Banking predictions

BACQUE FLORENT

LEOPOLD ARMAND

CHÂTEAU ROMAIN

Data presentation

Marketing campaign of a Portuguese banking institution

Phone calls

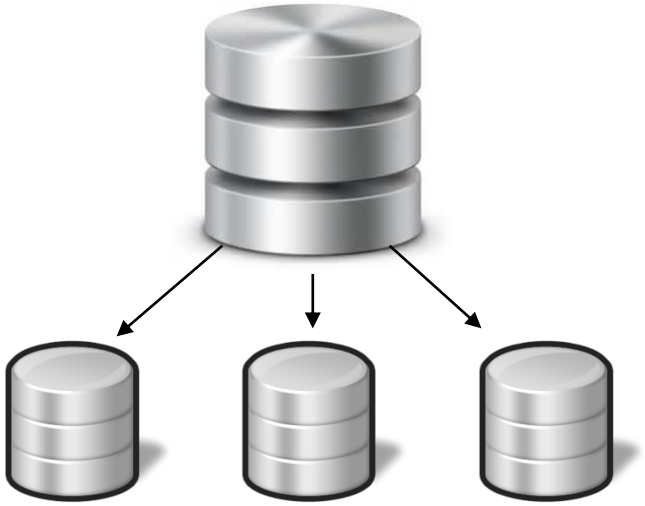
Age	Job	Marital status	Education	Default	Loan	Contact

Date	Duration	Number of contact	Days last contacted	Previous campaign	Has subscribed

=> Tasks repartition

Preparing to modeling

4521 rows

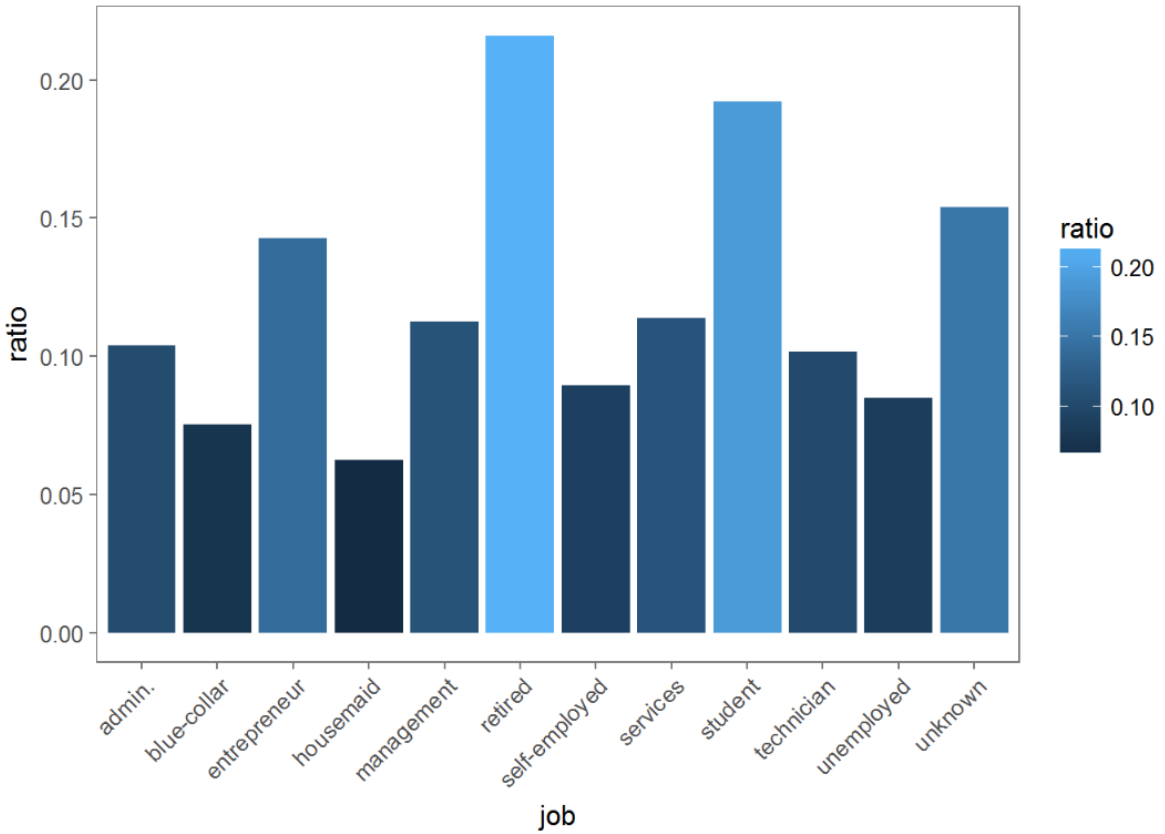


Train Train Test

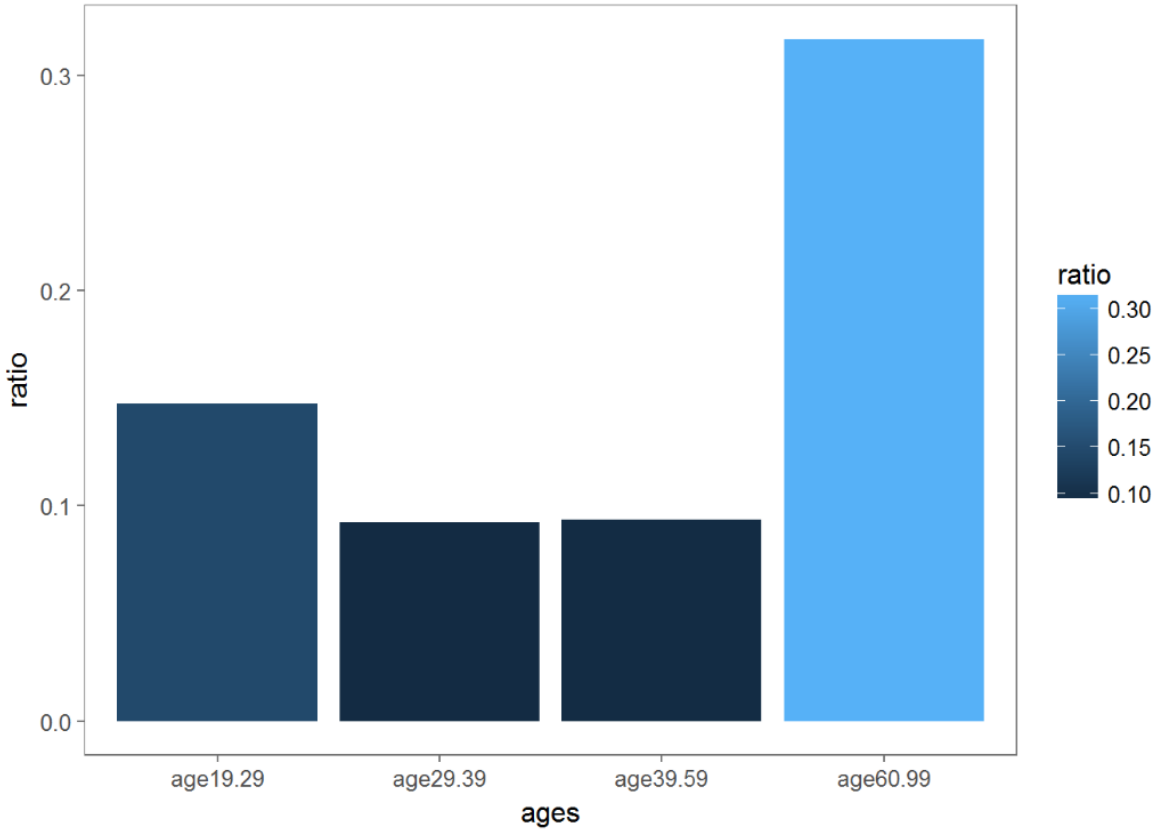
Overview

```
$ age      : int  30 33 35 30 59 35 36 39 41 43 ...
$ job      : Factor w/ 12 levels "admin.","blue-collar",..: 11 8 5 5 2 5 7 10 3 8 ...
$ marital  : Factor w/ 3 levels "divorced","married",..: 2 2 3 2 2 3 2 2 2 2 ...
$ education: Factor w/ 4 levels "primary","secondary",..: 1 2 3 3 2 3 3 2 3 1 ...
$ default  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
$ balance  : int  1787 4789 1350 1476 0 747 307 147 221 -88 ...
$ housing  : Factor w/ 2 levels "no","yes": 1 2 2 2 2 1 2 2 2 2 ...
$ loan     : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 1 1 1 2 ...
$ day      : int  19 11 16 3 5 23 14 6 14 17 ...
$ month    : Factor w/ 12 levels "apr","aug","dec",..: 11 9 1 7 9 4 9 9 9 1 ...
$ duration : int  79 220 185 199 226 141 341 151 57 313 ...
$ campaign : int  1 1 1 4 1 2 1 2 2 1 ...
$ pdays    : int  -1 339 330 -1 -1 176 330 -1 -1 147 ...
$ previous : int  0 4 1 0 0 3 2 0 0 2 ...
$ poutcome : Factor w/ 4 levels "failure","other",..: 4 1 1 4 4 1 2 4 4 1 ...
$ y        : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
```

Data inferences

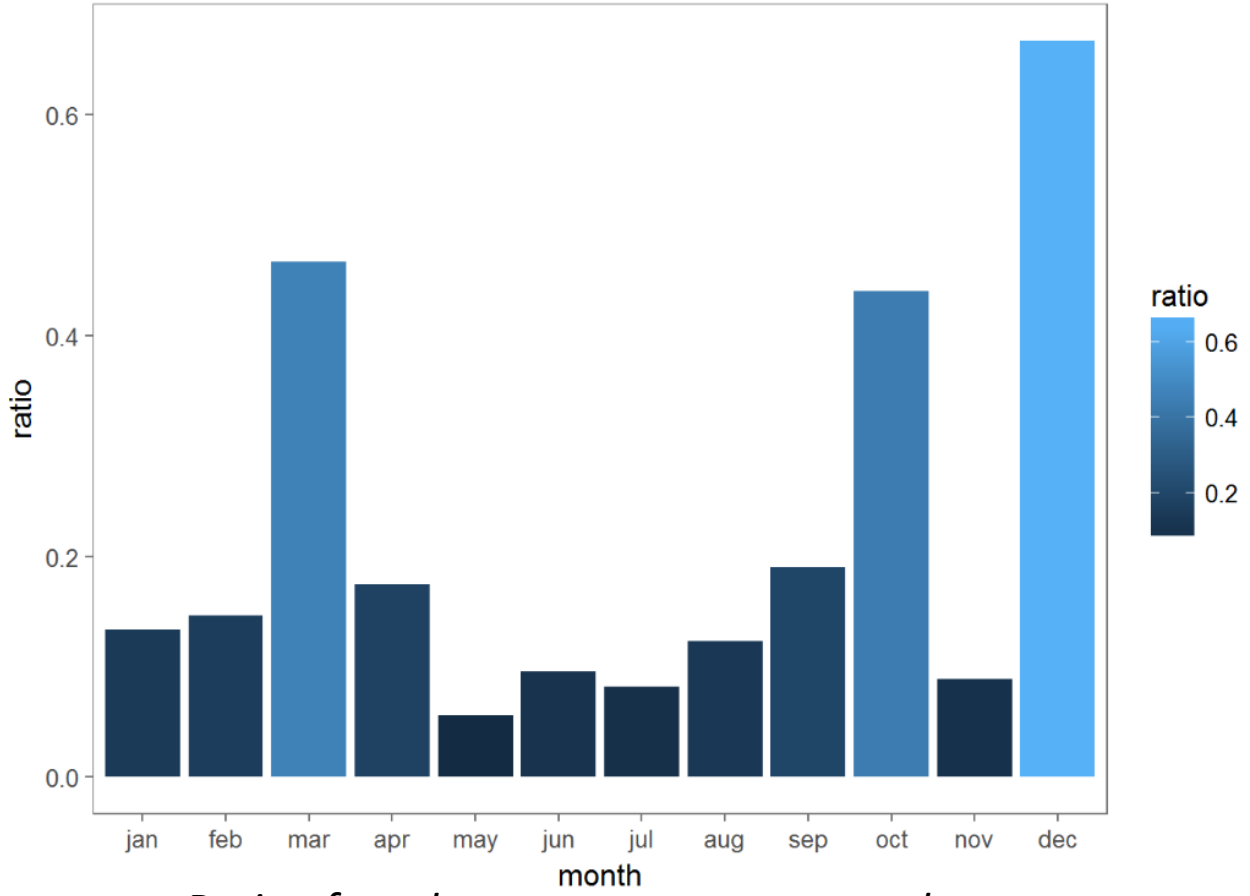


Ratio of product contractors per job



Ratio of product contractors per age range

Data inferences

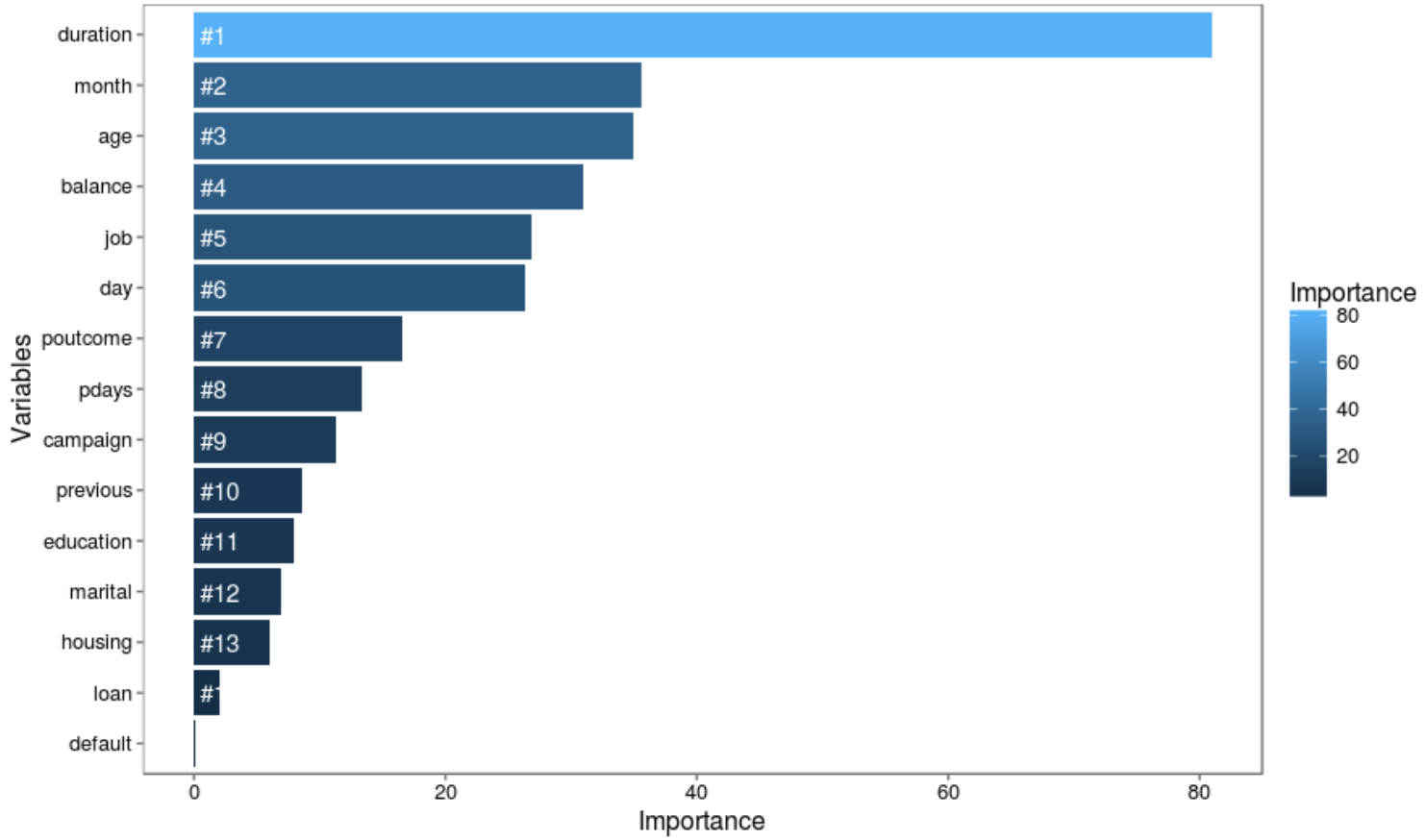


Ratio of product contractors per month

Choosing the best model

Model type	Accurcay (%)	AUC	Std. Dev
Logistic regression	91	0,945	0.0059
Classification tree	89	NA	0.0061
Naive Bayes	86	0,821	0.0090
Random Forest	90	0,664	0.0060
Stacking (Logistic Regression)	89	0,928	NA

Variable importance with RF



Variable importance ranking

Marketing campaign

Goal : see if a targeted advertising campaign could be usefull for the bank

–	Original dataset	Modified dataset
Number of subscription	159	619
Ratio of subscribers	0.1065684	0.4148794

Simulated campaign results

Conclusion

- Our analysis
  - month, balance, age, duration, job are the most relevant predictors

- Future work
  - Improve parameters of baseline classifiers
  - Use voting stacking system
  - Deeper analysis of each dataset parameter to highlight other axes of study