_test)
```

```
print(accuracy_score(y_test,y_pred))
```

```
0.007523825447249624
```

Q&A

1. What is the source of the Data.

A: The data for training is provided by client (ineuron) in form of csv and source of file.

2. What are types of data.

A: The combined of numerical values. There are no Null values.

3. What the complete flow you followed in this project.

A: Refer slide 4th and 5th slide for better understanding.

4.How are logs managed.
A: We are using different logs as per the steps that we follow in validation and modelling like file validation, Data insertion, Model Training, Prediction Log.

5. What are techniques were you using Data-Preprocessing.
A: * visualizing relation of independent variables with each other and output variables.
* Removing unwanted features.
* Checking and changing of distribution of data
* Cleaning the data

6. How training was done and model was used.

A: * Before diving the data in training and test set we performed pre-processing order to get better.

* As per the model the training and test are valid.

7. What are different stages of deployment.

A: * When model is ready we deployed it local environment, where UAT is perfomed.

* Then project uploaded in GitHub account.

* Deployed in AWS Cloud Platform.