

Predicting Telemarketing Success Rate for Banking Product Subscriptions

Predictive Modelling (2024 P3A)

Koen van Esterik
kd.vanesterik@student.han.nl

7th of July 2025

0. Abstract

1. Introduction

Voor elk commercieel bedrijf bestaat de behoefte om winst te maken. Dit om op basis daarvan investeringen te doen, groei te realiseren en uiteindelijk bestaan-szekerheid te garanderen. Naast deze behoefte bestaat ook een verplichting richting aandeelhouders. Deze verplichting bestaat uit het creëren van waarde om daarmee de aandelen van de aandeelhouders te vergroten.

Dit geldt ook voor investeringsmaatschappijen, die als voornaamste activiteit, investeren in andere bedrijven om daarbij waarde te creëren en winst te genereren. Het succes van een investeringsmaatschappij wordt bepaald door het vermogen om de juiste investeringsbeslissingen te nemen en daarmee de waarde van aandelen van aandeelhouders te maximaliseren.

Met dit als context richten we onze aandacht op de investeringsmaatschappij Blackrock, die een investering overweegt - om een Portugese bank over te nemen. Deze Portugese bank voert, naast reguliere financiële activiteiten, telemarketing campagnes uit om financiële producten te verkopen.

De meest robuuste telemarketing operatie bestaat uit campagnes om abone-menten voor bank depositos te werven. Ze gebruiken hiervoor gegevens van bestaande klanten om naar uit te bellen. Dit met wisselend succes, omdat het moeilijk te voorspellen is - hoe een klant gaat reageren op een mogelijk storend telefoongesprek. Veel van deze gesprekken zijn geregistreerd in een dataset, waarin is aangegeven of de prospect een bank deposito abonnement wil afnemen.

Hier ligt voor Blackrock een mogelijkheid om dieper strategisch inzicht te verkri-jgen met betrekking tot de overname van de Portugese bank. Dit door te

onderzoeken of winst gemaximaliseerd kan worden, door het telemarketing process te optimaliseren door een prospect selectieprocedure te implementeren.

De bijgehouden dataset kan dienen als input voor het onderzoek. Dit onderzoek zal gefocust zijn op de gesuggereerde selectieprocedure. Deze procedure wordt gebaseerd op een machine learning voorspellingsmodel. Dit voorspellingsmodel kan inzicht bieden, die Blackrock kan gebruiken om mee te laten wegen in de overname van de Portugese bank.

Dit onderzoek laat zich verwoorden als:

- Reduceer willekeurig gekozen telemarketing campagne prospects,
- Door de ontwikkeling van een selectieprocedure op basis van een voorspellingsmodel,
- Die beter presteert dan het uitbellen naar alle prospects,
- Om winstmaximalisatie te realiseren

Het voorspellingsmodel is gebaseerd op binaire classificatie, want we willen voorspellen of een prospect een bank deposito abonnement zal afnemen - ja of nee. Deze informatie is aanwezig in de eerder genoemde dataset en zal als trainings- en test-data dienen.

Binaire classificatie bestaat uit een begeleide leermethode om gegevens te categoriseren in één van twee mogelijke resultaten. Dit om op basis hiervan voorspellingen te doen op nieuwe, ongeziene gegevens. Bij het evalueren van de prestaties van het binaire classificatiemodel worden de volgende termen gebruikt:

- True Positive (TP): Het model voorspelt correct een positieve uitkomst.
- False Negative (FN): Het model voorspelt incorrect een negatieve uitkomst.
- False Positive (FP): Het model voorspelt incorrect een positieve uitkomst.
- True Negative (TN): Het model voorspelt correct een negatieve uitkomst.
- Thresholds: De bepaling waar voorspellingen van het model worden ingedeeld.

Deze termen worden gebruikt om verschillende standaard metrics uit te rekenen - bijv. accuracy, precision, recall, etc. Deze metrics dienen om een classificatiemodel te evalueren. In dit onderzoek gebruiken we deze termen, om te beoordelen welke de meeste winst oplevert. Daarbij stellen we een eigen metric voor die de meeste winst berekent op basis van de eerder genoemde termen (TP, FN, FP, TN en thresholds). We noemen deze metric de **Maximum Profit** (MP) metric.

Eerst bepalen we de voorspellingen voor elke threshold op basis van de kansberekeningen van het classificatiemodel:

$$y_{pred} = \sum_{i=1}^{thresholds} \begin{cases} 1 & \text{if } y_{probs} \geq thresholds_i \\ 0 & \text{otherwise} \end{cases}$$

waarbij:

- y_{pred} , vector met voorspellingen voor alle drempelwaarden;
- y_{probs} , vector met voorspelde probabilities;
- t , vector met alle drempelwaarden;

Daarna voeren we de voorspellingen in een confusion matrix om de TP's, FN's, FP's en TN's voor alle thresholds te bepalen:

$$\vec{tps}, \vec{fns}, \vec{fps}, \vec{tns} = \sum_{i=1}^{thresholds} \text{confusion_matrix}(y_{true}, y_{pred_i})$$

waarbij:

- y_{pred} , vector met voorspellingen voor alle drempelwaarden;
- y_{true} , vector met werkelijke waarden;
- $thresholds$, vector met alle drempelwaarden;
- \vec{tps} , vector met TP's voor alle drempelwaarden;
- \vec{fns} , vector met FN's voor alle drempelwaarden;
- \vec{fps} , vector met FP's voor alle drempelwaarden;
- \vec{tns} , vector met TN's voor alle drempelwaarden;

Vervolgens calculeren de winst voor alle drempelwaarden:

$$\vec{p} = r * \vec{tps} - c * (\vec{tps} + \vec{fps})$$

waarbij:

- \vec{p} , vector met de winst voor alle drempelwaarden;
- r , scalar met de opbrengst per succesvol gesprek;
- c , scalar met de kosten per gesprek;
- \vec{tps} , vector met TP's voor alle drempelwaarden;
- \vec{fps} , vector met FP's voor alle drempelwaarden;

Tot slot stellen we de maximale winst vast:

$$p = \max(\vec{p})$$

waarbij:

- p , scalar met de maximale winst;
- \vec{p} , vector met de winst voor alle drempelwaarden;

De MP metric introduceren we, om een verbinding te leggen tussen het technische gedeelte van het onderzoek en de business case. Dit met de gedachte dat standaard metrics vaak onvoldoende inzicht bieden met betrekking tot de strategische beslissingen die Blackrock doorgaans moet maken.

De MP metric zal uitkomst bieden bij het evalueren van alle mogelijke classificatiemodellen en zal een uiteindelijke model selecteren op basis van de maximale

winst. Vervolgens zal het geselecteerde classificatiemodel gebruikt worden om berekeningen en daarmee een vergelijking te maken met:

1. De huidige werkwijze waar alle prospects gebeld worden.
2. De voorgestelde werkwijze waar het classificatiemodel een voorselectie aan prospects maakt.

Deze vergelijking zal Blackrock (naast andere overwegingen) het diepere inzicht geven en uiteindelijk dienen om een gewogen besluit te maken met betrekking tot de overname van de Portugese bank.

2. Methodology

Dit onderzoek kan uitgevoerd worden met elke predictive data analysis tool setup, maar wij hebben onder andere de volgende systeemeisen gebruikt:

- Python 3.11
- PDM
- Jupyter
- Pandas
- Numpy
- Scikit-Learn

De beschrijving in de [repository] voor dit onderzoek geeft aan, hoe je dit onderzoek moet opzetten. Dit zodat het onderzoek zelf en de resultaten gevalideerd kunnen worden.

2.1. Dataset

De dataset voor het onderzoek laat zich als volgt beschrijven:

- tijdsreeks data
- 41,000+ instances
- 20 features
 - 10 numeriek
 - 10 categoriaal
- Target met binaire waarden: yes/no
- Geen missende waarden

De data is verzameld in de periode mei 2008 t/m november 2010. De structuur van de dataset is als volgt:

#	Column	Non-Null Count	Dtype
---	-----	-----	-----
0	age	41188 non-null	int64
1	job	41188 non-null	object
2	marital	41188 non-null	object
3	education	41188 non-null	object
4	default	41188 non-null	object

5	housing	41188	non-null	object
6	loan	41188	non-null	object
7	contact	41188	non-null	object
8	month	41188	non-null	object
9	day_of_week	41188	non-null	object
10	duration	41188	non-null	int64
11	campaign	41188	non-null	int64
12	pdays	41188	non-null	int64
13	previous	41188	non-null	int64
14	poutcome	41188	non-null	object
15	emp.var.rate	41188	non-null	float64
16	cons.price.idx	41188	non-null	float64
17	cons.conf.idx	41188	non-null	float64
18	euribor3m	41188	non-null	float64
19	nr.employed	41188	non-null	float64
20	y	41188	non-null	object

2.2. Data Cleaning

Afgezien dat er geen waarden missen in de dataset, is er wel een andere taak qua data cleaning vereist. De beschrijving van de dataset geeft aan dat de feature *duration* verwijderd moet worden. Dit om opmerkelijke voorsellingen te voorkomen.

Naar onze mening heeft dat te maken met mogelijke data leakage op het moment van trainen van een voorspellingsmodel. Dit met de gedachte dat een voorspelling van een gesprek vooraf niet gemaakt kan worden, als de *duration* van datzelfde gesprek nog vastgelegd moet worden.

Deze *duration* feature moet daarom verwijderd worden.

2.3. Feature Engineering

De dataset is opgebouwd als een tijdsreeks. Echter mist er een concrete timestamp in de data. Door de periodebepaling van de databeschrijving te gebruiken - samen met de waarden in de *month* feature, kan de het jaar per instance bepaald worden. Deze nieuw aangemaakte *year* feature zal meerdere inzichten qua data analyse mogelijk maken.

2.4. Train Test Split

Doordat de dataset een tijdsrijks is, moet er nadruk gelegd worden op hoe de data gesplits wordt mbt trainen en testen. Een model mag namelijk niet getest worden op data die chronologisch voor de data ligt waarop getraind is. Dit om nog een vorm van data leakage te voorkomen.

Om dit probleem op te lossen, is er gekozen voor configuratie die de volgorde van een tijdsrijks handhaaft. Dit zowel tijdens cross-validatie als model-evaluatie:

1. Er wordt geen gebruik gemaakt van de `shuffle` parameter bij de initiële train test split.
2. Tijdens cross-validatie wordt er gebruik gemaakt van een *rolling window* setup.

Op deze manier lossen we mogelijke data leakage mbt tijdsreeks data op.

2.5. Preprocessing

Een aantal taken aan preprocessing dienen uitgevoerd te worden, voordat modeltraining kan plaatsvinden.

1. Alle categorische waarden omvormen naar numerieke waarden.
2. Alle numerieke waarden omvormen naar normale distributies voor de features die dat niet zijn.
3. Alle numerieke waarden omvormen naar een range van nul naar één.

Deze stappen zijn van belang om goed presterend model te ontwikkelen.

2.6. Model Shortlist

De modellen die in aanmerking komen om geevalueerd te worden, voldoen aan de volgende condities:

- Het model moet voorzien zijn van een probability functie (om gebruik te kunnen maken van de MP metric).
- Het model mag niet langer dan maximaal een minuut trainen per data batch.

De modellen die voldoen aan deze condities zijn:

- AdaBoost
- Gradient Boosting
- K-Nearest Neighbors
- Logistic Regression
- Random Forest
- XGBoost

Wellicht bestaan er meer modellen die aan de bovenstaande condities voldoen, maar voor dit onderzoek beperken we ons tot deze shortlist.

2.7. Procedure

De procedure om het onderzoek uit te voeren bestaat uit de volgende stappen:

1. Hypertune alle modellen op basis van de MP metric als score.
2. Selecteer de best presterende parameters voor elk model.
3. Modelleer alle modellen door middel van rolling window cross-validatie.
4. Voorspel de probabilites van alle model door middel van rolling window cross-validatie.

5. Bereken de maximale winst voor alle modellen.
6. Evalueer de maximale winst voor alle modellen.
7. Selecteer het model met de hoogst maximale winst.
8. Gebruik de voorspellingen van het geselecteerde model om de huidige werkwijze met de voorgestelde werkwijze te vergelijken.

Wellicht heb je nu een kopje koffie verdiend. :-)

3. Exploratory Data Analysis

Voordat we de beschreven procedure van het onderzoek hebben uitgevoerd, hebben we tevens de dataset zelf onderzocht. Hierbij hebben we een aantal opmerkelijkheden ontdekt, die van belang zijn voor de prestaties van het uiteindelijke voorspellingsmodel.

3.1. Data Imbalance

De eerder beschreven *year* feature geeft de mogelijkheid om de data per jaar te categoriseren. Deze categorisatie laat de volgende twee opmerkelijkheden zien.

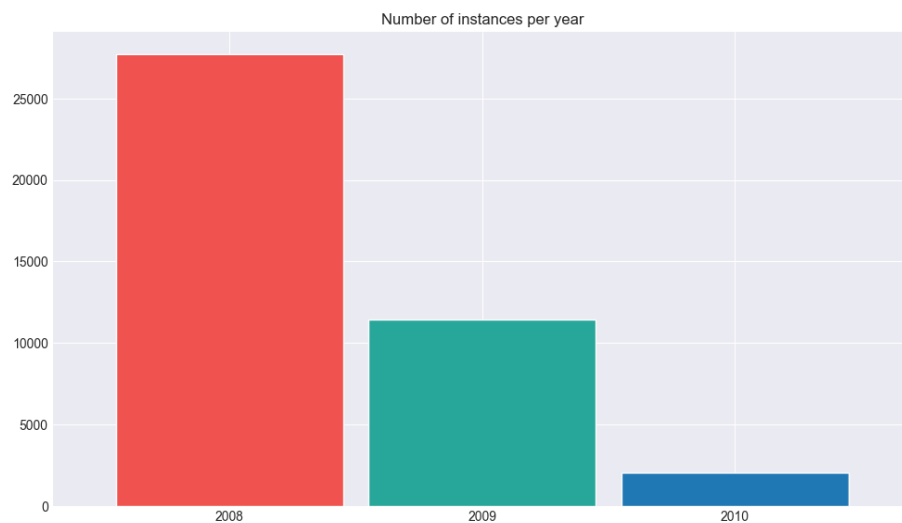


Figure 1: Number of instances per year

De bulk van de data is geconcentreerd in het jaar 2008, zoals figuur 1 aantoont. Daarnaast verschilt de verdeling van de target per jaar aanzienlijk, zoals figuur 2 aantoont.

Met beide imbalances moet rekening gehouden worden. Daarbij hebben we besloten om alleen data uit 2008 te gebruiken. Deze subset heeft namelijk de

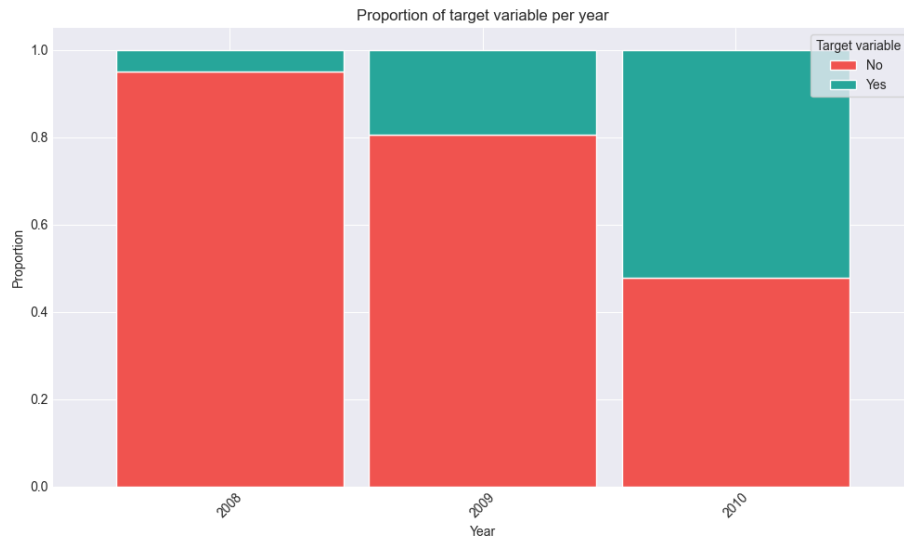


Figure 2: Proportion of target variable per year

meeste instances van alle jaren.

We weten niet waarom deze imbalance in de data aanwezig is. Een speculatie is dat er in 2008 een financiële crisis woedde en dat er daarom in dat jaar meer negatief is geantwoord dan andere jaren, op de vraag van de telemarketing campagne. Maar dit kunnen we niet verifiëren, omdat we geen toegang hebben tot de eigenaren van de dataset.

3.2. Approached vs Not-Approached

De *pdays* feature geeft volgens de databeschrijving het aantal dagen aan sinds het laatste contact binnen de lopende telemarketing campagne. Met de waarde 999 als uitzondering op die regel, want dit geeft aan dat de prospect nog niet is benaderd.

Deze mix van numerieke en categorische waarden binnen één feature zal een probleem opleveren bij het trainen van een voorspellingsmodel. Omdat een model dit contextuele onderscheid niet kan maken.

Figuur 3 toont aan dat de verdeling van benaderde ten opzichte van niet-benaderde prospects in het jaar 2008 marginaal is.

Om het probleem van de mix aan waarden binnen de *pdays* feature op te lossen, hebben we besloten om alleen te focussen op niet benaderde prospects. Dit omdat voor deze subset relatief meer data beschikbaar is.

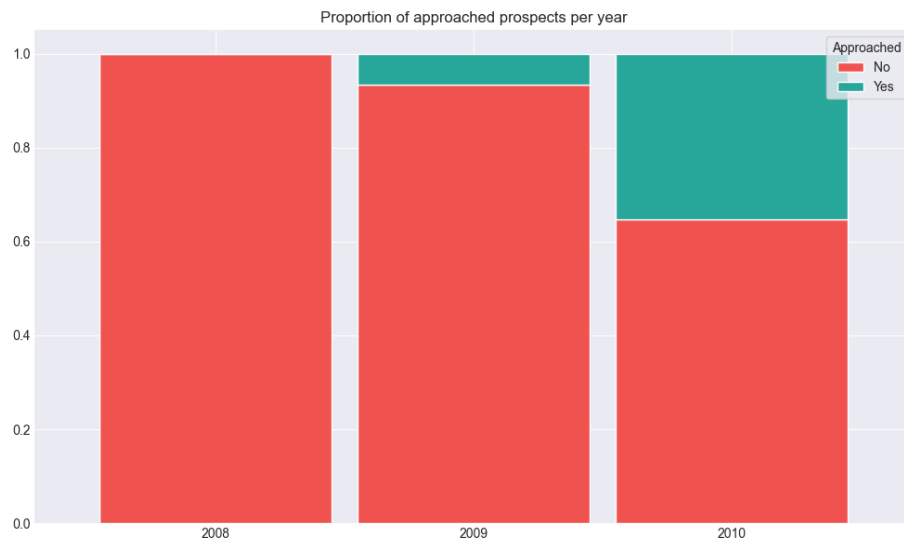


Figure 3: Proportion of approached prospects per year

4. Results

4.1. Metrics

- ROC
- Calibration
- Recall

Metric	AdaBoost	Neural Net	Random Forest
AUC	0.7465	0.7626	0.7739
ALIFT	0.2465	0.2626	0.2739

4.2. Model Evaluation

! Beschrijf alle evaluaties en waarom ze gekozen zijn:

- Confusion matrix
- Probability calibration
- Cost vs threshold analysis

4.3. Model Selection

4.4. ROI Calculations

5. Discussion

6. Conclusion

References