Banking Marketing

Project by Erlando Febrian

TABLE OF CONTENTS

Goal, Metric, and Objective

Goal

Increase conversion rate up to 14%

Metric

Conversion Rate

Objective

- Analyze factors that cause low conversion rate
- Predict customer will convert or not

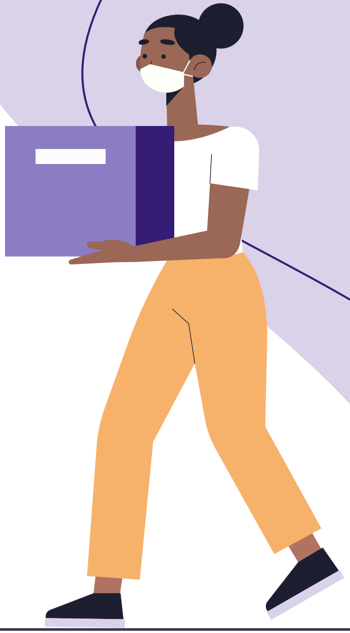
01
Project Background

Background, Metric, Objective and Goals

02

Exploratory Data Analysis

Data Exploration & Business Insights



Dataset Overview

1 Year Historical Data (41.188 Clients & 20 Features)

Bank Client Data

- Age
- Job
- Marital Status
- Education
- Default
- Housing
- Loan

Related to Last Contact

- Contact
- Month
- Day of Week
- Duration

Social & Economic Attribute

- Emp.Var.Rate
- Cons.Price.Idx
- Cons.Conf.Idx
- Euribor3m
- Nr.Employed

*) Detail Features Dictionary Written On Appendix

Dataset Overview

1 Year Historical Data (41.188 Clients & 20 Features)

Other Atributes

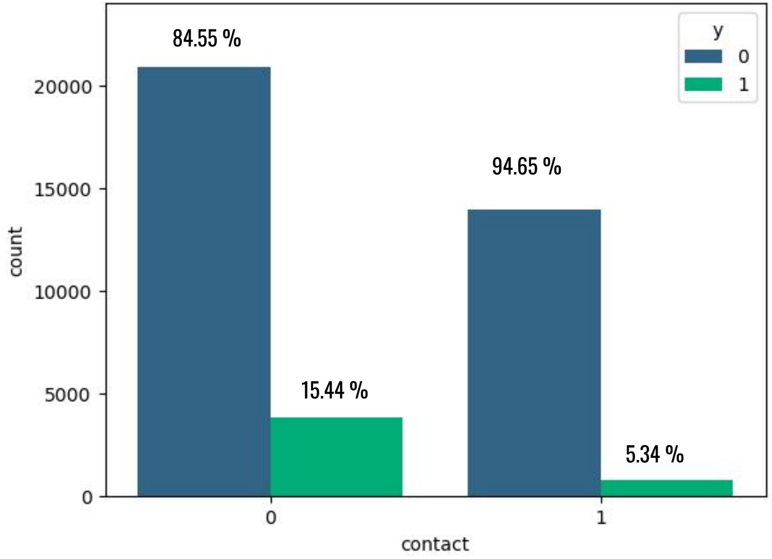
- Campaigns
- Pdays
- Previous
- Poutcomes

Y - Target Feature

Has the client subscribed a term deposit?

*) Complete Features Dictionary Written On Appendix

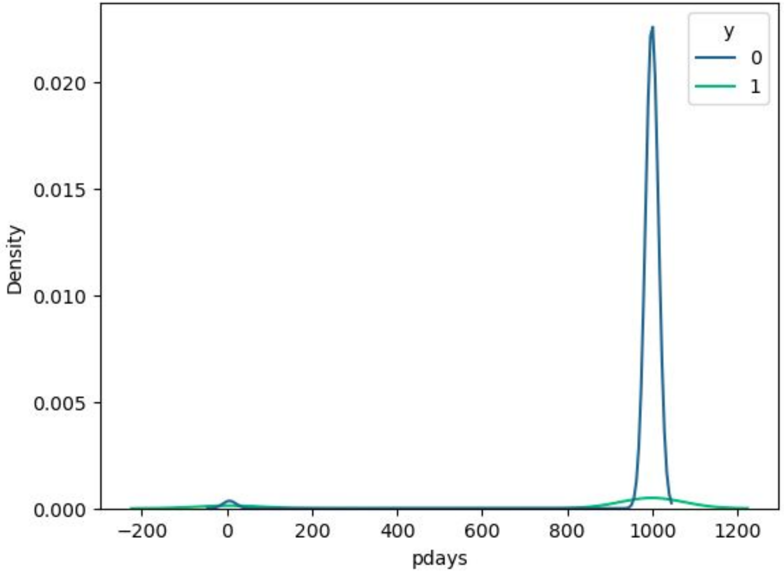
Data Exploration - Contact



Cellular users have slightly higher customers who subscribe term of deposit. It has also higher probability to subscribe (**15.44 %**)

*) contact → contact communication type. Cellular = 0, Telephone = 1

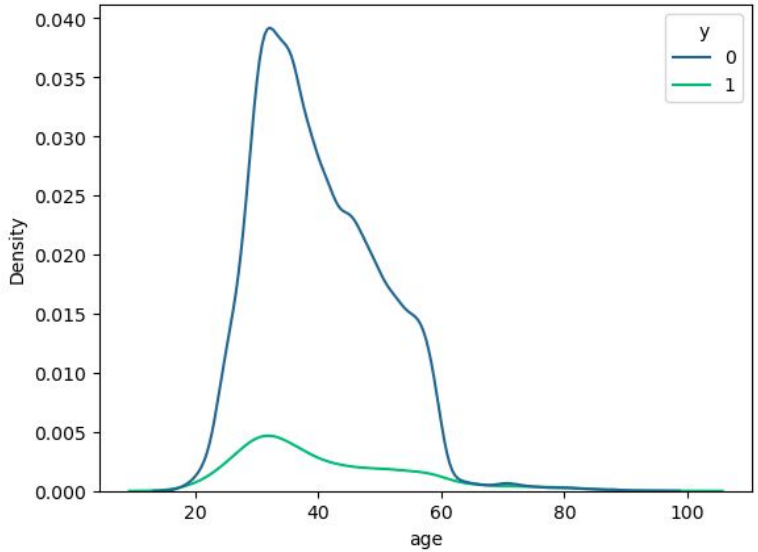
Data Exploration - Pdays



This feature has **high number of clients who was not previously contacted**. We will explore how this features affect target feature

*) **Pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

Data Exploration - Age



Our clients contains of people around 20-60 years old, but majority of them are in **30-40 years old.** In the next exploration, we will find how age affect to our target feature.

*) Age : How old our client is

03

Data Preprocessing

Data Pre Processing

1. Drop Duplicate Values

1784 values dropped.
Data is clean, no need to handling missing values

2. Feature Engineering

Engineered features:
Education, Default, Contact, Target Feature

3. Feature Encoding

Using One Hot Encoder:
As a result, there are 40 additional encoded features

4. Outlier Detection

No extreme/weird values found, no need to do outlier handling

Data Pre Processing

5

Feature Scaling

Using MinMaxScaler to scale X_train and X_test

6

Feature Selection

Using filter method: chi square, mutual info, quasi, univariate feature selection

7

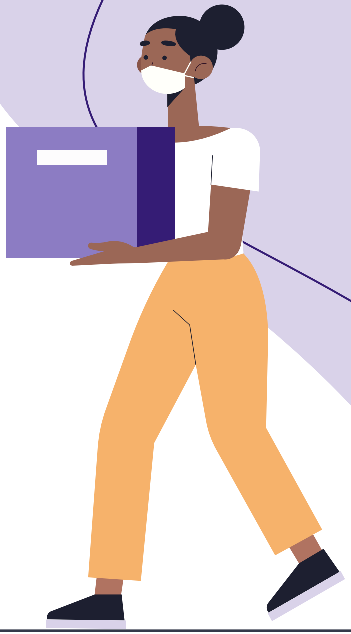
Handling Target Imbalanced

Using SMOTE with default sampling strategy (1:1)

04

Modeling

Basic Model, Hyperparameter Tuning, Feature Importance



Classification Model

	Model	Accuracy	Precision	Recall	F2-Score
0	XGBClassifier	0.89	0.51	0.36	0.38
1	AdaBoostClassifier	0.84	0.37	0.59	0.53
2	RandomForestClassifier	0.87	0.44	0.40	0.41
3	LogisticRegression	0.77	0.29	0.66	0.52
4	DecisionTreeClassifier	0.87	0.43	0.34	0.36
5	KNeighborsClassifier	0.87	0.43	0.47	0.47

Focus on Recall Score to minimize False Negative (Predicted **will not subscribe** term of deposit, but **actually subscribe**). We consider marketing cost that give customers who predicted as not subscribe is costly. Then, we will do hyperparameter tuning on:

- AdaBoost Classifier
- Random Forest
- Logistic Regression
- KNeighbors

Because they have high F2 Score

Hyperparameter Tuning

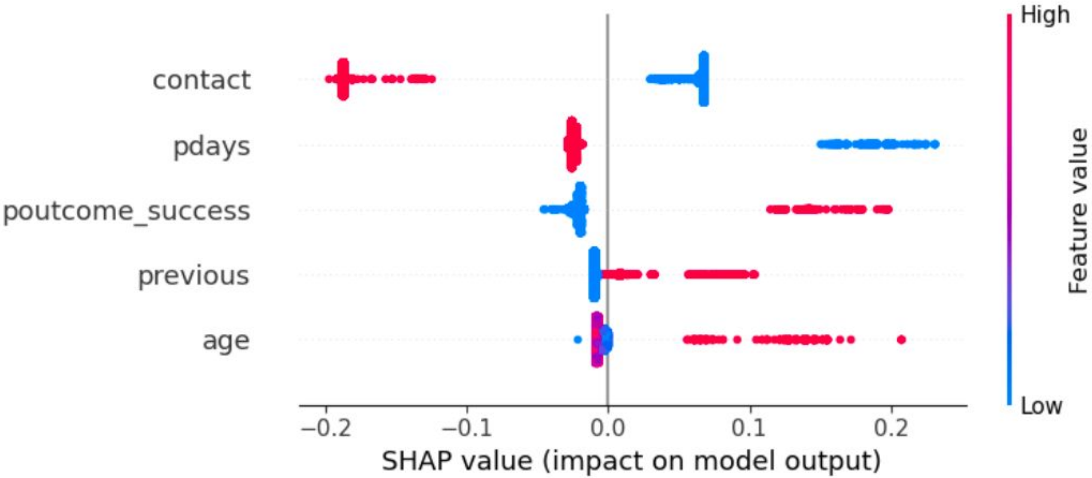
	Train Recall Score	Test Recall Score
Logistic Regression	0.83	0.83
Random Forest	**0.85**	**0.83**
Ada Boost	0.85	0.83
Kneighbors	0.59	0.59

We did 2x hyperparameter tuning, 1st aim to filter good predictive features. As a result we found top 5 predictive features, then 2nd hyperparameter focus on finding good recall score. Top 2 model are:

- AdaBoost Classifier
- Random Forest

Then we decided to choose **Random Forest** because it has **better importance features interpretation** rather than Adaboost.

Feature Importances



We can see that:

- Customer who contacted using cellular tend to be more subscribe to term of deposit
- Customers who are frequently contacted (recently contacted) tend to be more subscribe to term of deposit
- etc

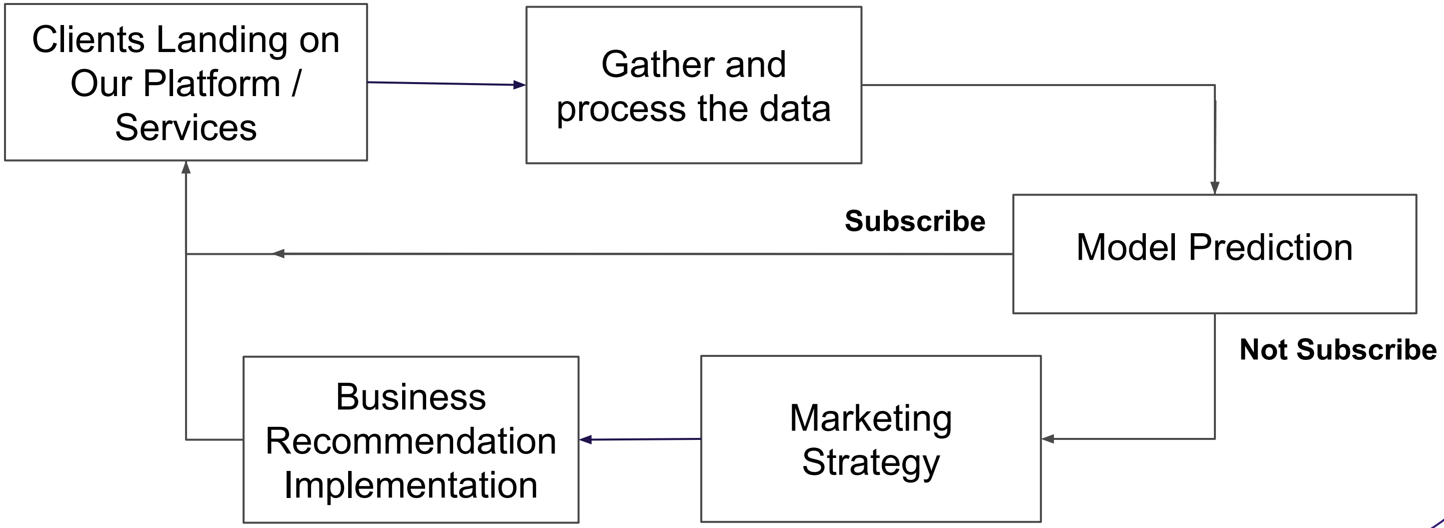
*) contact → Cellular = 0, Telephone = 1
*) pdays → number of days that passed by after the client was last contacted from a previous campaign
*) poutcome_succes → outcome of the previous marketing campaign
*) previous → number of contacts performed before this campaign and for this client

05

Business Insights and Recommendation

Business Simulation, Insights, and Recommendation

How Our Model Works?



Business Insights & Recommendation

Based On SHAP Values Feature Importance

Contact	Pdays	Poutcome Success	Previous	Age
For re-marketing purpose, **only target cellular users** to get more probability of subscribed term of deposit	**Keep inform / offer and contact our clients** to clients regularly using recommended platform (cellular)	To get more clients, For re-marketing purpose, **only target clients who previously accept our campaign**.	The **more contact performed** / make offers increase the client's probability to subscribe our product	**The older clients tend to be more subscribe term of deposit** rather than the younger one. For remarketing purpose, only target the older generation

*) For the next step, we will try to implement recommendation above using simulation

Business Simulation

Before

Without modeling, the Bank run blind marketing campaign and random strategy

Current Conversion Rate :
11.67 % / Year

After

With prediction model, the Bank run targeted campaign and segmented marketing strategy[1]

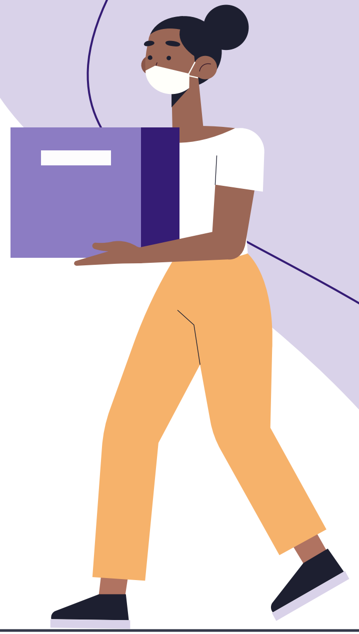
New Conversion Rate :
14.73 % / Year

After implementing the strategy, the bank can increase Conversion Rate up to **3.06%**

*) Strategy : target only cellular users, who had contacted recently, and they are around 30-50 years old

Thank You

Appendix



Features Dictionary

Bank Client Data

1. **Age** (numeric)

2. **Job** : type of job
   (categorical:'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')

3. **Marital** : marital status
   (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)

4. **Education**
   (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')

5. **Default**: has credit in default? (categorical: 'no','yes','unknown')

6. **Housing**: has housing loan? (categorical: 'no','yes','unknown')

7. **Loan**: has personal loan? (categorical: 'no','yes','unknown')

Features Dictionary

Related With the Last Contact of the Current Campaign:

8. **Contact**: contact communication type (categorical: 'cellular','telephone')
9. **Month**: last contact month of year (categorical: 'jan', 'feb', 'mar', …, 'nov', 'dec')
10. **Day of week**: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')

Other Attributes

11. **Campaign**: number of contacts performed during this campaign and for this client (numeric, includes last contact)
12. **Pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
13. **Previous**: number of contacts performed before this campaign and for this client (numeric)
14. **Poutcome**: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

Features Dictionary

Social and Economic Context Attributes:

8. **Emp.var.rate**: employment variation rate - quarterly indicator (numeric)
9. **Cons.price.idx**: consumer price index - monthly indicator (numeric)
10. **Cons.conf.idx**: consumer confidence index - monthly indicator (numeric)
11. **Euribor3m**: euribor 3 month rate - daily indicator (numeric)
12. **Nr.employed**: number of employees - quarterly indicator (numeric)

Target Feature

13. **Y** - has the client subscribed a term deposit? (binary: 'yes','no')