MACHINE LEARNING WORKSHOP

Defect Prediction on Production lines

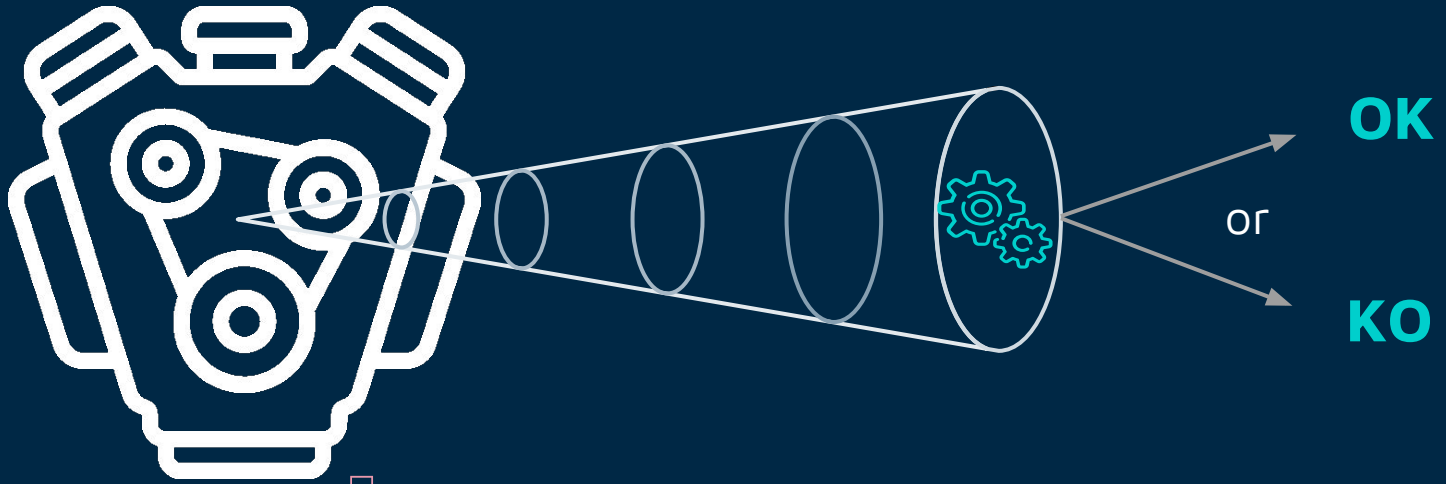
École d'ingénieurs
Télécom Physique Strasbourg

AIT BACHIR R., ALLEMAND F.
FLORET A., JARDOT C.

TABLE OF CONTENTS

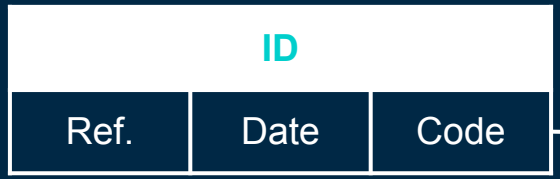
PRESENTATION OF THE CHALLENGE

STARTER ENGINE

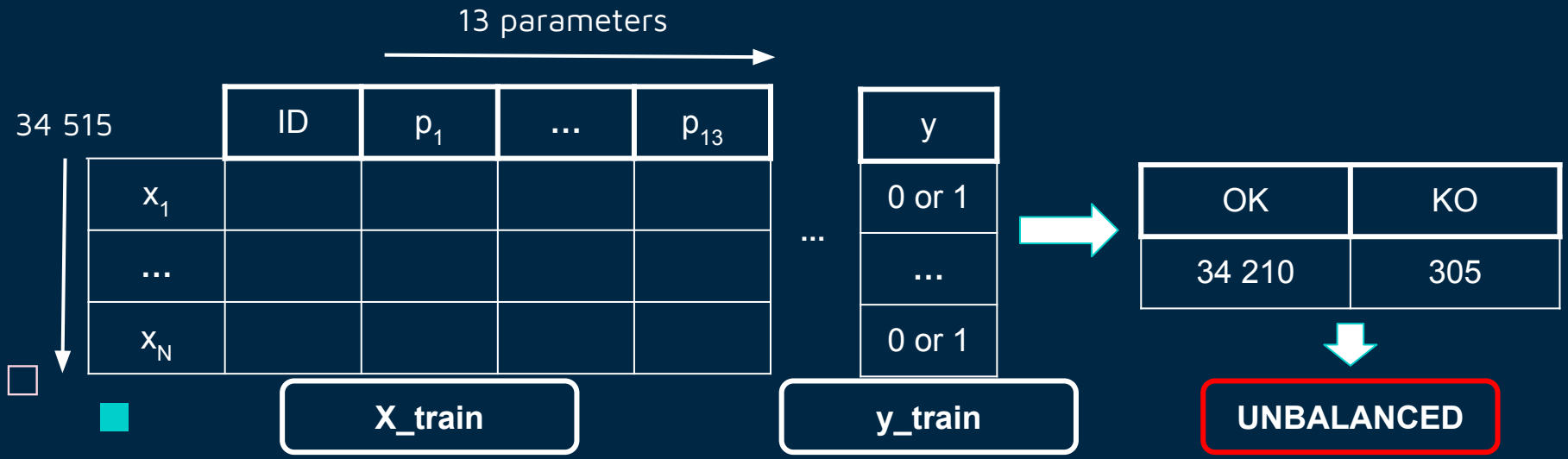


DATA ANALYSIS

Data description



I-B-XA1207672-190701-00494



DATA ANALYSIS

In-Depth Analysis

Feature “insertion cap” contains more than 50 % missing values

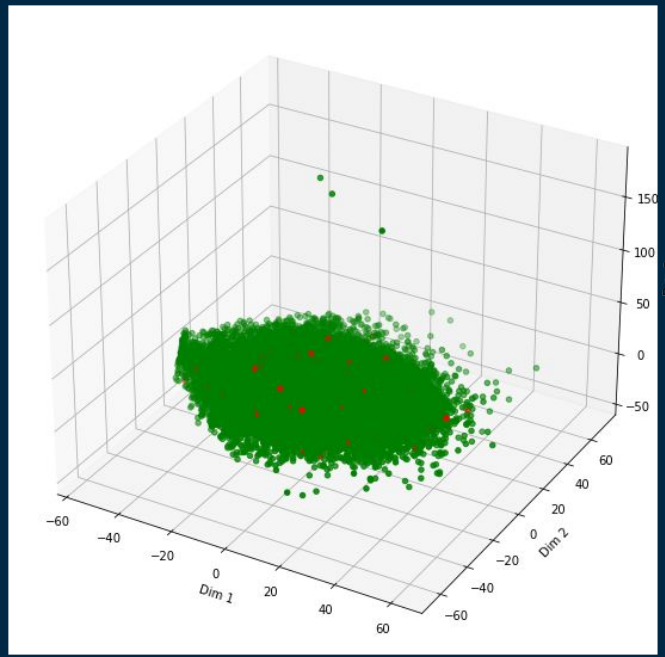
Not a relevant feature?

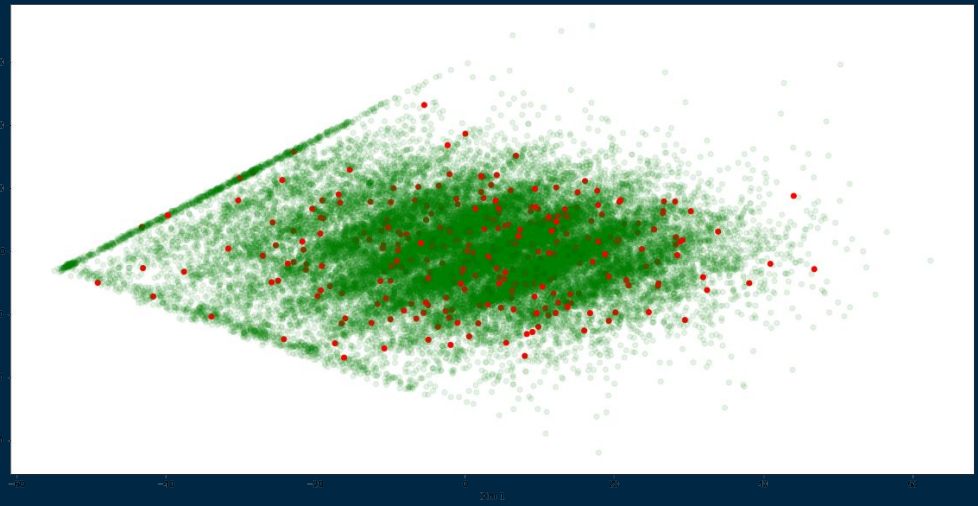
PCA

DATA ANALYSIS

In-Depth Analysis

Defectives Samples
Good Samples

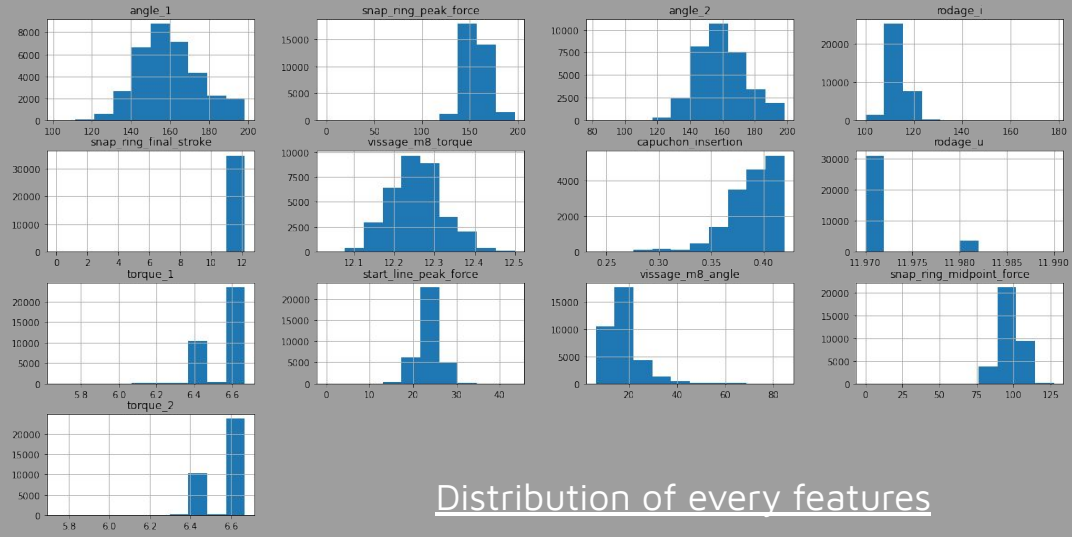




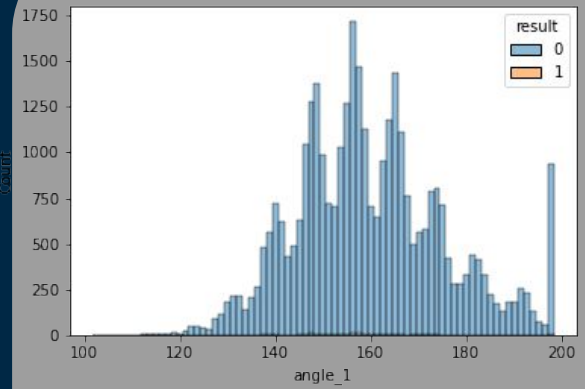
Projection in reduced spaced

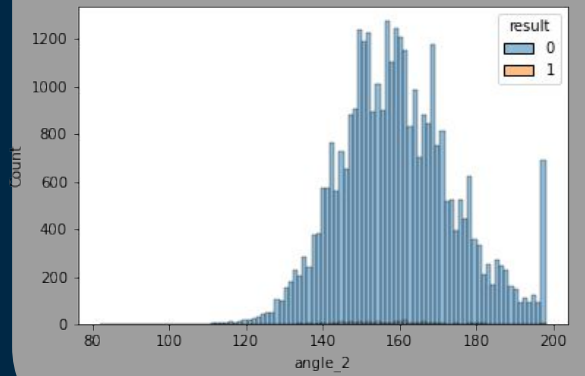
DATA ANALYSIS

In-Depth Analysis



Distribution of every features





DATA ANALYSIS

Resume

Solve the problem of unbalanced data

Deal with insertion cap feature

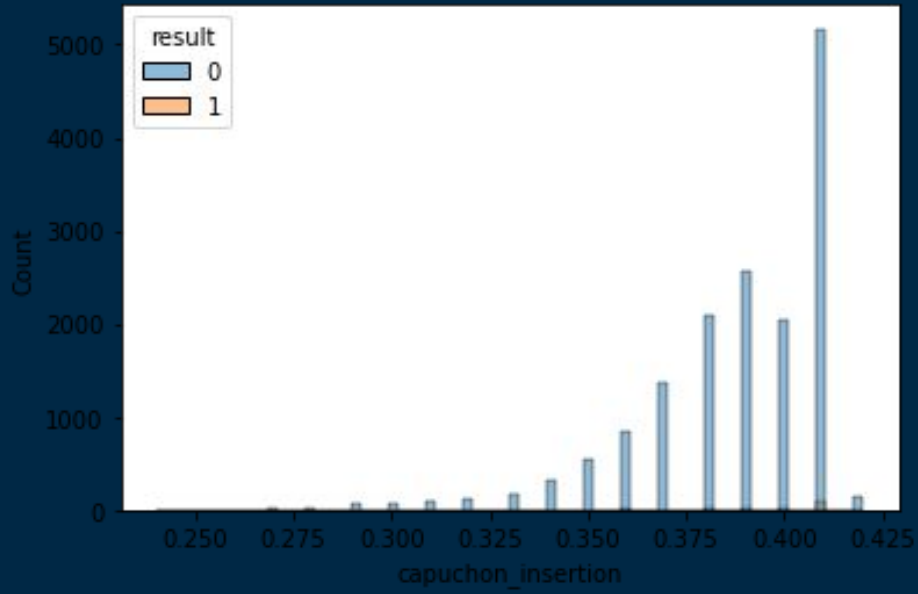
Found the best ML methods to fit a model to the data

DATA PREPARATION

Missing Values

id	0
angle_1	0
snap_ring_peak_force	0
angle_2	0
rodage_i	0
snap_ring_final_stroke	0
vissage_m8_torque	0
capuchon_insertion	18627
rodage_u	0
torque_1	0
start_line_peak_force	0
vissage_m8_angle	0
snap_ring_midpoint_force	0
torque_2	0

id	0
angle_1	0
snap_ring_peak_force	0
angle_2	0
rodage_i	0
snap_ring_final_stroke	0
vissage_m8_torque	0
capuchon_insertion	110
rodage_u	0
torque_1	0
start_line_peak_force	0
vissage_m8_angle	0
snap_ring_midpoint_force	0
torque_2	0



DATA PREPARATION

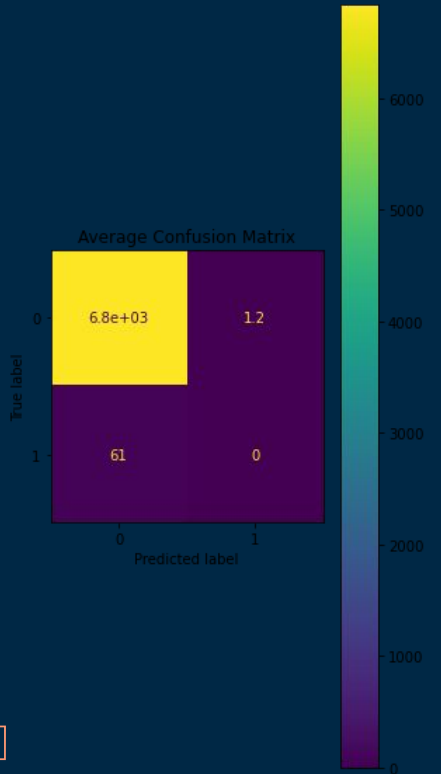
Balance Classes

Accuracy (cv 1)	0.990728669
Accuracy (cv 2)	0.991018398
Accuracy (cv 3)	0.991018398
Accuracy (cv 4)	0.991018398
Accuracy (cv 5)	0.991163262
Average Accuray	0.990989425
Accuracy Std. Deviation	0.000141938

KNN - Raw dataset

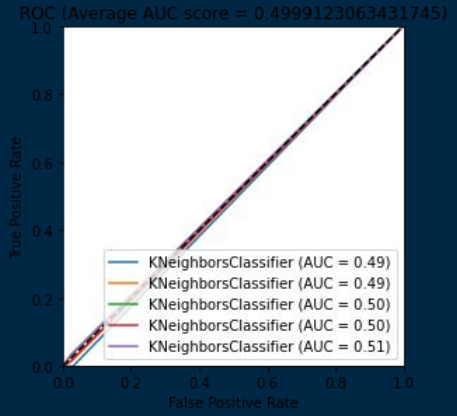
DATA PREPARATION

Balance Classes



Accuracy (cv 1)	0.990728669
Accuracy (cv 2)	0.991018398
Accuracy (cv 3)	0.991018398
Accuracy (cv 4)	0.991018398
Accuracy (cv 5)	0.991163262
Average Accuray	0.990989425
Accuracy Std. Deviation	0.000141938

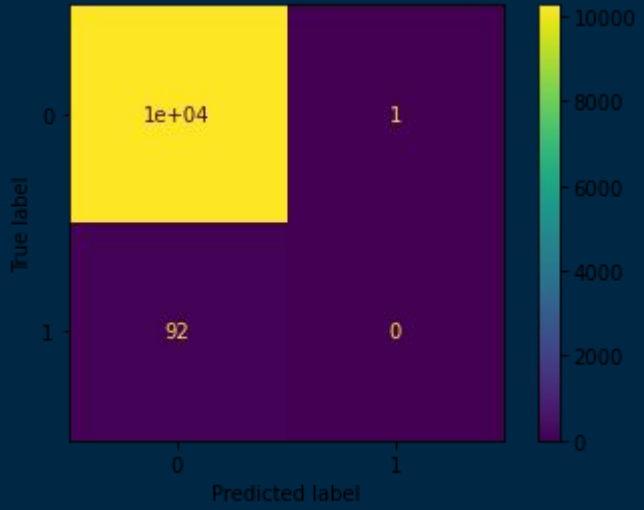
Precision (cv 1)	0.000000000
Precision (cv 2)	0.000000000
Precision (cv 3)	0.000000000
Precision (cv 4)	0.000000000
Precision (cv 5)	0.000000000
Average Precision	0.000000000
Precision Std. Deviation	0.000000000
Recall (cv 1)	0.000000000
Recall (cv 2)	0.000000000
Recall (cv 3)	0.000000000
Recall (cv 4)	0.000000000
Recall (cv 5)	0.000000000
Average Recall	0.000000000
Recall Std. Deviation	0.000000000
F1 (cv 1)	0.000000000
F1 (cv 2)	0.000000000
F1 (cv 3)	0.000000000
F1 (cv 4)	0.000000000
F1 (cv 5)	0.000000000
Average F1	0.000000000
F1 Std. Deviation	0.000000000



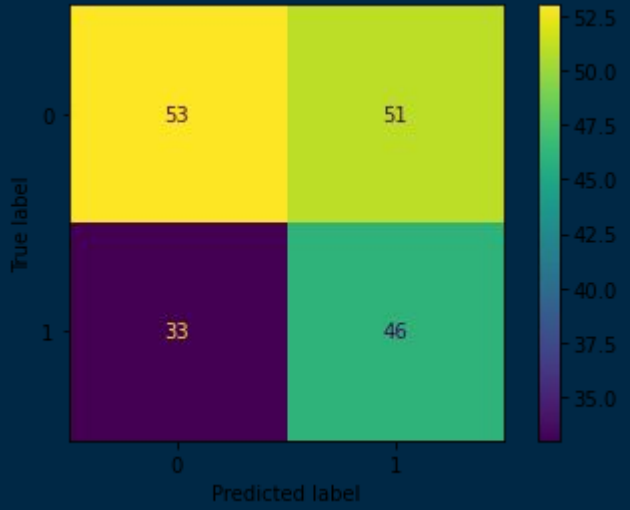
KNN - Raw dataset

DATA PREPARATION

Balance Classes



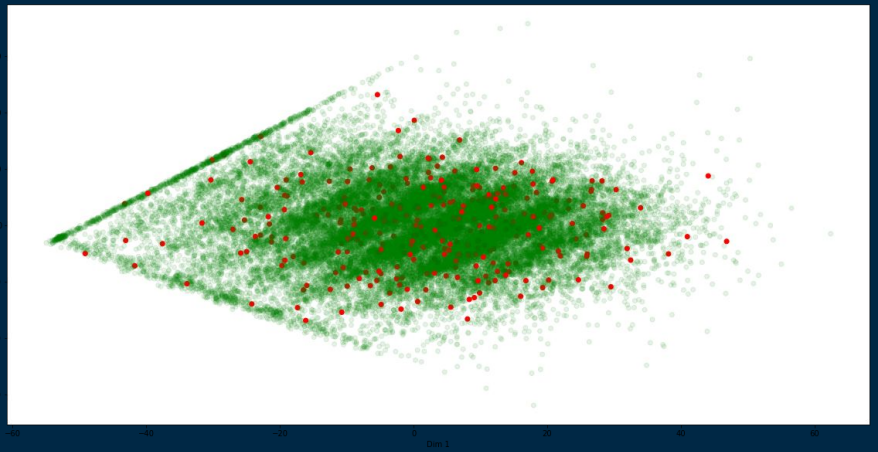
KNN - Raw dataset



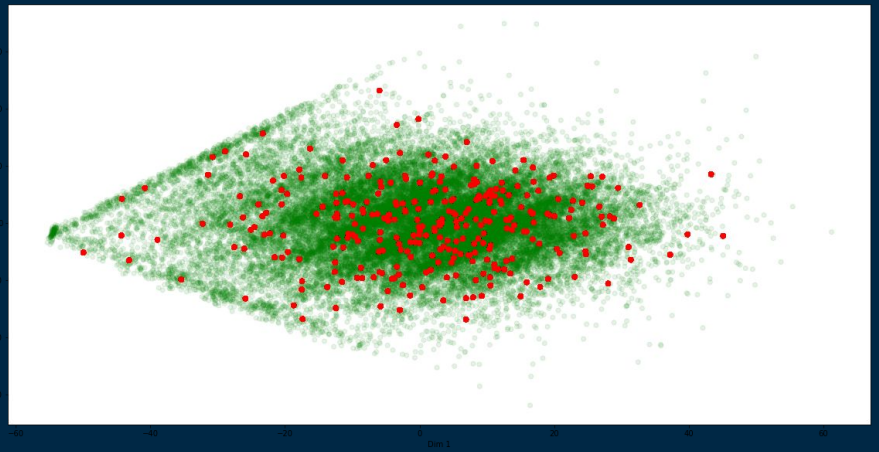
KNN - Balanced dataset (valid items removed)

DATA PREPARATION

Balance Classes



Unmodified Dataset



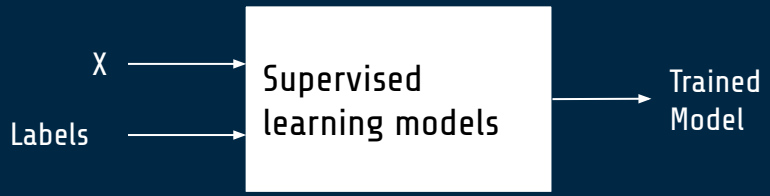
Balanced dataset (valid items removed)

MODEL SELECTION

Two types of algorithms:

Supervised learning

- → k-Nearest Neighbours
- → Naïve Bayes Classifier
- → Random Forest
- → Multilayer Perceptron



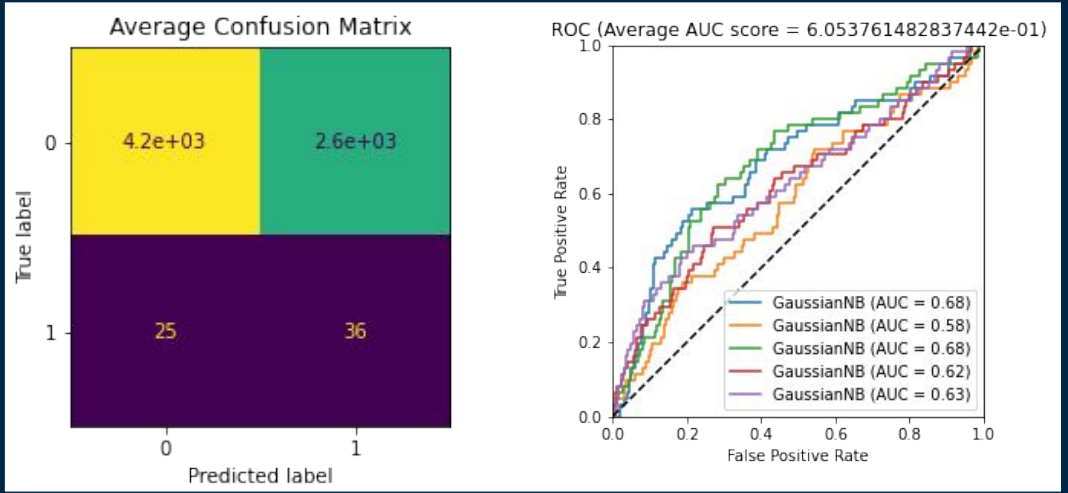
Unsupervised learning - novelty detection

→ One-Class SVM

MODEL SELECTION

Metrics used to evaluate models:

· Precision
· Recall
→ F1-Score
· Area under ROC curve (metric requested by Valeo)



→ Model Chosen: Naive Bayes Classifier for its relatively high recall and area under ROC curve

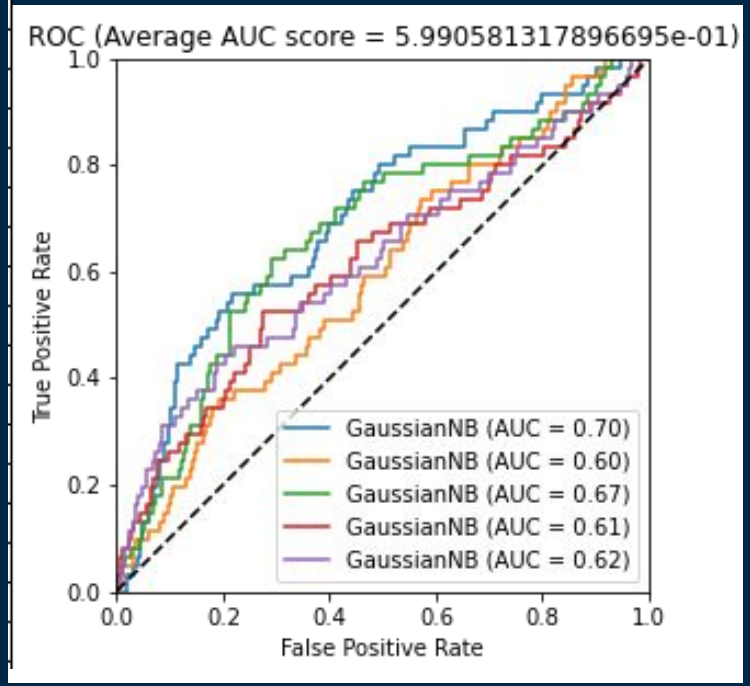
MODEL FINE TUNING

Model Chosen : the Naive Bayesian classifier

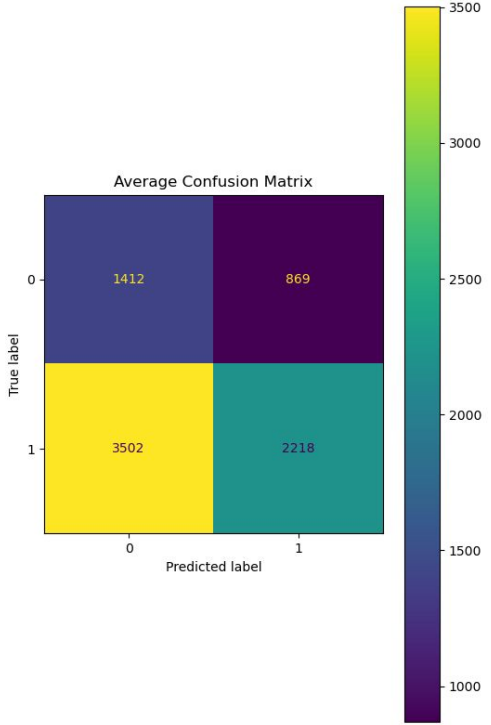
-> Using Grid Search

Var_smoothing ≈ 1e-7

Accuracy (cv 1)	0.643343474
Accuracy (cv 2)	0.667390989
Accuracy (cv 3)	0.650441837
Accuracy (cv 4)	0.679849341
Accuracy (cv 5)	0.687816891
Average Accuray	0.665768506
Accuracy Std. Deviation	0.016880892



Final Naive Bayes Classifier Evaluation



Accuracy	0.453693288
Precision	0.718496923
Recall	0.387762238
F1	0.503690246

CONCLUSION

Results

- → Classification is a difficult task
- → Good understanding of the situation
- → Critical thinking
- → Testing

How can we get better results?

- → Learn more about the dataset with a specialist from Valeo
- → By getting even more data to train deeper models
- → Changing metrics (F1-score depends on class imbalance)

THANK YOU

Do you have any questions ?