Data Glacier

Your Deep Learning Partner

Bank Marketing Campaign

Nuri Öztürk

Chitra S Chaudhari

27-March-2023

Agenda

- Executive Summary
- Data Understanding
- EDA
- Data Transformation
- Data Dependency
- Model Building
- Model Results
- Recommendations

Executive Summary

Client:

ABC Bank wants to sell its term deposit product to customers Before launching the product they want to predict which clients are most likely to subscribe to a term deposit. In this way, the bank wants to save time and money by running the marketing campaign more effectively and successfully.

Problem description:

- Create an ML model for the bank that will shortlist customers who are more likely to subscribe the term deposit product.
- This will allow the marketing team to target those customers more efficiently.

EDA

The analysis has been divided into three parts:

- Data Understanding
- Univariate and Bivariate Analysis
- Model Recommendations

Data Understanding

- Datasets:
  - bank_additional_full : 21 features and 41119 observations
  - bank_additional: 21 features and 4119 observations
  - bank_full: 18 features and 45211 observations
  - bank: 18 features and 4521 observations

- As mentioned in data description, the addition of the five new social and economic attributes (made available here) lead to substantial improvement in the prediction of a success. So We used bank_additional_full dataset.

Data Exploration

Proportion of people who agreed to a term deposit (positive class) compared to the people who did not is 11.3%. That tells us that data seems unbalanced

Clients Subscribed to Term Deposit

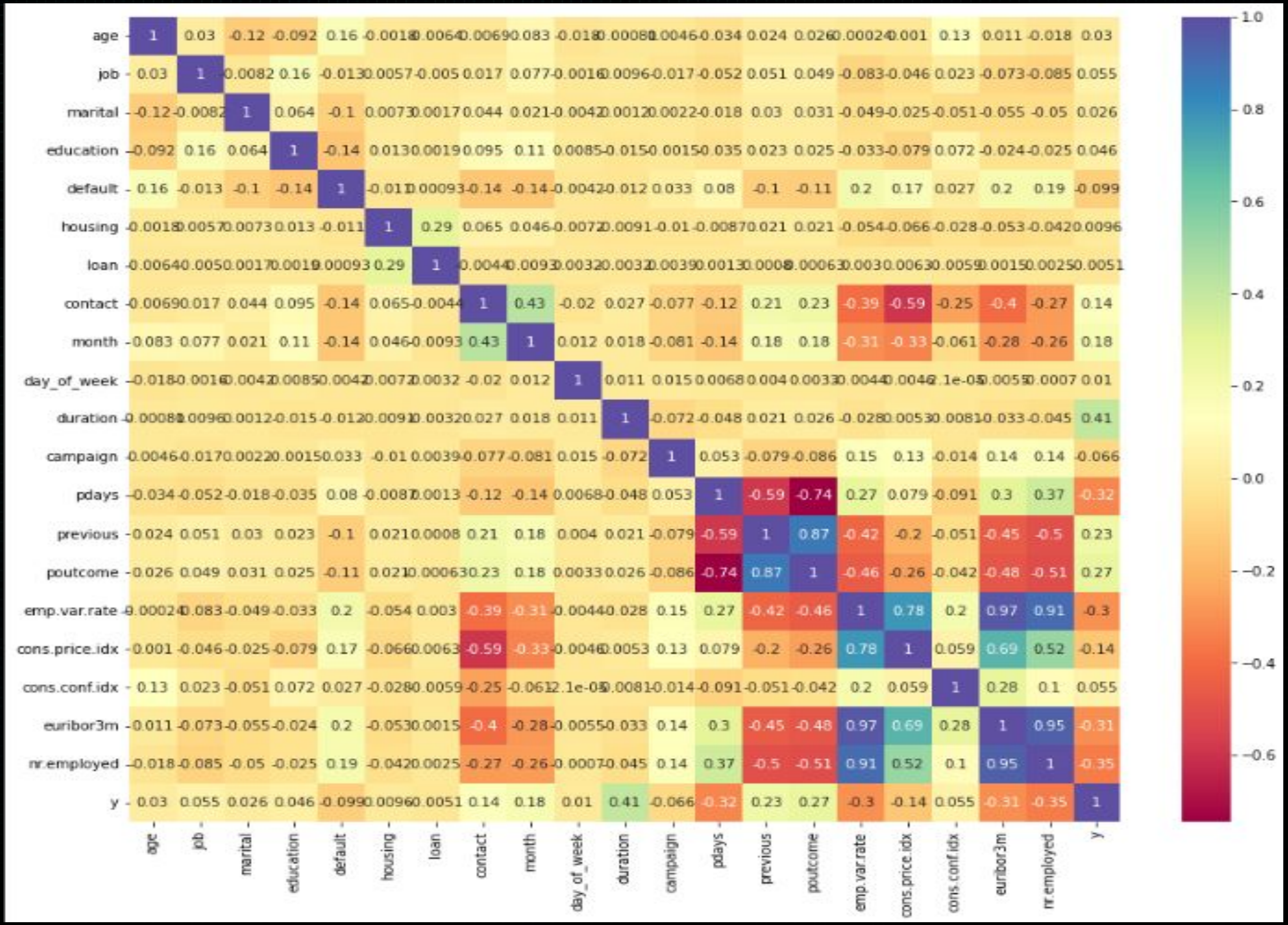
Y	
no	36,548
yes	4,640

Data Transformation

- Timeline of Observations: May 2008 to November 2010
- There are several missing values in some categorical attributes, all coded with the “Unknown” label.
- Duplicate rows are dropped from dataset
- Outliers are not removed as they will help with model generalization
- Encoding Categorical features as machine learning algorithms can only read numerical values.

Data Dependency

When we checked for correlation between features after transformation, we observed no features are highly correlated or inversely correlated with outcome.



Model Building

- ➔ We will use predictive ML model to help us identify potential customers who may subscribe to deposit term plan.
- ➔ In order to build predictive model, we begin by splitting the dataset into training and testing sets in 80% and 20% respectively.
- ➔ We choose to test out the following set of models since we don't know yet what algorithms will do well on this dataset.
  - ◆ Linear Algorithms : Logistic Regression
  - ◆ Nonlinear Algorithms: Support Vector Machines (SVM)
  - ◆ Ensemble Methods: Random Forest, Gradient Boosting (XGBoost)

Model Results

- **Key features with most importance:**

The top 3 key features or attributes that helped in the prediction of the class variable were:

1. nr.employed
2. euribor3m
3. emp.var.rate

Other than these 3 duration and the categorical feature month and poutcome were also held a lot of importance in our case study

Model Results

Here is the summarized chart for different model results:

	Model	Accuracy	Precision	Recall	f1 score	ROC_AUC
0	Logistic Regression (with Duration)	0.854905	0.432347	0.871140	0.577888	0.861978
1	SVC (linear)	0.844463	0.415430	0.894569	0.567376	0.866292
2	SVC (rbf)	0.846285	0.420360	0.919063	0.576872	0.877991
3	Random Forest (n=200)	0.913429	0.672256	0.469649	0.552978	0.720092
4	XGBoost	0.912458	0.636591	0.541001	0.584917	0.750629

Model Results

Since it is unbalanced dataset we also tried running the models with balancing dataset. Here is the summarized chart:

	Model	Accuracy	Precision	Recall	f1 score	ROC_AUC
0	Logistic Regression (with Duration)	0.873384	0.878947	0.874346	0.876640	0.873355
1	SVC (linear)	0.880927	0.869960	0.903665	0.886492	0.880245
2	SVC (rbf)	0.884698	0.857281	0.930890	0.892570	0.883314
3	Random Forest	0.884159	0.839450	0.958115	0.894866	0.881943
4	XGBoost	0.868534	0.862385	0.885864	0.873967	0.868015

Model Recommendation

What type of ML Model to use?

We can see from model results charts SVM (support vector machine) with rbf kernel provide strong ROC_AUC score upto 88%. And it is best compared to the other ones. Therefore we should consider that model for production.

Recommendations

- More target based campaigning can help increase the overall effectiveness of this marketing campaign.
- Looking at customer base, the age groups of 26-40 and 41-60 have a higher proportion. These groups present a profitable target for marketing team.
- According to job category, focus on Admin, Blue-collar, Technician, Services and management professions.
- Students and senior citizens respond better to this proposal so we can also focus on those specific categories.
- It is also possible to use other means of communication besides cellular.
- Take into account the best time to contact potential clients.

Thank You

Group Name: LimitLess Team

Name: Chitra S Chaudhari
Email: chitraksonawane@gmail.com
Country: USA
Specialization: Data Science

Name: Nuri Öztürk
Email: ozturknuri8@gmail.com
Country: Turkey
Specialization: Data Science

Batch Code: LISUM 17
Github Repo link: https://github.com/ChitraChaudhari/Bank-Marketing-campaign