Data Driven Contact Strategy

Machine Learning applied to Marketing

LUISS x Deloitte.

Fabiana Caccavale
Marco Amadori
Lorenzo Meloncelli

INDEX

LUISS

GOALS

To estimate the likelihood of positive responses to a specific campaign (cross-selling/solution).

To analyze customers' contactability.

To monthly assign each customer to a campaign and a contact channel.

LUISS

Project overview

START

DATASET

1 EXPLORATORY DATA ANALYSIS

related to

"SOLUTIONS"

"COMMODITY_DUAL"

2 PROPENSITY MODEL

GOAL: to identify features that

COMMODITY_DUAL = 0
&
COMMODITY_DUAL = 1

SOLUTIONS = 0
&
SOLUTIONS = 1

have in common

HOW?

1 MEAN ANALYSIS
2 RANDOM FOREST

Once the important features are identified, they are used to train and test 3 machine learning algorithms (whose outputs were converted to probability values) and a logistic regression in order to predict the probability of a customer's positive response to a specific campaign.

The model with the best performance is chosen to predict the propensity

3 ELIGIBILITY MODEL

- Risk churn
- Leaving customers
- Privacy consensus = NO

Creation of 6 datasets according to the **GENERAL RULE** (those customers whose last contact dates back to at least N months ago are eligible) to which campaign and cross-campaigns rules are applied:

1 Cross_Selling_DEM.csv
2 Cross_Selling_SMS.csv
3 Cross_Selling_TLS.csv
4 Solution_DEM.csv
5 Solution_SMS.csv
6 Solution_TLS.csv

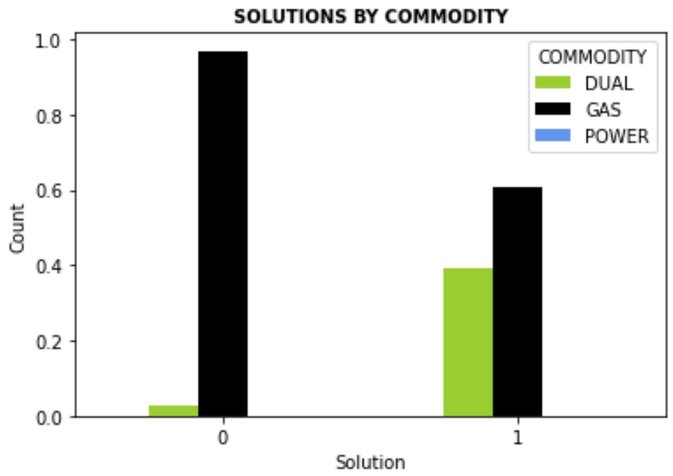
- no email validated
- no phone validated
- no phone validated

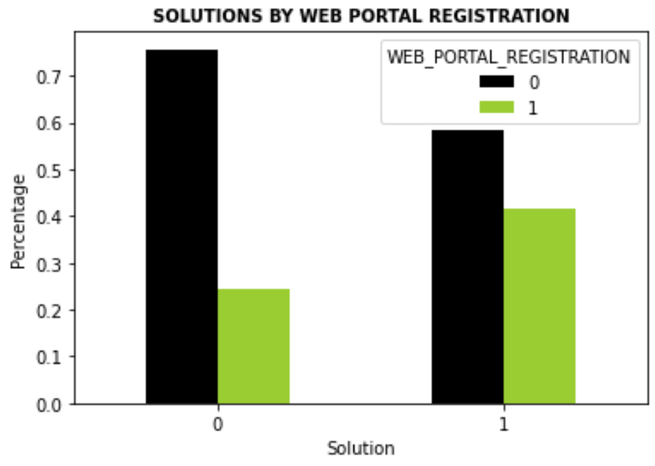
CONTACT STRATEGY

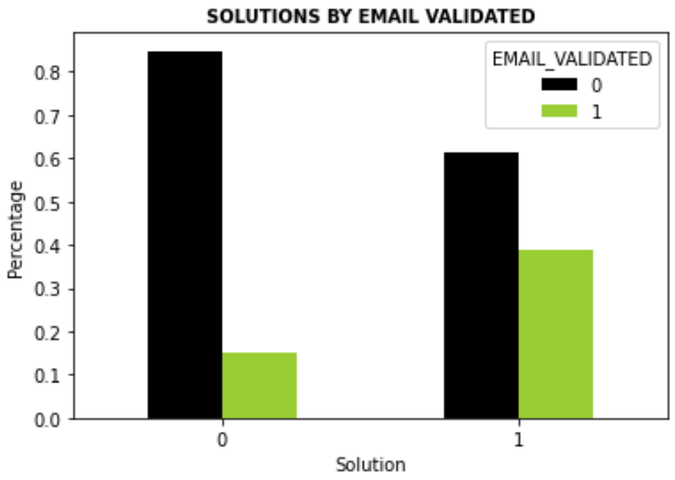
- budget assumptions

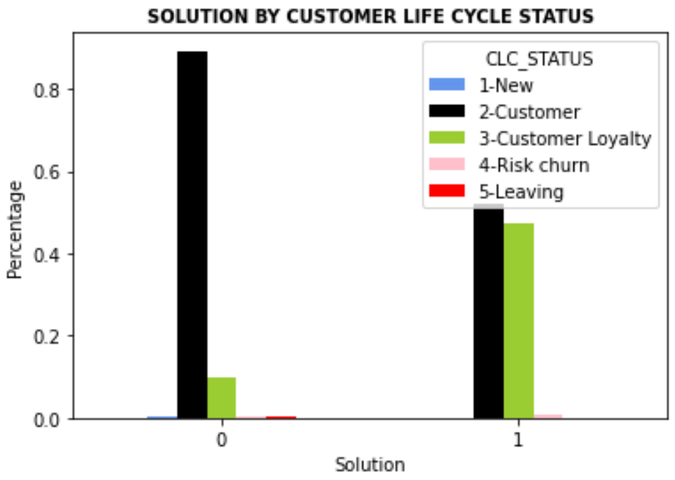
LUISS

Data comprehension (SOLUTIONS)



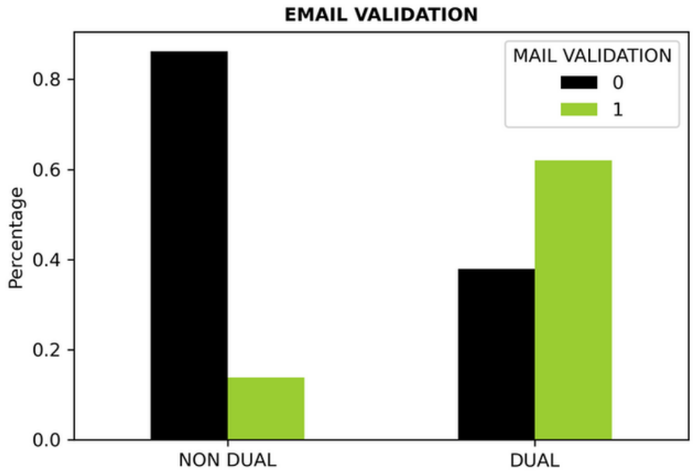


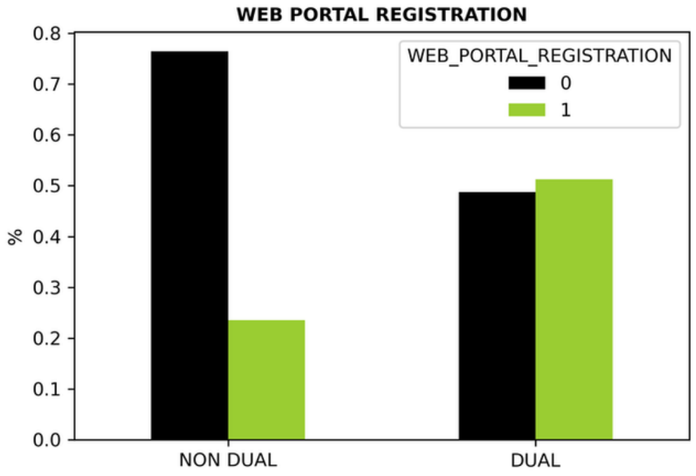


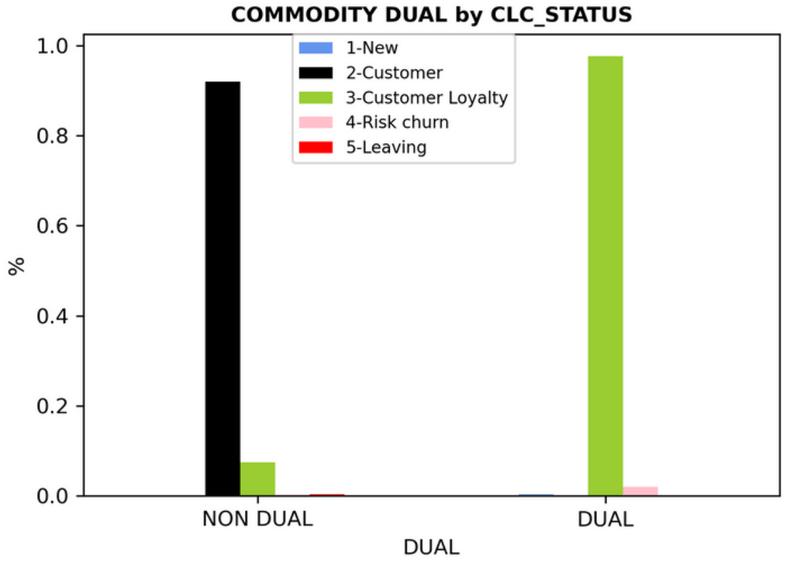


LUISS

Data comprehension (COMMODITY_DUAL)

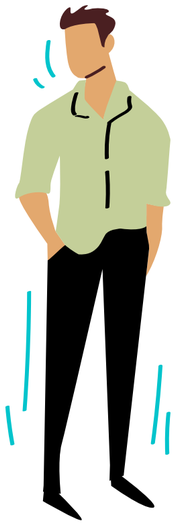






LUISS

The typical customer:



"COMMODITY_DUAL" = 1

- 60% are registered to the web portal
- Loyal customers
- 60% validate their mail
- Area: NORTH

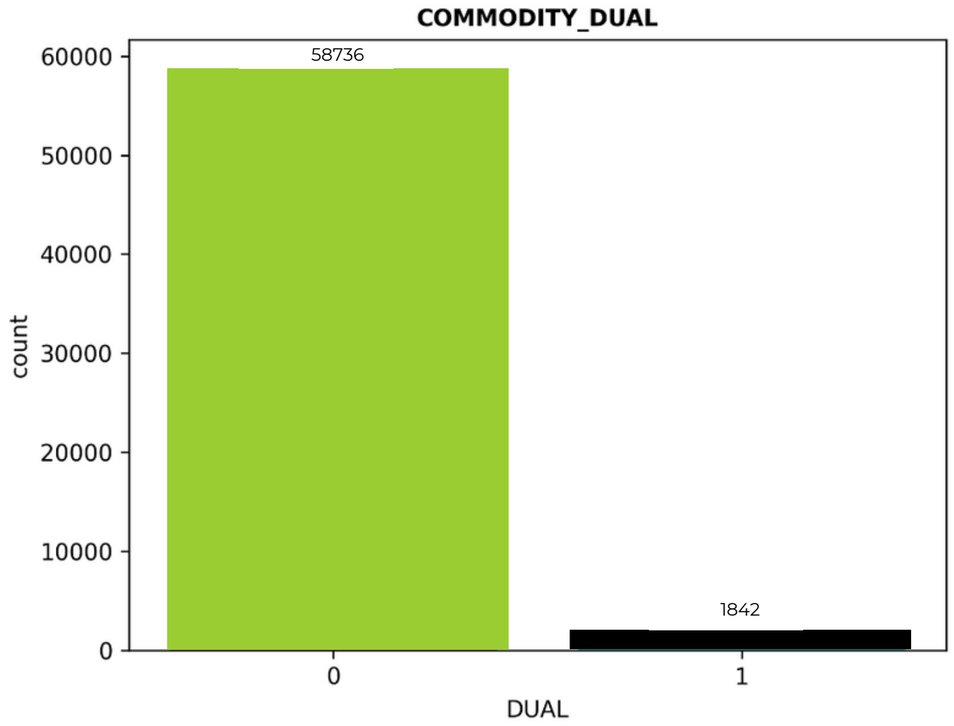
"SOLUTIONS" = 1

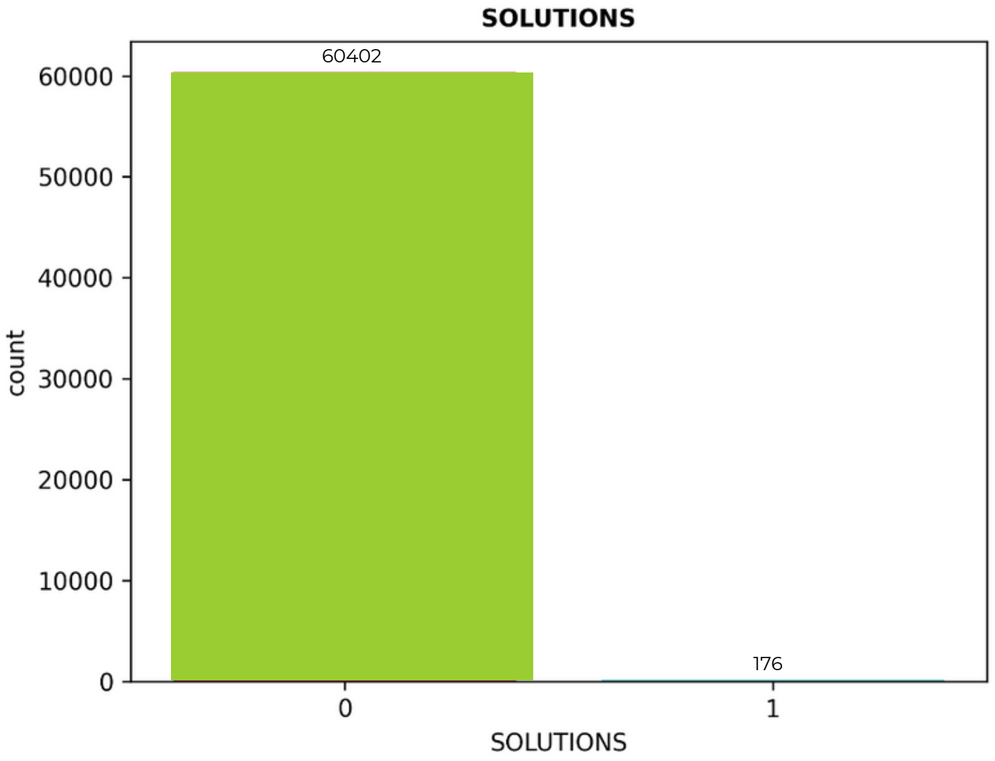
- 40% are registered to the web portal
- Loyal customers
- 40% validate their mail
- Dual or gas contract
- More frequent inbound contacts
- Registered to the web portal

LUISS

Data comprehension:

Dataset imbalance





LUISS

Data preparation

Main issues:

Missing values

Outliers

Dataset imbalance

- Missing values were removed.
- Outliers were replaced with the appropriate median.
- To overcome the problem of dataset imbalance, we randomly selected observations from the majority class and deleted them from the training dataset.

LUISS

Propensity model

GOAL: to develop a model that would have helped to select which customers are most likely to purchase a particular contract.

To do that, features that could have most influenced consumers' choice were identified by comparing customers who had already purchased the contract to those who had not.
This was done through:

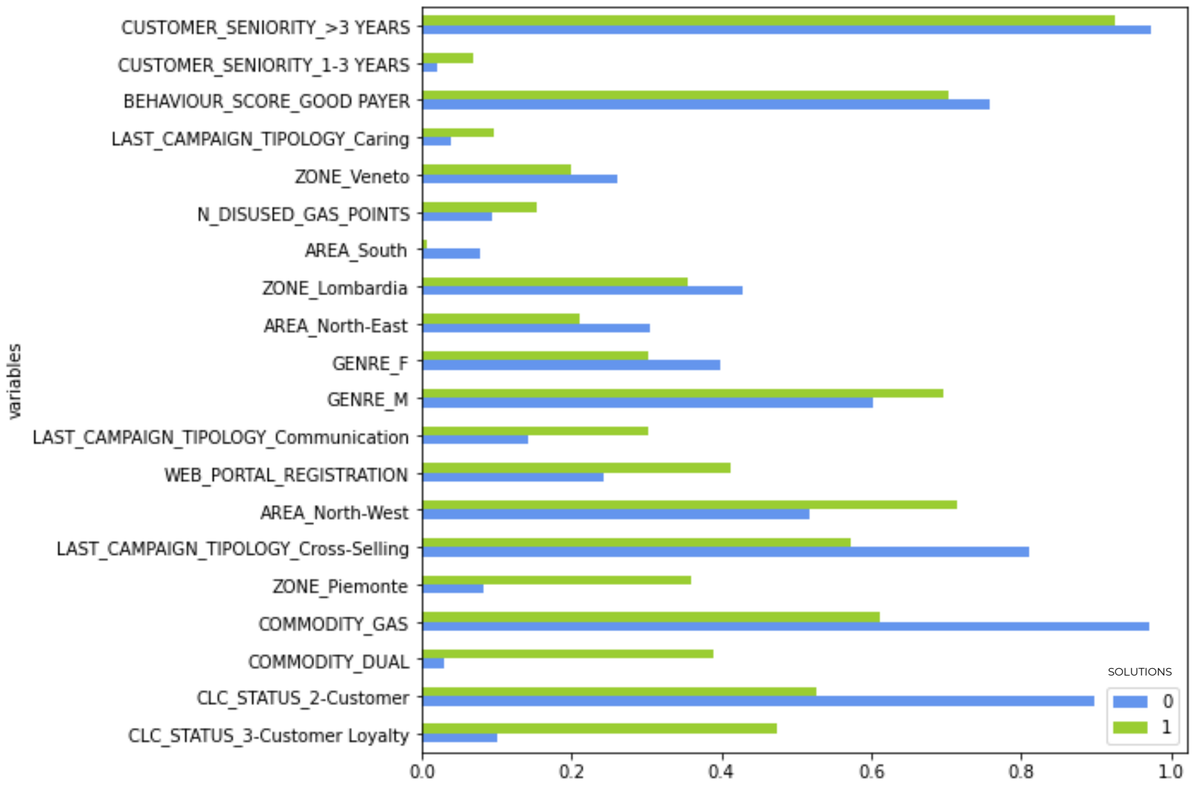
- Mean analysis
- Random Forest feature importance

LUISS

Propensity model

Features selection: mean analysis

"SOLUTIONS"

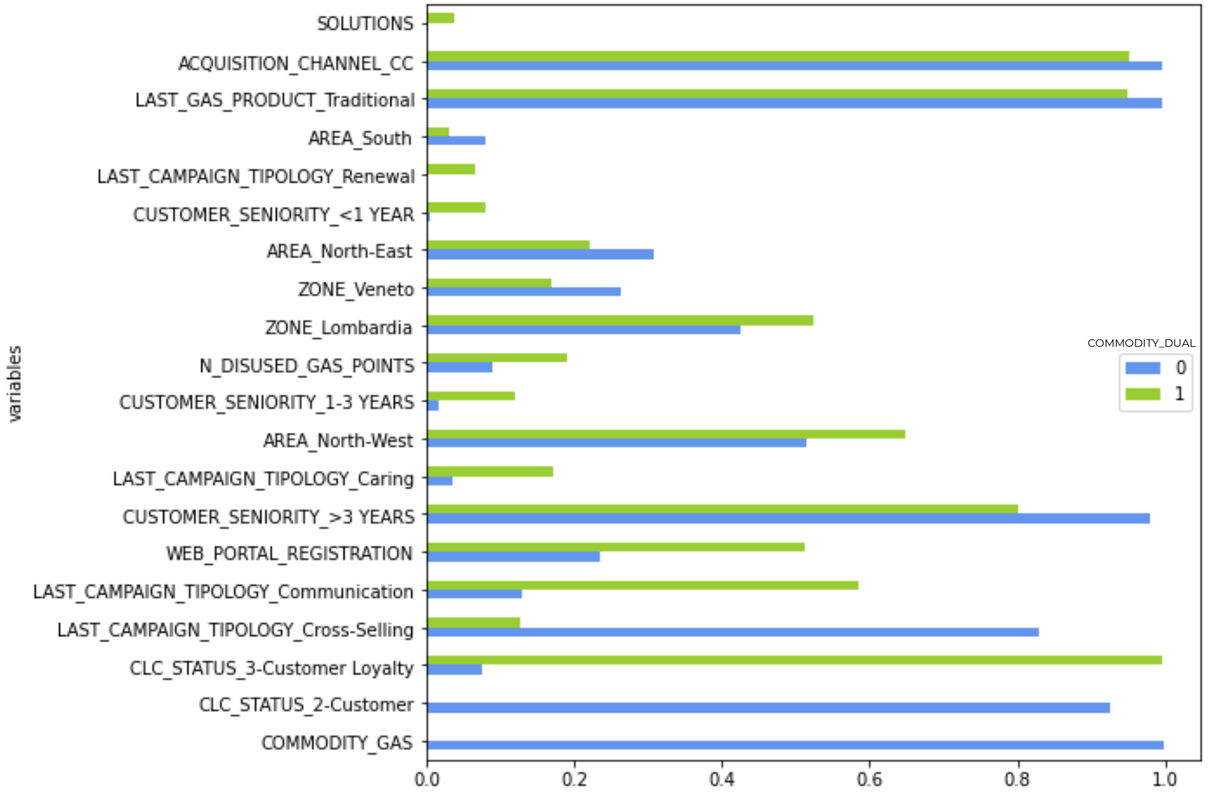


LUISS

Propensity model

Features selection: mean analysis

"COMMODITY_DUAL"

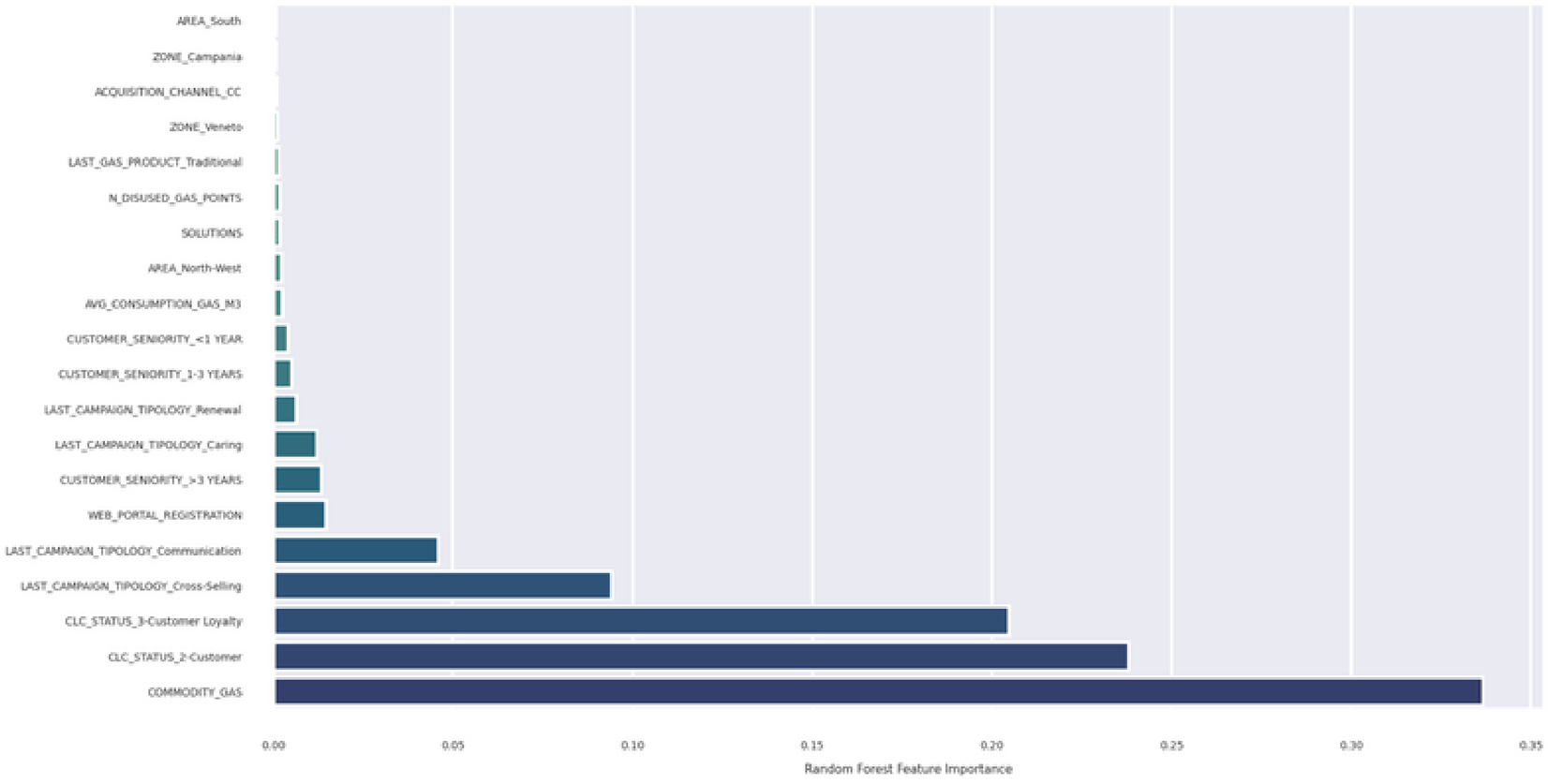


LUISS

Propensity model

Features selection: Random Forest features importance

y: COMMODITY_DUAL

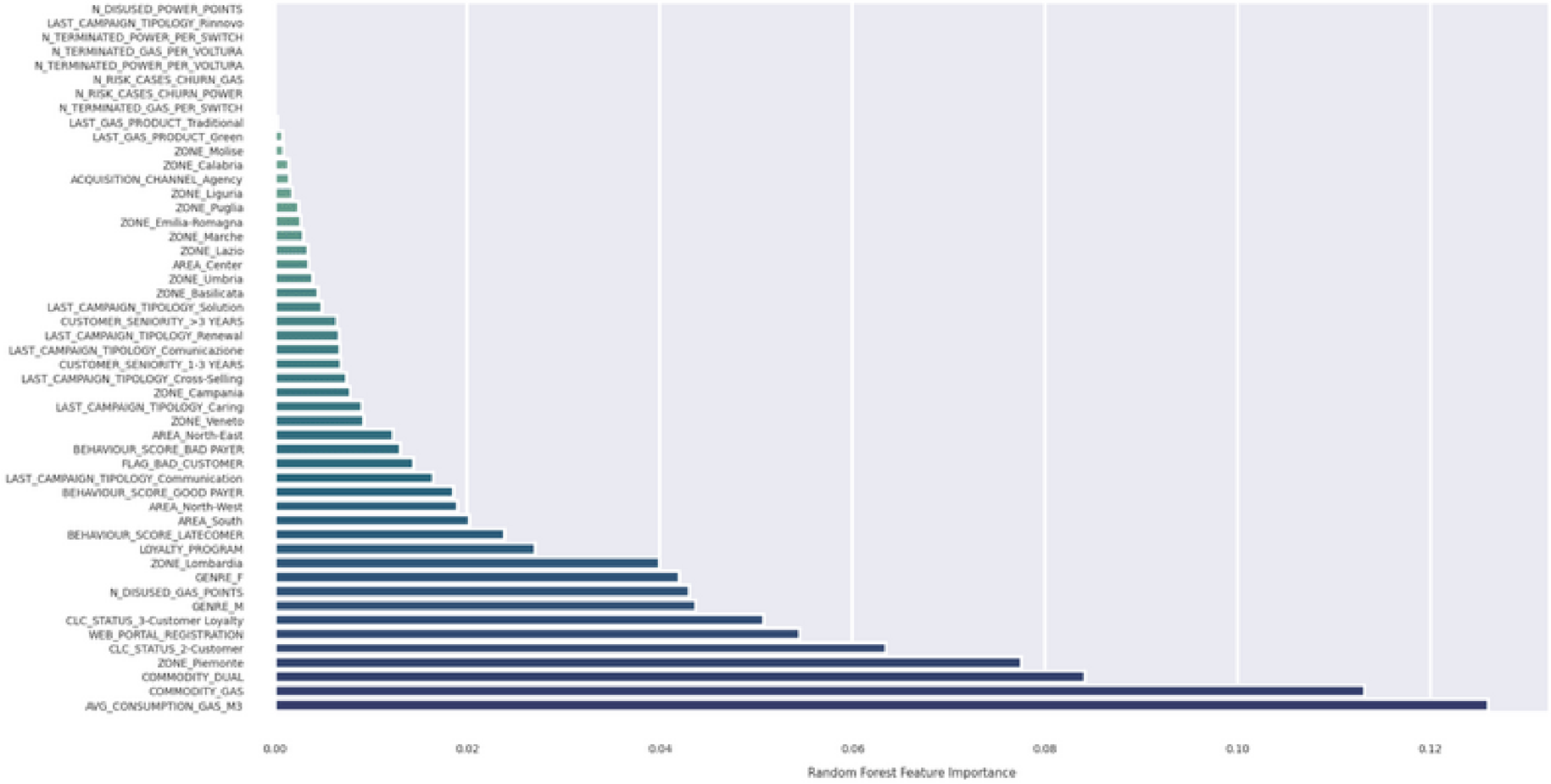


LUISS

Propensity model

Features selection: Random Forest features importance

y: SOLUTIONS



LUISS

After removing the variables that lacked an economic meaning together with the strongly correlated ones, the resulting selected features were the following:

SOLUTIONS

['AVG_CONSUMPTION_GAS_M3', 'COMMODITY_DUAL', 'ZONE_Piemonte', 'WEB_PORTAL_REGISTRATION', 'AREA_North-West', 'CLC_STATUS_3-Customer Loyalty', 'BEHAVIOUR_SCORE_GOOD PAYER', 'LOYALTY_PROGRAM', 'AREA_SOUTH', 'ZONE_VENETO', 'LAST_CAMPAIGN_TIPOLOGY_Caring', 'AREA_North-East', 'LAST_CAMPAIGN_TIPOLOGY_Cross-Selling','CUSTOMER_SENIORITY_>3 YEARS', 'CUSTOMER_SENIORITY_<1 YEAR', 'ACQUISITION_CHANNEL_CC', 'BEHAVIOUR_SCORE_BAD PAYER', 'AREA_CENTER']

COMMODITY_DUAL

['CLC_STATUS_3-Customer Loyalty', 'AREA_North-West', 'WEB_PORTAL_REGISTRATION', 'LAST_CAMPAIGN_TIPOLOGY_Cross-Selling', 'N_DISUSED_GAS_POINTS', 'LAST_CAMPAIGN_TIPOLOGY_Caring', 'SOLUTIONS', 'CUSTOMER_SENIORITY_>3 YEARS', 'LAST_CAMPAIGN_TIPOLOGY_Renewal', 'CUSTOMER_SENIORITY_1-3 YEARS', 'AVG_CONSUMPTION_GAS_M3', 'LAST_GAS_PRODUCT_Traditional', "COMMODITY_DUAL"]

LUISS

How is the propensity measure captured?

Y → SOLUTIONS
Y → COMMODITY_DUAL

X → important features

Used models
- K-Nearest Neighbors
- Random Forest
- Support-Vector Machine
- Logistic Regression
- Ensemble method

MODEL → trained/tested on NON ELIGIBLE customers

predicts

ELIGIBLE → 0 / 1

probability
0.3
0.7

} 1 → ELIGIBILITY MODEL

PLATT SCALING

LUISS

Models' performance metrics

COMMODITY_DUAL

	Accuracy	Precision	Recall
Random Forest	0.89	0.83	0.97
KNN	0.86	0.81	0.94
SVM	0.89	0.83	0.97
Logistic Regression	0.89	0.83	0.97
Ensemble method	**0.89**	**0.83**	**0.98**

Most important evaluation metric when the cost of false negative is high (the cost of NOT contacting someone who is likely to give a positive response to the campaign is higher than the cost of contacting a customer who is not willing to sign the contract.

LUISS

Models' performance metrics

SOLUTIONS

	Accuracy	Precision	Recall
Random Forest	0.68	0.72	0.57
KNN	0.67	0.68	0.62
SVM	0.68	0.67	0.69
Logistic Regression	0.65	0.65	0.63
Ensemble method	**0.7**	**0.68**	**0.71**

LUISS

Eligibility model

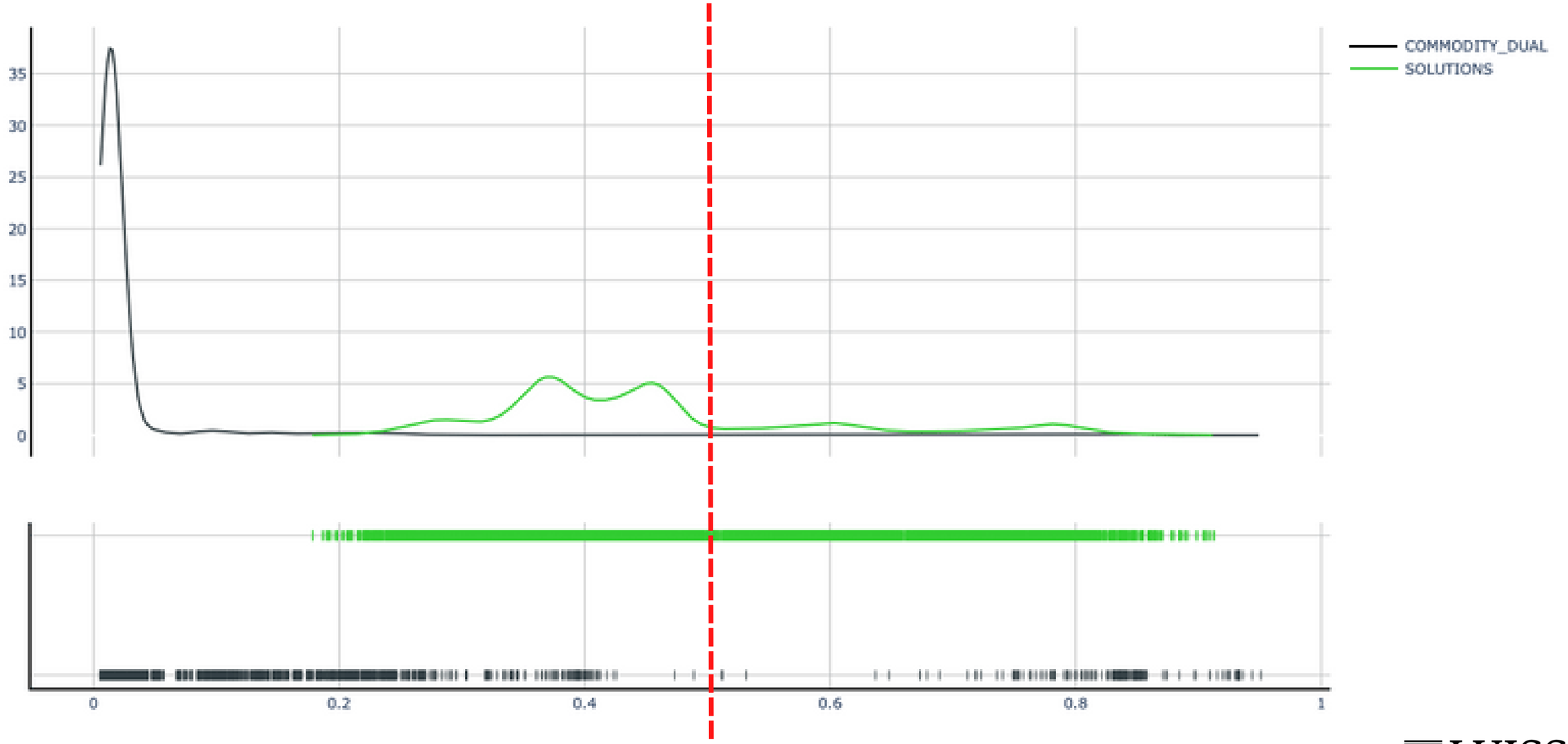
1. Customers who are at **risk of being churned**, who are **leaving** and who did not give the **privacy consensus** will not be contacted and, thus, they are deleted from the dataset.

2. Six datasets were created according to the **general rule** (*those customers whose last contact dates back to at least N months ago are eligible*) and, according to the type of channel, those customers who did not validate their phone number or their e-mail are deleted.

3. **Campaign** and **cross-campaigns rules** are applied to design the monthly contact strategy.

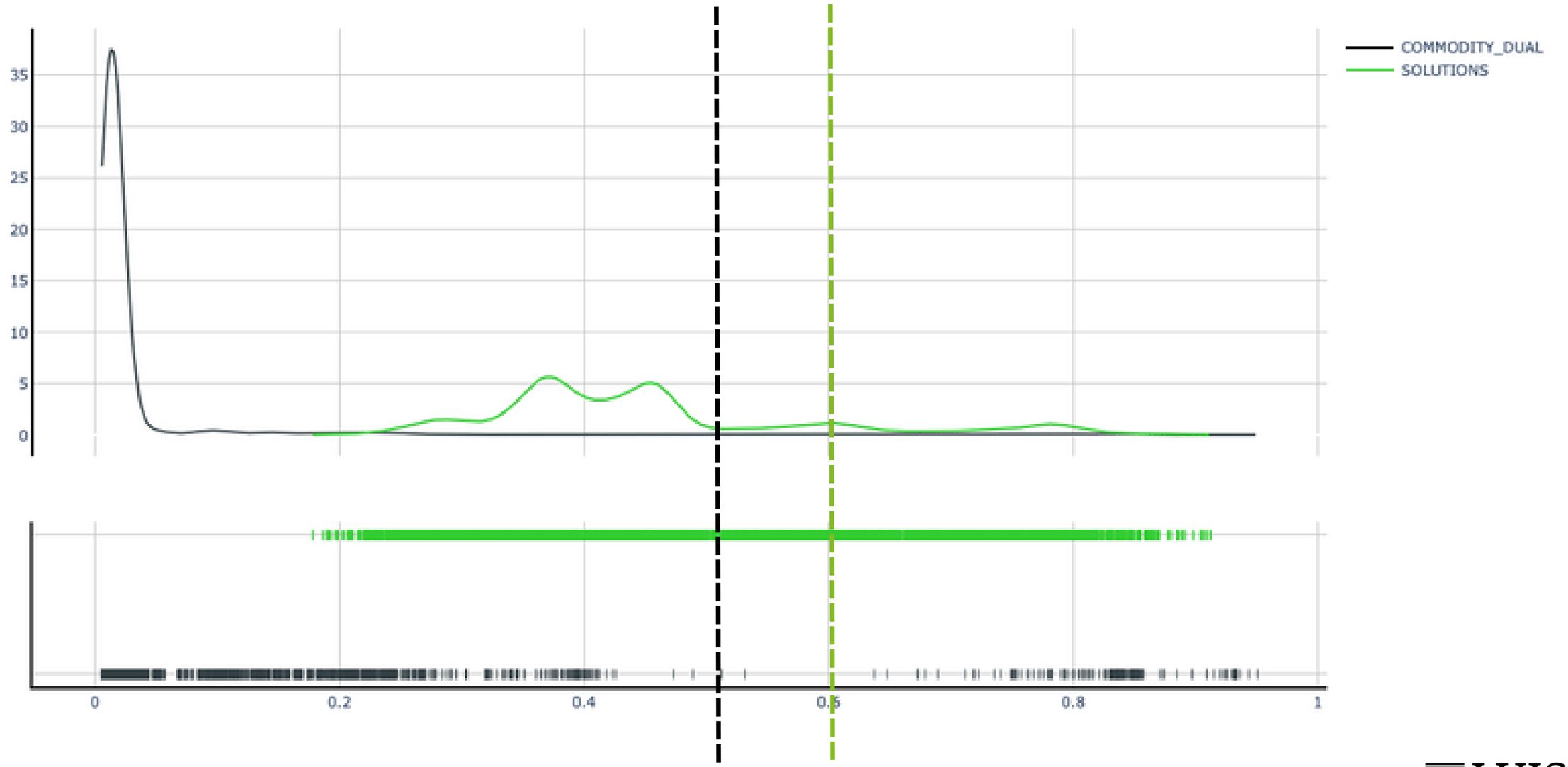
LUISS

MODEL FLEXIBILITY



LUISS

MODEL FLEXIBILITY



LUISS

Model Flexibility

No budget

- Efficiency Oriented
- Only the most propense are selected

≈ 5000€

Defined Budget

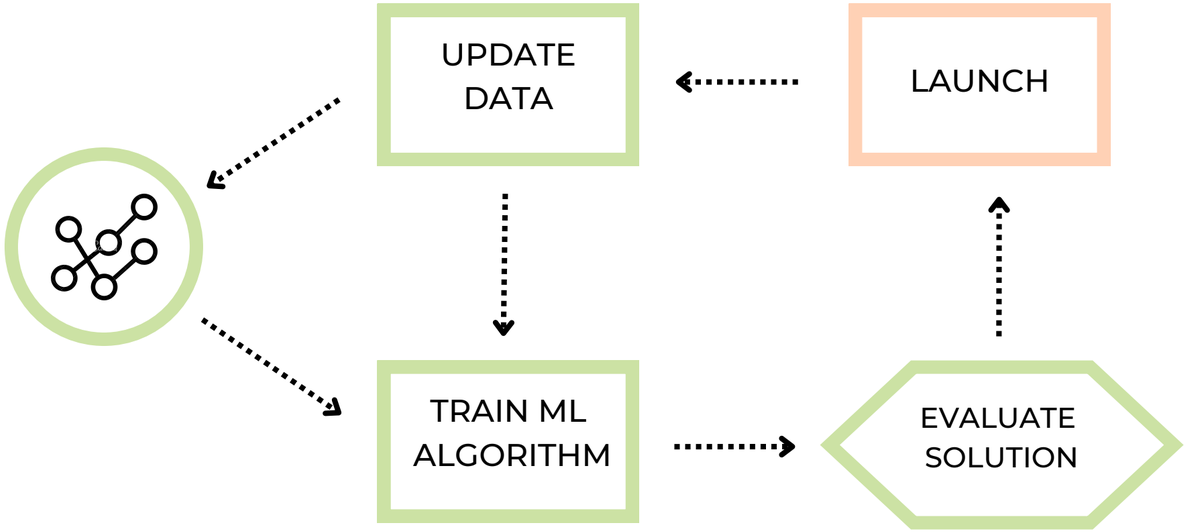
- Allows more variability
- Also customers with a lower probability may be contacted

LUISS

Conclusions and future developments

We have been able to **optimize** the marketing strategy of the company **prioritizing** the contact of a client based on their **propensity to sign the contract**

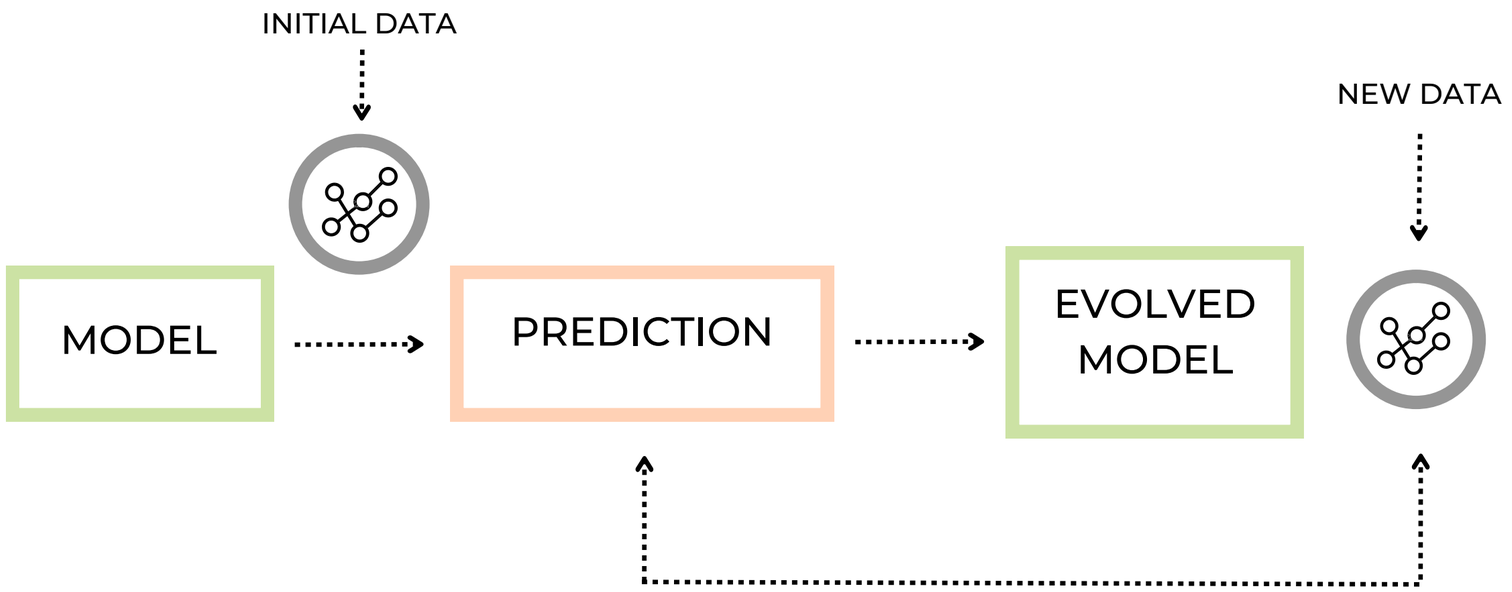
The **resources and time saved** may give the company a **competitive advantage** with respect to the competitors that still do not follow this **data-driven approach**.



LUISS

Conclusions and future developments

ONLINE LEARNING



LUISS

Our Team

Fabiana Caccavale

Lorenzo Meloncelli

Marco Amadori

Thank you for the attention!