

Reduce maintenance cost through predictive techniques

Background

Company (3D Technologies) has a fleet of devices transmitting daily aggregated telemetry attributes.

Predictive maintenance techniques are designed to help determine the condition of in-service equipment in order to predict when maintenance should be performed. This approach promises cost savings over routine or time-based preventive maintenance, because tasks are performed only when warranted.

Goal

You are tasked with building a predictive model using machine learning to predict the probability of a device failure. When building this model, be sure to minimize false positives and false negatives. The column you are trying to predict is called **failure** with binary value 0 for non-failure and 1 for failure.

Code

We are looking for you to show off your coding skills using either Python or R.

Data

Download link: http://aws-proserve-data-science.s3.amazonaws.com/device_failure.csv

Metadata	
columns	description
date	date in YYYY-MM-DD format
device	device id
failure	non-failure is 0, failure is 1
attribute1 - attribute9	daily aggregated telemetry

Compute

The dataset is small enough to work with on your personal computer but you are welcome to use your \$10 AWS credit to spin up an EC2 instance. If you use AWS services to create a working solution, please use the provided coupon and shut down the services to avoid any recurring charges. You may restart the services during the phone screen if you wish to use that for any walkthroughs.

Report

Please return a converted PDF document from Markdown displaying your code and thought process including the following parts:

Introduction

Introduce this model, what's the purpose of this model and what's the potential end user or applications.

Materials & Methods

What kind of algorithm/statistics theory has been applied in your approach? Have you thought of any others? What's the pros and limitations?

Result

The result of your predictive model.

Discussion

What does the result mean? Any improvements or achievements?

Source code

Evaluation

Work Sample (70)

1. Clean Python or R code (15)
2. Data exploration through summary stats and visualizations (10)
3. Creation of novel features (5)
4. Splitting the data into train and test (10)
5. Cross validation (5)
6. Hyperparameter tuning (5)
7. Comparing performance of multiple machine learning algorithms (10)
8. Minimize false positives instead of accuracy or misclassification error (5)
9. Performance on test dataset (5)

Presentation (30)

1. Clarity of explanation (15)
2. Focus on business outcome (15)