FSM Online Internship Report

Project Title: Remaining Usable Life Estimation(Bearing Dataset)

Domain: Machine Learning

Phase1: Exploratory Data Analysis

Submitted by:

Akanksha Singh

Thapar Institute Of Engineering & Technology

Patiala, Punjab - 147004

Under Mentorship of:

Devesh Tarasia

IITD-AIA Foundation for Smart Manufacturing

[1-June-2022 to 31-July-2022]

Remaining Usable Life Estimation (Bearing Dataset)

About Dataset:

The experiments from which the data set is derived involved three different loading conditions. Condition 1 had seven ball bearings operated at 1800 rpm with 4000 N radial load. Complete run-to-failure data for algorithm training are provided for two of the bearings, and truncated data for algorithm testing were provided for the other five. Condition 2 featured seven ball bearings operated at 1650 rpm with 4200 N radial load. Of the seven bearings, complete run-to-failure data for training were provided for two, and data for testing were provided for five of the bearings. Condition 3 featured three bearings operated at 1500 rpm with a 5000 N radial load. Data from two of the bearings were provided for training and data from the other bearing was provided for testing.

Data collected was segregated into the Learning and Test Dataset. For this project, the Learning Set includes operational data (time-series of horizontal and vertical measurements) from 6 different bearings, more exactly, 2 bearings for each of 3 groups exposed to different operating conditions. In each Learning Data Set, the unknown bearing was run for a variable time until failure. The lengths of the runs varied, with the minimum run length of 5,150 sec and the maximum length of 28,030 sec. The Test Set included operational data from bearing3 from the first operating conditions. In each Test Data Set, the unknown bearing was run for a variable time until the failure, but researchers have got only truncated time series.

a) Two accelerometers were mounted on the bearing housing to measure vibration in the vertical and horizontal directions.
b) Vibration measurements were performed every 10 sec at 25.6 kHz sampling rate and 0.1 s duration; hence, each observation contained 2560 points. (2,560 measurements per 0.1 sec).
c) The Learning and Test Sets also included temperature measurements (with a period 0.1 sec), but only for 10 bearings from all 17.

For the time being, we will use acceleration vibrational data only and will neglect have used temperature data. The reason for the same will be explained in the upcoming study.

```
/content/drive/MyDrive/Dataset/Learning_set   ['Bearing2_2', 'Bearing1_1', 'Bearing3_1', 'Bearing1_2', 'Bearing2_1', 'Bearing3_2']   []
total file  0
acc files :  0  | temp files :  0

/content/drive/MyDrive/Dataset/Learning_set/Bearing2_2   []   ['acc_00002.csv', 'acc_00008.csv', 'acc_00004.csv', 'acc_00007.csv', 'acc_00001.csv', 'acc_00006.csv', 'acc_00009.csv', 'a
total file  797
acc files :  797  | temp files :  0

/content/drive/MyDrive/Dataset/Learning_set/Bearing1_1   []   ['acc_02271.csv', 'acc_02270.csv', 'acc_02240.csv', 'acc_02234.csv', 'acc_02260.csv', 'acc_02275.csv', 'acc_02247.csv', 'a
total file  3269
acc files :  2803  | temp files :  466

/content/drive/MyDrive/Dataset/Learning_set/Bearing3_1   []   ['acc_00013.csv', 'acc_00028.csv', 'acc_00036.csv', 'acc_00009.csv', 'acc_00004.csv', 'acc_00007.csv', 'acc_00029.csv', 'a
total file  604
acc files :  515  | temp files :  89

/content/drive/MyDrive/Dataset/Learning_set/Bearing1_2   []   ['acc_00013.csv', 'acc_00026.csv', 'acc_00027.csv', 'acc_00045.csv', 'acc_00036.csv', 'acc_00010.csv', 'acc_00039.csv', 'a
total file  1015
acc files :  871  | temp files :  144

/content/drive/MyDrive/Dataset/Learning_set/Bearing2_1   []   ['acc_00031.csv', 'acc_00075.csv', 'acc_00061.csv', 'acc_00074.csv', 'acc_00059.csv', 'acc_00072.csv', 'acc_00026.csv', 'a
total file  1062
acc files :  911  | temp files :  151

/content/drive/MyDrive/Dataset/Learning_set/Bearing3_2   []   ['acc_00605.csv', 'acc_00620.csv', 'acc_00627.csv', 'acc_00639.csv', 'acc_00606.csv', 'acc_00597.csv', 'acc_00612.csv', 'a
total file  1637
acc files :  1637  | temp files :  0
```

Fig1. Snippet of no. of files in each Bearing folder

Data Storing:

Processing of each file every time when data is required is time-consuming as well as a hectic process. So, to make easier our work easier, all datafiles in each dataset are merged together. At the end, we get 6 Bearing Dataset files.

```
file merge successful Bearing2_2  csv file created
Bearing2_2 .pkz file created
file merge successful Bearing1_1  csv file created
Bearing1_1 .pkz file created
file merge successful Bearing3_1  csv file created
Bearing3_1 .pkz file created
file merge successful Bearing1_2  csv file created
Bearing1_2 .pkz file created
file merge successful Bearing2_1  csv file created
Bearing2_1 .pkz file created
file merge successful Bearing3_2  csv file created
Bearing3_2 .pkz file created
```

Fig2. Snippet of 6 Bearing folders after merging files

```
Bearing1_1 files with not 600 datapoints 1
Bearing3_1 files with not 600 datapoints 2
Bearing1_2 files with not 600 datapoints 1
Bearing2_1 files with not 600 datapoints 1
```

Fig3. Snippet of Bearing with not 600 datapoints of temperature dataset

HOW is our Dataset a Time-Series?

Time-series data is a set of observations i.e. data usually collected at discrete and equally spaced time intervals. Here, in our collected dataset, vibration signals are collected every 10 seconds which is a different time and equal time intervals. Hence our data is time-series data.

WHY 'PICKLE' FOR STORAGE INSTEAD OF CSV

Pickle is 80 times faster alternative. It's also 2.5 times lighter and offers functionality every data scientist must know. Storing data in the cloud can cost you a pretty penny. Naturally, you'll want to stay away from the most widely known storage format – CSV – and pick something a little lighter. Pickling isn't limited to datasets only.

Why not use Temperature Dataset

- Data in temperature measurement files are either semicolons or commas or dot-separated. This created problem in combining the files into one CSV file and then converting them to a pickle file. [But later on, the issue was resolved]
- The number of data points in a few bearing datasets was different Fig3
- The data points were not sufficient enough to train the model. It can be clearly observed in Fig 1 that in Bearing2_2 and Bearing3_2 there are 0 temp files.
- All the graph have shows almost constant temperature change at the end except Bearing1_1

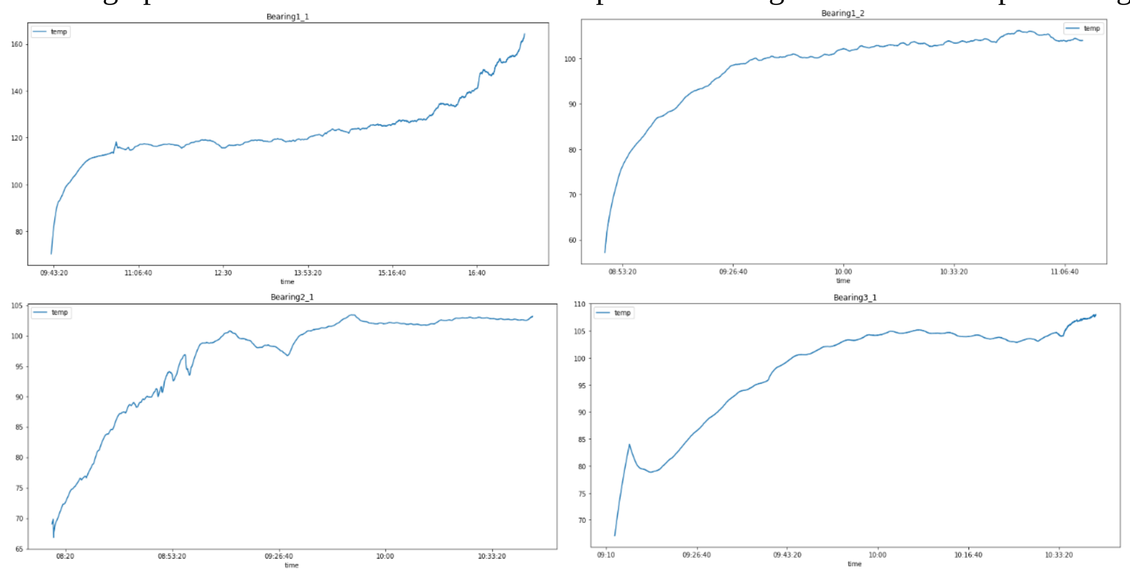

Fig4. Time-Series Temperature Signal Plot of Bearings

```
7175680 : Bearing1_1
True

2229760 : Bearing1_2
True

2332160 : Bearing2_1
True

2040320 : Bearing2_2
True

1318400 : Bearing3_1
True

4190720 : Bearing3_2
True
```

Fig5. No. of datapoints in each Bearing dataset for Data Acquisition

Data Pre-processing:

In each bearing dataset, there are 6 columns

`['hour', 'minute', 'second', 'microsecond', 'horiz accel', 'vert accel']`

Since our dataset is time-series so time column was created by merging first four column and then the time column was converted to DateTime format

Now there are 3 columns `['time', 'horiz accel', 'vert accel']`

Data Exploration:

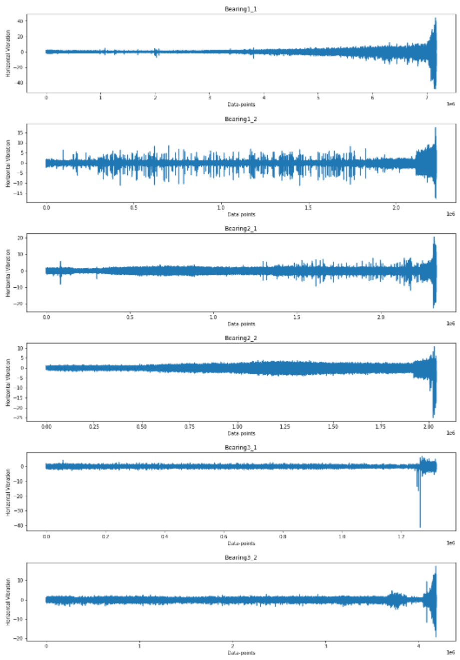

Fig6. Time-Series Horizontal Vibration Plot of Bearings

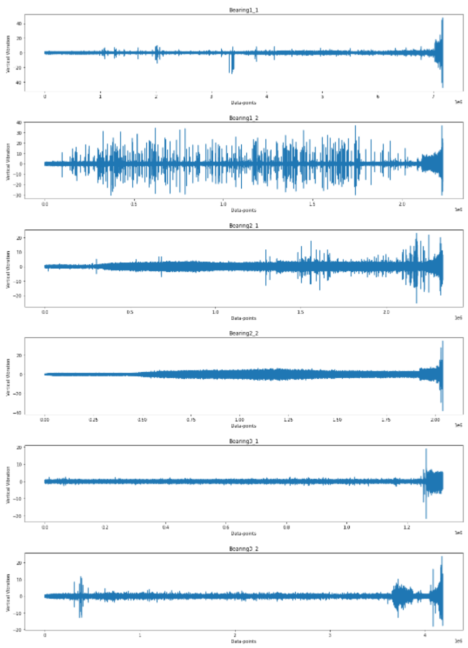

Fig7. . Time-Series Horizontal Vibration Plot of Bearings

From the plot in Figure5, Bearing1_1 it is clear, that until the last 1/3 of the full time-series the trendability of the vibration behavior is absent. Similarly in the remaining plots. Given this “non-trendability” in the data, it is impossible to easily observe any trends in the data seeking for possible correlation to the degradation in the bearing. The reason for the loss of trendability is the abnormal conditions of bearing working – large radial force applied on the bearing (4000 N in the first conditions, 4200 N in the 2nd operating conditions & 5000 N in the third operating conditions). Due to such abnormal conditions, the bearings’ time lives were only 1 to 7 hours. We can also observe that the measured data is very noisy, it is impossible to use them directly. Therefore, the first task is to perform de-noising. There are 2 different ways to perform de-noising: Trendability dependent de-noising & Trendability independent de-noising.Most of the data-driven RUL predicted methods are oriented for Trendability Statistics. Unfortunately, for the non-trendable and non-periodical statistics other wide-known prognostics models are not applicable.

Getting no good result from the time-series plot of raw data, next the goal is to squeeze the long sequence of large data points to extract some meaningful features out of it. So using this series data we get features such as Max Value, Min. Value, Standard Deviation, RMS Value, Kurtosis Value, Skewness Value, Crest Factor & Form Factor. So these 9 time domain features will get extract out of this (no. of rows in each bearing) datapoints

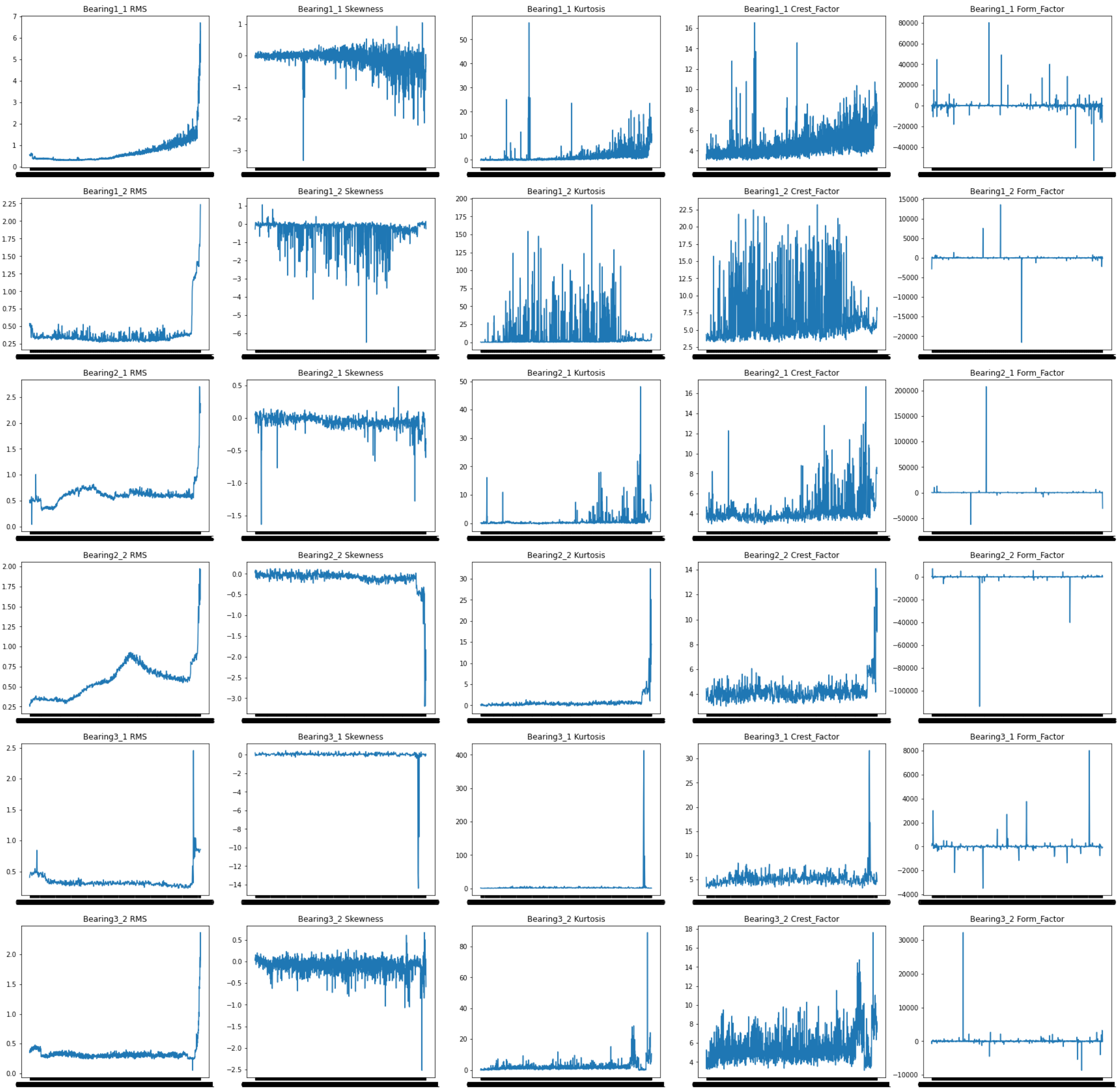

Fig8. Plots of Time Domain Features of Horizontal Acceleration Data

Fig8 Shows the values and trends of 'RMS','Skewness','Kurtosis','Crest_Factor','Form_Factor' for 6 bearing instances under the operating conditions of 1800 rpm & 4000N, 1650 rpm & 4200 N, 1500 rpm & 5000 N (2 each).

They were calculated for every time cycle. However, it is apparent that not a single parameter demonstrates consistent patterns across all 6 instances as bearing faults progress. Thus, it is difficult to derive the RUL estimation model based on these statistical parameters.

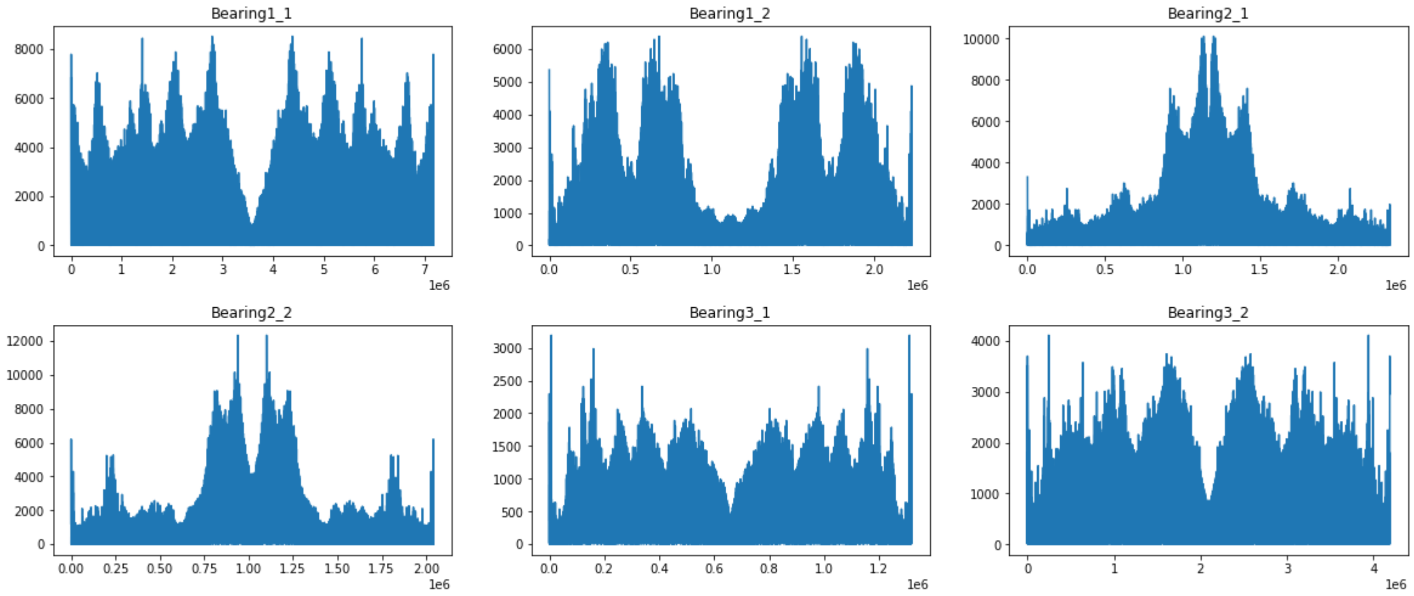

Fig9. Plots of Fast Fourier Transform of Horizontal Acceleration Data

IMPLEMENTATION:

1. The entire exploratory analysis is performed in a Collaboratory platform using python programming
2. For easy data access & faster data retrieval the CSV files were converted to pickle format
3. The Data Files in each dataset are very combined and stored in a pickle format. As a result, we get 6 total datasets for preprocessing. Total number of rows in each data set = no. of files in each dataset * 2560
4. Date time columns were combined into time column and converted to datetime format
5. Plots of Raw Data
6. Time-Domain Features were calculated and graphs were plotted

Code Implementation :

Data Acquisition
https://colab.research.google.com/drive/1Fn6XKVbQKvGmO9H2xZhRweTtaBH9jHm5?usp=sharing

Data Preprocessing
https://colab.research.google.com/drive/16Pw5r-WN3psy-hVIZWhDRB0NvnPcmopd?usp=sharing