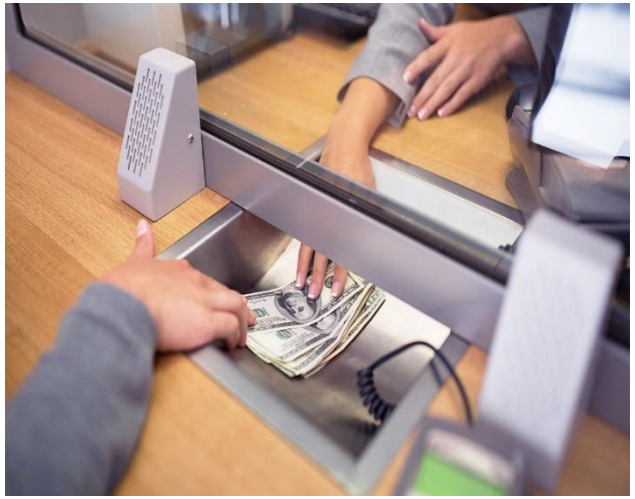
Direct marketing campaigns (phone calls) of a Portuguese banking institution

- **Executive Summary:**

- The main focus of this project is to Predict whether the client will open a term deposit (variable y) using sociodemographic, social and economic context attributes and account attributes.

- Different supervised classification models were used and the most important features for the high score model (Ridge) explored are:
  - ❑Consumer price index, Communication type cellular
  - ❑Contact Month Jul, Consumers confidence Index
  - ❑Age, Job technician



- The results show that the Logistic regression model (using Ridge algorithm) is the best model for prediction.

Objective:

- To predict whether potential consumers will put deposits.
- To find out what machine learning model can predict the target variable better than other used ones
- to find out what features affect consumer decision to put deposits

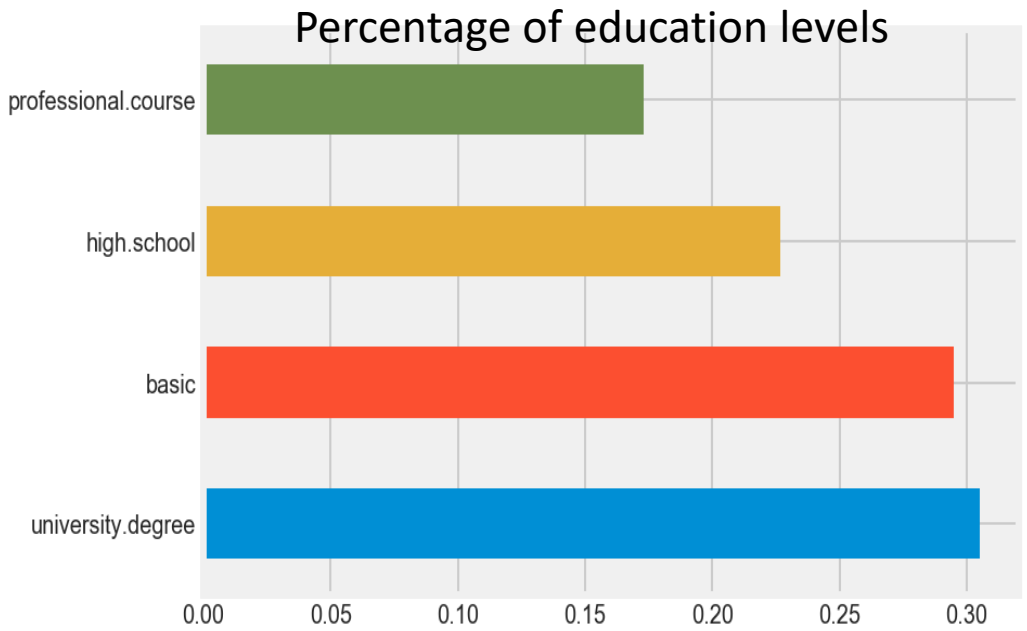
Dataset Overview

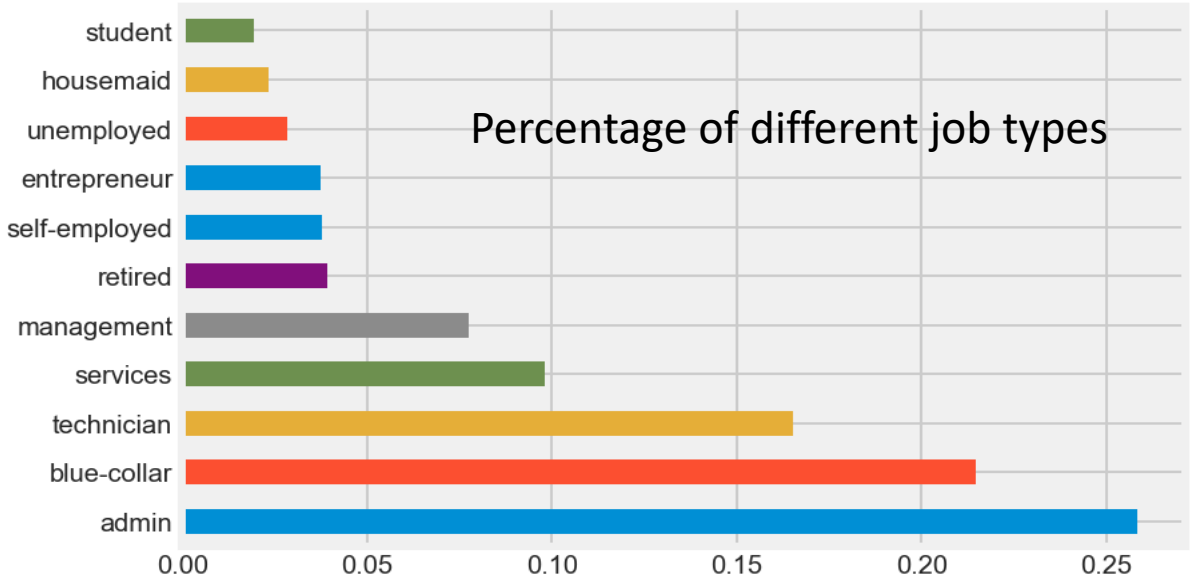
Feature Variables (21)		
Age	communication type	outcome of the previous marketing campaign(poutcome)
Job	month	Employee variation rate
Marital	last_contact_day	consmer_price_indx
Education	last_contact_duration	Consumer confidence index
Have_credit by default	Number of contacts with client	euribor 3 month rate
Housing loan	Number of days client was contacted in prev_campaign	employed staff rate
Personal loan	no_contct_bef_campaign_wth_samepersn	

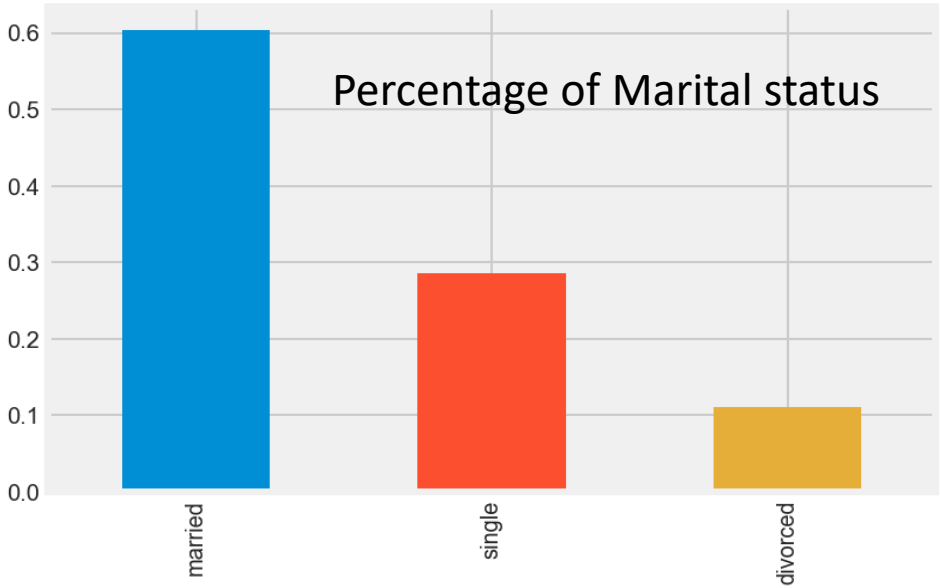
Target is 'open deposit account or not'.

Dataset Size → 2999 observations with 21 feature.

Exploratory Data Analysis Part1

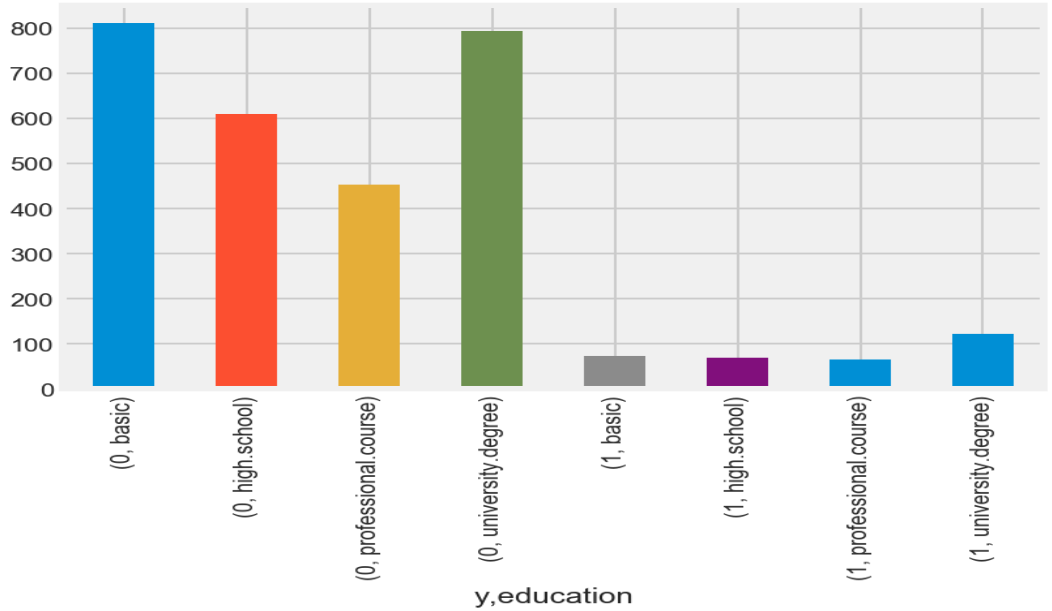


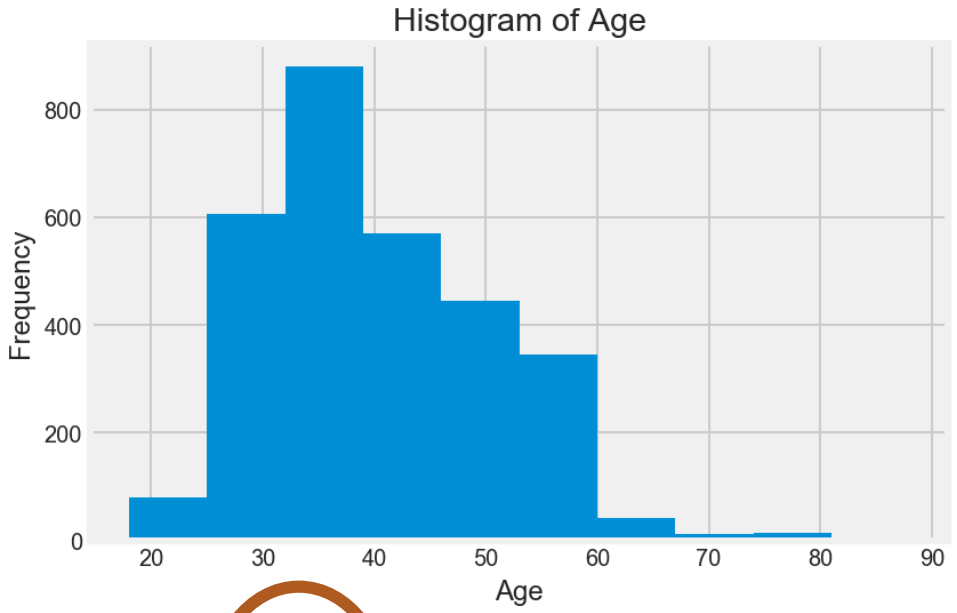




- Majority of potential customers are either highly educated or have basic level of education
- About 50% are either have admin (26%) or blue-collar jobs(22%).
- More than 60% are married

Exploratory Data Analysis2



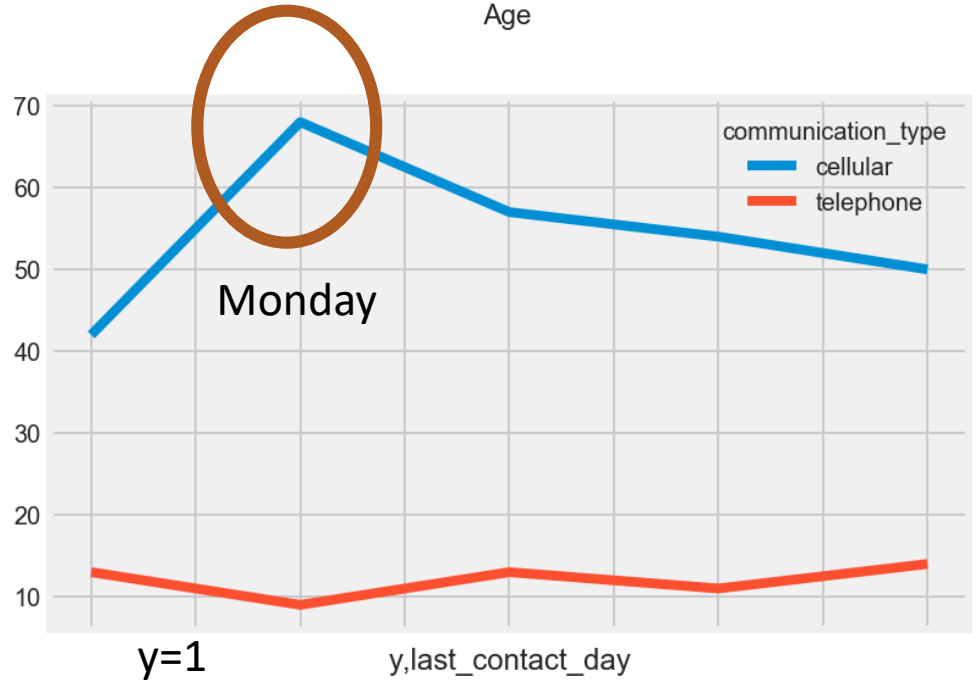


- Majority of customers were age bet ween 30-40(Gen Y)

Char:
- Majority of them use the web and are mobile web user.
- They seek out info and engage in two-way brand conversation.

No wonder that majority of them use cellular phones

Majority of contacts who put deposits for cellphone users were contacted on Mondays



Feature Engineering Part 1 - Dealing with Multicollinearity

- Used VIF Variance Inflation Factor to detect multicollinearity.
  - The Variance Inflation Factor (VIF) is a measure of *collinearity among predictor variables* within a multiple regression.
  - Benchmark: If the VIF is between *5-10*, multicollinearity is likely **present** and you should consider dropping the variable.

- Columns deleted
  - nr_employees (Number of employees)
  - euribor3m (Euro rate for last 3 months)
  - Previous_campaign_outcome (previous campaign outcome)

Feature Engineering Part 2 - Dealing with Imbalanced Dataset

Problem –

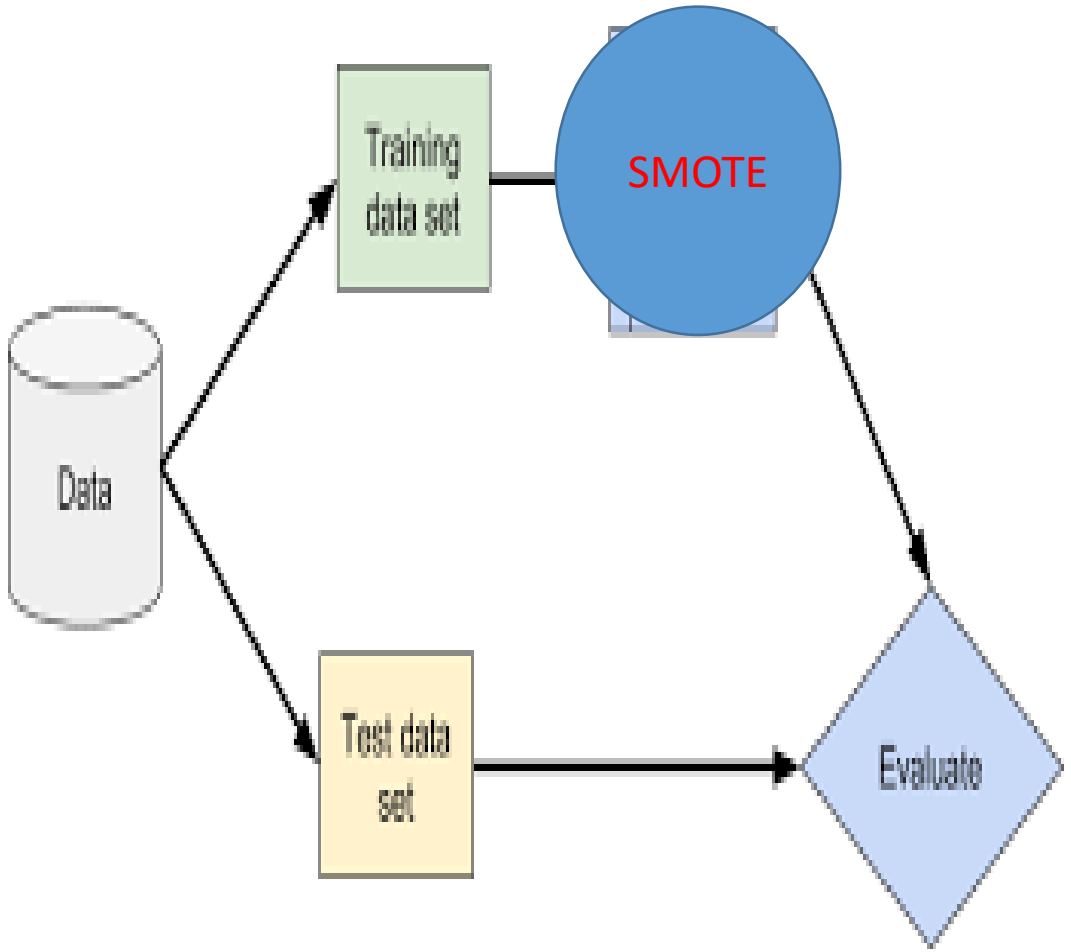
Imbalance dataset:

88%: Didn’t put a deposit.

In other words the baseline accuracy was 88%

Solution – The word is SMOTE !!!

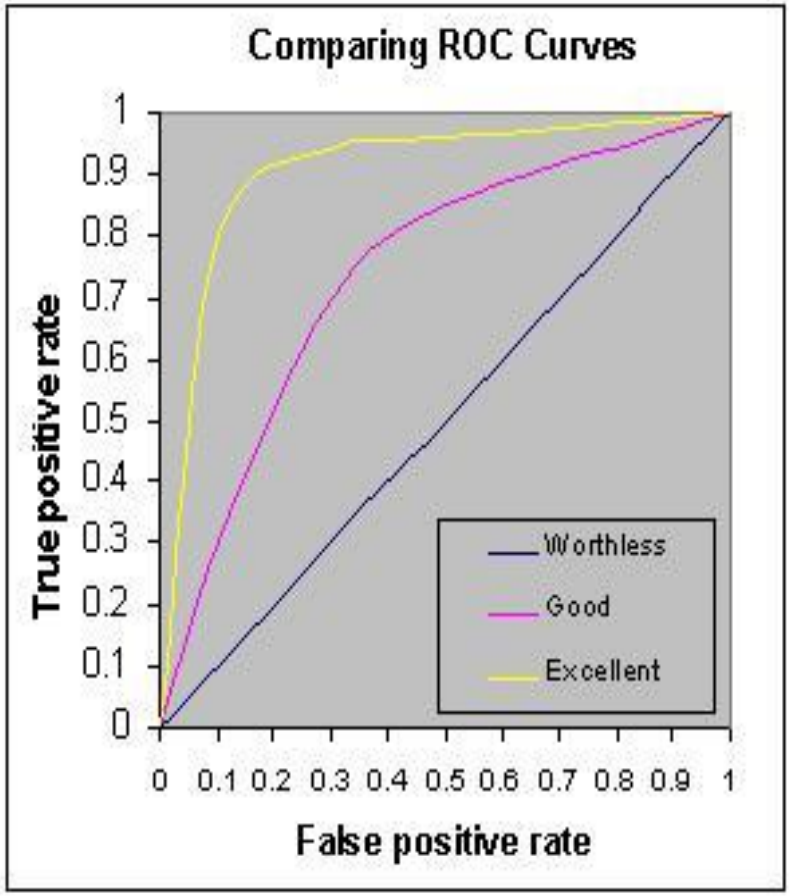
Modeling Workflow



- Train-test split on your original data, for myself. 0.1:0.9.
- On the train data set, upsampling to make it a 50-50 split because of class imbalance.
- GridsearchCV on the training data set for all the learning algorithms.
- Check the baseline on the training set.
- Put in your test set on your best_estimator for each learning algorithm.
- Used the AUCROC score here to comapare models

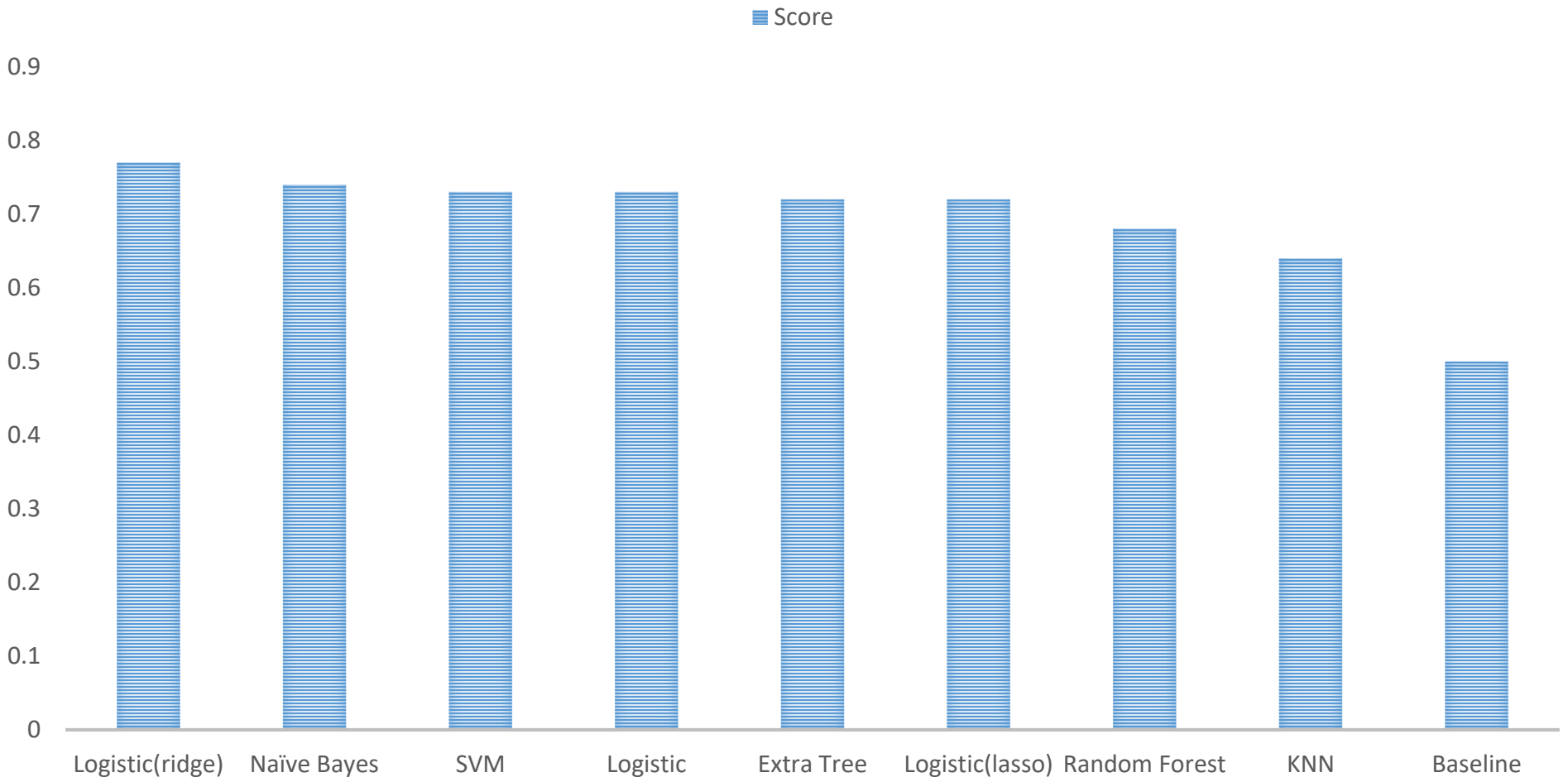
Model Evaluation and Metrics

- Accuracy score is influenced by imbalanced dataset.
- Used AUROC
  - Not influenced by imbalance dataset.
  - The true positive rate is **plotted** in **function** of the false positive rate



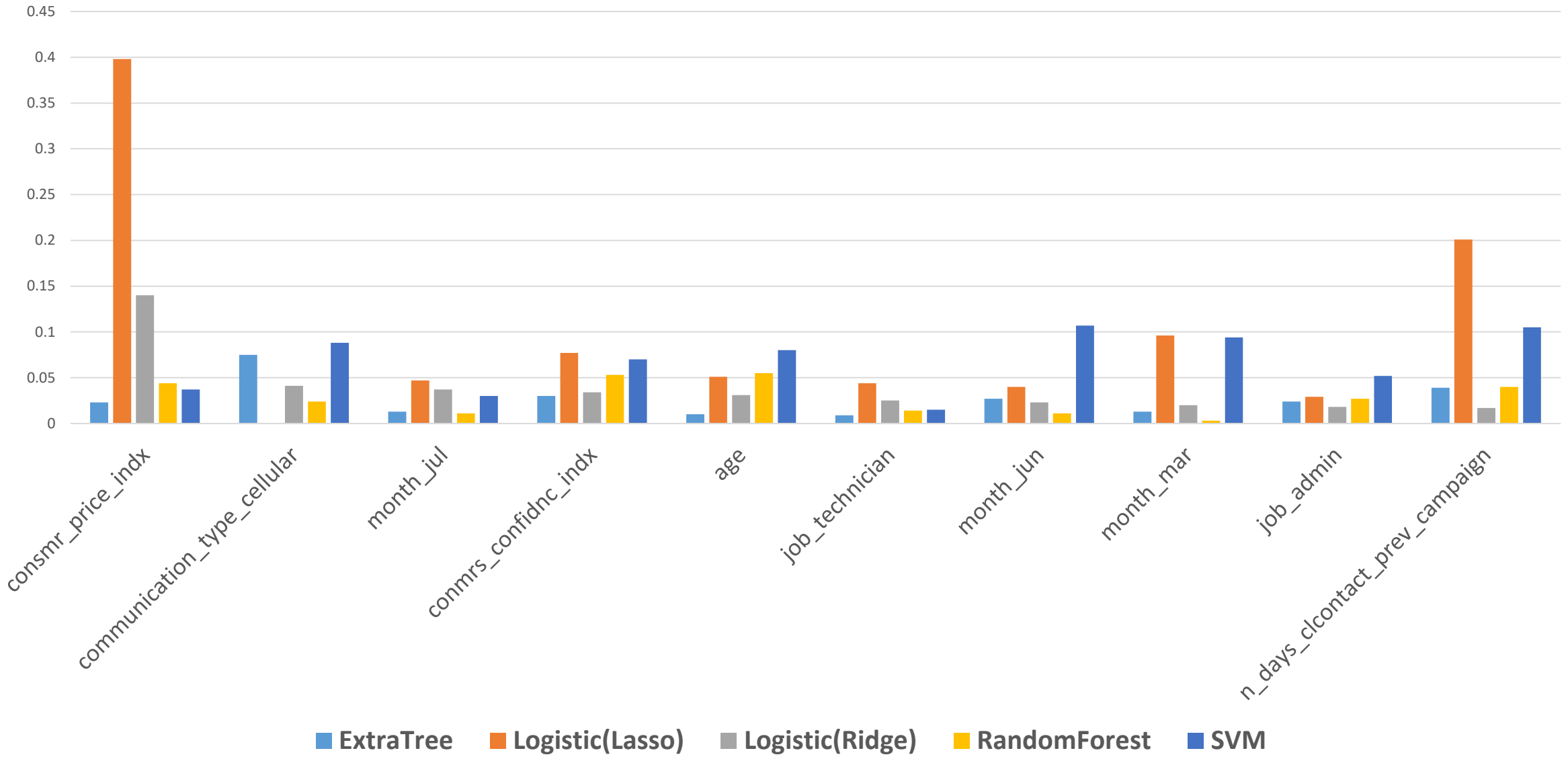
For Demo only, not model results

Various Model Results



Baseline=0.5

Feature _Importance



Important features

No	Ridge	SVM	ExtraTree	Lasso	Random Forest
1	Consumer price index	month_jun	communication_type_cellular	Consumer price index	age
2	Communication type cellular	n_days_clcontact_prev_campaign	n_days_clcontact_prev_campaign	n_days_clcontact_prev_campaign	conmrs_confidnc_indx
3	Contact Month Jul	month_oct	conmrs_confidnc_indx	month_mar	consmr_price_indx
4	Consumers confidence Index	month_mar	housing_loan_no	conmrs_confidnc_indx	n_days_clcontact_prev_campaign
5	age	communication_type_cellular	marital_single	housing_loan_no	marital_single
6	Job technician	age	month_jun	age	housing_loan_no

Conclusion

- **Economic factor (Macro economical factor) :**
  - ❖The Consumer Price Index (CPI) is a measure that examines the weighted average of prices of a basket of consumer goods and services
- **Social: Communication type: cellular**
  - ❖Target this group through social media in platforms they are active
- **Month-July:**
  - ❖ Is the time people may receive the end of financial year
- **Economic factor(Macro economical factor): Consumer confidence**
  - ❖ The degree of optimism that consumers are expressing through their activities of savings and spending.


Thank
you