**Approach Document**

**Predictive Maintenance of machines in Industry**

* **Problem Description**

A major problem faced by businesses in asset-heavy industries such as manufacturing is the significant costs that are associated with delays in the production process due to mechanical problems. Most of these businesses are interested in predicting these problems in advance so that they can proactively prevent the problems before they occur which will reduce the costly impact caused by downtime.

* **Data Sources**

Common data sources for predictive maintenance problems are

1. Failure history: The failure history of a machine or component within the machine.
2. Maintenance history: The repair history of a machine, e.g. error codes, previous maintenance activities or component replacements.
3. Machine conditions and usage: The operating conditions of a machine e.g. data collected from sensors.
4. Machine features: The features of a machine, e.g. engine size, make and model, location.
5. Operator features: The features of the operator, e.g. gender, past experience.

The data for this example comes from 4 different sources which are real-time telemetry data collected from machines, error messages, historical maintenance records that include failures and machine information such as type and age.

* **Feature Engineering**

The first step in predictive maintenance applications is feature engineering which requires bringing the different data sources together to create features that best describe a machines’ health condition at a given point in time.

## Label Construction

## The prediction problem for this example scenario is to estimate the probability that a machine will fail in the near future due to a failure of a certain component.

## More specifically, the goal is to compute the probability that a machine will fail in the next 24 hours due to a certain component failure (component 1,2,3 or 4). In the following, labelling is done by labelling all the feature records that fall into the 24 hours’ window before a failure due to component 1, component 2, component 3 and component 4 as comp1, comp2, comp3 and comp4 respectively.

## The rest of the records are labelled as "none" indicating, there is no failure within the next 24 hours.

## Training, Validation and Testing

## For predictive maintenance problems, a time-dependent splitting strategy is often a better approach to estimate performance which is done by validating and testing on examples that are later in time than the training examples.