Machine Learning

Defect Prediction on Production Line

Team:
AIT BACHIR Romuald
ALLEMAND Fabien
FLORET Arthur
JARDOT Charles

Contents

List of Figures

1 Introduction

The goal of this project is to develop an AI based solution for a real world industry problem faced by Valeo [1].

Valeo [2] is an industry leading French automotive supplier. In order to stay competitive, the company wants to develop a system that is able to identify defects on products before testing.

Four files containing relevant data are available. The data are mainly values measured during production on different mounting stations as well as additional measures performed on test benches.

The target is to find the best prediction: Output = f(inputs)

2 Data Analysis

Data is contained in four different comma-separated values (csv) files: inputs and output for both training and testing. We will take a first look a the data in order to understand what it represents, then we will proceed to an in-depth analysis of said data.

2.1 Data Description

The very first line of each file correspond to the headers. The following lines contain informations about a given product where each product is identified by a unique code:

ID = PROC_TRACEINFO = It's a unique code given to the product.
Example: I-B-XA1207672-190701-00494.

- XA1207672 is the reference.
- 190701 is the date: here 01st of July of year 2019.
- 00494 is the unique code given to the product, whatever it happens, the product will have this id number frozen forever.

This number is increased by 1 each time we process a new product, every 12 seconds. So for example: I-B-XA1207672-190701-00495 is the next product.

Input features are measures (numerical values) collected on different assembly stations with the sensors or devices connected to Programmable Logic Controllers which are storing all of them to keep the full quality traceability. Examples : OP070_V_1_angle_value, OP120_Rodage_I_value...

This is the result value of OP130 (test bench). Value 0 is assigned to OK samples (passed) and value 1 is assigned to KO samples (failed). This is the combined result of multiple electrical, acoustic and vibro-acoustic tests.

2.2 In-Depth Analysis

Let us put the testing data aside for now and take a closer look a the training data.

The training data consists of 13 measures (inputs features) made during production line on just shy of 35000 products. This measures correpond to various forces, angles, screwing torques... By taking a look at the distribution for each feature (Figure 1) we can clearly see that every one of them has a different shape.

We can see that there might be some outliers...

One feature requires more attention. capuchon_insertion is the only feature containing missing values. In fact, more than 50% of the sample population are missing this value.

The training data also contains the result value of the test bench for each product tested on the production line. As expected, there are only a few of defective items. Although this is good for the company, it is not so good for us as less than 1% of the training population are defective items. The effectives of the two classes are very unbalanced which could really impact our models.

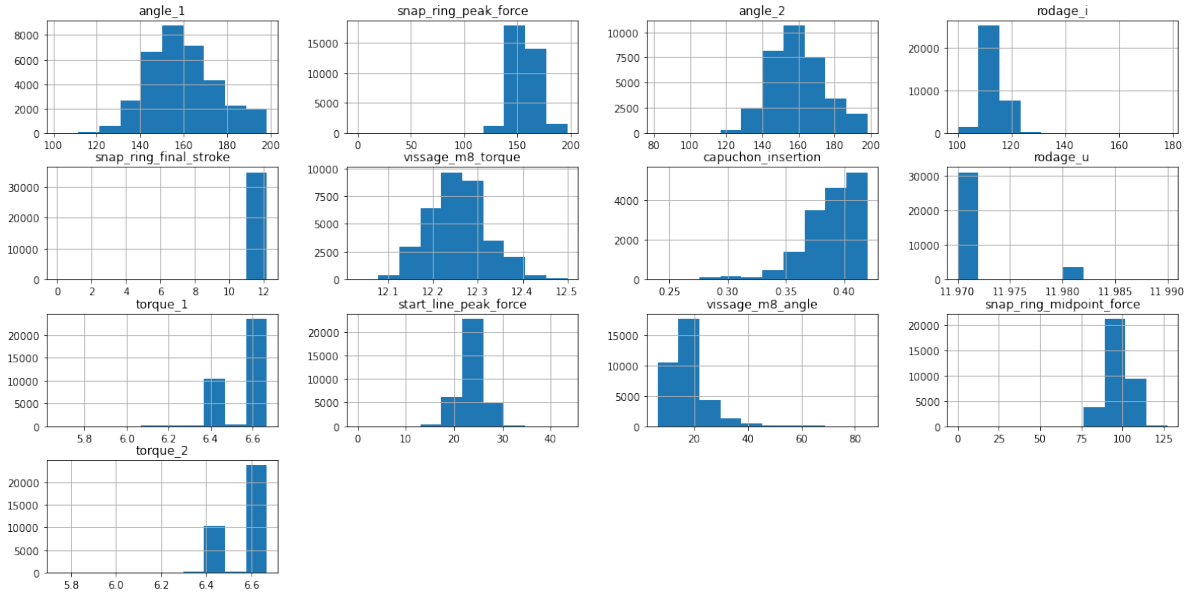

Figure 1: Distributions of input features

3 Data Preparation

4 Model Selection

5 Model Fine Tuning

6 Conclusion

References

[1] Defect prediction on production line by valeo. `https://challengedata.ens.fr/challenges/36`. Accessed: 2023.

[2] Valeo. `https://www.valeo.com/fr/`. Accessed: 2023.