Campaign Uplift:
empik recruitment assesment

Maciej Kasztelanic | kasztelanicmaciej@gmail.com

Project Goal

The project aims to enhance targeting efficiency of marketing campaign by using advanced analytics to identify customers with the highest potential gain from being contacted.

Problem

1

Campaign Effectiveness

Current outreach campaign has a low conversion rate

2

Targeting Strategy

Targeting people at random is ineffective

3

Customer Segmentation

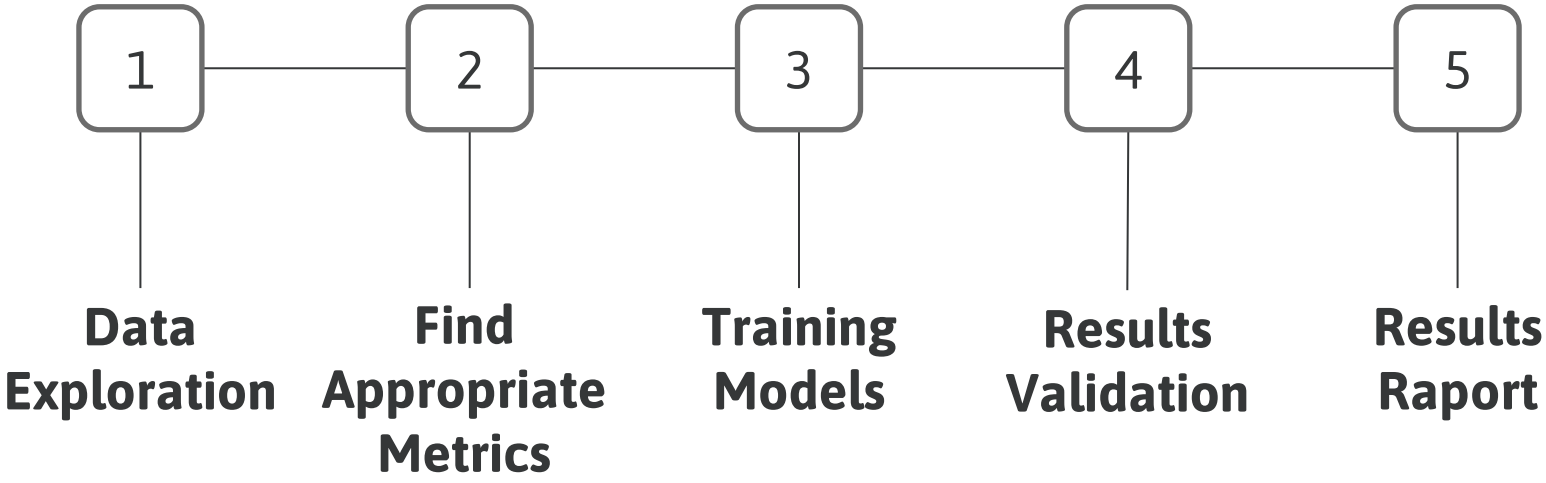
Calling people, without knowing how well they suit to the product

4

Resource Allocation

Spending budget on individuals who are not interested in the product

Uplift modelling: **Problem approach**



*all steps are presented in attached codes

Data Exploration

There are over **41.000** observations, where **12.5%** of people subscribed

Nearly **60%** of people were not in campaign group

The dataset column are of three categories

Demographics

- Age
- Job
- Marital
- Education
- Loan data

Previous Contact

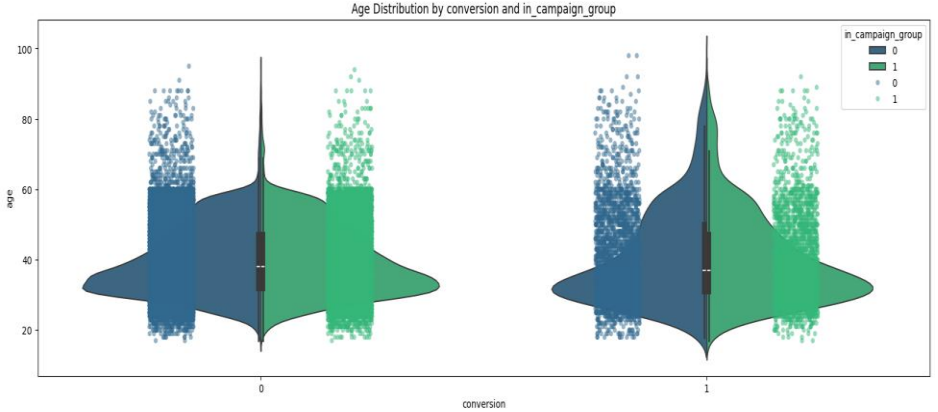
- Contact device
- Month and Day
- Call Duration
- Campaign

Economic indicators

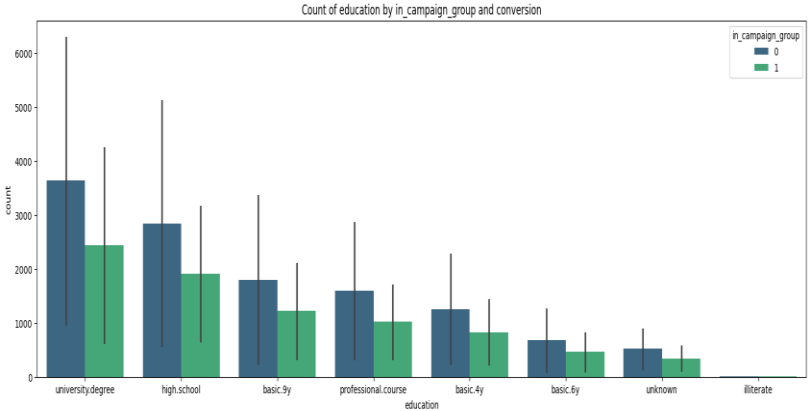
- CPI and CCI
- EURIBOR 3M
- Number of employees
- Employment variation rate

Data Exploration: **Demographics**

The majority is **under 40 y.o** | higher proportion of **older people** among the subscribers



The majority has at least a high school diploma

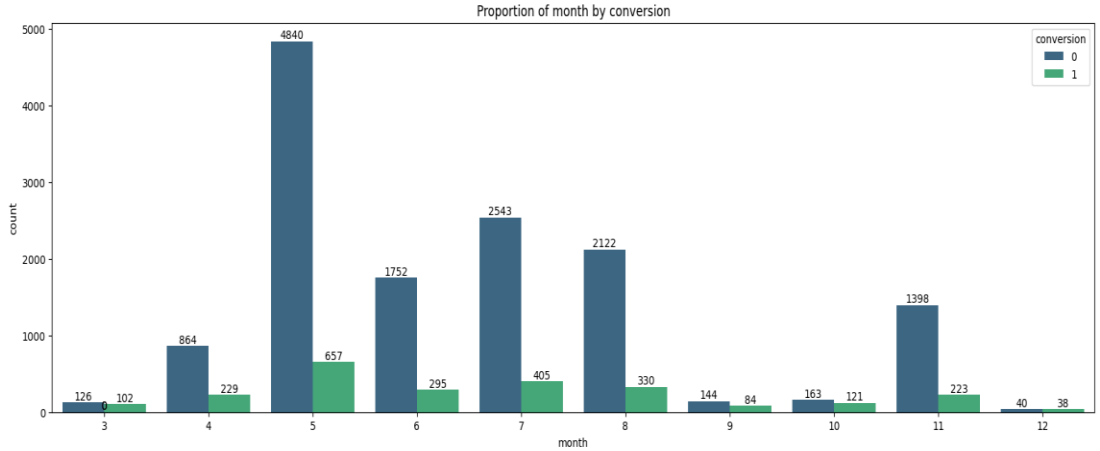


Key takeaways:

- Young people (under 25 y.o) and Elderly people (over 60 y.o) have the highest conversion rates
- Students or Retired people were most likely to subscribe
- People with any type of loans are less likely to subscribe to term deposit

*all graphs and analytics are presented in attached codes

Data Exploration:
Previous Contact



Key takeaways:

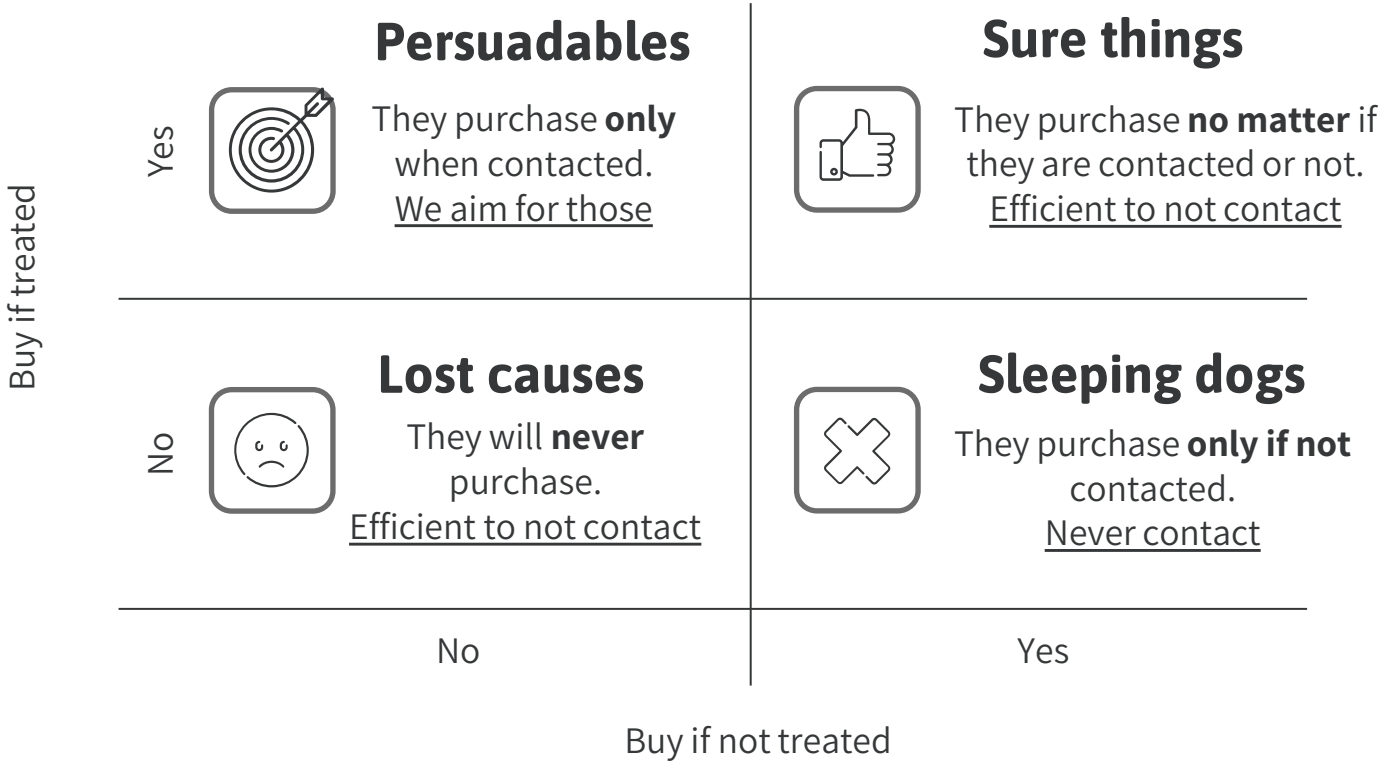
- Highest number of contact was in May, while it is the month with the lowest conversion rate (**11.9%**)
- Contact day of the week did not influence the conversion
- The best performing months were March, September and December

*all graphs and analytics are presented in attached codes

Uplift modelling - term definition

Uplift is a measurable increase in key performance indicators, such as sales or customer engagement, resulting from a marketing campaign or intervention. It quantifies the positive impact of marketing efforts by comparing outcomes with a baseline control group

Uplift modelling: **Client types**



Uplift modelling: **Evaluation Metrics**

	Description
Uplift at k	Evaluates directly the model's effectiveness in identifying the most responsive customers. Measures incremental response rate among top k customers
AUUC score	Evaluates the overall performance of the uplift model by measuring the area under uplift curve. Provides a view of the model's ability to differentiate between customers who will respond positively and those who won't
QINI score	Evaluates the model's ability to generate uplift. It is a cumulative gain from targeting customers based on model predictions versus a random baseline

*for more insights on metrics see codes

Uplift modelling: **Used Models**

S-Learner

- Meta-Learner
- Single model
- Estimates both treatment and control effects

T-Learner

- Meta-Learner
- Two separate models for treatment and control group
- Each model estimates only results for the group it was trained on

Uplift modelling: **Model Training & Validation**

To ensure comperability and validity of results, each model was set up and evaluated using following pipeline:

1. **Uplift models** S-Learner, T-Learner
2. **Classifying models** Logit, XGBoost, CatBoost, SVM
3. **Cross validation** Cross validate all models (50 folds)
4. **Metrics an Plotting** Evaluate models performance (AUUC, QINI, Uplift@k)

*pipeline and its functions are in codes

Uplift modelling: **Best Model**

SVM with radial kernel

AUUC Score: 0.29

We expect a moderate cumulative uplift across the entire popualtion

QINI Score: 0.423

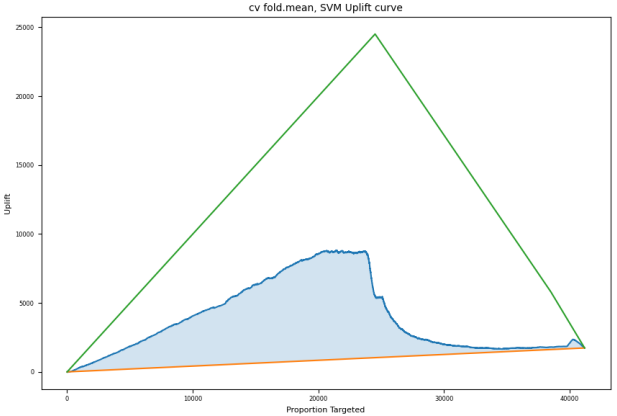
We expect strong cumulative uplift in the top segments of the population

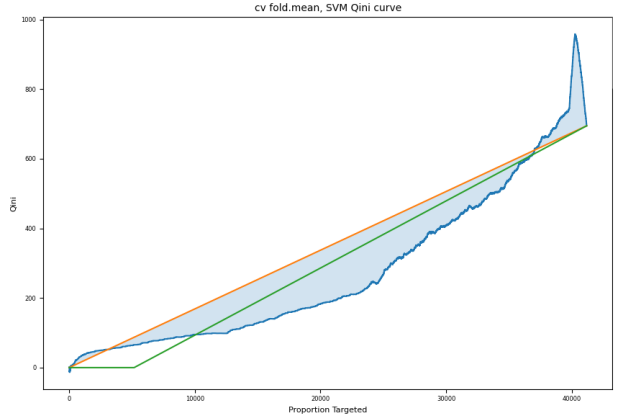
Uplift at 30%: by group – **0.118**, overall – **0.423**

- By group: We expect an increase in subscription rate for the contacted individuals compared to the control group within top 30%;
- overall: We expect a substantial increase in conversion rates between the contacted and control groups

*more in depth metrics analysis in codes

Uplift modelling: **Uplift & QINI curves**





AUUC/QINI curve

Perfect Curve

Random

SVM model was the only model which AUUC and QINI curves were shaped somewhat similar to the Perfect curves.

- The model reflects the diminishing returns with targeting more of the population
- The shape of QINI curve indicates that model performs well when identifying individuals with highest uplift

Future campaign improvement proposition

Data collection

Gather additional data points of information like: account balance, occupation, gender

Contact and Marketing

Focus the campaign during months with higher effective subscription rate

Model improvements

Further investigate model performance and its predictive capabilities

Thanks!

All codes can be viewed at:
https://github.com/kasztelmacko/uplift_ml/tree/master

Maciej Kasztelanic
kasztelanicmaciej@gmail.com | +48 531 260 986