PROJECT: Payment Plan Prediction

Eduardo Gil González-Madroño

Date: 30th September, 2020

Contents

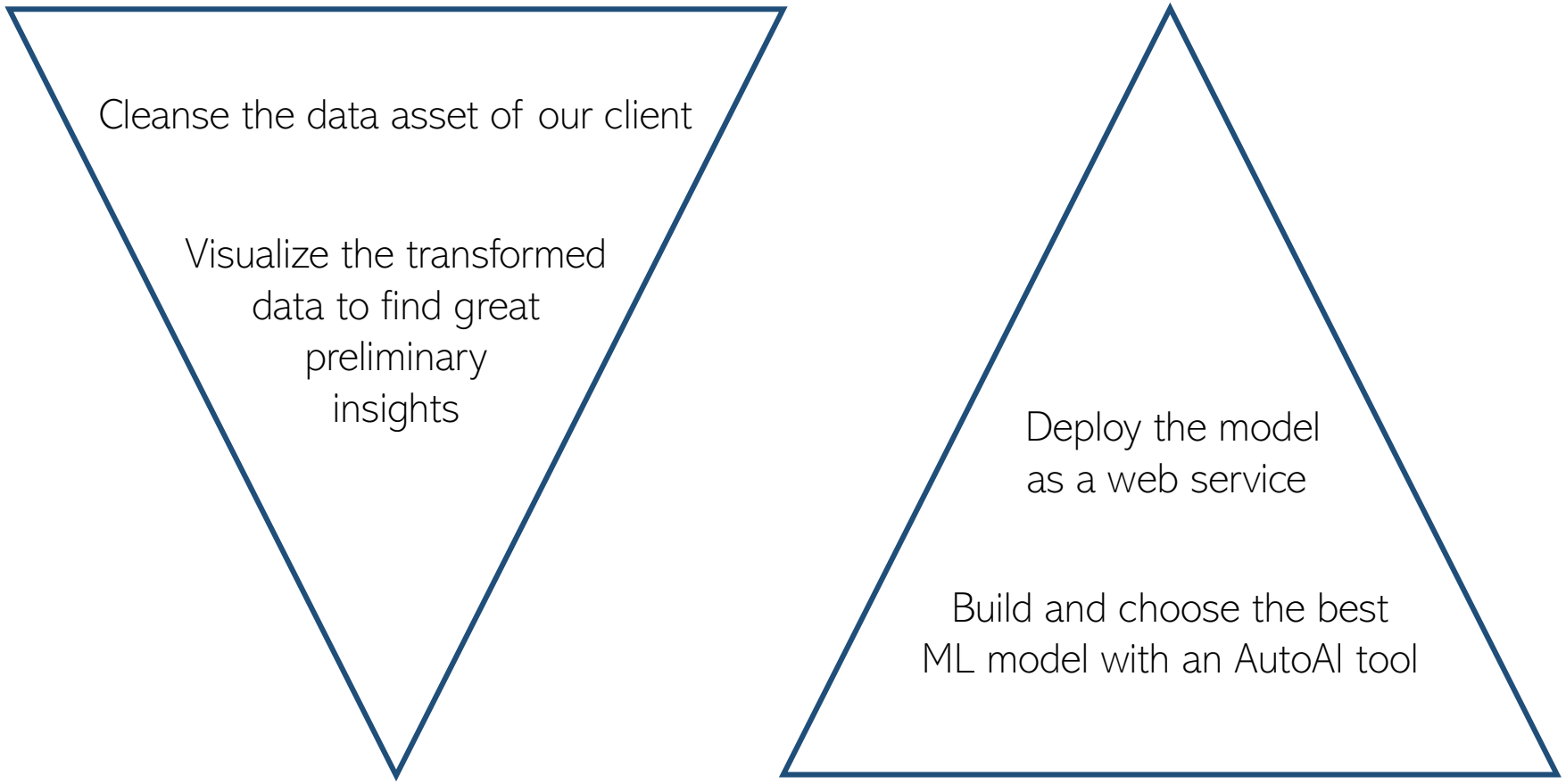
The Business Concern

Our client is concerned about the capability for some customers to complete their payments on time due to general and global situation of COVID-19. They **want to help** their customers avoid missing payments.

The Accounts department struggles to identify customers who might need a renovated payment plan.

So, our goal is to help them identify those customers at risk to address their needs properly

2 Main Cores for the project



Baseline of our client

With the existing customer insight tool embedded in the firm the **Accounts department can identify only 10% of customers** who will miss a payment.

The Accounts department struggles to identify customers who might need a renovated payment plan.

So, our goal is to help them identify those customers at risk to address their needs properly

The Dataset

The existing amount of data provided consist of a dataset with the following dimensions:

- 6790 observations → Payments
- 47 predicting features:
    - Categorical Data: 30%
    - Numerical Data (Non Binary): 25%
    - Binary Data: 30%
    - Non–Usable Data (Ids, Names, Last Names, Phone Numbers…): 15%
- 1 target feature: MISSED PAYMENT → Binary

Main Findings of Data Quality

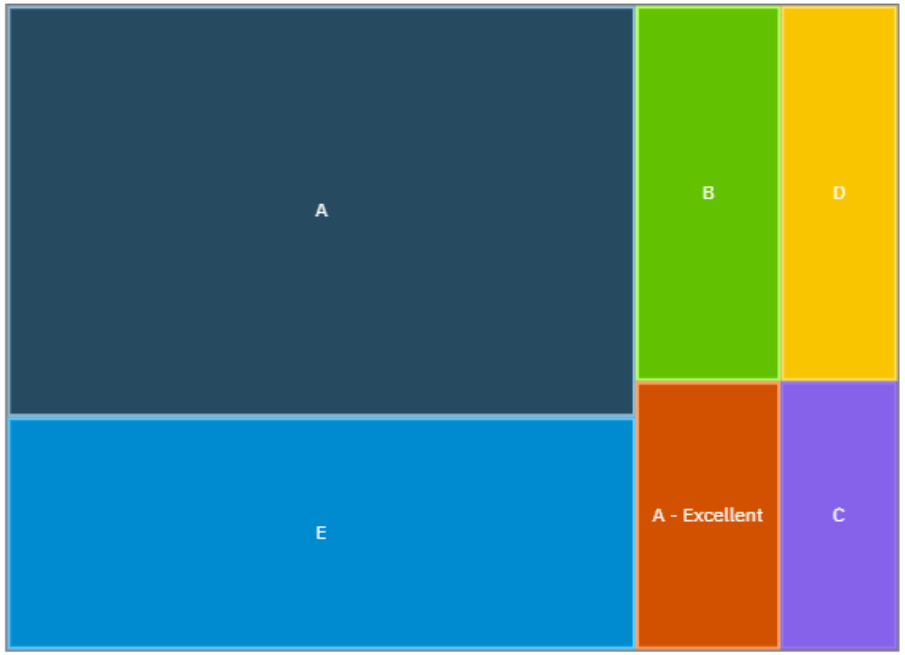
FEATURE: CREDIT HISTORY

· It´s our first bet for being the most helpful feature to predict if a customer will fail to pay during the current month.

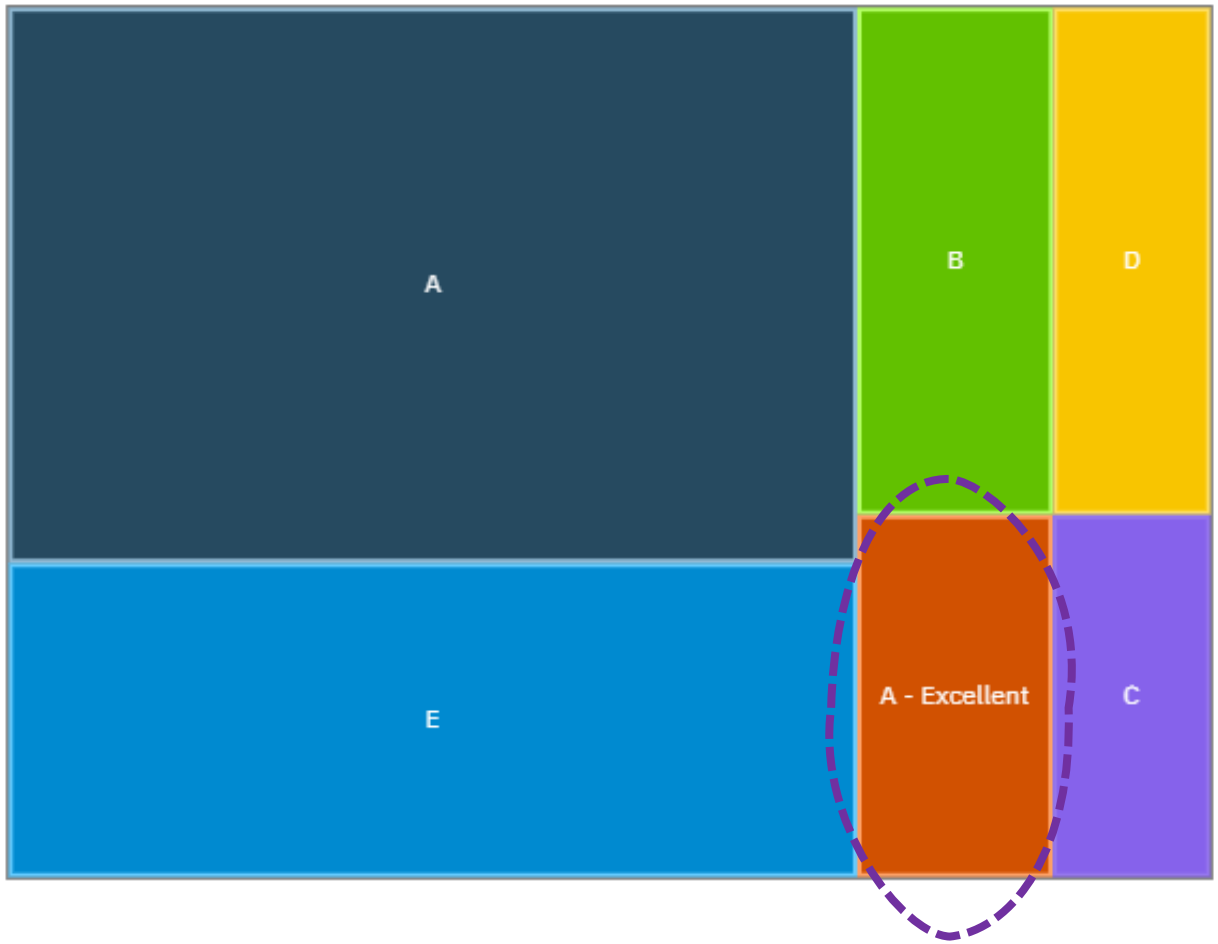
· It´s Good that we have majority of our customer base being A and B credit customer.

· However… we found something strange in the data…

Could you tell what it is?

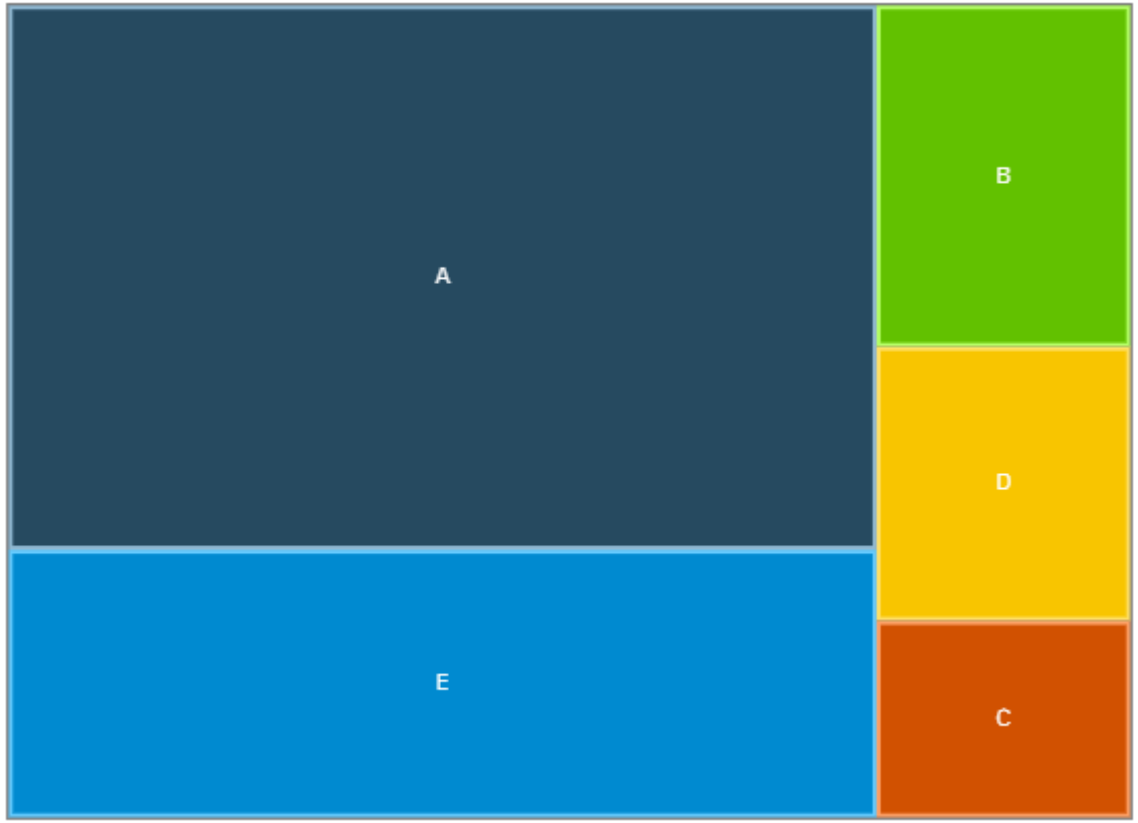


Credit History Feature: Poor Data Q



Credit History Feature: Poor Data Q

Sometimes, the key is in the basics…

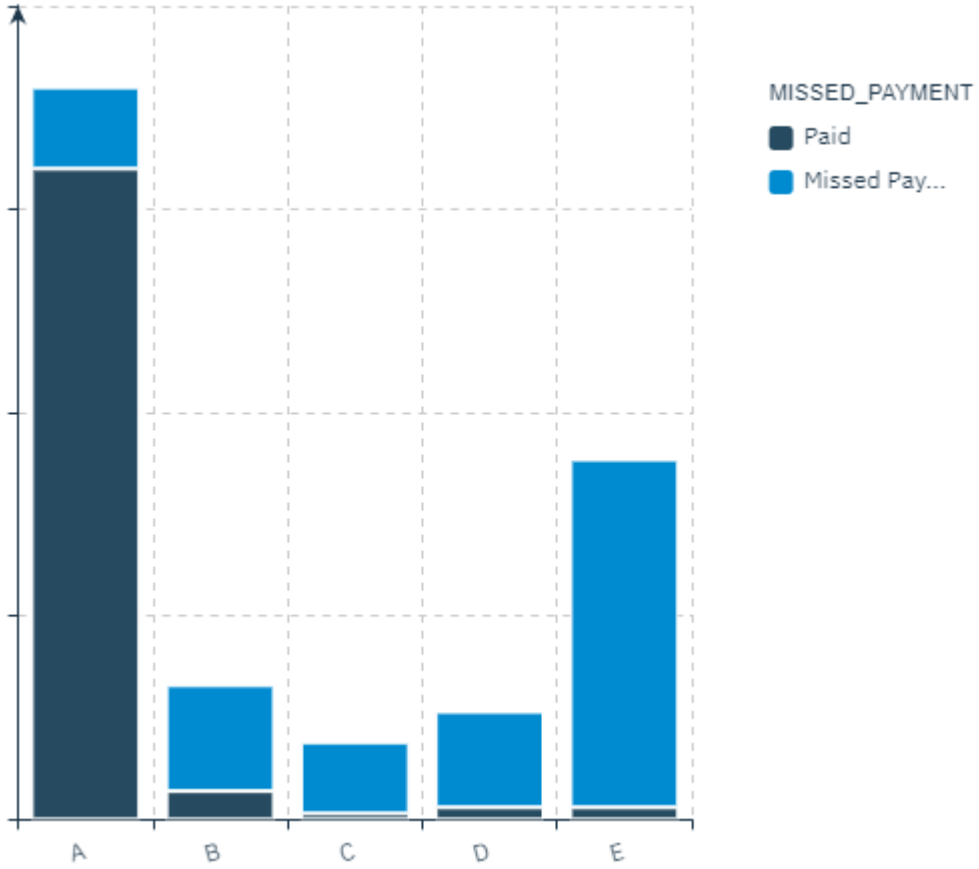


Main Findings of Data Quality

FEATURE: CREDIT HISTORY and MISSED PAYMENT

· So it´s visible that the segment proclive

to miss payment is E as suspected.



Translating the concern into Analytics Language

The goal of the project is to build a model that is capable of detecting customers who will fail payment with a better accuracy than the current Accounts department effectivity ~10%

Translating the concern into Analytics Language

The goal of the ML Analytical part is to build a model that is capable of detecting customers who will fail payment with a better accuracy than the current Accounts department effectivity ~10%

Basis of ML Modelling

The project will be based on tuning a Classifier Algorithm to work over the binary feature MISSED_PAYMENTS.

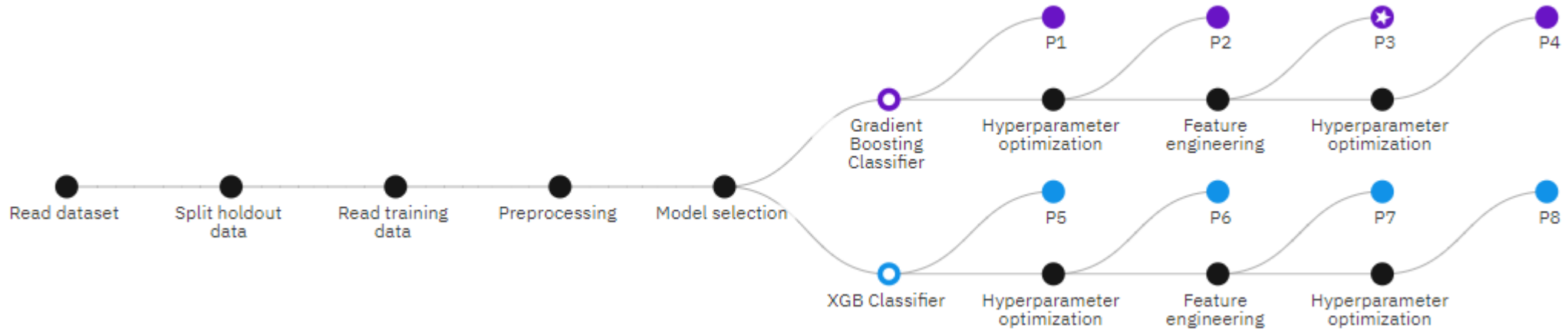
Classifier Candidates:

·· Gradient Boosting Classifier

·· XGB Classifier

Basis of ML Modelling

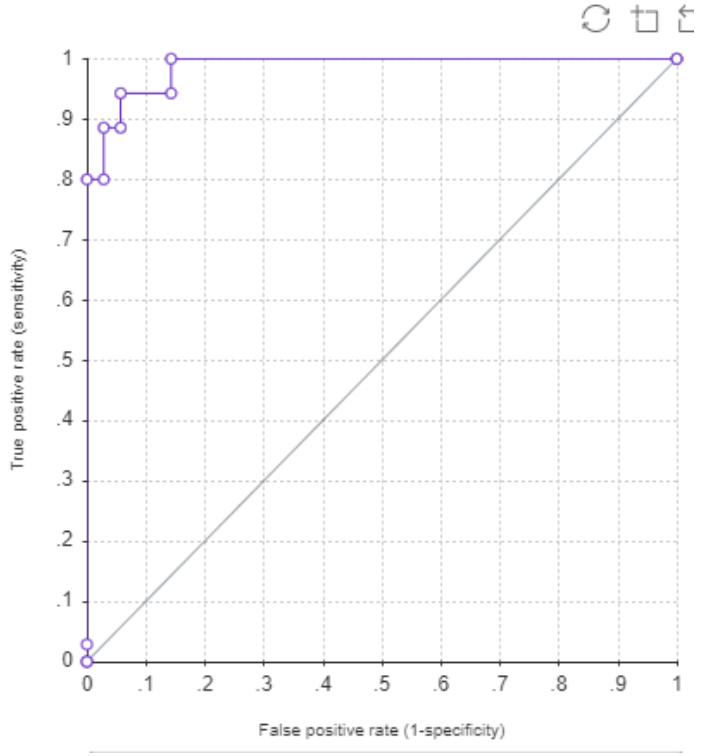
8 Pipelines will be generated and run:



Result Summary for 'Naked' pipelines (P1, P5)

P1: 'Naked' Gradient Boosting Classifier

ROC Curve



Model Accuracy

0.900

Scores: 3–Fold CV

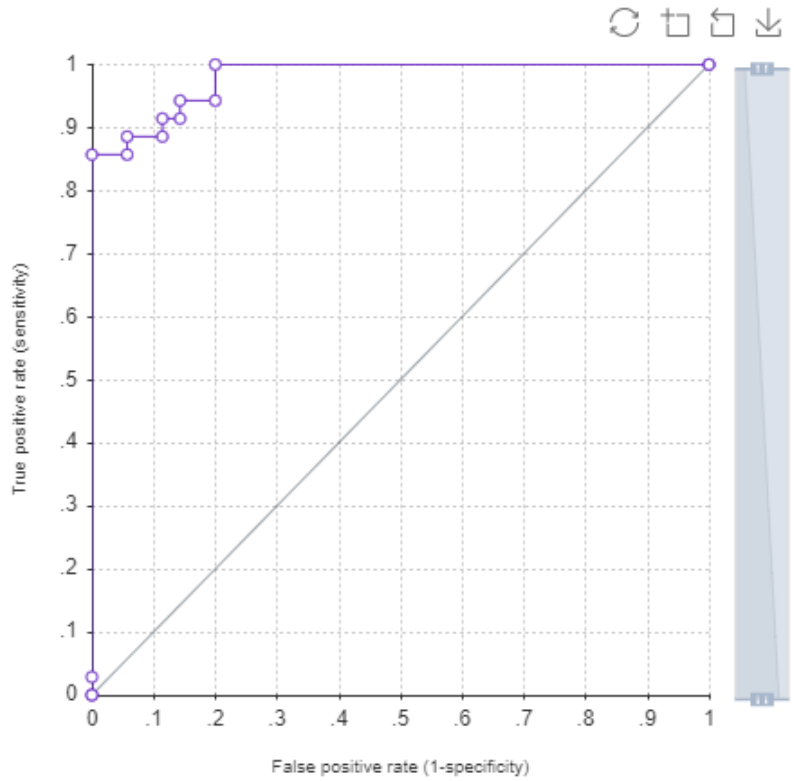
Model Evaluation Measures

	Holdout Score	Cross Validation Score
Accuracy	0.900	0.896
Area Under ROC Curve	0.985	0.951
Precision	0.938	0.902
Recall	0.857	0.887
F_1 Measure	0.896	0.894
Average Precision	0.986	0.955
Log Loss	0.197	0.286

Result Summary for 'Naked' pipelines (P1, P5)

P5: 'Naked' XGB Classifier

ROC Curve



Model Accuracy

0.900

Scores: 3–Fold CV

Model Evaluation Measures

	Holdout Score	Cross Validation Score
Accuracy	0.900	0.888
Area Under ROC Curve	0.981	0.945
Precision	0.967	0.898
Recall	0.829	0.874
F_1 Measure	0.892	0.886
Average Precision	0.982	0.937
Log Loss	0.239	0.296

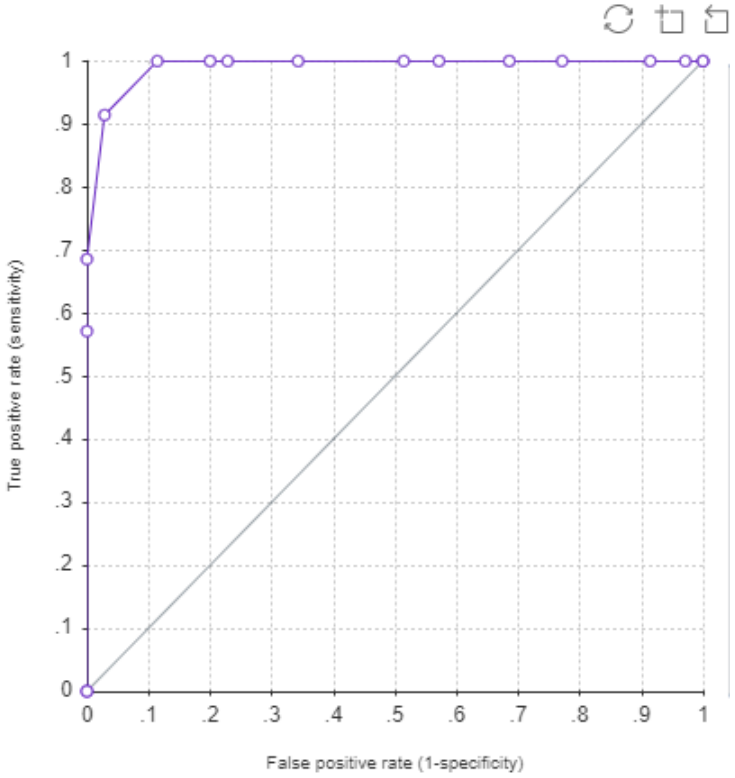
Result Summary for best pipeline (P2)

Model Accuracy 0.943

P2: Gradient Boosting Classifier (w/ Hyperparameter Optimization)

Scoring: 3-Fold CV

ROC Curve



Model Evaluation Measures

	Holdout Score	Cross Validation Score
Accuracy	0.943	0.909
Area Under ROC Curve	0.991	0.957
Precision	1.000	0.907
Recall	0.886	0.910
F_1 Measure	0.939	0.908
Average Precision	0.986	0.953
Log Loss	0.213	0.272

Project Results

· A model based on Gradient Boosting Classification was done and deployed as a web service to provide the Accountants department the necessary vision to **identify nearly 82% of customers at risk** of failing payment in order to provide them with ad-hoc solutions that safeguard their finances and the ones of the company

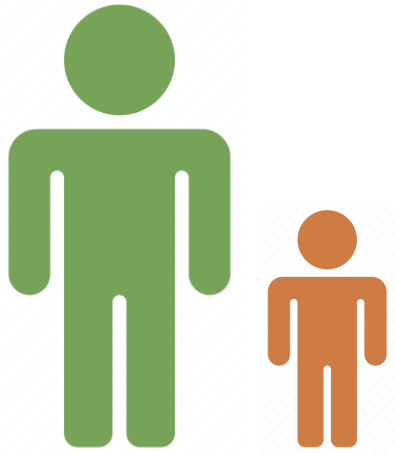
Project Results

· At the beginning

Fail to pay and identified

Fail to pay and identified

· After model deployment



Q&A

Thank You

www.profesorDATA.com