8

APPLICATION OF BIG DATA ANALYTICS

Izzatdin Abdul Aziz, Amir Ali

The use of machine learning in engineering applications varies depending on the problem's complexity. Nonetheless, the machine learning foundation is essentially the same. Then comes data prep, modelling, and outcome validation. We said before in this book that data science is the synthesis of three domains: computing, statistics, and applications. This chapter describes the procedures used to apply data analytics to a real-world oil and gas issue, predicting machine failure based on known past failure instances. Time To Fail is the core of failure (TTF). This chapter's case study is based on an offshore gas compressor. Field managers and offshore and onshore operation engineers with extensive expertise in managing vital equipment in oil and gas platforms, particularly gas compressors, assist in this task. This chapter assumes the dataset has been thoroughly pre-processed. Pre-processing data was discussed previously in this book's chapter. In this case, the data preparation stages are given briefly.

8.1 BACKGROUND

In this case study, the subjected equipment is a gas turbine-powered centrifugal natural gas compressor. This case study refers to two states of equipment. Normal (running) and Failed (failed) (unplanned shutdown condition). Many sensors are mounted on rotating machinery to ensure reliability and safety. Figure 8.1 show a typical unit schematic flow diagram.

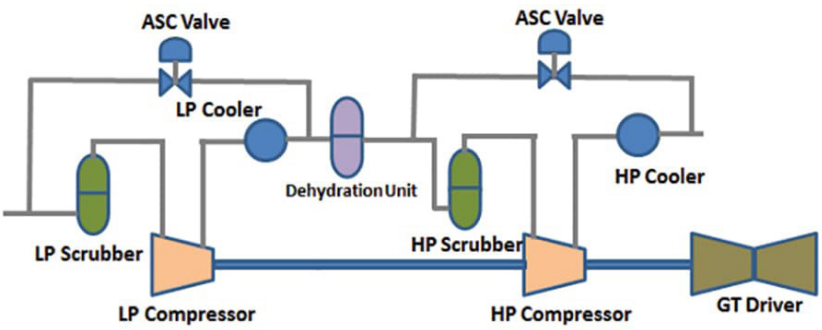

Figure 8.1 Schematic diagram of centrifugal compressor

Classification and regression analytics are used to forecast failure. Pressure, vibration, stress, flow, and temperature sensors are used to monitor equipment health (s). These sensors capture numerical data in time series. An item of equipment may be in two states: failure or normal. Time series prediction of failure is a regression issue. Sensors are usually represented as Tags, which is an attribute or column inside a dataset. The Figure 8.2 shows an example of sensors in a compressor system.

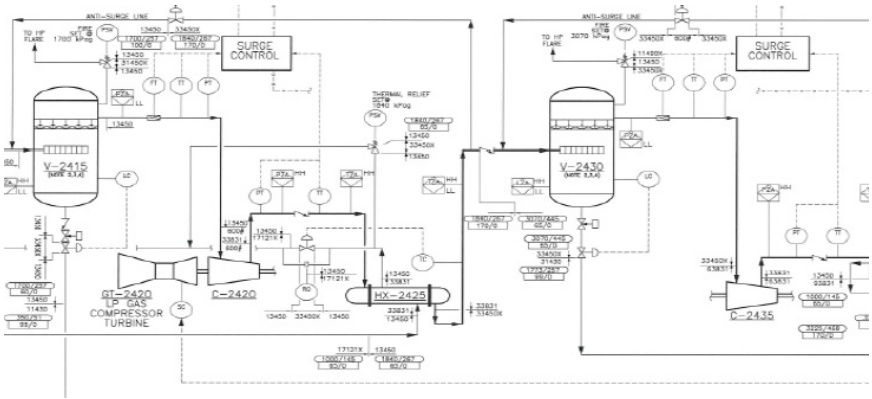

Figure 8.2 Typical sensors for centrifugal compressor module

This chapter outlines the approach used to deal with a large dataset from an oil and gas firm. There were 7000 tags or columns per 300 billion rows of data gathered over two years from a centralised database. Each tag has a Safe Operating Limit (SOL) table that shows the lowest and maximum thresholds for a sensor.

This historical dataset analyses the health of a gas compressor on an offshore oil rig that has had multiple failures and trips over the last two decades. The gas compressor in this case