Contents

Predicting the potential maintenance opportunity of Machines

Problem Description

Predictive maintenance of machinery is very critical from both cost and safety perspectives. Cost of serving/replacing a perfectly working part can be significant and can add huge financial burden to manufacturers and customers besides causing machine down time. On the other hand, cost of not-maintaining a failed part can result in human safety disasters. Thus, predicting events related to maintenance based on past events till failure helps avoid breakdown and achieve full useful life and thus availability of systems. The data on failures or trends about degradation of a machine over time can help in predicting a potential failure based on patterns and helps estimate time to a failure based on the progression of identified faults. Working backwards the data can help develop early warning signals to plan for maintenance which includes corrective actions

A locomotive manufacturing company observes failures in the engines of several locomotives of their fleet. The data captured through sensors and inspection of previous fleet is given. The data depicts the scheduled and unexpected services done for each machine component. Unexpected services include error, repair, replacement etc and related data is provided with time stamps. You are expected to create an analytical and modelling framework to predict the major possible future events/ action points for each MachineID ie., like any “ComponentRepair”(major), “ComponentReplacement”, “NoIssue” (Minor error or normal status) , in the next one month for each machine. And you are also required to obtain the actionable top 10 data insights (patterns) for each of “ComponentRepair”, “ComponentReplacement” class levels using the tree based algorithms.

About Data:
The data provides the information on the condition of the various components of the machine including the data collected through sensors. The data depicts the scheduled and unexpected services that took place for each machine as historical data.

Every single machine is observed at various time stamps, to record the details of its operating condition like any error, repair, scheduled maintenance, replacement etc for each component of that specific machine.

Objective:

You are expected to create an analytical and modelling framework to predict the future events/ action points based on known past events/action points for each MachineID ie., what would be the possible major action point, like any “ComponentRepair”, “ComponentReplacement”, “NoIssue” , in next one month for each machine. And also obtain the actionable top 10 data insights (patterns)

for each of “ComponentRepair”, “ComponentReplacement” classes using the tree based algorithms.

Details of the datasets

The datasets are provided as cited below for the analysis:

1. **Train & Test Data:**
   - **“Train.csv” & “Test.csv”**
   - These files consist of the MachineID & ActionPoint Which is the target attribute.
   - Train.csv has the target attribute “ActionPoint” also, whereas Tes.csv doesn’t have as it has to be predicted.

Note: the datasets cited below consists of the data related to train and test data MachineIDs

2. **MachineDetails data :**
   - “MachinesDetails.csv”
   - This file depicts the MachineID, model, it’s service period for both train and test MachineIDs

3. **Minor Error information:**
   - “ComplaintsLog.csv”
   - This file consists of the details about minor errors that took place with each machine at various time stamps. Machines are still operational though the errors occurred.

4. **Different Services Details:**
   - “ComponentServiceLog.csv”
   - This file depicts the timestamp, MachineID, ServiceType, ComponentAttended etc.

5. **Replacement Details:**
   - “ComponentReplacementLog.csv”
   - This file depicts the timestamp, MachineID and which component got replaced.

6. **Operating Conditions Details:**
   - “OperatingConditionsData.csv”
   - This file depicts the timestamp, MachineID, readings of different sensors at each point of time.

Note: For analysis, consolidate/aggregate all the datasets cited above (1) to (6)

Main Tasks:

1. Exploratory Data Analysis using visualizations in R Notebook or Jupiter notebook format (Use only Train data for this task)
2. You are expected to build a framework that predicts the ActionPoint ("ComponentRepair" or "ComponentReplacement" or "NoIssue" in target attribute "ActionPoint").
3. *You are expected to build a framework that generates the patterns for "ComponentReplacement" & "ComponentRepair" on target attribute "ActionPoint", using any decision tree algorithms only.* ***(Use only train data for this)***
4. *Viva*

Note: Use the aggregated/consolidated train dataset for model building and tuning the model and then apply the model on test data for obtaining the predictions.

Submission 1 (16th September 2018 by 6 p.m.) :

i. Preprocessing of Data

1. Data preparation for model building

ii. Visualization

1. Train data should be used for data analysis and visualizations.

iii. Documents to be submitted into Grader tool are cited below:

Visualization :

1. Commented Code developed for Preprocessing, visualisation with the name "submission.R" or "submission.ipynb" as the case may be.
2. Report on your understanding about the problem, data analysis based on the visualizations you made and pre- processing/feature engineering/aggregations steps required for model building with the name "Report.doc".

Submission 2 (18th September 2018 by 11.30 pm):

i. Model Building and predictions

i. Train data should be used not only to build the model but also to tune and conclude the model for submission
ii. Test data should be used for evaluation of model
iii. Test data is totally unseen data and hence it does not have the target attribute, "ActionPoint".
iv. The predictions obtained for test dataset should be uploaded to Grader tool with the name "predictions.csv".

ii. Documents to be submitted into Grader tool are cited below:

Visualization & Modelling:

1. Commented Code developed for Preprocessing, visualisation and modelling with the name "submission.R" or "submission.ipynb" as the case may be.
2. Upload the predictions obtained on the Test dataset to Grader tool with the name "predictions.csv".

Submission 3 (21st September 2018 by 8 p.m.):

i. Generating the patterns using any decision tree algorithm only:

i. Traindata should be used for generating the patterns
ii. Generate the patterns for Replacement & Repair levels (ie., "ComponentReplacement" & "ComponentRepair" levels in the target attribute) in data and also compute evaluation metrics for patterns

ii. Documents to be submitted into Grader tool are cited below:

a. Improved versions of Submission 1 & 2

i. Final Commented R code / python code, predictions, patterns with additional efforts and improvements made during the week

b. Pattern Extraction

i.Upload the top 10 patterns for Repair & Replacement - ActionPoint (ie., "ComponentReplacement" & "ComponentRepair" levels in the target attribute) to the Grader tool with the name "patterns.csv"

c. Viva Presentation

i.Final presentation for viva with the name "viva.ppt"

Note: Please follow the naming convention for the submission files as cited above only.

Error Metrics:

- Consider "F1-statistic" for "ComponentReplacement" level of Target attribute as error metric and tune the model accordingly.
- Consider appropriate error metric for deciding the top 10 patterns for "ComponentReplacement" & "ComponentRepair" levels of target attribute