Campaign Analysis

April, 2021

The Campaign

Help marketing area to improve the performance of the campaign to get profit

The new campaign, sixth, aims at selling a

new gadget to the Customer Database

The sample results

The sample campaign had a high investment and financial losses due to low conversion

Total cost: 6.720MU

Revenue :3.674MU

Profit: -3.046MU

The success rate of the campaign was 15%

Objective

The study objective is to improve target customer using a predict model

The objective of the study is to build a

predictive model that will produce the

highest profit for the next direct

marketing campaign

Customer understanding

1- Data Analysis
2- Customer segmentation
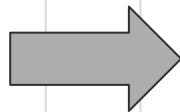3- Predictive model

Customer understanding

1- Data Analysis

Step 1: Data Analysis

Understanding and preparing database

The first step was to check database:

- check the information

- check for null values

- check for outliers

- set cleaned database to work

After database check :

- Null values and outliers for *Income* variable

- Drop outliers

- Replace null values for average values

- Database with 2.229 observations

ifood

Step 1: Data Analysis

Histogram analysis to create hypothesis

With histogram analysis it was possible to create some hypothesis to check:

H1: Customer age can evaluate the campaign acceptance. Older customers accept more campaigns.

H2: Marital status can evaluate the campaign acceptance. Married customers accept more campaigns.

H3: Customers with better recency are most likely to accept a campaign.

Step 1: New variables

New variables were created to help the analysis

- Age: customer age based on the customer year of birth

- Acceptance score: 0-6 to say how many offers teh customer accepted being 0 for customer who didn't accept any campaign and 6 to customers who accepted all the six campaigns.

- Lifetime: time (in years) since customer enrollment with the company.

- TPV: customer total paid value (total spent in products)

Customer understanding

2- Customer segmentation

Step 1: Hypothesis check

H1: Customer age can evaluate the campaign acceptance. Older customers accept more campaigns.

Customers with 41 years old or more are the ones who accepted at least one campaign.

Customers under 30 didn't accept 5 campaigns while customers between 61 and 70 years old are most likely to accept 5 campaigns.

H2: Marital status can evaluate the campaign acceptance. Married customers accept more campaigns.

Married customers are most likely to accept more campaigns, but the average TPV for the past 2 years loses for customer with other marital status.

Customers with marital status *Together* and *Single* are good targets: once they have almost the same average TPV and also have a good acceptance score for the campaigns.

H3: Customers with better recency are most likely to accept a campaign

Customers with < 30 days since last purchase are most likely to accept campaigns. Customers with recency above 90 days are not good targets to accept campaigns.

Step 2: Customer segmentation

K-Means algorithm was used for customer clusterization
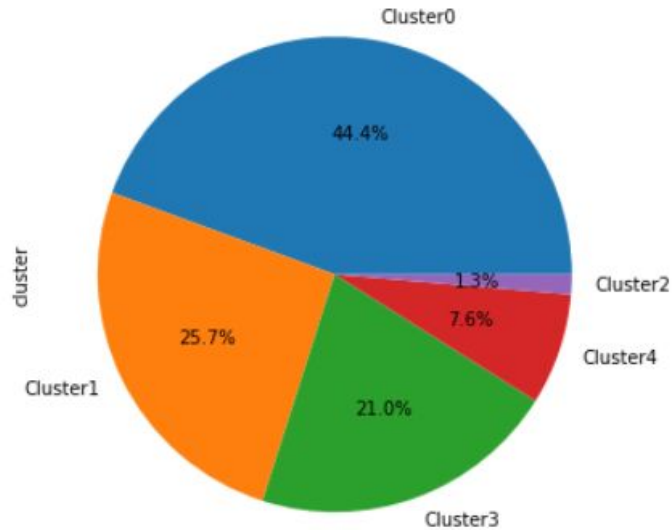
- After preparing data for clustering, the K-Means model returned 5 different clusters based on customer behavior.4
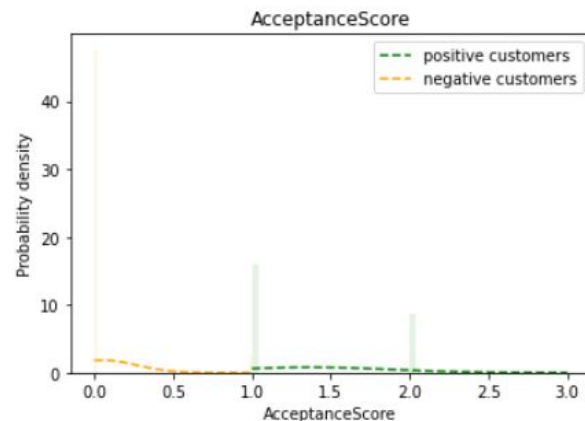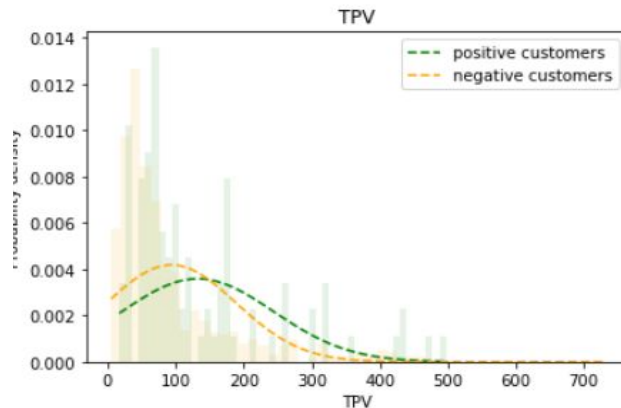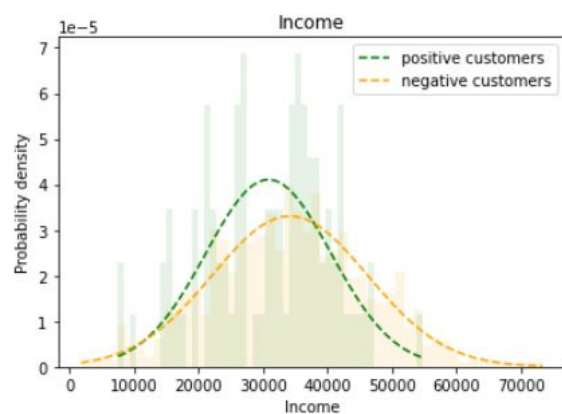- 90% of the customers are in cluster 0,1 and 3.

Step 2: Customer segmentation

Split database into positive and negative customers

Positive targets (customers who bought the new gadget in the last campaign)

Negative targets (customers who did not buy the new gadget in the last campaign).

Step 2: Cluster 0

- Income between 10k and 60k
- 30 and 50 years old, single or divorced who spend more in Gold products and buy products online.
- The average TPV for the past two years is between 100 and 150
- Recency of 30 days on average.
- Accepted at least one campaign

Step 2: Cluster 1
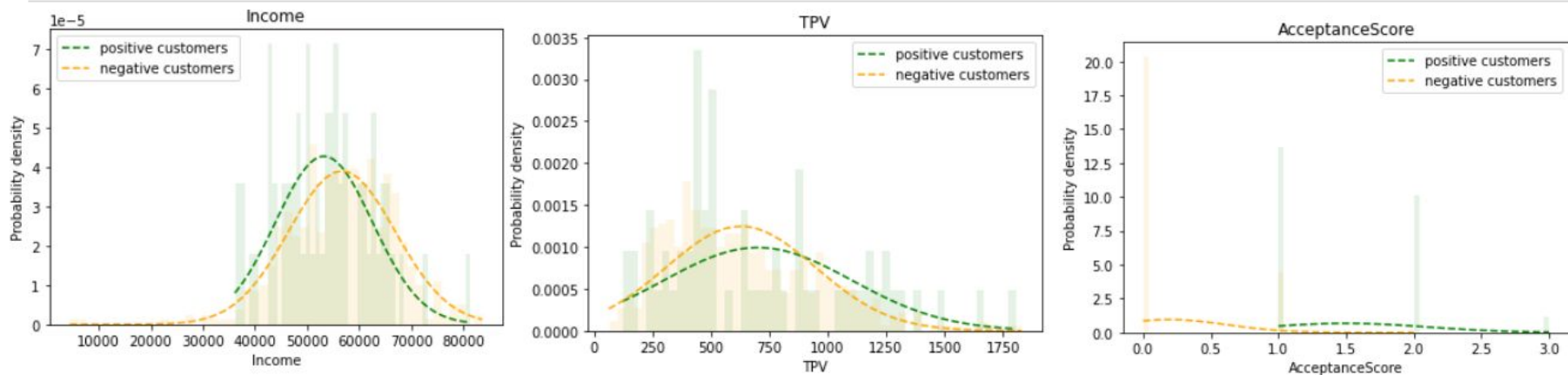
- Age above 51 years old, most of them married or together who spend more money in wine.
- They spent 750 on average in the past two years and the they usually come back to another purchase within 30 days.

Step 2: Cluster 3
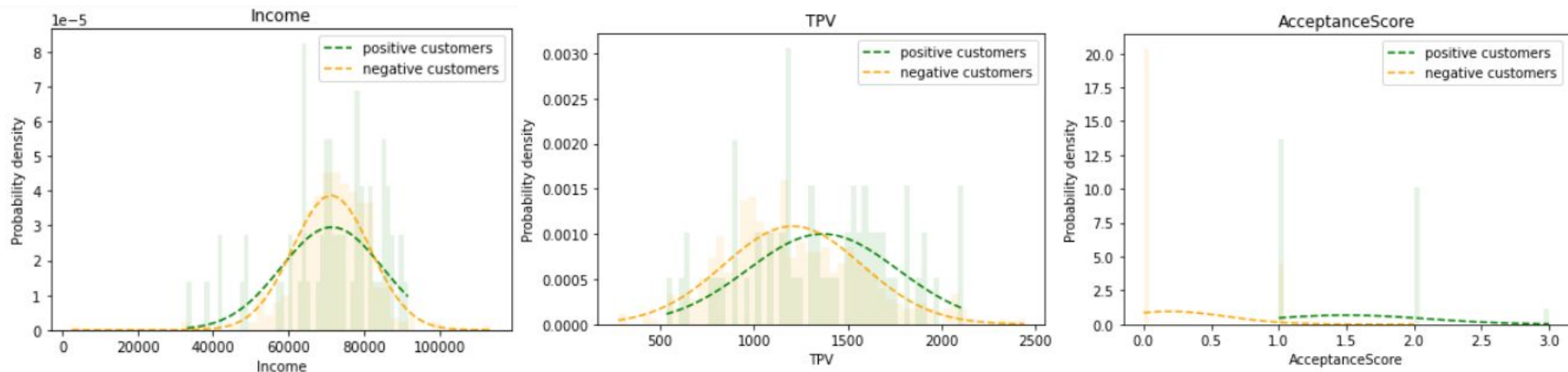
- Age between 51 and 60 years old most of them are married, together, divorced.
- They spend money specially on wine, fish and fruits.
- TPV, around 1300 in the past two years and they also have accepted at least one previous campaign.
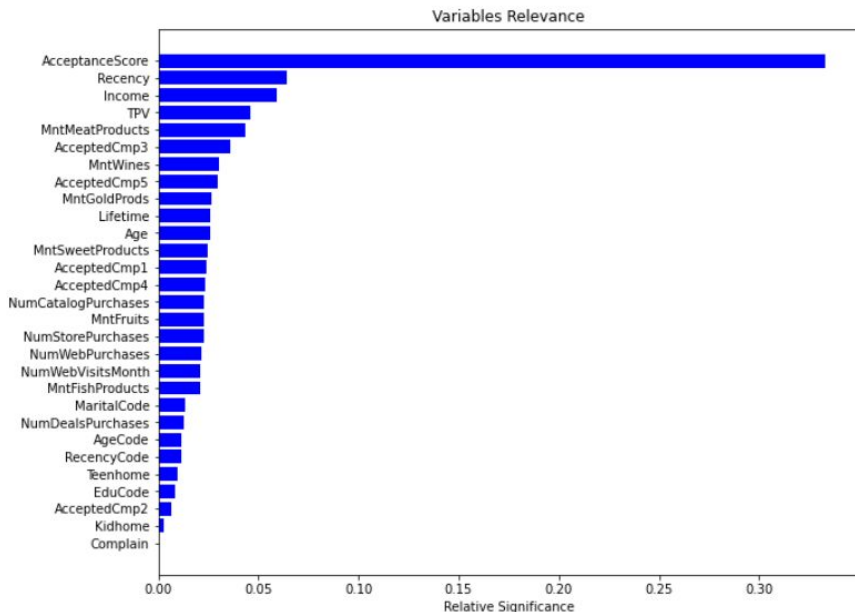
Customer understanding

3- Predictive model

Step 3: Predictive Model

Step 3: Predictive Model

Identify variables relevance by Random Forest model

Variables Relevance

Random forest importance function were used to determine variables relevance.

Based on the results and previous analysis the selected variables to the models are

1. Income
2. Recency
3. TPV
4. MaritalCode
5. AgeCode
6. MntWines
7. MntFruits
8. MntMeatProducts
9. MntFishProducts
10. MntSweetProducts
11. MntGoldProds

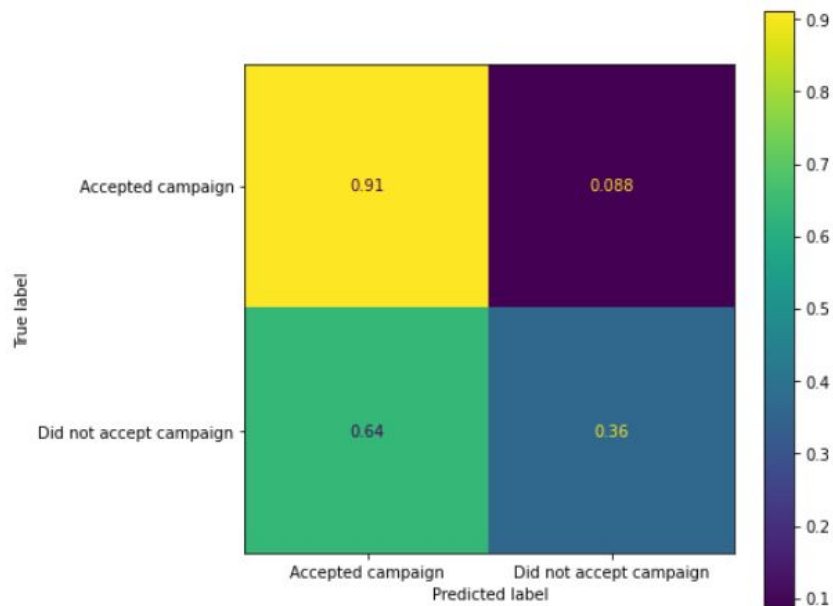ifood

Step 3: Predictive Model

Choose Random forest as the best model

Models Tested	Models Accuracy
Logistic Regression	0.72
K-Nearest neighbors (KNN)	0.74
Decision Tree	0.80
Random Forest	0.82

ifood

Step 3: Random Forest model

Confusion matrix shows a high chance to indicate False Positive customers

The confusion matrix is useful for measuring the type of model erros:

True Positive (TP): Model predict that that I would accept the campaign and actually did.

True Negative (TN): Model predict that that I would not accept the campaign and actually did not.

False Positive (FP): Model predict that that I would accept the campaign and in fact I did not (Type 1 Error).

False Negative (FN): Model predict that that I would not accept the campaign and in fact I did (Type 2 Error).

Conclusion

Random Forest model was chosen to help marketing area to find customer who will accept the campaigns whit 82% of accuracy.

That means that the model will provide a customer who will accept the campaign with 82% of accuracy and the same racional is valid for customer who will not accept.

If this model could help in the same scenario it would tell us previously that 82% of the customers who didn't accept the campaign would do that so the company would not spend money with these customers:

-        if we knew that 82% of 1.895 customers would not accept the campaign we would send the campaign only to 342 customers plus the 334 customer who accepted.

-        we could save 1.026MU (342 x 3) and the total amount spent would be 2.028(676 x 3).

So in that scenario, the company would spent 2.028MU and the revenue would be the same 3.674MU and the company would make a profit of 1.646MU.

We can also consider that this model could help with the conversion model increasing the revenue.

Thanks!

Luisa Raniro
luisarasera@gmail.com