ML BENCHMARK

PRESENTATION

IÉSEG
SCHOOL OF MANAGEMENT

MEET THE TEAM

Nithesh Ramanna

Juan José Medina

Pedro Román I.

IÉSEG
SCHOOL OF MANAGEMENT

OVERVIEW

IÉSEG SCHOOL OF MANAGEMENT

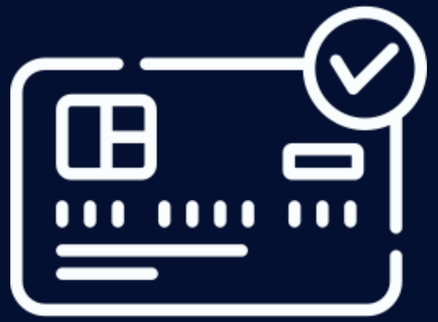
PROBLEM DEFINITION

> About one in five U.S. adults have a credit card balance that is higher than the amount of money in their emergency savings accounts.[1]

1- (2018, 09). 10 Surprising Credit Card Debt Facts. Sofi. Obtenido 04, 2022, de https://www.sofi.com/learn/content/10-surprising-credit-card-debt-facts/?__cf_chl_tk=teBh0wjbb89BWHFHDrLSDtqmH1O3u0BraVfN2c.tEQI-1649588890-0-gaNycGzNCKU

IÉSEG SCHOOL OF MANAGEMENT

PROBLEM DEFINITION



A financial institution wants to build a classification model to predict the likelihood of credit card default for its customers and identify the key factors that determine this behavior.

IÉSEG SCHOOL OF MANAGEMENT

EXPLORATORY DATA ANALYSIS

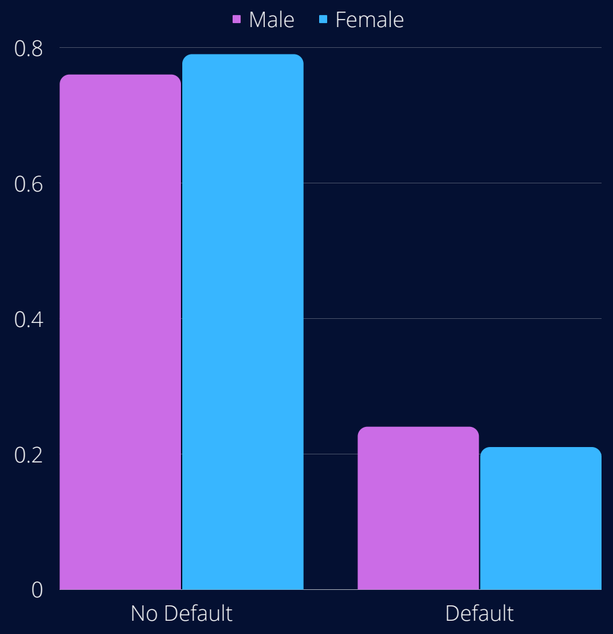
30,000	OBSERVATIONS
66.66%	TRAIN OBSERVATIONS
33.33%	TEST OBSERVATIONS
77.93%	NO DEFAULTS
22.07%	DEFAULTS

23	PREDICTORS
14	NUMERICAL PREDICTORS
9	CATEGORICAL PREDICTORS

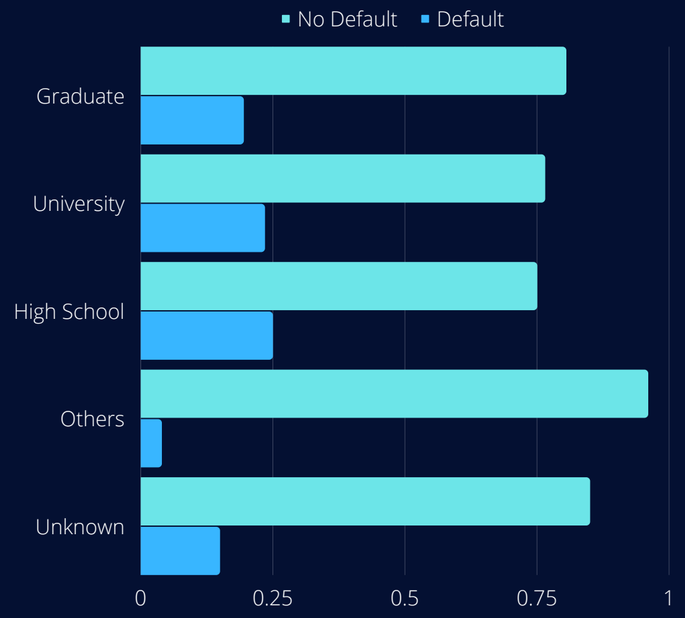
IÉSEG SCHOOL OF MANAGEMENT

EXPLORATORY DATA ANALYSIS

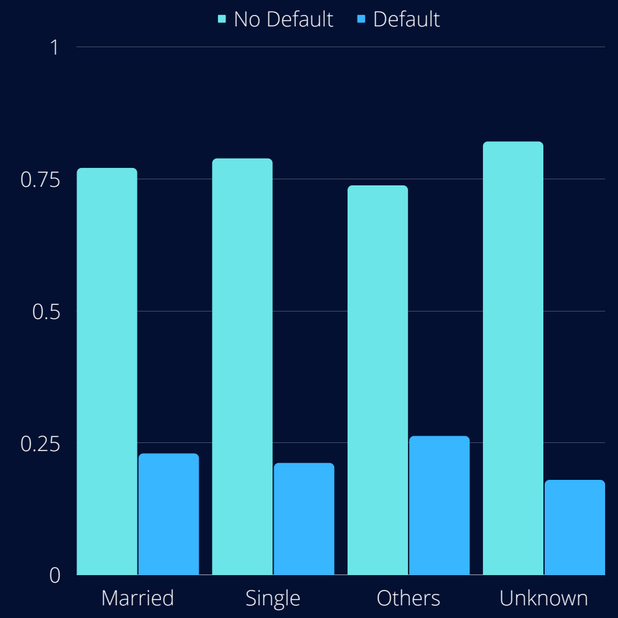
GENDER VS TARGET DISTRIBUTION



EDUCATION VS TARGET DISTRIBUTION



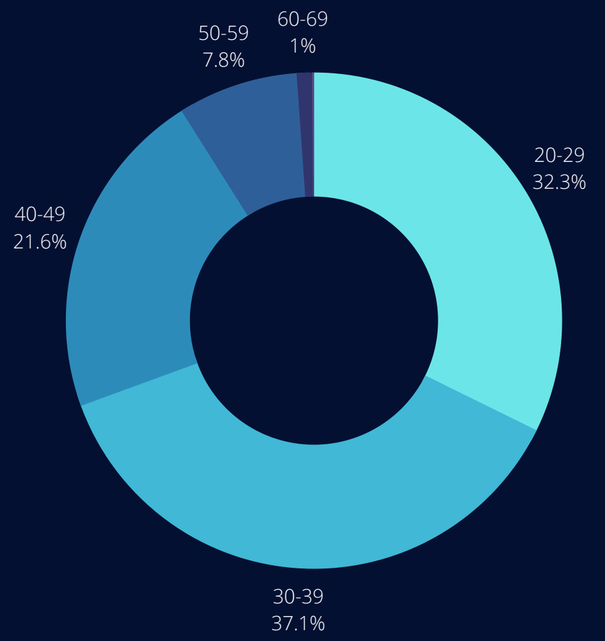
MARRIAGE VS TARGET DISTRIBUTION



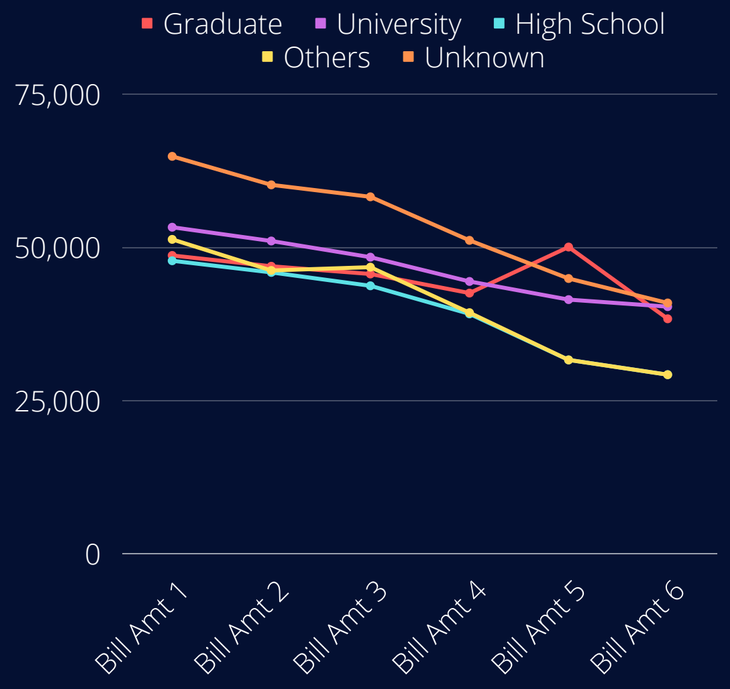
IÉSEG SCHOOL OF MANAGEMENT

EXPLORATORY DATA ANALYSIS

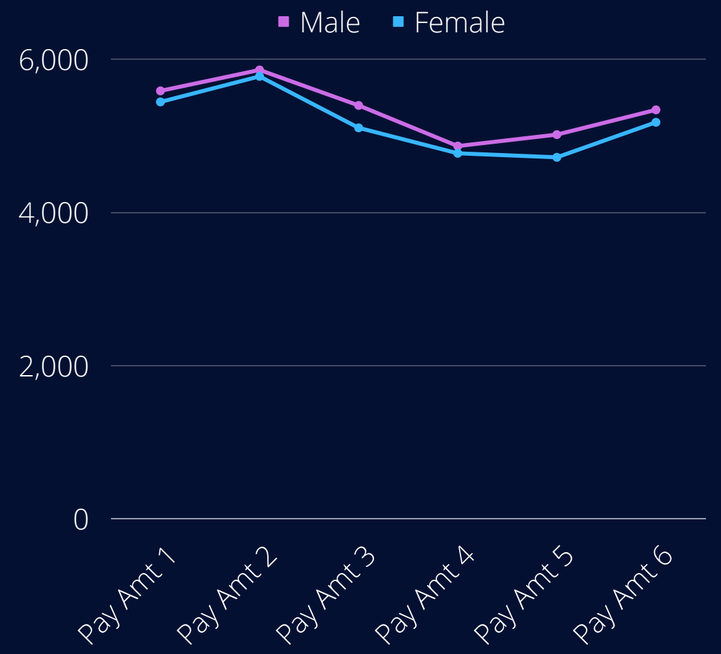
AGE CATEGORIES DISTRIBUTION



EDUCATION VS BILL AMOUNTS



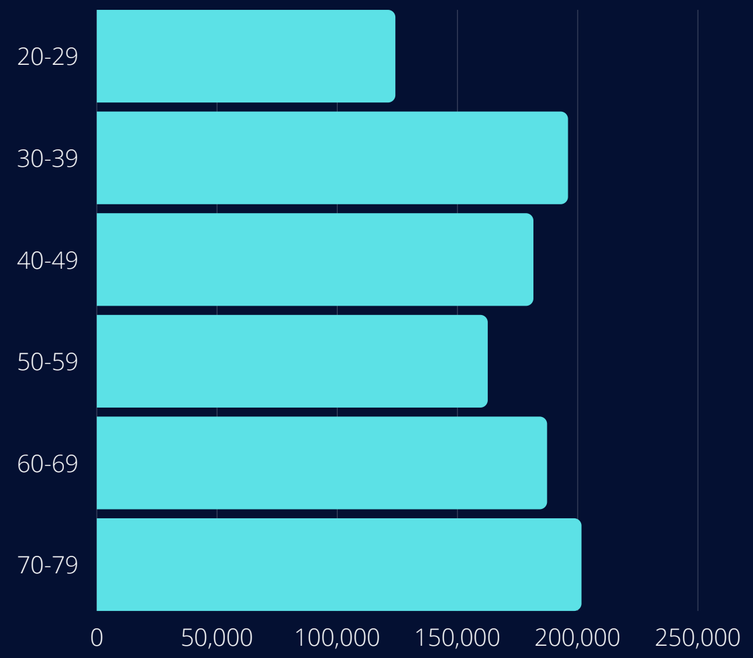
GENDER VS PAY AMOUNTS



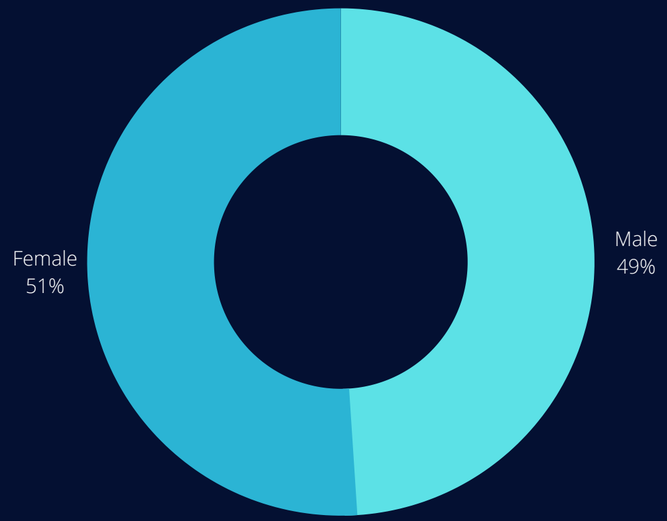
IÉSEG SCHOOL OF MANAGEMENT

EXPLORATORY DATA ANALYSIS

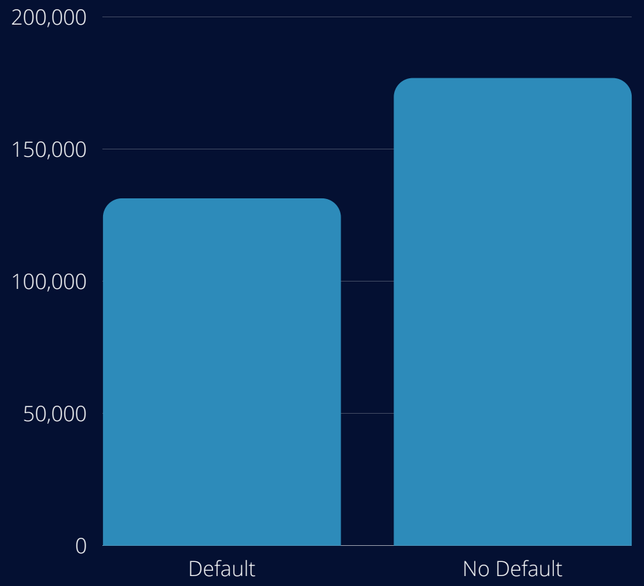
LIMIT BALANCE PER AGE CATEGORY



LIMIT BALANCE PER GENDER



LIMIT BALANCE AVERAGE PER TARGET



IÉSEG SCHOOL OF MANAGEMENT

DATA PROCESSING

25%

NULL VALUES THRESHOLD

DROP VARIABLES OVER THRESHOLD

EDUCATION FEATURE

MERGE MULTIPLE UNKNOWN CATEGORIES

MARRIAGE FEATURE

MERGE MULTIPLE UNKNOWN CATEGORIES

IÉSEG SCHOOL OF MANAGEMENT

DATA PROCESSING

FILLING NULL WITH -1 IN CATEGORICAL VARIABLES

CREATING INCIDENCE RATES PER PAYMENT VARIABLE

ONE HOT ENCODING MARRIAGE, SEX, EDUCATION

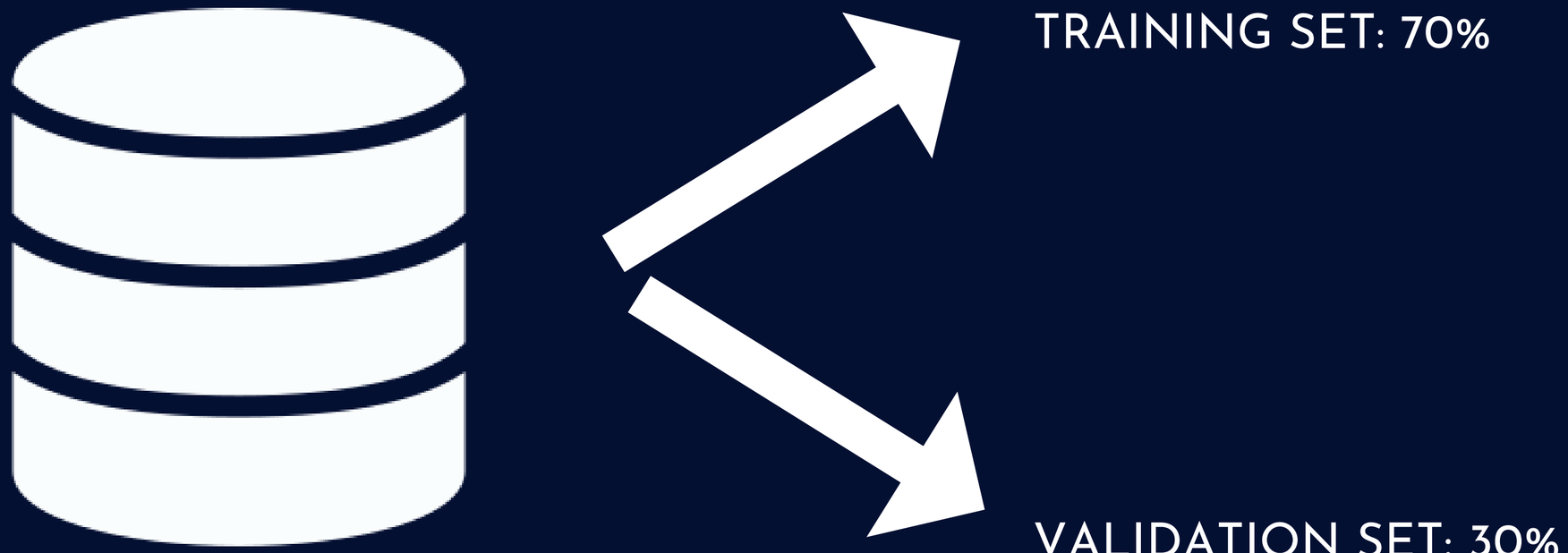
MISSING VALUE IMPUTOR USING MEAN IN NUMERICAL VARIABLES

WINSORIZING OUTLIERS ON NUMERICAL CATEGORIES

NORMALIZING NUMERICAL VARIABLES

IÉSEG
SCHOOL OF MANAGEMENT

MODELING

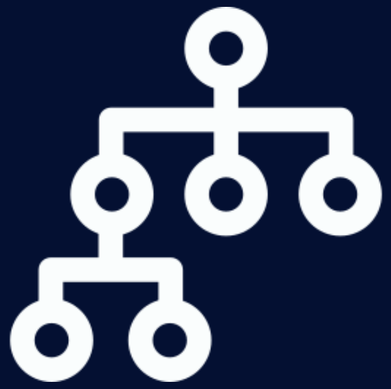


IÉSEG SCHOOL OF MANAGEMENT

MODELING

LOGISTIC REGRESSION

AUC: 0.77



DECISION TREES

AUC: 0.69

RANDOM FOREST

AUC: 0.79

K NEAREST NEIGHBORS

AUC: 0.73

IÉSEG SCHOOL OF MANAGEMENT

MODELING

NEURAL NETWORKS

AUC: 0.78

GRADIENT BOOSTING

AUC: 0.79

SV CLASSIFIER

AUC: 0.77

HYBRID MODEL

AUC: 0.80

IÉSEG SCHOOL OF MANAGEMENT

EXPERIMENTAL SETUP

HYBRID MODEL

SUBSAMPLE: 0.6
MIN_CHILD_WEIGHT: 1
MAX_DEPTH: 10
LEARNING RATE: 0.1
GAMMA: 5
TREE METHOD: GPU_HIST
COLSAMPLE_BYTREE: 0.6

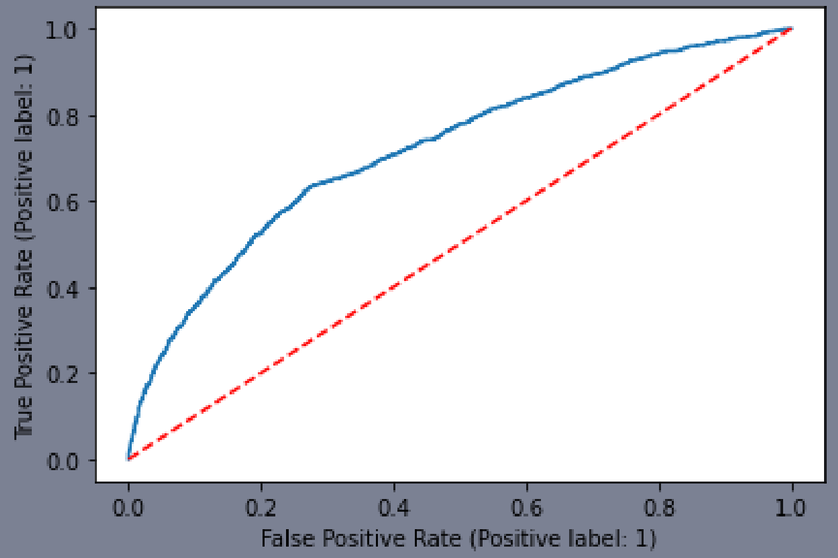
CRITERION: GINI
N_ESTIMATORS: 200
MAX_DEPTH: 10
MAX_FEATURES: AUTO

IÉSEG
SCHOOL OF MANAGEMENT

RESULTS

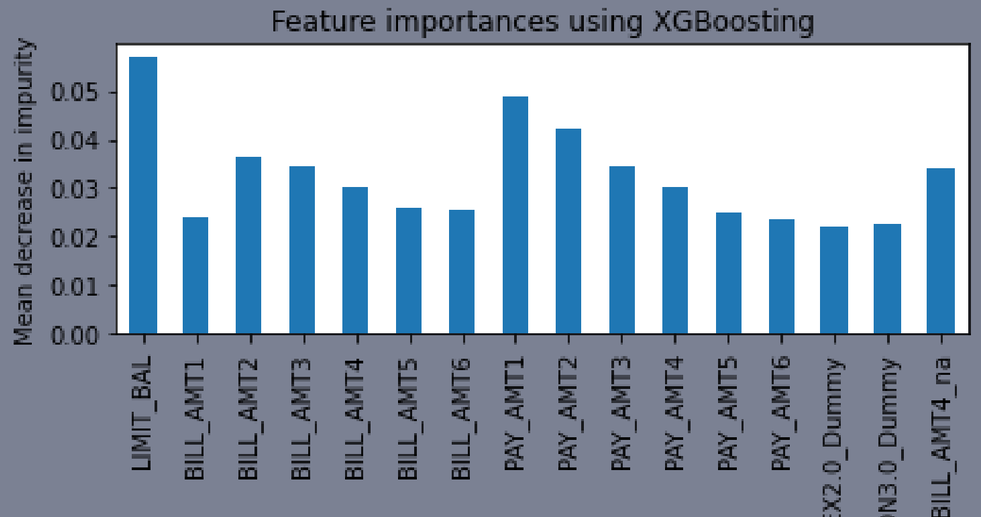
ACCURACY TRAIN: 0.80
ACCURACY VAL: 0.788

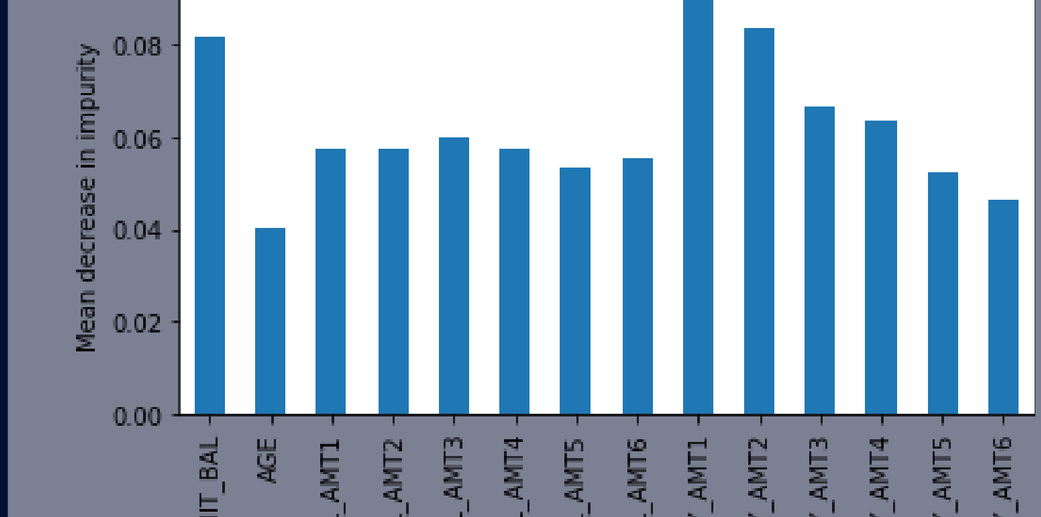
AUC TRAIN: 0.86
AUC VAL: 0.73



IÉSEG SCHOOL OF MANAGEMENT

RESULTS



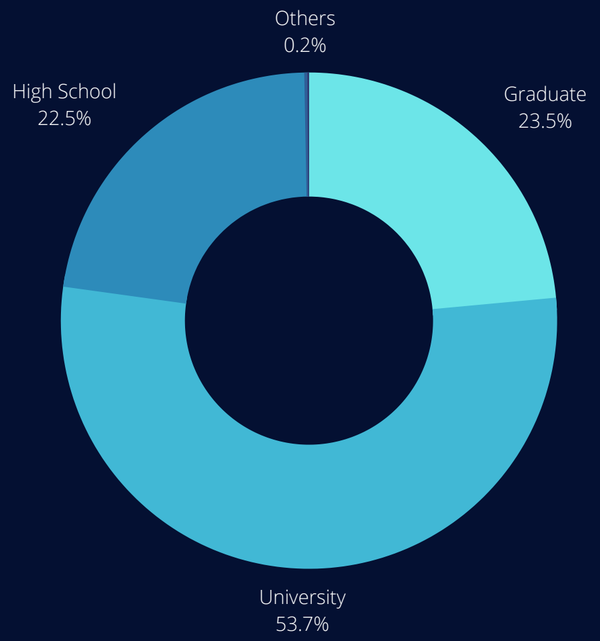


IÉSEG SCHOOL OF MANAGEMENT

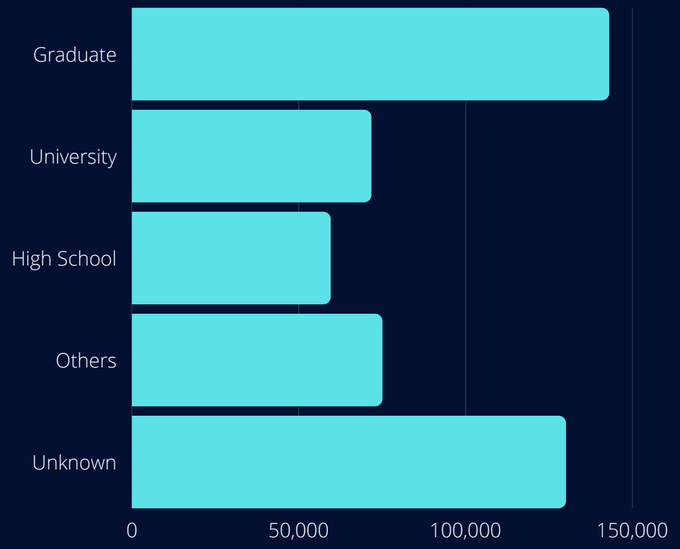
PROFILING

TOP 10% PROFILES WITH HIGHEST PROBABILITY TO DEFAULT

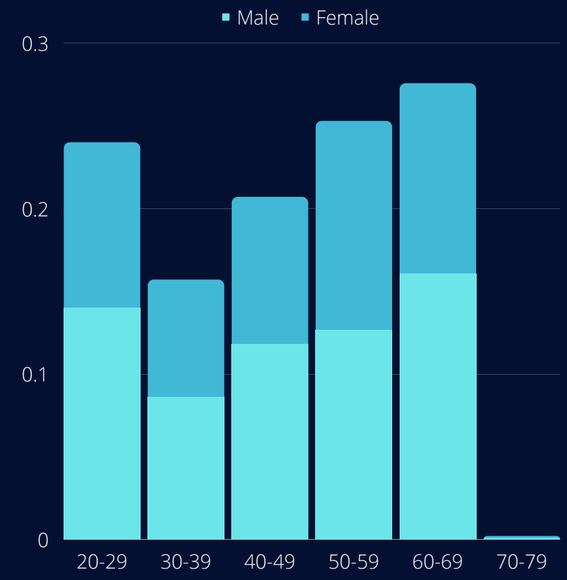
EDUCATION DISTRIBUTION



LIMIT BALANCE PER EDUCATION



AGE CATEGORY DEFAULT RATE

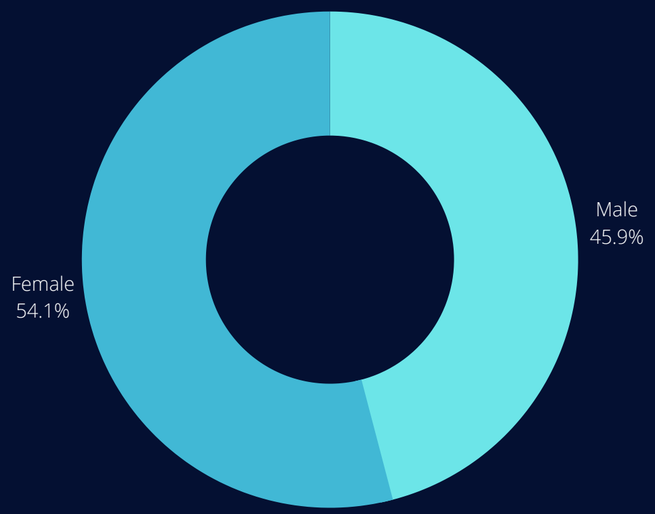


IÉSEG SCHOOL OF MANAGEMENT

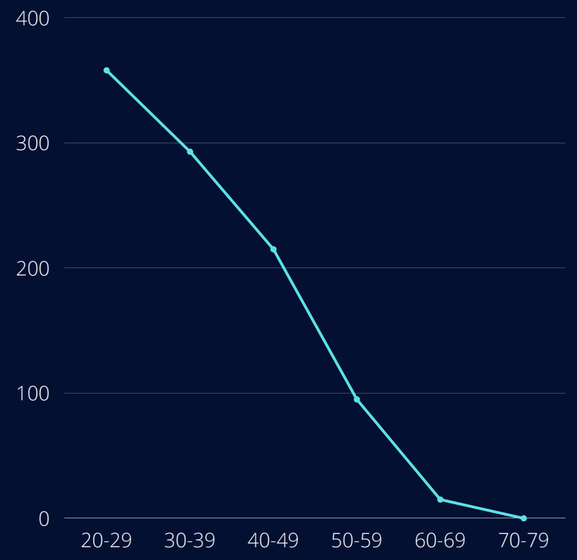
PROFILING

TOP 10% PROFILES WITH HIGHEST PROBABILITY TO DEFAULT

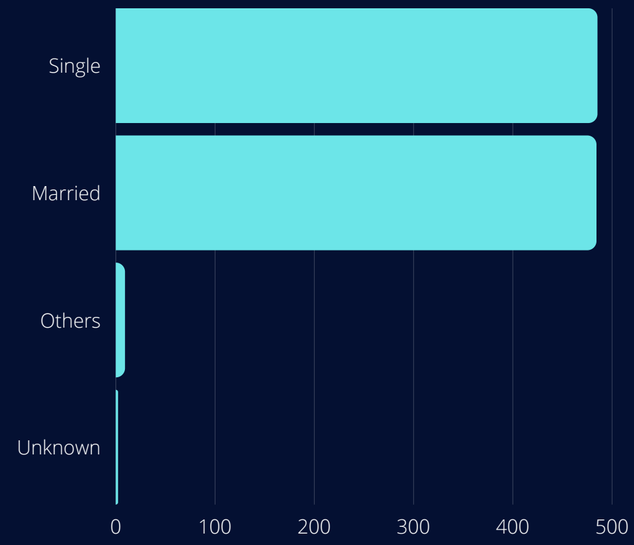
GENDER DISTRIBUTION



AGE CATEGORY DISTRIBUTION



MARRIAGE DISTRIBUTION



IÉSEG SCHOOL OF MANAGEMENT

Q&A SESSION

IÉSEG
SCHOOL OF MANAGEMENT

ML BENCHMARK

PRESENTATION

IÉSEG
SCHOOL OF MANAGEMENT