Bank Marketing Campaign Analysis

Presented by Flora Ching Yi Yip

AGENDA

Introduction	• Project Objective • Target Audience • Models Developed
Dataset	• Dataset Overview • Variables Description
Model Development	• Logistic Regression • Classification Tree • Bagging & Random Forest
Conclusion	• Models Evaluation • Insights • Recommendations

PROJECT INTRODUCTION

Objective

- To identify customer segments that are most likely to subscribe to a term deposit
- To provide insights for developing suitable marketing strategies

Target Audience

- Banks, financial institutions

Models Developed

- Logistic Regression
- Classification tree
- Bagging
- Random Forest

Language Used

- R

What is a term deposit?
It is a fixed-term investment at banks where money would be locked up for a period of time and customers can withdraw their funds after the term ends.

DATASET OVERVIEW

Dataset Description

- It contains information about a marketing campaign of a financial institution
- Source: UCI Machine Learning Repository

Variables

- 17 (7 numerical, 10 categorical)

Response Variable

- “deposit”: has the customer subscribed to a term deposit?

Observations

- 11,162

VARIABLES DESCRIPTION

Customer Information

- 1 - age: (numeric)
- 2 - job: type of job (categorical)
- 3 - marital: marital status (categorical)
- 4 - education: (categorical)
- 5 - default: has credit in default? (categorical)
- 6 - housing: has housing loan? (categorical)
- 7 - loan: has personal loan? (categorical)
- 8 - balance: Balance of the individual. (numeric)

Marketing Campaign Details

- 9 - contact: contact communication type (categorical)
- 10 - month: last contact month of year (categorical)
- 11 - day: last contact day of the week (categorical)
- 12 - duration: last contact duration, in seconds (numeric)
- 13 - campaign: number of contacts performed during this campaign and for this client (numeric)
- 14 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric)
- 15 - previous: number of contacts performed before this campaign and for this client (numeric)
- 16 - poutcome: outcome of the previous marketing campaign (categorical)
- 17 - deposit - has the client subscribed a term deposit? (categorical)

1

LOGISTIC REGRESSION

LOGISTIC REGRESSION

- Original Model: fits all predictors
- Refined Model: fits only significant predictors
  - job, marital, education, balance, housing, loan, contact, month, duration, campaign, poutcome

VARIABLES THAT HAVE A SIGNIFICANT CORRELATION WITH THE RESPONSE VARIABLE

Variables related to the customer's information

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)      -7.870e-01  2.685e-01  -2.931 0.003383 **
age              -7.645e-04  3.204e-03  -0.239 0.811405
jobblue-collar   -3.314e-01  1.043e-01  -3.179 0.001480 **
jobentrepreneur  -3.955e-01  1.762e-01  -2.244 0.024818 *
jobhousemaid     -4.757e-01  1.911e-01  -2.490 0.012790 *
jobmanagement    -2.683e-01  1.075e-01  -2.496 0.012563 *
jobretired        2.972e-01  1.474e-01   2.017 0.043693 *
jobself-employed -4.298e-01  1.618e-01  -2.657 0.007886 **
jobservices      -2.835e-01  1.205e-01  -2.352 0.018673 *
jobstudent        5.907e-01  1.763e-01   3.351 0.000805 ***
jobtechnician    -1.567e-01  9.935e-02  -1.578 0.114676
jobunemployed    -1.169e-01  1.673e-01  -0.698 0.484879
jobunknown       -3.942e-01  3.446e-01  -1.144 0.252691
maritalmarried   -1.800e-01  8.566e-02  -2.101 0.035607 *
maritalsingle     7.670e-02  9.853e-02   0.778 0.436313
educationsecondary 2.053e-01  9.281e-02   2.212 0.026969 *
educationtertiary 4.631e-01  1.093e-01   4.236 2.27e-05 ***
educationunknown  2.640e-01  1.506e-01   1.753 0.079628 .
defaultyes       -8.455e-03  2.215e-01  -0.038 0.969556
balance           2.799e-05  8.516e-06   3.287 0.001012 **
housingyes       -7.001e-01  6.217e-02 -11.260  < 2e-16 ***
loanyes          -5.019e-01  8.381e-02  -5.988 2.13e-09 ***
```

Positive Correlation

- Job: student, retired
- Education: secondary, tertiary
- Balance

Negative Correlation

- Marital Status: married
- Housing: the customer has housing loan
- Loan: the customer has personal loan

VARIABLES THAT HAVE A SIGNIFICANT CORRELATION WITH THE RESPONSE VARIABLE

Variables related to marketing campaign

contacttelephone	-5.330e-02	1.080e-01	-0.494	0.621644	
contactunknown	-1.555e+00	9.665e-02	-16.091	< 2e-16	***
day	3.741e-03	3.541e-03	1.056	0.290813	
monthaug	-8.185e-01	1.109e-01	-7.378	1.61e-13	***
monthdec	1.373e+00	3.706e-01	3.706	0.000211	***
monthfeb	-1.675e-01	1.277e-01	-1.311	0.189690	
monthjan	-1.239e+00	1.671e-01	-7.414	1.22e-13	***
monthjul	-9.824e-01	1.122e-01	-8.753	< 2e-16	***
monthjun	2.854e-01	1.327e-01	2.151	0.031511	*
monthmar	2.030e+00	2.289e-01	8.868	< 2e-16	***
monthmay	-6.584e-01	1.068e-01	-6.165	7.03e-10	***
monthnov	-9.556e-01	1.207e-01	-7.915	2.47e-15	***
monthoct	1.080e+00	1.762e-01	6.128	8.92e-10	***
monthsep	9.350e-01	1.994e-01	4.688	2.75e-06	***
duration	5.469e-03	1.244e-04	43.978	< 2e-16	***
campaign	-9.119e-02	1.362e-02	-6.696	2.14e-11	***
pdays	-8.934e-05	4.300e-04	-0.208	0.835407	
previous	1.731e-02	1.421e-02	1.218	0.223360	
poutcomeother	8.847e-02	1.331e-01	0.665	0.506246	
poutcomesuccess	2.227e+00	1.416e-01	15.731	< 2e-16	***
poutcomeunknown	-2.768e-01	1.376e-01	-2.012	0.044259	*

Positive Correlation

- Months: March, June, September, October, December
- Duration
- Campaign outcome: success

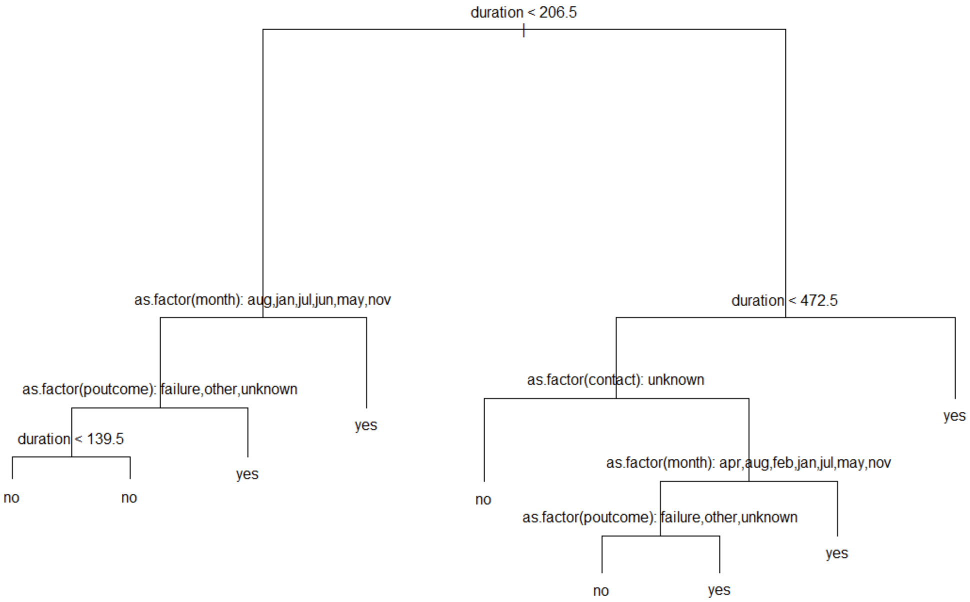
Negative Correlation

- Months: January, May, July, August, November
- Campaign (number of contacts performed during this campaign)

2

CLASSIFICATION TREE

CLASSIFICATION TREE



- 10-fold cross-validation was performed
- Classification tree was pruned to its optimal size 9

3

BAGGING & RANDOM FOREST

BAGGING & RANDOM FOREST

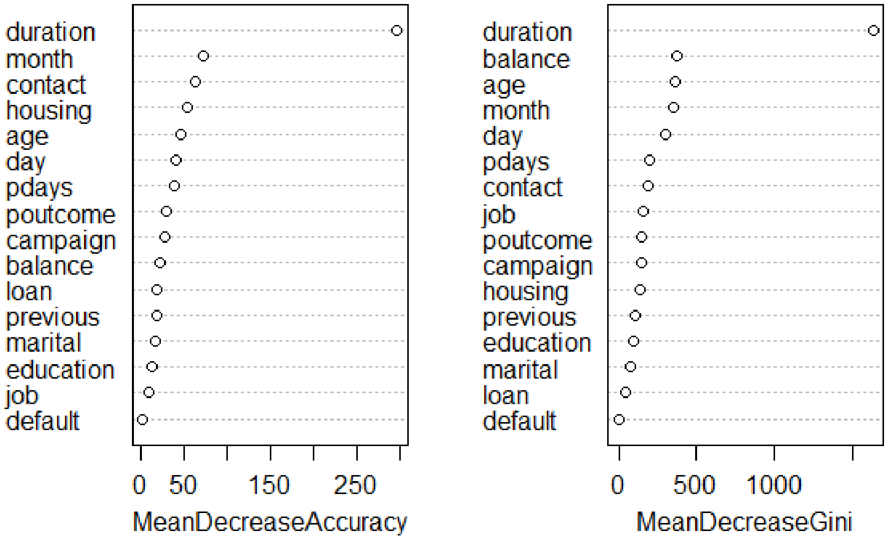
Bagging

- Used all 16 predictors in each tree

Random Forest

- Used only 4 predictors in each tree

Importance of Each Variable in Random Forest



CONCLUSION

MODEL EVALUATION

Model	Recall	Accuracy
Logistic	78.12%	82.40%
Bagging	86.85%	84.37%
Random forest	87.42%	84.24%
Classification tree	78.59%	79.76%

- A confusion matrix was created for each model
- Banks wouldn't want to miss out potential customers who will actually subscribe to a term deposit; therefore, recall was calculated to find the optimal model
- The best model with highest recall: **Random Forest**

INSIGHTS

Who are more likely to subscribe to a term deposit?

- Customers who are students or retired, have at least secondary education level, have a larger amount of balance
- They also have spent more time conversing with the sales representatives during calls

The best time to launch marketing campaigns

- March, June, September, October, December

RECOMMENDATIONS

Do's

- Allocate more budget to target the customer segments that are more likely to subscribe to term deposit
- Launch marketing campaigns during the months of March, June, September, October, and December
- Encourage the sales representatives to increase the duration of their calls with customers, such as inviting them to do short surveys

Don'ts

- Avoid contacting the same customers too many times: the more a sales representative calls a customer, he or she will be more likely to decline to subscribe to term deposit

THANK YOU!

Check out my other projects here: https://www.florayip.com

chingyiycy@gmail.com

https://github.com/flora-yip

http://www.linkedin.com/in/florayip