**Predictive Maintenance: Failure Prediction Model**

Ibrahim Asiri

**Abstract**

The goal of this project was to use machine learning models to predict the failure and failure types for machinery. The dataset is about group of process parameters (features) like temperature and torque…etc. and failure and failure types of columns were my targets. After simple exploratory data analysis and resolving imbalance by over-sampling, logistic regression and random forest models were built to predict failures and classify failure modes. The results were acceptable for logistic regression despite features collinearity, but random forest model was outstanding for failure predictions. For failure types, logistic regression was having clear issue for two classes (six classes in total) while random forest was coming back with 99% F1-score. The model needs more complicity to it but for now it did what is intended to, and more checks were done by cross validation (Kfolds=10) and return similar results.

**Design**

This project originates from UCI Machine Learning Repository for California university as referenced below in the data. The dataset is about group of process parameters (features) like temperature and torque…etc. and “failure” and “failure types” columns were my targets. Predicting failure and classify failure types via machine learning models (logistic regression, random forest). The results may help to reduce the down time for the machine, spare parts inventory strategy and optimize Maintenace cost by predicting the failures.

**Data**

The dataset consists of 10 000 data points stored as rows with 14 columns:

**Table

Description automatically generated**

### **Acknowledgements**

UCI : [**https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset**](https://archive.ics.uci.edu/ml/datasets/AI4I+2020+Predictive+Maintenance+Dataset)

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

**Algorithms**

*Feature Engineering*

1. Failure and failure type columns to be taken out from features all the time to avoid leakage in the model.
2. Converting categorical features (Type, Failure type) by LabelEncoding.
3. Over-sampling using Smote would be required since the data is imbalanced (96.5% no failure ,3.5% failure happened)
4. Outliers in rotational speed and torque

*Models*

logistic regression and random forest models were built to predict failures and classify failure modes. The results were acceptable for logistic regression despite features collinearity, but random forest model was outstanding for failure predictions. For failure types, logistic regression was having clear issue for two classes (six classes in total) while random forest was coming back with 99% F1-score. The model needs more complicity to it but for now it did what is intended to, and more checks were done by cross validation (Kfolds=10) and return similar results.

*Models Evaluation:*Table

Description automatically generated

Table

Description automatically generated

**Tools**

* Numpy and Pandas for data manipulation
* Imblearn for oversampling
* Scikit-learn for modeling
* Matplotlib and Seaborn for plotting

**Communication**

Predictive maintenance goes beyond time-scheduled maintenance to condition based action to predict the likelihood of future failures by applying machine learning and data analytics to reduce asset failures and their costs. This is trending in the industrial community to use sensors to monitor the live data of equipment and machinery, then continuously evaluates it against historical trends to predict failure before it occurs.

The major issue is still about the data storage and accessibility since most of these data is commercial secrets for licensers or even the manufacturers of the equipment. You can find these files on [my GitHub](https://github.com/asirim00/Machine-Predictive-Maintenance-modeling-) along with other files.