

02935 Introduction to Applied Statistics Project Description

Emil Njor

December 23, 2021

The topic of my PhD project is on-device Predictive Maintenance (PdM). This involves running machine learning inference on tiny microcontrollers to predict machine failure. I have currently been working on the PhD project for about 4 months.

I plan to use this course to acquire the skills required to compare the results of my work to the works of others in my field.

PdM predictions can be split into two categories. The first is a binary prediction, which uses input to predict whether or not there is an imminent or current failure. The second is a continuous prediction, which uses input to give a prediction of the Remaining Useful Life (RUL) of a machine.

I believe that the second option makes room for some more interesting statistics, as the first is only concerned with binary variables. I have identified a data set called CMAPSSData that belongs in the second category. This data set contains simulated sensor readings from turbofan engines as they degrade towards failure.

A training set is provided where the engines are run to failure, and a test set is provided where data stops before reaching a failure. It is then the job of PdM models to predict the RUL of the engines from the point that the data stops. The CMAPSSData data set contain the actual RUL for the test cases.

There are some interesting considerations in processing the results of PdM models. An example is that a PdM model should predict a lower RUL than a higher as we want to avoid failures, especially in safety-critical systems.

In this project, I want to compare the RUL predictions for the CMAPSSData data set from two PdM models. The models will not be my own, as I have yet to get results from my models. See the readme.txt and dataset in the same folder as this PDF for details.

The research question will be which model performs most appropriate to the situation.