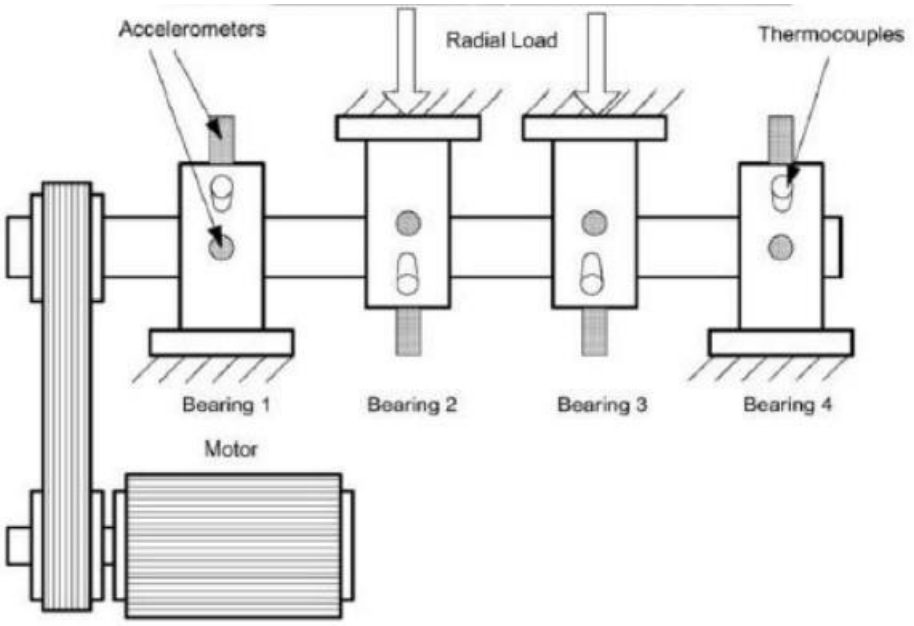
Bearings anomaly detection in a motor

1 Use Case

The case of use is attributable to an motor of any type and a eventually axile that transmits the movement. The motor under certain stress or bad working will can broke and a prevenction of this event can be good for decrease an eventually more damage and manage a future imminent or lately maintenance.

2 Architectural Components Overview



NSF I/UCR Center for Intelligent Maintenance Systems with support from Rexnord Corp. in Milwaukee, WI.

3.1 Data Source

The data was generated by the NSF I/UCR Center for Intelligent Maintenance Systems (IMS – www.imscenter.net) with support from Rexnord Corp. in Milwaukee, WI.
Four bearings were installed on a shaft. The rotation speed was kept constant at 2000 RPM by an AC
motor coupled to the shaft via rub belts. A radial load of 6000 lbs is applied onto the shaft and bearing
by a spring mechanism. All bearings are force lubricated.

Rexnord ZA-2115 double row bearings were installed on the shaft as shown in Figure 1. PCB 353B33
High Sensitivity Quartz ICP accelerometers were installed on the bearing housing (two accelerometers
for each bearing [x- and y-axes].
Sensor placement is also shown in Figure 1. All failures occurred after exceeding designed life time of
the bearing which is more than 100 million revolutions. The files totally are 2156, produced each one every 10 minutes.

3.1.1 Technology Choice

The data sets are included in the data packet (IMS-Rexnord Bearing Data.zip). Each data set describes a test-to-failure experiment. Each data set consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of 20,480 points with the sampling rate set at 20 kHz. The file name indicates when the data was collected. Each record (row) in the data file is a data point. Data collection was facilitated by NI DAQ Card 6062E. Larger intervals of time stamps (showed in file names) indicate resumption of the experiment in the next working day. The files are in TXT and need to be renamed for clear number of sequences, this will be doing with a batch script in Windows Powershell. The data will be formatted as CSV with title of every field adding a comma after every measure and this will be doing in a Python Notebook, for a better management with a dataframe.

3.1 Streaming analytics

No streaming analytics is used because it is used a pre-created data in laboratory.

4.1 Data Integration

The historical datasets are supervised by searching any invalid values, irregular cardinalities and missing values. The data appear have a clean scheme: every filed is spaced only by a TAB character. The data is complete with no NULL values. The (PCA) to reduce the dimensions of the descriptive features is useless because there are only 2 field to analyse for bearing and these are essential. The final result is a field of a serial unique number for row, a field for a file number, and a field of X and Y movement.

4.1.1 Technology Choice

The platform where it operates is Jupyter Notebook and IBM for Data Pak for simplicity and ease of use. For formatting and modify data it is installed some packages that help this work: Python Pandas, Python Numpy.

4.1.2 Justification

Due to the simple nature of historical data in csv-format, it is sufficient to use popular libraries in terms of data cleansing and integration.

5.1 Data Repository

The data repository for convenience of time is loaded from local for a quick test from different scenario. In future it is possible to add frameworks for large dataset for an object-storage.

5.1.1 Technology Choice

In a Jupiter Notebook we can choose after Data Integration if it can save CSV file of a entire dataset or work only with a dataframe.

5.1.2 Justification

For a very low time available

6.1 Discovery and Exploration

In this part attempts to get an general view of the data and find some anomalies or similarities form data. For helping this purpose, it can be using the FFT (Fast Fourier Transformation) that transform the domain of the data for every axis of the data and plot with a Box Plot for a comparison with MatplotLib.

6.1.1. Technology Choice

MatplotLib with FFT data.

7.1 Actionable Insights

It can be hard to observe anomalies with a view of a 20000 samplings windows and Machine Learning and Deep Learning as self-learning algorithms can deal with this classification problem. As a baseline, we use a multi-layer (two) system of Machine Leaning with start layer as LTSM with two inputs and a dense layer that exits the two output. For Replacement Optimization Algorithm the choice is Adam and the Mean Absolute Error for calculating the loss. In this system is not provided the dataset of a future value and for this we provide in input and the same dataset for output for a reference. The system is deliberately with few neurons and timesteps for a better visualization of the loss in the graph.

7.1.1 Technology Choice

The technology choice follows the classic configuration as Keras for setting the ML and Tensorflow for some features.

7.1.2 Justification

It is used this configuration and technology for this example because it doesn't use a large dataset

8.1 Applications / Data Products

The application of this system is for a anomalies detector that indicate through the MAE a problem within the bearings on advancing of time. In a normal situation with a dataset that indicate a bearing in a good condition the loss is very low because the ML system have find a "path" to rebuild the output (the same of input) after some time of training but if we add after some consolidated time a data of broken bearing we see that the error change and grow because the system have to adapt to the new output, and this indicate an anomalies.

8.1.1 Technology Choice

MAE and MatplotLib for identifying an anomaly and dataframe for splitting the data from healthy and faulty for see the differences from apply alone or together

9.1 Security, Information Governance and Systems Management

It doesn't need some particular attention for security because is stand-alone system and no connected to any external system, the data also are taken from a local system. It needs to repeat some times the experiment for a good choice of a quantity of neurons to apply and timesteps.