DATA SCIENCE PORTFOLIO

MUTHIA AISYAH PUTRI

medium.com/@muthiaaisyahputri26

https://github.com/muthiaap

Muthia Aisyah Putri

Completed Bachelor of Mathematics Education from Jakarta State University. Interested in studying mathematics, academics, and data. Currently pursuing a lot in the field of data science by attending bootcamps and several courses.

EDUCATION

Universitas Negeri Jakarta

Mathematics Education

YEAR : 2017 - 2022
GRADE : 3.73/4.00

WORKING EXPERIENCES

2020

SMAN 100 Jakarta

Mathematics Teacher

Taught mathematics to 10th graders.

2021 - 2022

ZENIUS EDUCATION

Mathematics Tutor

Taught mathematics to high school students through learning videos and tests as one of the products of Zenius Education.

2022 - PRESENT

QUIPPER

Mathematics Specialist

Carried out quality control and supervision of learning products produced by Quipper Indonesia, especially in mathematics.

SKILLS AND PROFICIENCY

Math and Statistics

Python Coding

SQL Database

Data Visualization

Machine Learning

DATA SCIENCE PROJECT

01
Predicting Telco Customer Churn

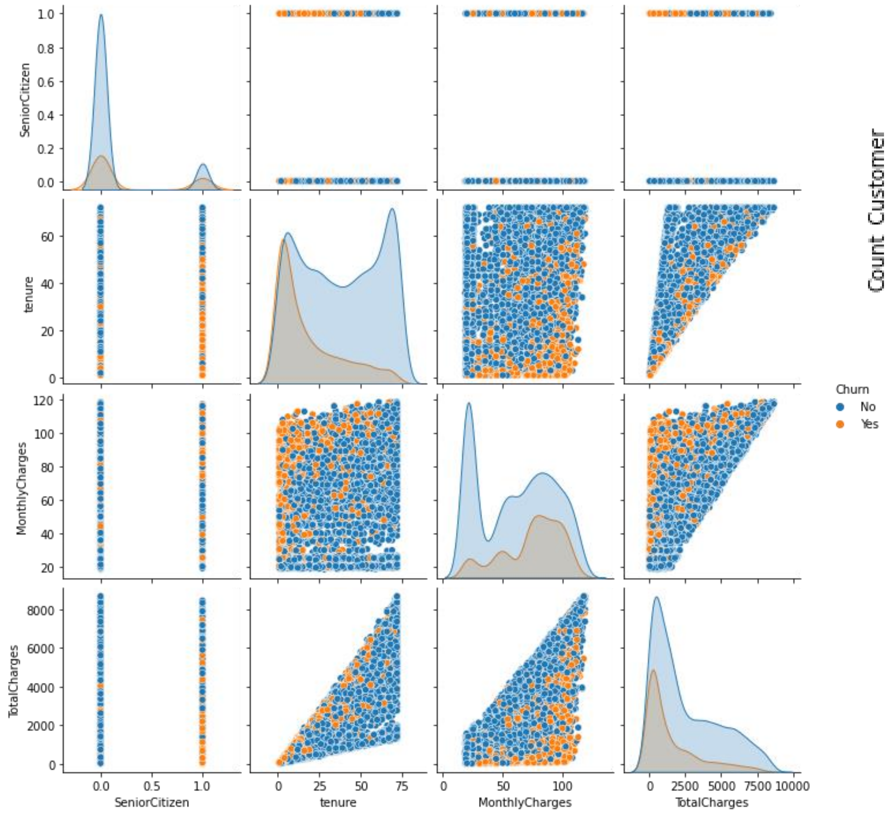
02
Airline Customer Value Analysis

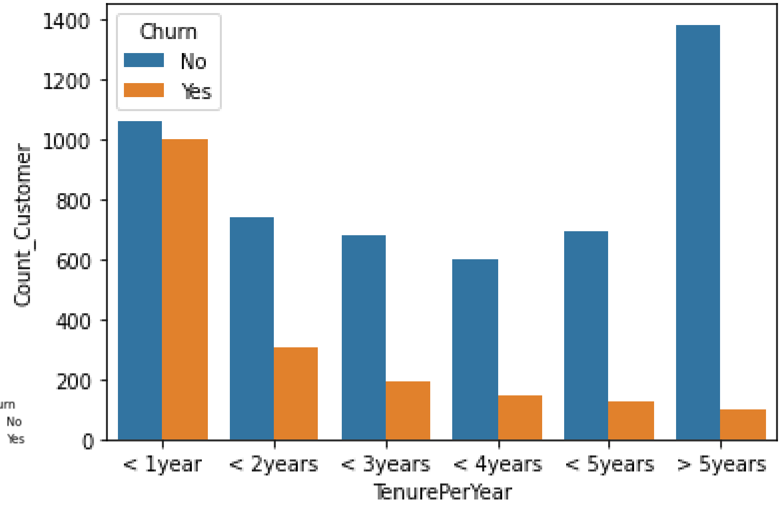
03
Analyzing Bank Customers on Credit Approval Predictions

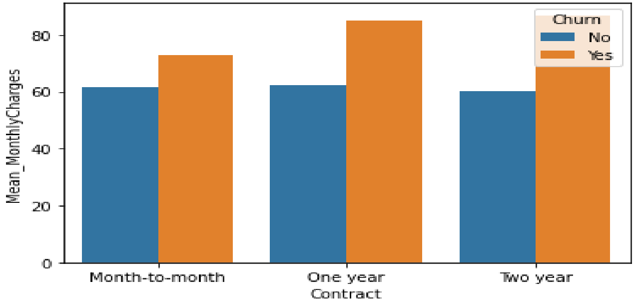
Predict behavior to retain customers. Analyze all relevant customer data and develop focused customer retention programs.

01

Predicting Telco Customer Churn





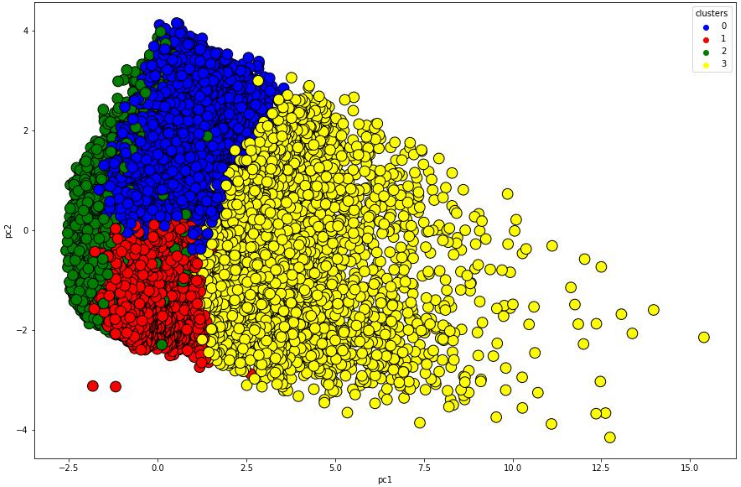


Predicting Telco Customer Churn

Segmenting customers at an airline company. Based on the flight activity records of all customers and transactions, customers can be classified to determine the best treatment.

02

Airline Customer Value Analysis



Cluster 0. New customers (around 27 months) with the least average discounts (around 68%).
Cluster 1. Customers with the least number of flights per year (around 2 times) but they are old customers (around 75 months).
Cluster 2. Customers who haven't made a flight for a long time and classified as new customers (around 48 months).
Cluster 3. Customers with the farthest total flight distance (around 50000km) with the biggest number of flights per year (around 10).

	index		FFP_TIER		SEG_KM_SUM		LAST_TO_END		avg_discount		Meet_Time		Flight_Year	
	mean	median	mean	median	mean	median	mean	median	mean	median	mean	median	mean	median
clusters														
0	29943.472302	28715.0	4.001599	4.0	16097.952246	11772.0	107.308427	79.0	0.680202	0.688594	27.104427	24.936857	7.872241	6.440044
1	29165.071797	27664.0	4.002426	4.0	16706.744709	12114.5	94.697842	72.0	0.697932	0.703143	75.256094	74.810571	2.340385	1.774039
2	45139.589926	47536.0	4.001538	4.0	6129.582044	4574.0	456.384120	447.0	0.714648	0.725000	48.343690	42.678494	3.613595	1.948999
3	7094.329903	3911.0	5.224200	5.0	58984.892629	50854.0	27.604172	12.0	0.767737	0.763938	62.477917	63.475636	10.913414	9.235332

Airline Customer Value Analysis

Analyzing personal customer information criteria for a bank based on predictions that credit applications will be approved. Some of the most meaningful information can be identified by predicting whether credit applications from customers will be approved.

03

Analyzing Bank Customers on Credit Approval Predictions

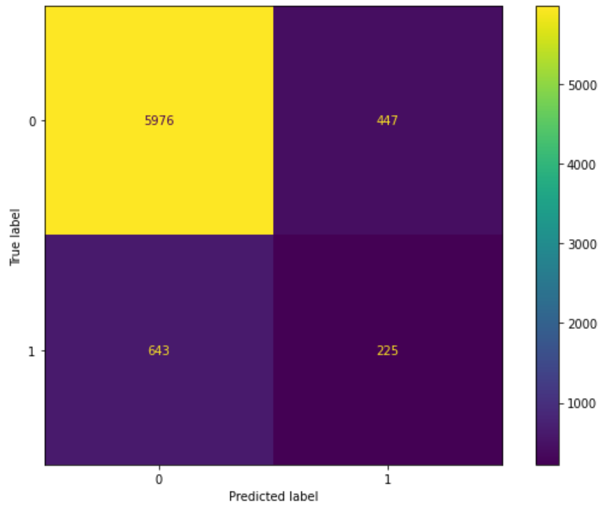
11,8%

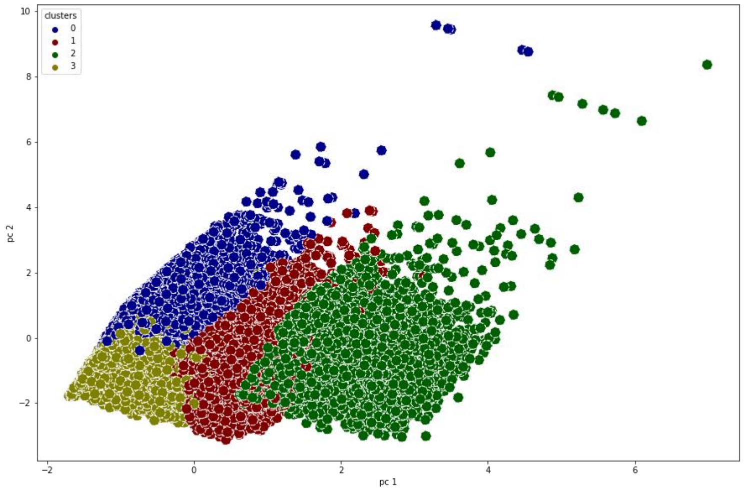
Predicted to be risk

88,2%

Predicted not at risk

	Method	Precision	Recall	F1-Score	Accuracy	AOC
0	LogisticRegression	0.158200	0.534562	0.244146	0.605953	0.601110
1	KNeighborsClassifier	0.195835	0.455069	0.273830	0.712659	0.618404
2	DecisionTreeClassifier	0.250549	0.262673	0.256468	0.818681	0.578652
3	RandomForestClassifier	0.336377	0.254608	0.289836	0.851461	0.712241
4	XGBClassifier	0.153948	0.804147	0.258423	0.450555	0.657442





Analyzing Bank Customers on Credit Approval Predictions

CONTACT ME

EMAIL

muthiaaisyahputri26@gmail.com

PHONE

(+62) 821 1171 5878

LINKEDIN

www.linkedin.com/in/muthiaap/

Analyzing Bank Customers on Credit Approval Predictions

MUTHIA AISYAH PUTRI

OUTLINE

1. Data Feature

2. Machine Learning Model

3. Customer Analysis

4. Conclusion

INTRODUCTION

World Situation

Technological Understanding

Application Record

Personal information of the customer

Credit Record

Credit activity for the last 5 years

01

DATA FEATURE

Source: https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction?datasetId=426827&sortBy=voteCount

DATA FEATURE

Application Record

438557 Customer

Credit Record

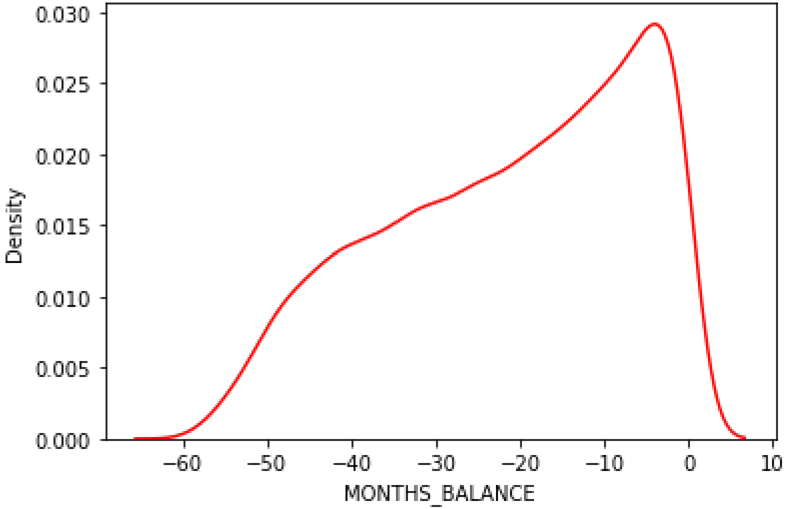
1048575 activity from
36457 customer

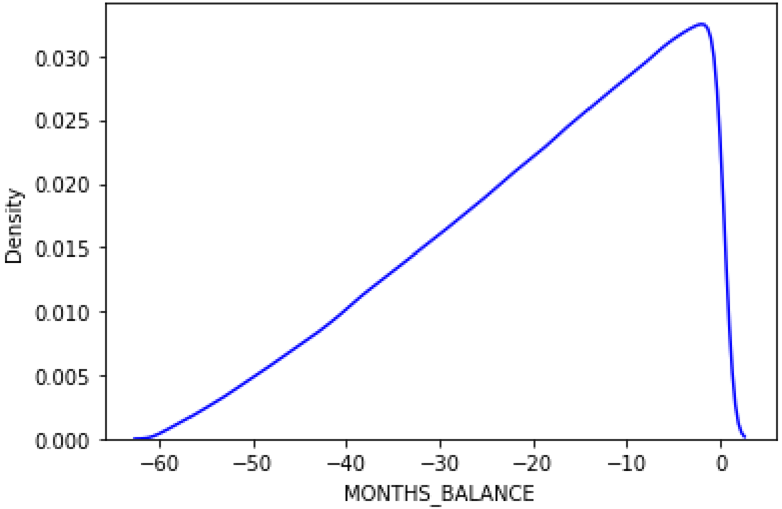
402100 customer

Offers can be given to customers by giving a small interest rate or offers related to reward points for the first transaction.

DATA FEATURE

How is the increase in the number of customers who make credit from time to time?





Overdue Customer

All Customer

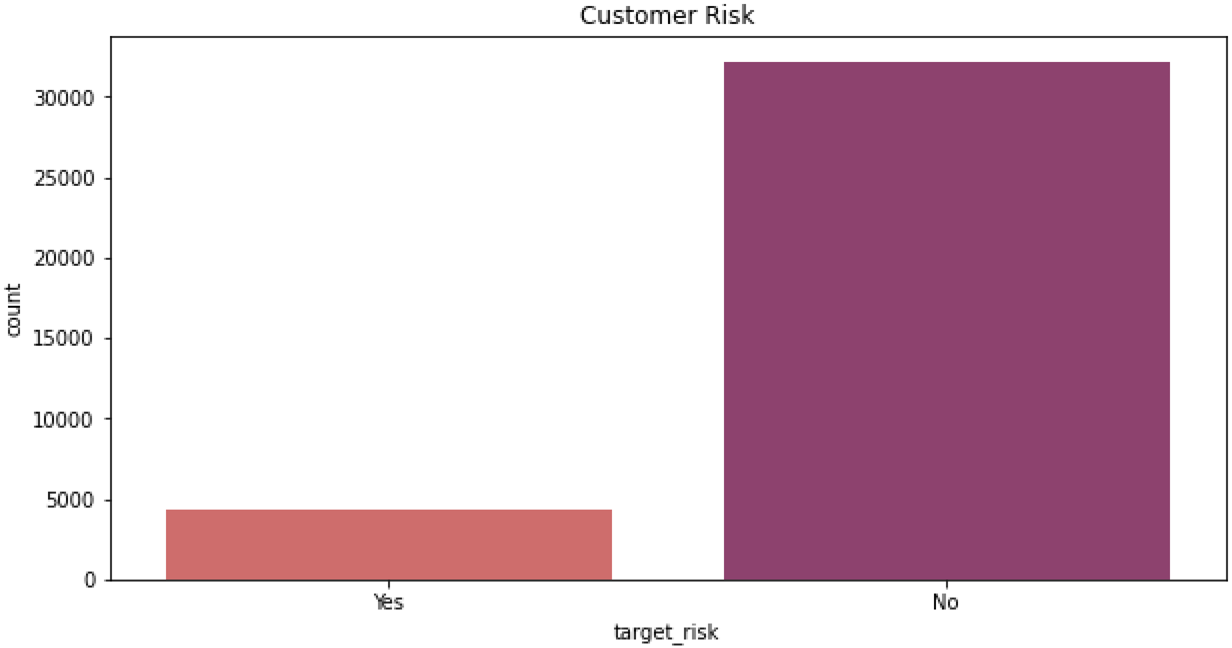
DATA FEATURE

How many Customers make credit most often and pay on time within 5 years?

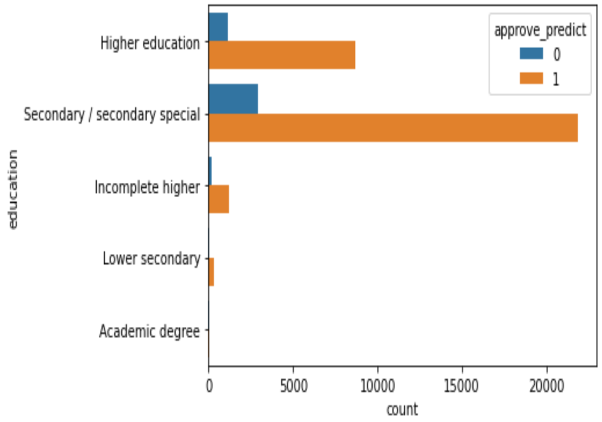
	MONTHS_BALANCE		status	
	count	max	count	max
ID				
5001730	61	0	61	0
5002160	61	0	61	0
5002165	61	0	61	0
5002171	61	0	61	0
5002287	61	0	61	0
...	...	...	...	...
5143482	61	0	61	0
5145767	61	0	61	0
5146385	61	0	61	0
5148524	61	0	61	0
5148819	61	0	61	0

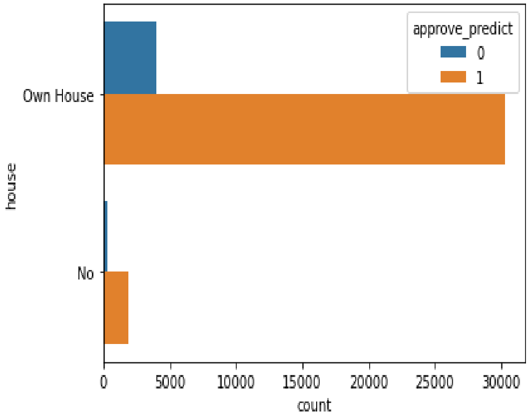
180 rows × 4 columns

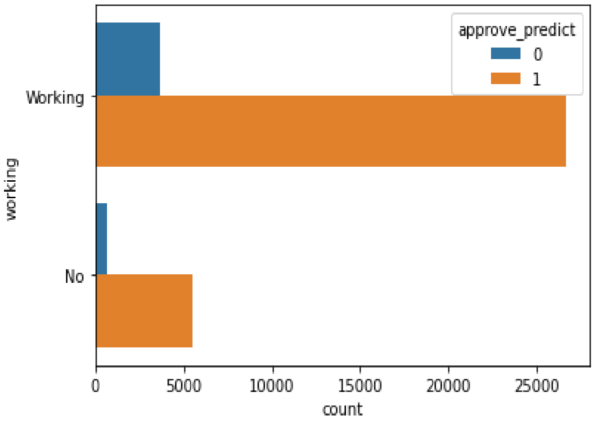
DATA FEATURE



DATA FEATURE







02

MACHINE LEARNING MODEL

DATA FOR MODELLING

11,8%

Predicted to be risk

88,2%

Predicted not at risk

80% Train Data

20% Test Data

ACCURACY FOR THE PREDICTING

	Method	Precision	Recall	F1-Score	Accuracy	AOC
0	LogisticRegression	0.151318	0.562212	0.238456	0.572487	0.589022
1	KNeighborsClassifier	0.147237	0.475806	0.224884	0.609519	0.578105
2	DecisionTreeClassifier	0.161409	0.353687	0.221661	0.704293	0.550195
3	RandomForestClassifier	0.213064	0.315668	0.254410	0.779728	0.647052
4	XGBClassifier	0.164163	0.352535	0.224012	0.709231	0.617414

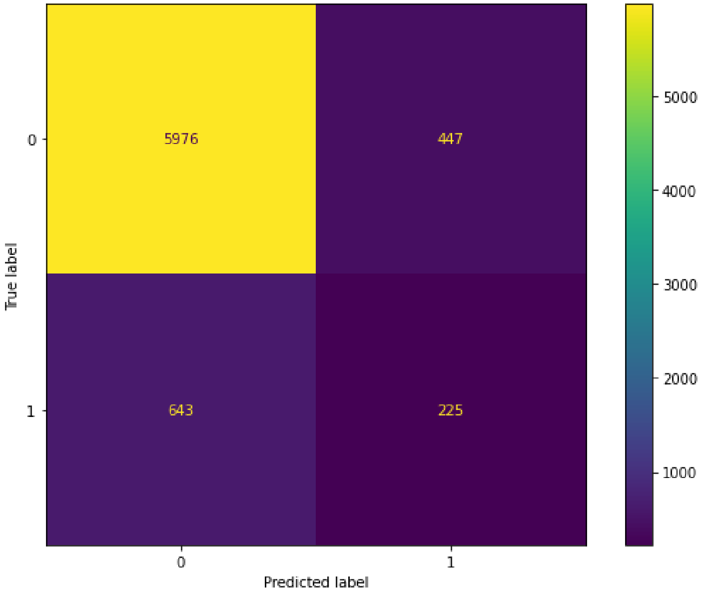
Oversampling only

Sampling and oversampling

	Method	Precision	Recall	F1-Score	Accuracy	AOC
0	LogisticRegression	0.158200	0.534562	0.244146	0.605953	0.601110
1	KNeighborsClassifier	0.195835	0.455069	0.273830	0.712659	0.618404
2	DecisionTreeClassifier	0.250549	0.262673	0.256468	0.818681	0.578652
3	RandomForestClassifier	0.336377	0.254608	0.289836	0.851461	0.712241
4	XGBClassifier	0.153948	0.804147	0.258423	0.450555	0.657442

ACCURACY FOR THE PREDICTING

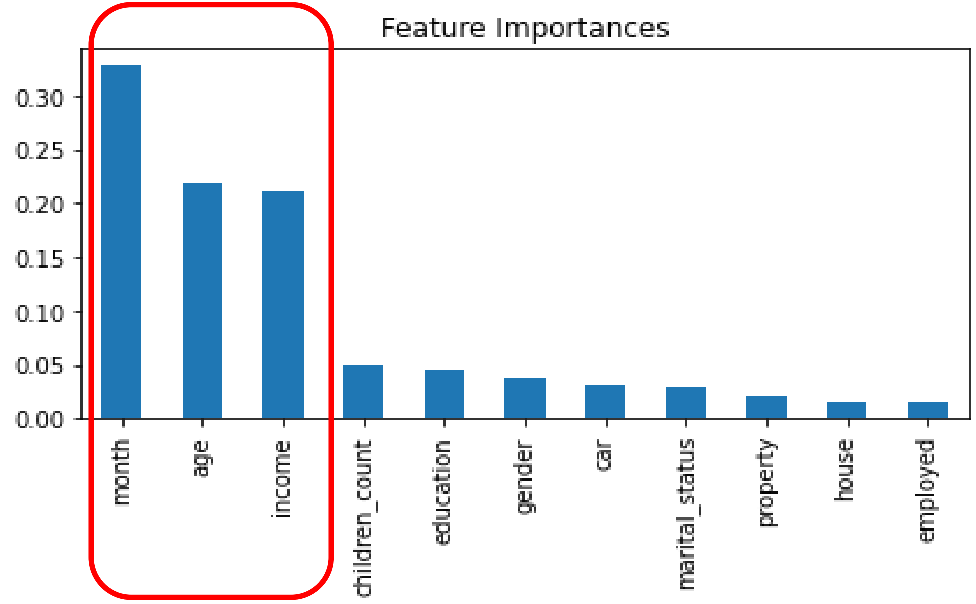
Oversampling only



	precision	recall	f1-score	support
0	0.90	0.93	0.92	6423
1	0.33	0.26	0.29	868
accuracy			0.85	7291
macro avg	0.62	0.59	0.60	7291
weighted avg	0.84	0.85	0.84	7291

FEATURE IMPORTANCES

Oversampling only



03

CUSTOMER ANALYSIS

ANALYSIS BY THE CLUSTERS

	month		income		age		target_risk	
	mean	median	mean	median	mean	median	mean	median
clusters								
0	-7.061689	-5.0	202575.534070	180000.0	34.597019	34.0	0.0	0.0
1	-36.292813	-35.0	191317.580982	171000.0	43.047370	42.0	0.0	0.0
2	-20.148053	-18.0	193415.684542	171000.0	42.312660	41.0	1.0	1.0
3	-8.553359	-6.0	158555.244187	135000.0	55.358108	56.0	0.0	0.0

ANALYSIS BY THE CLUSTERS

0

- 13487 Customers
- 75% age <40
- High Income
- Time < 3 years
- Not at risk

1

- 8613 Customers
- Random age
- Middle Income
- Time > 2 years
- Not at risk

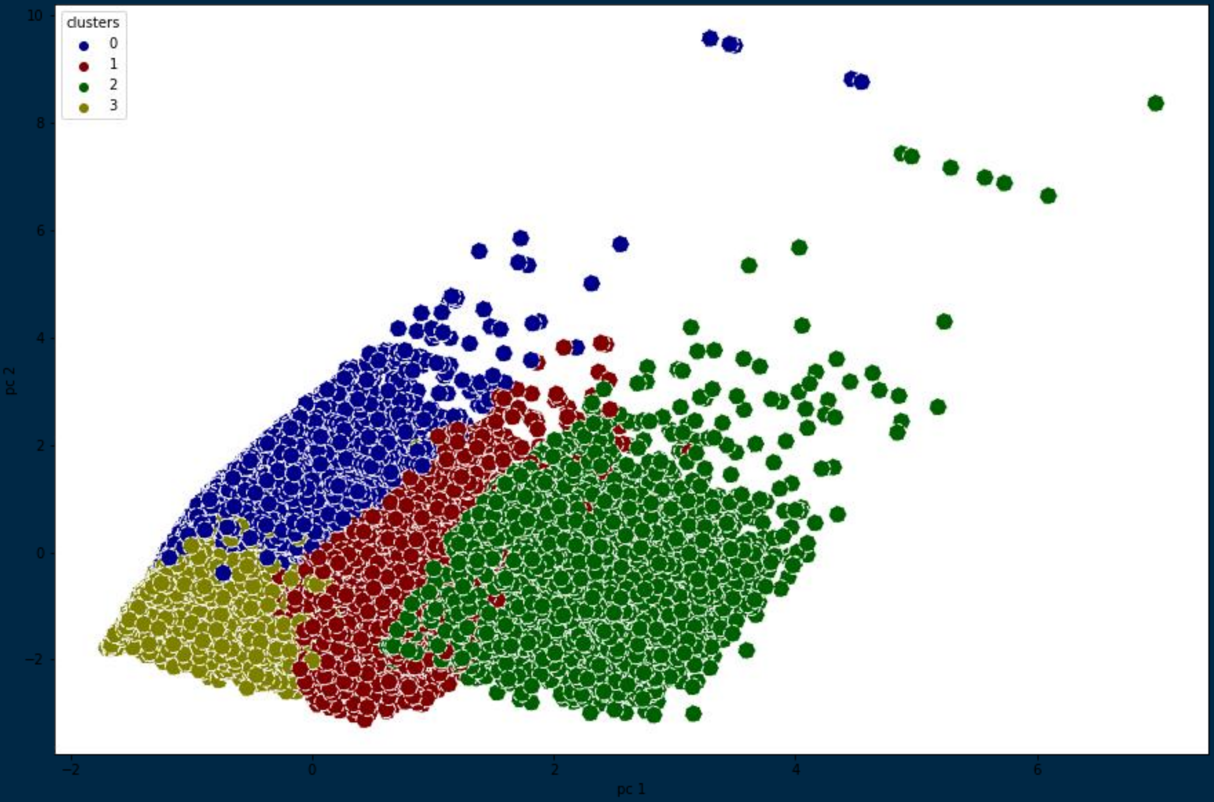
2

- 4289 Customers
- Random age
- Middle Income
- 75% time < 3 years
- At risk

3

- 10064 Customers
- Age > 40
- Low Income
- Time < 3 years
- Not at risk

ANALYSIS BY THE CLUSTERS



04

CONCLUSION

INSIGHT

offers for high-income customers

Offer for first transaction

Offer for those who frequently do credit.

Impose late fees

THANK YOU

MUTHIA AISYAH PUTRI