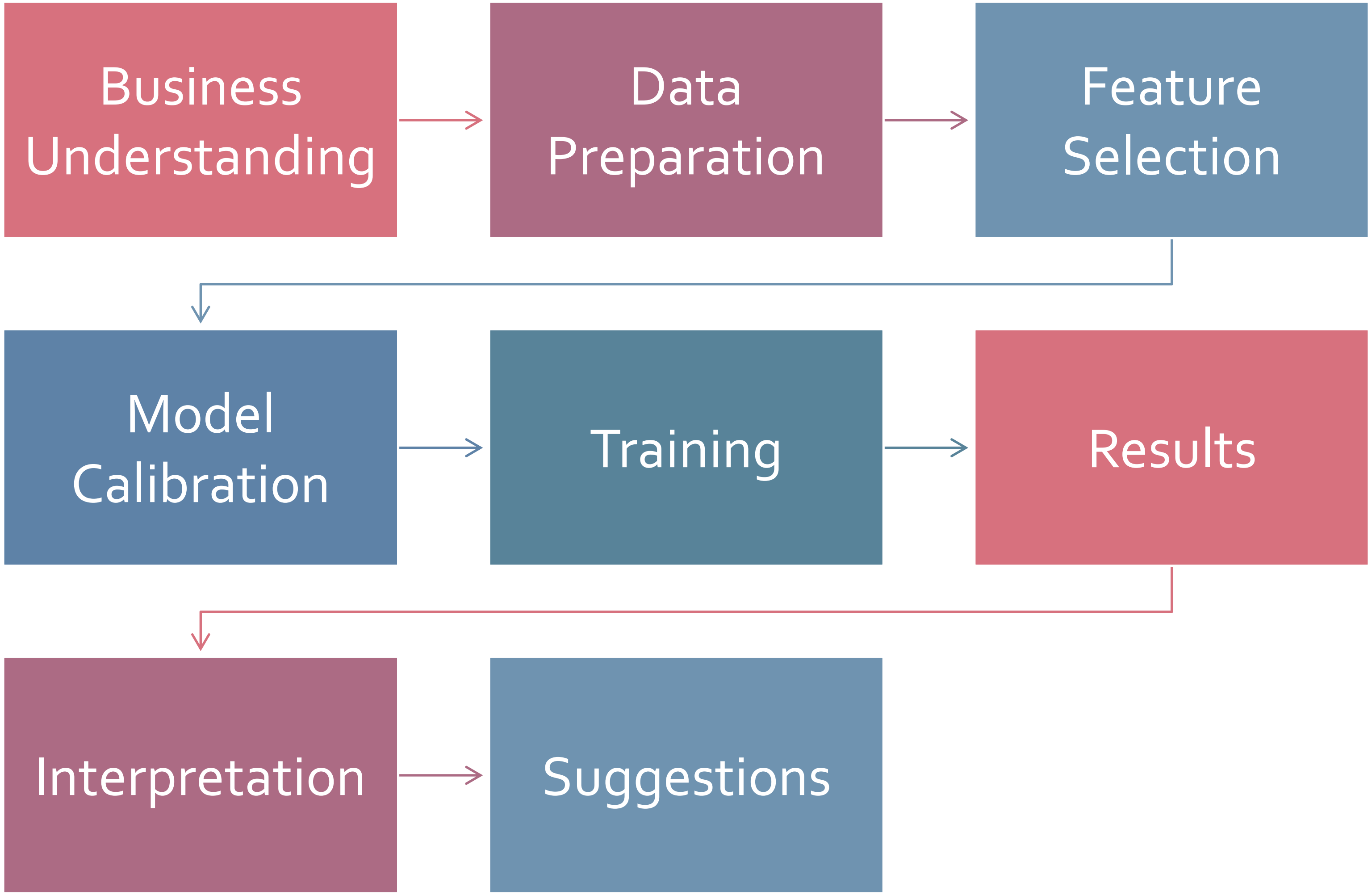


ABC Credit Scoring

TEAM B:
Anup Satyal , Ignacio Mouawad, Meng-Chen Hung, Esther Chaelin Lee, Eduardo Cort, Juliana Villaveces, Varun Raja

Business Understanding

Revenues

Spanish laws (artículo 38.2 de la Ley 38/2003 and Ley 6/2018) have restricted how much banks can earn from default payments.

How Banks make money from defaults on credit card payments :

3% → 9% + Commissions → 25%

Legal interest of money

Banks by law are allowed to charge 3 times that interest

Including commissions, this comes to ~25%

In addition to the fees on default, the bank also charges a fixed annual fee of between 35 to 45 euros to credit card holders.

Cost

In the case of default, the bank would lose the defaulted amount + 25% . In our case, while the median default amount is 150 euros, and the most often defaulted amount is 10 euros, it should be noted that we have not been provided information as to the timeline of defaults or whether these actually apply to credit card defaults. Thus, any inferences from this field is not possible.

Data Preparation

Approach:

We initially discussed the data preparation tasks each group member performed and then chose steps that every member agreed upon. To provide interpretability and consistency, all columns were renamed and all text variables were lower-cased

DEDUPLICATION : REMOVED 1913 ROWS FROM THE SAMPLE DATASET (19% OF SAMPLE).

DISCRETIZATION: THE AGE, MARITAL STATUS, SALARY, CHANNEL AND PRIOR LOAN PAYMENT WERE BINNED

MISSING VALUES: AS LONG AS MISSING VALUES WERE LESS THAN 5% OF THE SAMPLE DATASET THEY WERE REMOVED.

IMPUTATION: THE OPTION OF IMPUTING A CONSTANT (-1) WAS CHOSEN. BOTH THE **ACTIVE LOANS** AND **ACTIVE_MORTGAGES** COLUMNS HAD TO BE TRANSFORMED (THROUGH DISCRETIZATION) INTO CATEGORICAL VARIABLES

Feature Selection & Model Calibration

Feature Selection:

Gender

ONLY -23% CORRELATION BETWEEN GENDER AND CREDIT_PAYMENT. UNSURE OF WHETHER IT MAKES BUSINESS SENSE.

Default Amount

AS LONG AS THE PERSON HAS DEFAULTED, WE NOTICED THAT THE VALUE WAS CLASSED AS SUB-PAR AND HENCE IT DOESN'T TELL US ANYTHING USEFUL.

Model Calibration:

- The initial dataset was imbalanced with a target variable of 73% paid(0) and 27% unpaid (1). A class rebalancing was carried out after which our target variable was balanced to a ratio of 51:49.
- We divided a train: test set with ratios of 60:40, 70:30 and 80:20.
- To allow for proper cross-validation, 10 k-folds were calibrated.

Results

From the bank's perspective, it should want to increase its market share from increased credit card fees = **HIGH TRUE PAID**

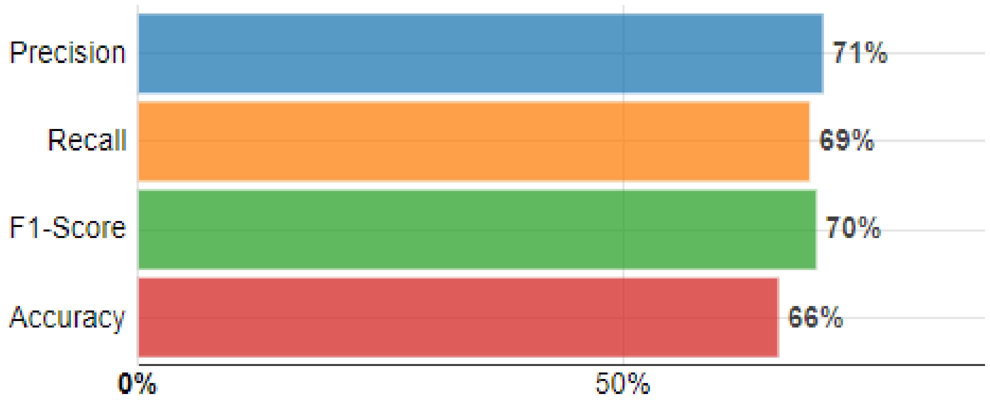
At the same time, a high default rate would take it out of business = **HIGH TRUE UNPAID**

The metric that gives us both is Accuracy (TP & TN/Total) which is why we choose to maximise our accuracy.

Threshold (cut-off) 0 1 0.550

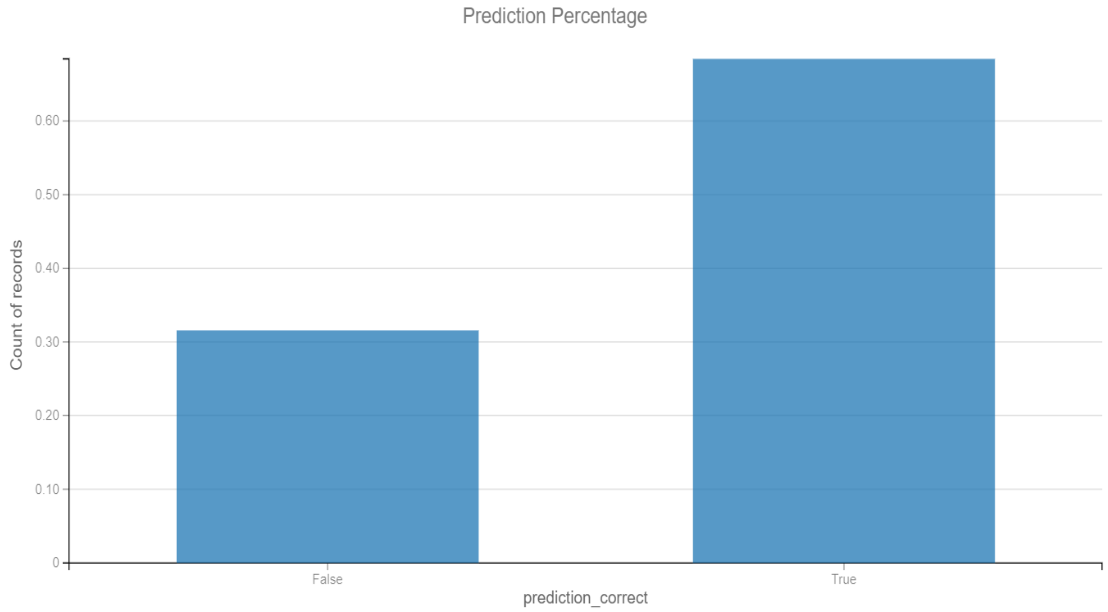
Display: Record count

	Predicted 1	Predicted 0	Total
Actually 1	43823	19508	63331
Actually 0	18268	29428	47696
Total	62091	48936	111027



Interpretation

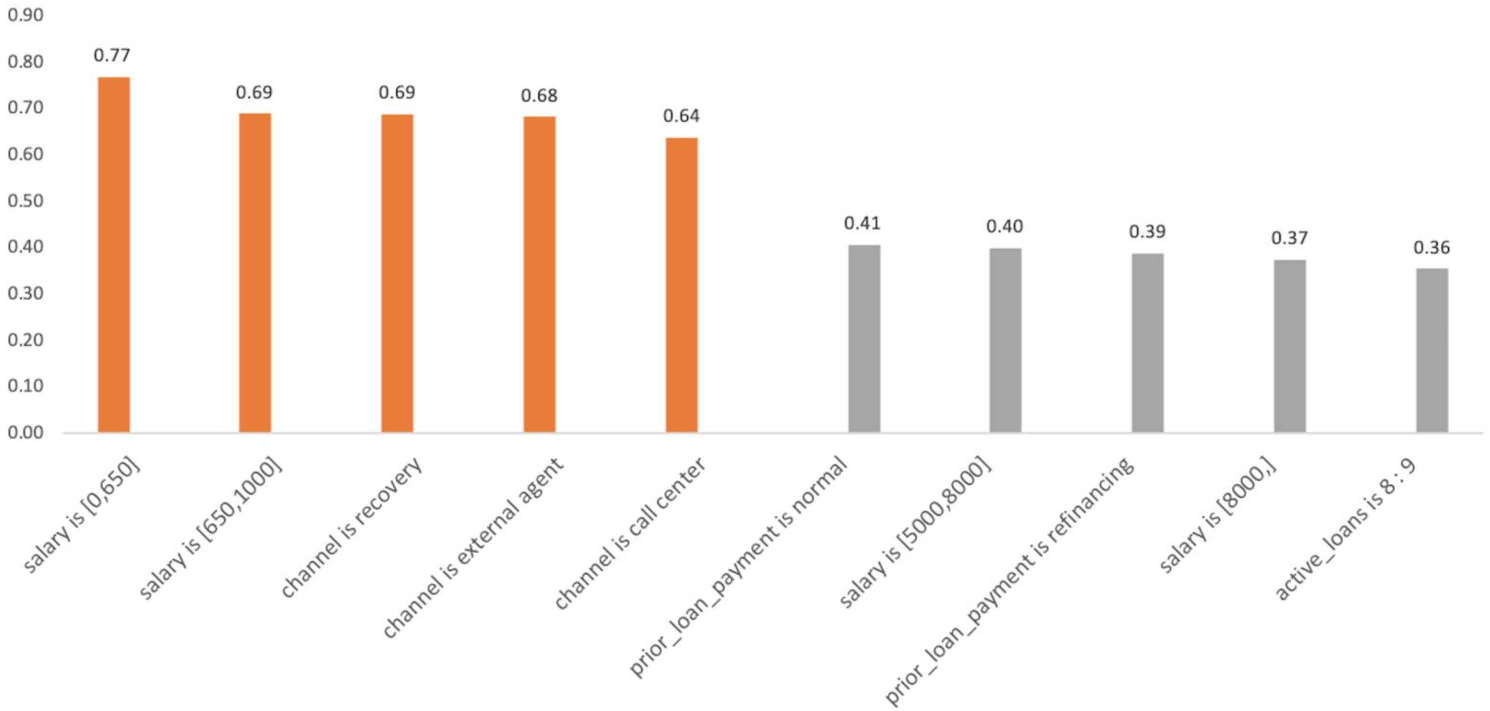
The model classifies true paids and true unpaids correctly 68% of the time, which is better than flipping a coin. Out of 10000 new customers, 6,800 would be classified correctly. If we assume 35 euros of credit card fees we would generate 238,000 euros in revenue.



The dataset applies only to the people that were accepted for credit cards. The model therefore assumes that a 100% of the people that applied were eventually accepted for credit cards.

Interpretation

- The chart below shows the exponentialized coefficients for the top (in orange) and bottom 5 (in grey) variables.
- The coefficients show those variables that are more (or less) important in determining whether the target will be 1 (unpaid).
- As illustrated by the graph, and as can be intuitively gauged salary range is clearly an important variable. In addition, channel is also a standout variable.



Suggestions

- Table 1 below shows the profit that the bank may expect to generate if the dataset was balanced, which includes the loss of the false negatives.
- The value for the false negatives is imputed as 12.5, which is 25% + the mode for illustrative purposes. As mentioned before, as it is an uncertain estimate and should not be considered to be an estimate of the credit defaults.
- To get a more accurate measure of potential profits, the bank should get a accurate estimate of credit defaults per customer

Cost matrix

If model predicts 1	and value is 1	the gain is	0	×	22231	=	0.00
	but value is 0	the gain is		×	9203	=	0.00
Model predicts 0	and value is 0	the gain is	35	×	14613	=	511,455.00
	but value is 1	the gain is	-12.5	×	9750	=	-121,875.00
	Average gain per record		**6.98**	×	55797	=	389,580.00

- In addition there is an opportunity cost for the false positives, which does not have a cashflow impact but does have an economic cost. The bank should also consider imputing a value in order to get a better estimate of economic profits.

Cost matrix

If model predicts 1	and value is 1	the gain is	0	×	22231	=	0.00
	but value is 0	the gain is	-35	×	9203	=	-322,105.00
Model predicts 0	and value is 0	the gain is	35	×	14613	=	511,455.00
	but value is 1	the gain is	-12.5	×	9750	=	-121,875.00
	Average gain per record		**1.21**	×	55797	=	67,475.00