Tutorial Task 3: Analytics Challenge

Group No. 12

Names

Rimu Patil
Tom Vincent
Rajeeva Abhignya Pamidikalwa
Krishna Madhav
Prajwal Bengaluru Amarnath
Roja Raj

Task Description (Intro)

This task revisits the Case Study "Kruse GmbH". It covers the phases of an Industrial Use Case:

a) Use Case Understanding and Planning
b) Data Understanding: Exploration, Preprocessing and Features
c) Modelling

The accompanying dataset is provided in PANDA and can be downloaded there. It also includes the use case and dataset description (see below).

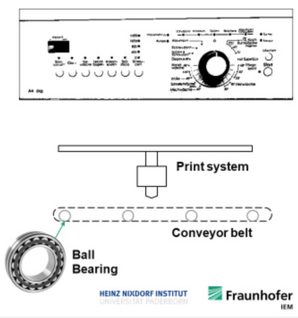
The tasks are detailed in the boxes below.

Kruse GmbH: Detect Product Faults with a Smart Quality System

Use Case Description

Kruse GmbH is struggling with increasing quality problems and inefficiencies in maintenance in production of their highest-selling product, the washing machine. The front panels are particularly affected by the quality problems. Embossing with subsequent printing by a stamp results in a reject rate of 5-10%. Another problem is the frequent stoppage of the conveyor belts. This has resulted in a very expensive regular replacement of the bearings.

After consultation with an engineer, it is suspected that the vibrations of the ball bearings in the conveyor belt are related to the quality condition of the product being manufactured.
Vibration data with corresponding time stamps are recorded by a Machine Database, historical data is available in the MES System. In addition, there are about 1500 production logs in the ERP System that show which product was manufactured at which time and the quality state (ok, nok = not ok) of the printing on the products is known from the quality logs in the QMS.

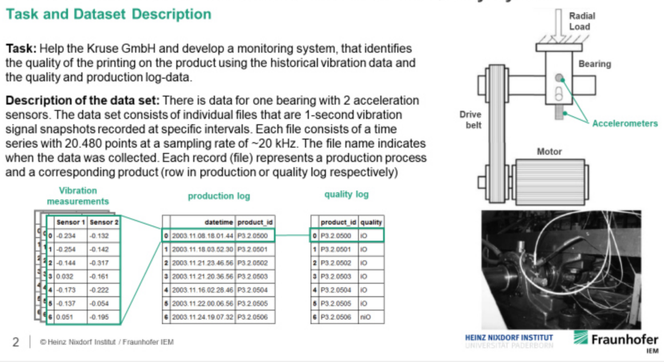


Kruse GmbH: Detect Product Faults with a Smart Quality System

Task and Dataset Description

Task: Help the Kruse GmbH and develop a monitoring system, that identifies the quality of the printing on the product using the historical vibration data and the quality and production log-data.

Description of the data set: There is data for one bearing with 2 acceleration sensors. The data set consists of individual files that are 1-second vibration signal snapshots recorded at specific intervals. Each file consists of a time series with 20.480 points at a sampling rate of ~20 kHz. The file name indicates when the data was collected. Each record (file) represents a production process and a corresponding product (row in production or quality log respectively)



Vibration measurements

	Sensor 1	Sensor 2
0	-0.234	-0.132
1	-0.254	-0.142
2	-0.144	-0.317
3	0.032	-0.161
4	-0.173	-0.222
5	-0.137	-0.054
6	0.051	-0.195

production log

	datetime	product_id
0	2003.11.08.18.01.44	P3.2.0500
1	2003.11.18.03.52.30	P3.2.0501
2	2003.11.21.23.46.56	P3.2.0502
3	2003.11.21.20.36.56	P3.2.0503
4	2003.11.16.02.28.46	P3.2.0504
5	2003.11.22.00.06.56	P3.2.0505
6	2003.11.24.19.07.32	P3.2.0506

quality log

	product_id	quality
0	P3.2.0500	IO
1	P3.2.0501	IO
2	P3.2.0502	IO
3	P3.2.0503	IO
4	P3.2.0504	IO
5	P3.2.0505	IO
6	P3.2.0506	nIO

a) Use Case Understanding and Planning

Business to Analytics Objectives

TASK DESCRIPTION

Describe the business objective for your use case and translate it into an analytics objective and define the expected output of the analytics model. Use the template to the right.

Business objective

The main business objective is to increase the quality of the product so as to decrease the rejection rate of the machines and also to have proper maintenance of the equipment to decrease the frequent stoppage of the production line so that the maximum output is achieved.

Analytics objective

Predict about the quality of print on the front panel. For this we will try to generate features by statsitical understanding and then use this features for classification tasks.

The analytics is performed using the historical vibration data from the acceleration sensors in the production process and also makes use of production log and quality log data

Output of the model

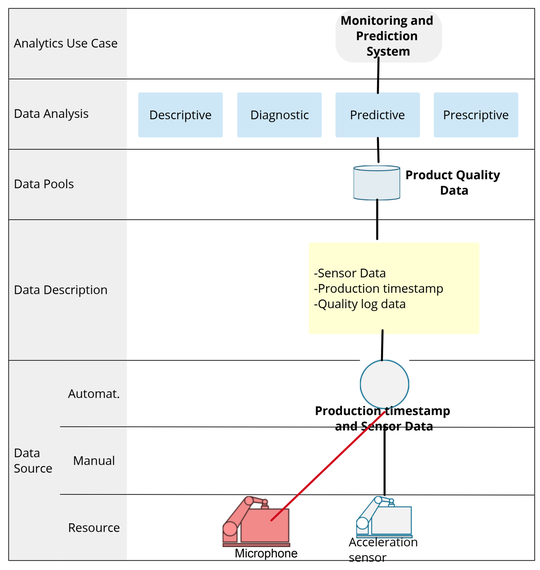
Accurately predicting the quality of the product.

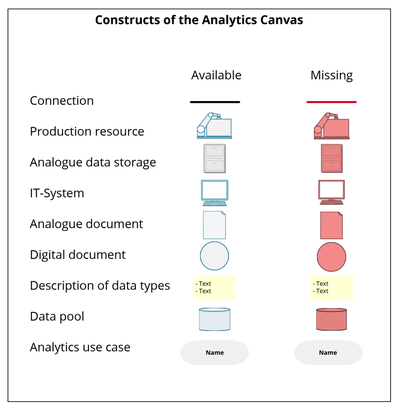
Analytics Canvas

TASK DESCRIPTION

Fill in the Analytics Canvas according to the Use Description. Use the icons for "available" in the template.

Make suggestions of how to further improve the Use Case by adding "missing" icons.



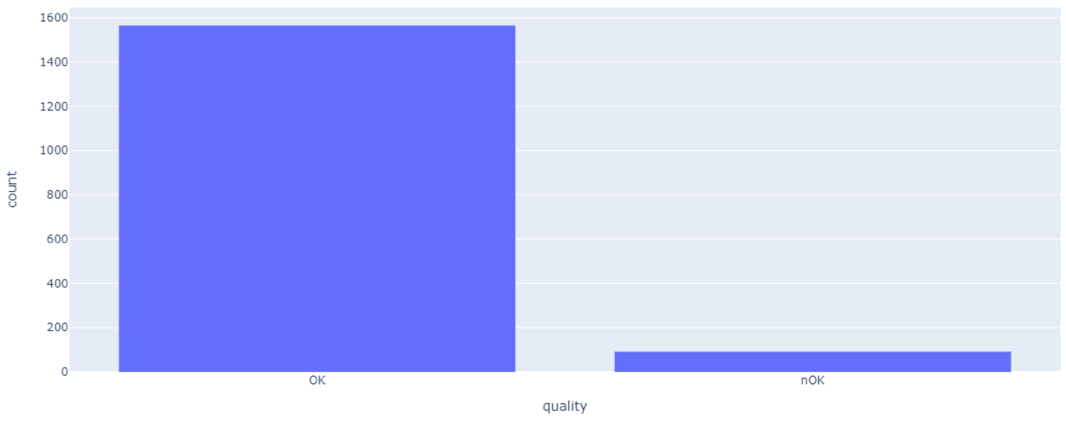


b) Data Understanding: Exploration, Preprocessing, Features

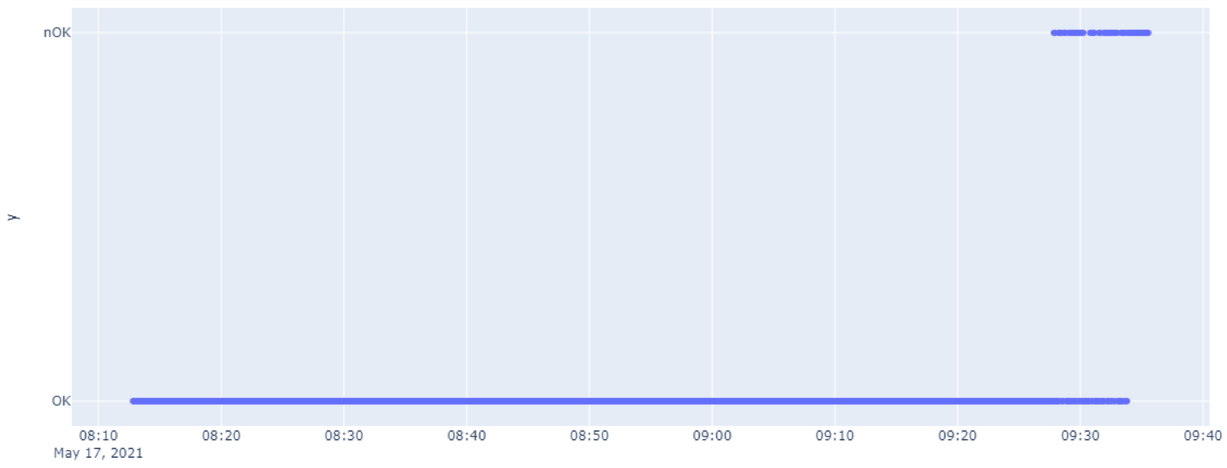
Data Visualization and Exploration

TASK DESCRIPTION

Visualize the data and paste the important visualizations into this field. Highlight and explain your insights and expected challenges with the dataset.



Insights : The above dataset corresponds to count of product which are OK and nOK.



The above visualization is plotted as Time (X- axix) and Condition of product (Y-axis).
Insights : This visualization states the approx time from which the printing device starts producing nOk products.

Challenges : For the time around 9:30, the presented dataset shows some overlapping product quality (OK and nOK)

Data Preprocessing and Feature Extraction

Which preprocessing is necessary for the data and why?

Method of preprocessing	Explanation
Data Integration	Combining of Production log, quality log data.
Data Reduction	In order to maximize storage efficiency, attributes based on timestamp data were created, however frequencies related to sensor data were not stored in the merged data.
Importing libraries	The libraries makes it easier to manipulate the data in python.
Data Transformation	Sensor data is used to generate new attributes for features generation needed for Visualization.

Which features are useful and why?

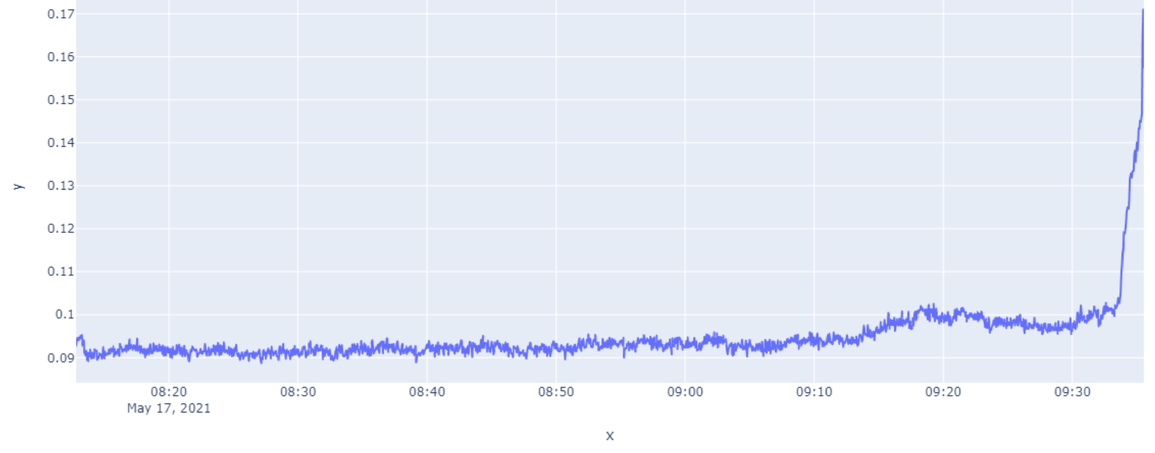
Feature	Explanation
Mean	This value gives the average of all the values in sensors data
Standard Deviation	This value is most useful in our use-case as it tells about the amount of devaition from meanv value. This devaition tells if ball bearing have started to malfunction.
Min and Max	The min and max of a attribute will help us to plot lowest and high frequency recorded for sensors. It lets us visaulize sudden spike in dataset
Median	It gives the value separating the upper half and the lower half.

TASK DESCRIPTION

Plan your Preprocessing and Feature Extraction with the tables provided above. Perform Preprocessing and Feature Extraction on the data accordingly and add visualizations below. You can put them next to your insights of the Exploration phase, if fitting.

The above visualization is plotted as Time (X- axix) and Condition of product (Y-axis).
Insights : This visualization states the approx time from which the printing device starts producing nOk products.

Challenges : For the time around 9:30, the presented dataset shows some overlapping product quality (OK and nOK)



Insights: The visualisation corresponding to Standard Deviation of the sensor data that clearly indicates the time from which quality of product goes from OK to nOK state.

c) Modelling

Modelling

TASK DESCRIPTION

Which methods are suitable to model your analytics case? List possible models to the right and explain your choice. Train your model and past visualisation of results below.

Define Suitable Models and Explain your reasoning.

Model	Explanation
Logistics Regression	It is a supervised learning algorithm to predict categorical independent variable such as true false or Ok and not ok
Random forest	Random forest is used as it fits for the use-case and helps in classification problem. Predication can be classiifed as OK and nOK
Decision Tree	The purpose of employing a decision tree is to develop a training model that can use basic decision rules inferred from training data to predict the class or value of the target variable.

```
c = classification(X,y)
c.logistic_regression()
```

```
Performing modelling for Logistic Regression

precision 1.0
 recall 0.9427710843373494
 f1 0.9705426356589147
 accuracy 0.9427710843373494
```

```
c.decision_tree()
```

```
Performing modelling for decision tree

precision 0.9840255591054313
 recall 0.9655172413793104
 f1 0.9746835443037976
 accuracy 0.9518072289156626
```

```
c.random_forest()
```

```
Performing modelling for Random forest

precision 0.9840255591054313
 recall 0.9716088328075709
 f1 0.9777777777777777
 accuracy 0.9578313253012049
```

Model Evaluation and Improvement

TASK DESCRIPTION

Evaluate your model and fill in the table below.
How could you improve your model? Add the improvements in the analytics canvas using the red symbols.

Define indicators on which you can validate the performance of your model. Explain the results.

Indicator	Explanation
F1	The harmonic mean of precision and recall is identified using the F1 value
Recall rate	In the total number of all positive samples in the data set, the proportion of correctly classified positive samples is obtained by calculating recall rate
Accuracy	It corresponds to number of correct prediction for correct ouput.

Our model could be improved if we use a third sensor. According to IGFD_CNN model, if microphone sensor data is fused with the acceleration sensor data, we get a better accuracy.