ESTIMATING CLIENTS' NEEDS

BUSINESS CASE 2

R. Aitkaliyev,
A. Frabetti,
A. Iacovelli,
L. Marcosignori,
M. Rizzo

Fintech Final Project
21/22
Prof. Raffaele Zenti

The Workflow

0
Introduction
Datasets and goals

01
Data exploration
Definition of the ideal target audience: 1-1

02
PCA
PCA on subsets 1-0, 0-1, 0-0

03
Clustering
Clustering and defining transition clusters

04
Classifiers
Finding the best classifier for our target variables

05
Test
Testing the classifiers and using QDA classifier to predict transition clusters

06
Recommender System

0 Introduction

DATASETS → GOAL

1) NEEDS OF CLIENTS

2) PRODUCTS TO RECOMMEND

RECOMMENDATION SYSTEM

0 Introduction

DATASETS

1) NEEDS OF CLIENTS

2) PRODUCTS TO RECOMMEND

FEATURES

- ID
- Age
- Gender
- Family Members
- Financial Education
- Risk Propensity
- Income
- Wealth

RESPONSE

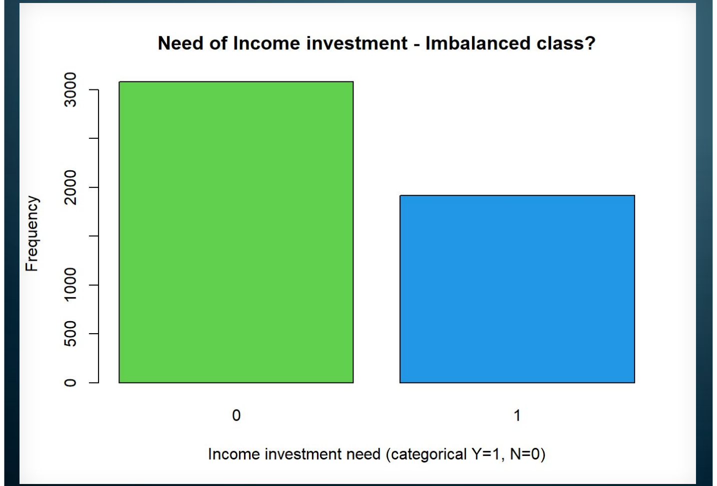
- Accumulation Investment
- Income Investment

PRODUCTS

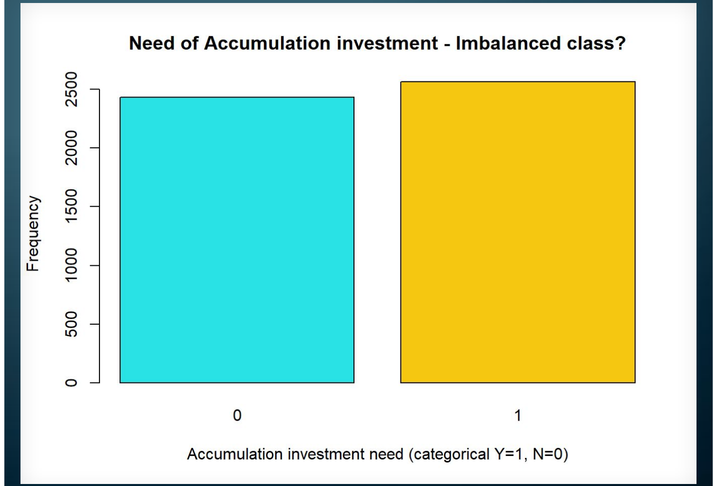
- ID Product
- Type
- Risk

01 Data Exploration

Looking at univariate distributions, we check for **Imbalanced Classes.**
Response variables:



Imbalanced



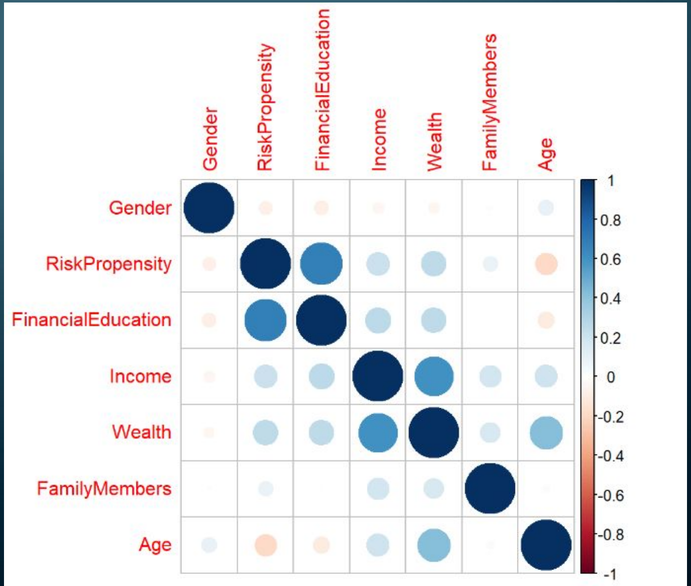
Balanced

01 Data Exploration

Correlation

The variables among which we see a stronger **correlation** are:

- **Financial Education** and **Risk Propensity** (0.68)
- **Income** and **Wealth** (0.60)



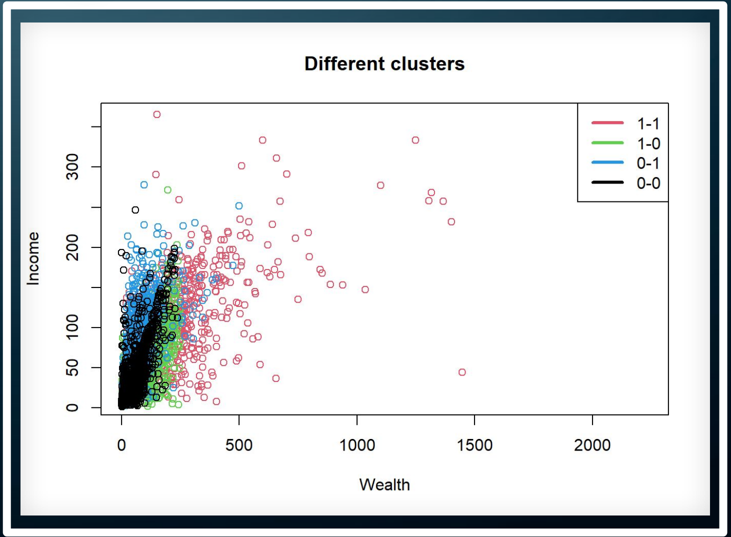
01 Data Exploration

Splitting the dataset

We proceed to split the clients in four subsets taking in account Income and Accumulation propensity to investments (responses):

- **Customers 1-1** (Both investments):
- **Customers 1-0** (Income investment only):
- **Customers 0-1** (Only accumulation investment)
- **Customers 0-0** (No investment)

From the plots we notice that the difference between the subsets is significantly outlined by the responses **Income** and **Wealth**.



01 Data Exploration

Customers inclined to both types of investment

THE IDEAL TARGET:
1 - 1 CUSTOMERS

≃ 60 years old	High Income	High Wealth

01 Data Exploration

Comparison:

IDEAL CLUSTER vs REMAINING CUSTOMERS

Attribute	11	01/10/00	Differences
Age	61.34	53.73	Older in 1-1
Gender	0.48	0.5	Same
FamilyMembers:	2.6	2.5	Same
FinancialEducation	0.45	0.41	Same
RiskPropensity	0.39	0.36	Same
Income vs	87.83	56.8	Richer in 1-1
Wealth	177.45	72.95	More Wealth in 1-1

01 Data Exploration

Logistic Regression to understand significative variables

Having split the data into train and test set (80%/20%), we proceed by performing two logistic regressions to **predict** the variables **Income Investment** and **Accumulation Investment.**

In the prediction of the variable Income Investment play a key role:

- The **Income** variable: negatively correlated;
- The **Wealth** variable: positively correlated;
- The **Age** variable: positively correlated.

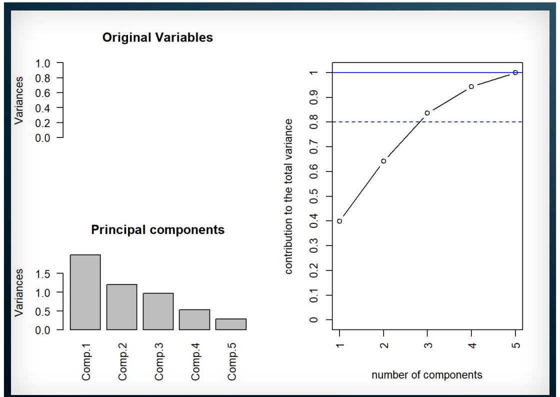
In prediction of the variable Accumulation Investment play a key role:

- The **Income** variable : positively correlated;
- The **Age** variable: negatively correlated.
- The variable **Financial Education**: positively correlated.

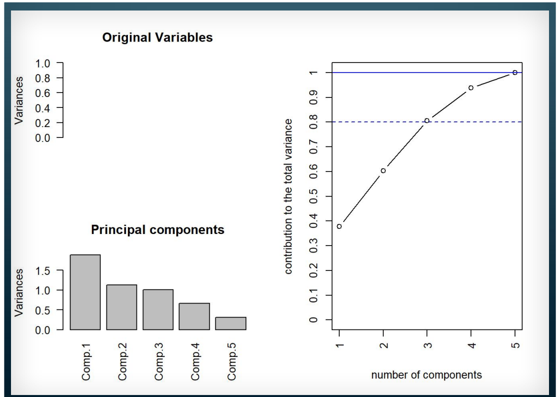
02 PCA

We perform a **PCA** for the subsets 1-0, 0-1 and 0-0 on the training data.

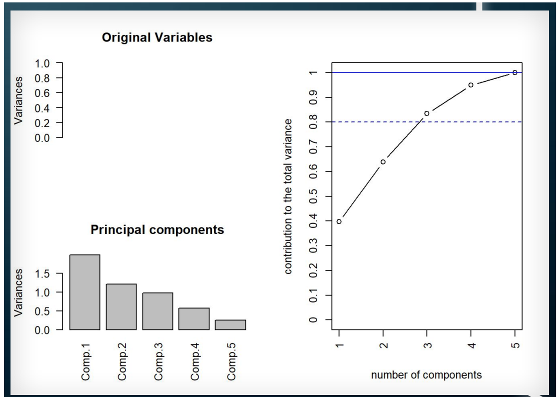
For all the subsets above, 3 principal components explain **more than 80% of the variability in the data**, so we decide **keep only 3** of them.



PCA 1-0

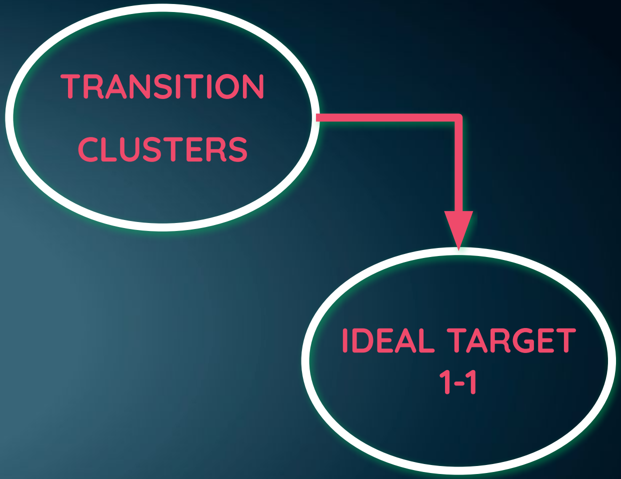


PCA 0-1



PCA 0-0

03 Clustering



- ➔ We perform clustering on the first 3 principal components of each of these 3 subsets: this will allow us to obtain certain clusters that we will find to be very similar to our ideal cluster (1-1). We call these clusters: "**Transition clusters**," since if customers fall here, it means that they can be identified as ideal customers (1-1).

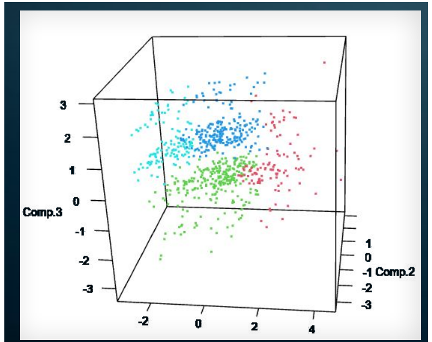
- ➔ These clusters will be useful in later stages when we try to predict new labels on a test dataset: the moment the predicted data fall into the transition clusters, they will be classified as 1-1 data. This will allow our ideal target to expand, while also **correcting possible errors in the classifiers**.

Looking for Transition Clusters

Data 10

Hierarchical clustering
Euclidean Distance
Ward linkage

Number of clusters = 4

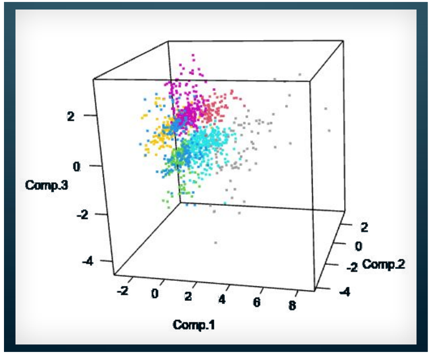


Transition cluster: **1st group**
Here, Income and Wealth are greater than the average ones of the ideal target.

Data 01

Hierarchical clustering
Manhattan Distance
Ward linkage

Number of clusters = 7

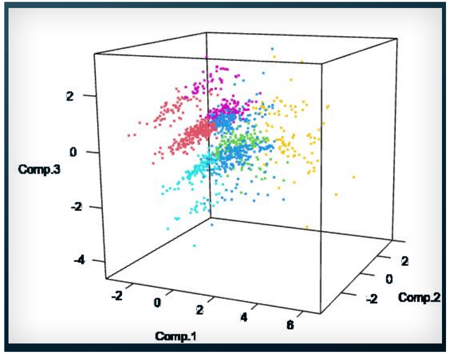


Transition cluster: **7th group**
Here, Income is greater than the average one in the ideal target, while Wealth is quite similar.

Data 11

Hierarchical clustering
Euclidean Distance
Ward linkage

Number of clusters = 6



Transition cluster: **6th group**
Here, Income and Wealth are greater than the average ones of the ideal target.

04 Classifiers

Income Investment prediction:

Logistic Regression

Accumulation Investment prediction:

Ensemble tree (Bagged)

Prediction for Income and Accumulation Investment response variables

Metric's choice

→ Since we are interested to match the persona to the right product, we are interested on labeling in the right way the Income and Accumulation Investment by minimizing the mistakes.

→ For this reason, in our opinion, among all these measures, the best metrics are represented by the F1 and FBeta score since they are computed respectively as the harmonic and weighted mean (Beta = 1.5) of the model's precision and recall.

Accuracy

Precision

Recall

F1 score
FBeta

Income Investment Classifier

Logistic Regression

Confusion Matrix

Predicted class

True class

	0	1
0	460	123
1	143	274

Mis.Error = 0.27

Precision = 0.66

F1 = 0.67

FBeta = 0.68

Accumulation Investment classifier

Misclassification cost

How to adjust the Cost Matrix?

➔ **Increasing the cost for a FN**: 1 → 2

We consider that we would rather recommend a financial product to an uninterested client than the opposite.

Predicted class

True class		0	1
	0	0	1
	1	2	0

In this way we are dealing with a possible **missed opportunity** for profit.

Accumulation Investment Classifier

Ensemble Tree (Bagged)

Confusion Matrix

Predicted class

True class	0	1
0	367	96
1	111	426

Mis.Error = 0.21

Precision = 0.82

F1 = 0.80

FBeta = 0.80

05 Test

After applying the classifiers we have 2 types of responses:

Ideal target:
1-1

Others:
1-0, 0-1, 0-0

We use a **QDA classifier** (Quadratic Discriminant Analysis) to figure out whether the clients are in transition clusters. If yes, they will be sent to the ideal 1-1 cluster.

06 Recommender System

New variables:

- ➔ We define **new variables** that may be useful in the process of deciding which product to recommend.
- ➔ Each variable defined is actually intended to be **dual**, in fact it will have different parameters (alpha) depending on whether one is considering income or accumulation type investment.

New products:

- ➔ **FINANCIAL COURSE**: we propose a discounted Financial course (issued by the bank to which we are hypothetically selling this project) that will push clients to invest, increasing their financial education as well as their risk propensity .
- ➔ **FINANCIAL ADVISOR**: For clients with already above-average financial knowledge, the persuasion process will be more delicate, and may not go through a basic finance course. So, the bank (again the one we are selling this project to..) will be able to offer a free consultation with an expert, which we believe can be an incentive to trust the bank and its recommendations.

Economic availability =

$$\frac{(\alpha)Wealth + (1-\alpha)Income}{log(FamilyMembers + 1)}$$

Economic Availability

Two different variables:

- Income investments: **α** = 0.6
- Accumulation investments: **α** = 0.4

Investment Propensity =

$$(\beta)\ \mathit{Risk\ Propensity}\ +$$
$$(\gamma)\ \mathit{Financial\ Education}\ +$$
$$(1 - \beta - \gamma)\ \mathit{Economic\ Availability}$$

Investment Propensity

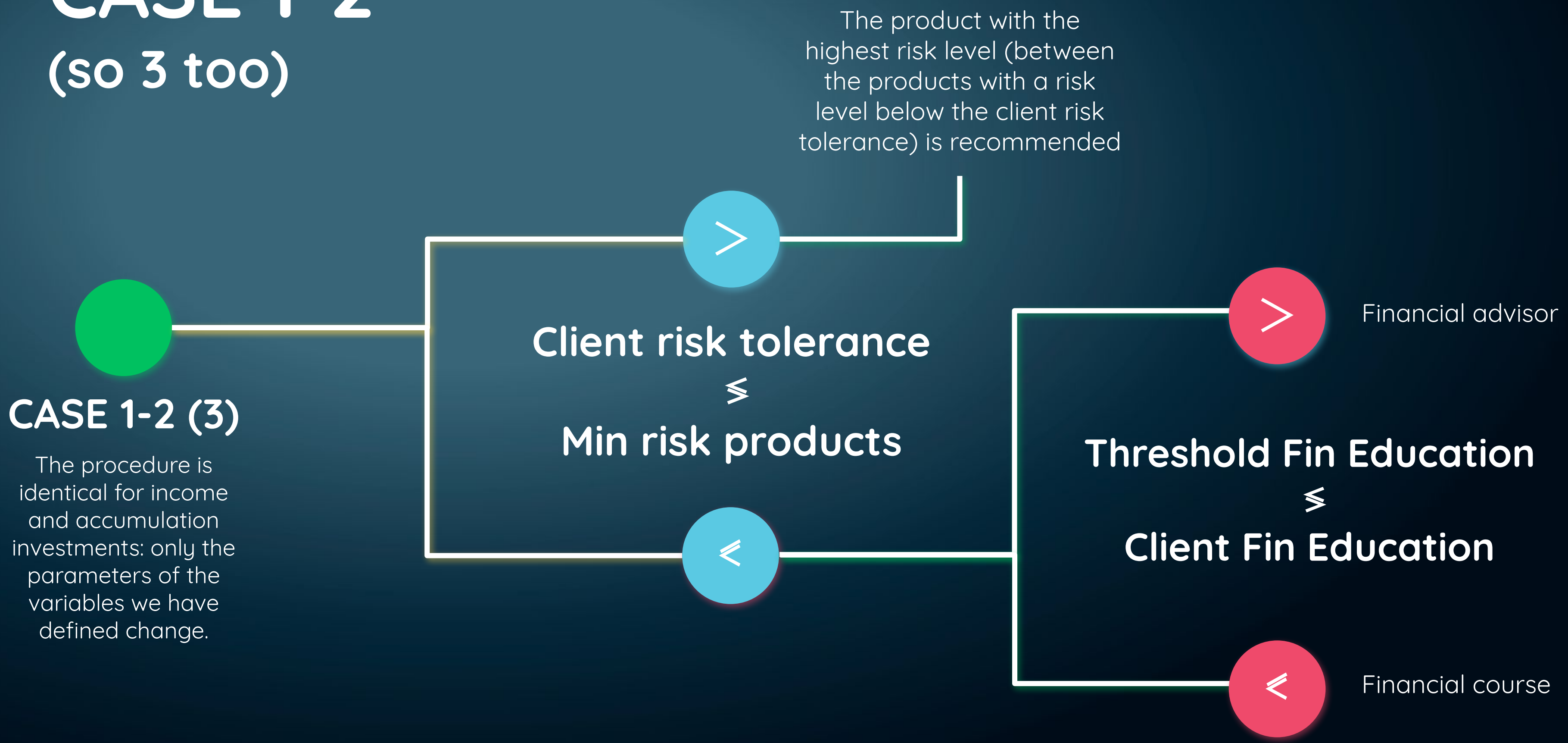
Two different variables, depending on the **α** of the variable *economic availability*!

06 Recommender System

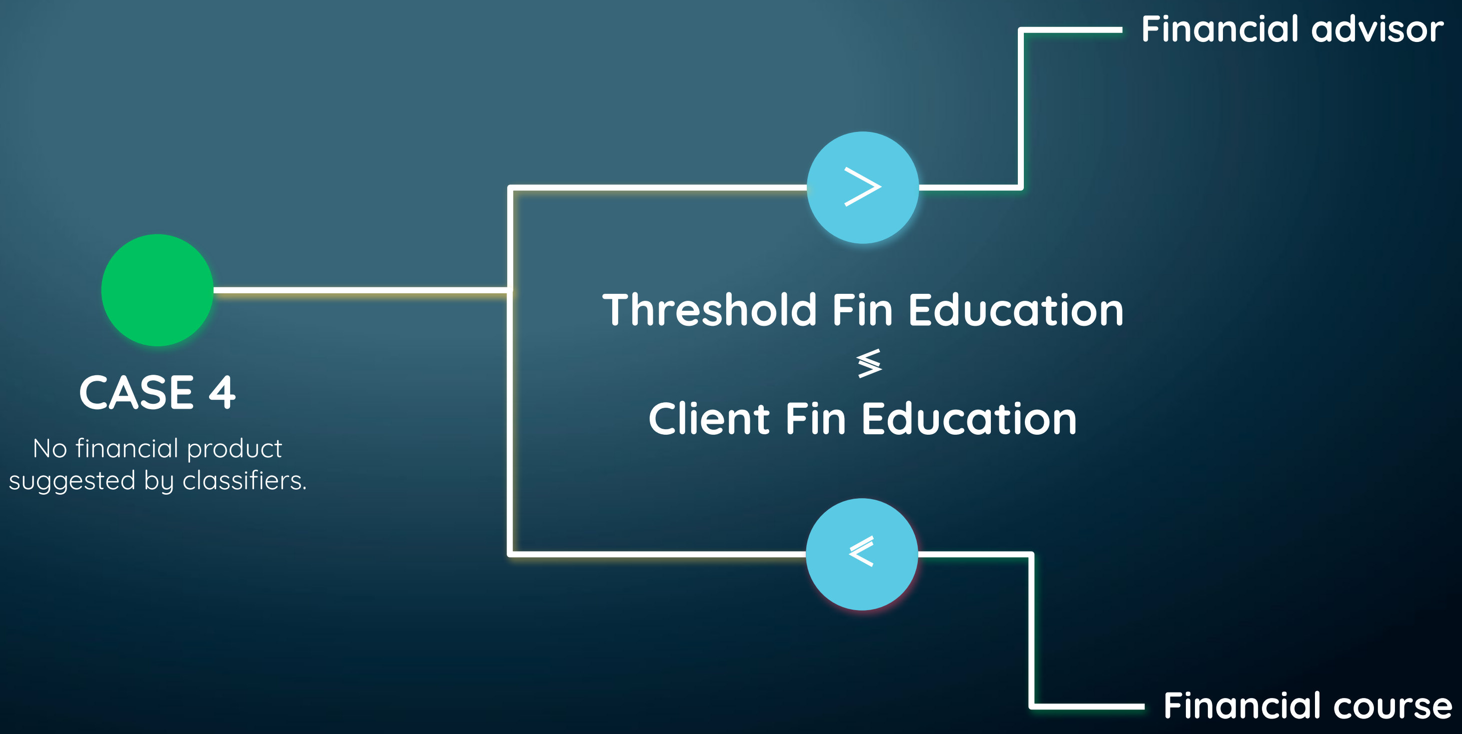
Four different cases:

	Income Investment	Accumulation Investment
CASE 1	✔	✘
CASE 2	✘	✔
CASE 3	✔	✔
CASE 4	✘	✘

CASE 1-2
(so 3 too)



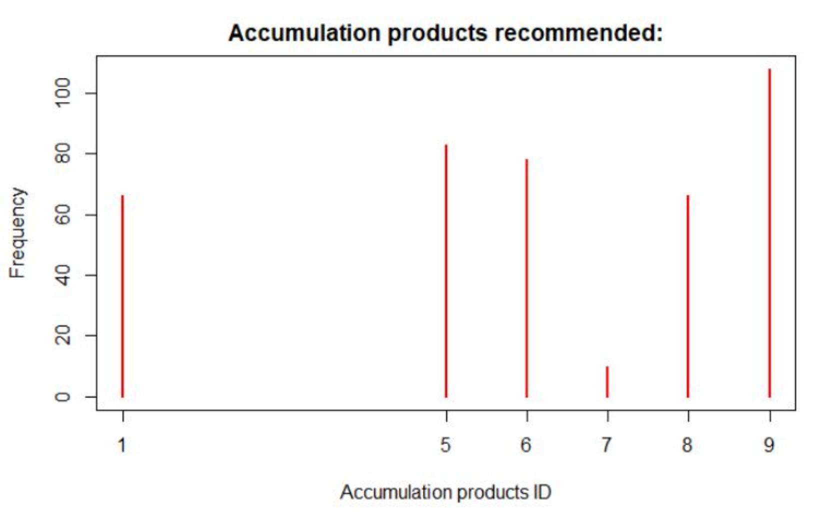
CASE 4



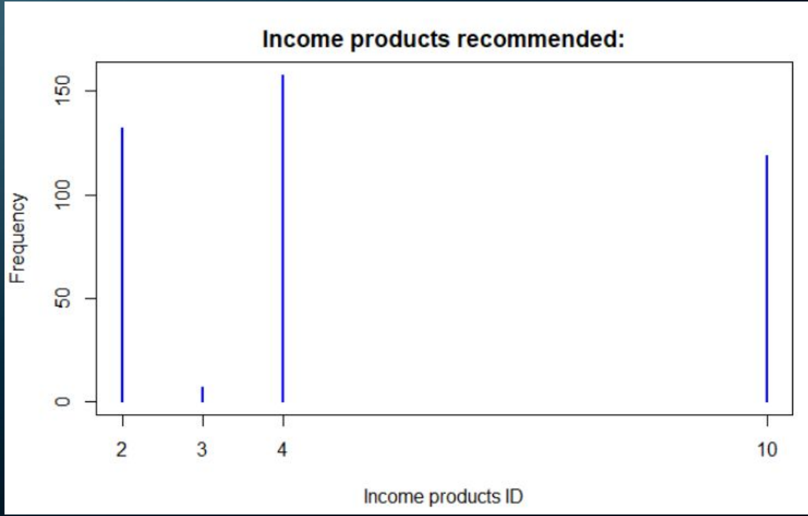
CASE 4
No financial product suggested by classifiers.
$>$
Threshold Fin Education
$\lessgtr$
Client Fin Education
$\leqslant$
Financial advisor
Financial course

Recommended products (Investment)

Accumulation products



Accumulation products

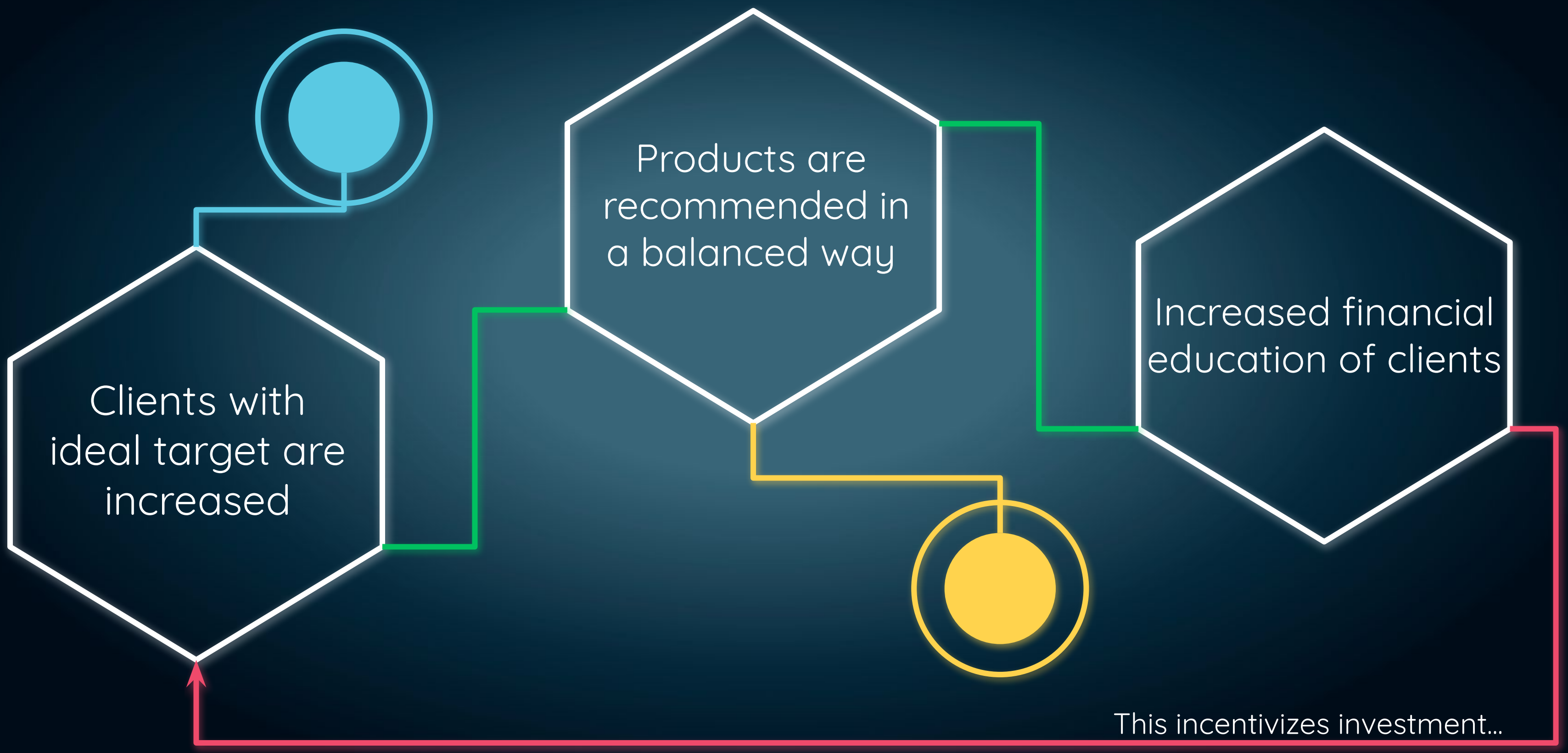


Recommended products
(No Investment)

12: FINANCIAL COURSE ID

13: FINANCIAL ADVISOR ID

Results achieved



Thanks for your attention!

marco4.rizzo@mail.polimi.it
lucrezia.marcosignori@mail.polimi.it
alessandro.frabetti@mail.polimi.it
angelica.iacovelli@mail.polimi.it
ramazan.aitkaliyev@mail.polimi.it

Do you have any *questions*?