Project 2 – Term Deposit Prediction Statistics for BA II

This data relates to telemarketing phone calls to sell long-term deposits. Within a campaign, the agents make phone calls to a list of clients to sell the product (outbound) or, if meanwhile the client calls the contact-center for any other reason, he is asked to subscribe the product . The purpose of our project is to be able to predict a successful contact (the client subscribes to the product). The rest variables are potential candidates for examining the variable under study. We are going to employ 3 methods and compare them.

- About the Data:

Input variables:

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')

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')

11 - duration: last contact duration, in seconds (numeric).

- # other attributes:

  12 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

  13 - 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)

  14 - previous: number of contacts performed before this campaign and for this client (numeric)

  15 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

  # social and economic context attributes

  16 - emp.var.rate: employment variation rate - quarterly indicator (numeric)

  17 - cons.price.idx: consumer price index - monthly indicator (numeric)

  18 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)

  19 - euribor3m: euribor 3 month rate - daily indicator (numeric)

  20 - nr.employed: number of employees - quarterly indicator (numeric)

  Output variable (desired target):

  21 - SUBSCRIBED - has the client subscribed a term deposit? (binary: 'yes','no')

The three predictive models implemented are:

- ## Random Forest

  AUC = 60.43 %

  - Accuracy: 0.90
  - Sensitivity: 0.97
  - Specificity: 0.21

- ## K-NN Model

  AUC = 59.08 %

  - Accuracy: 0.90
  - Sensitivity: 0.98
  - Specificity: 0.17

- ## Cart Tree

  AUC = 59.71 %

  - Accuracy: 0.90
  - Sensitivity: 0.99
  - Specificity: 0.12

Conclusion

According to the above measures and especially the AUC resulting from Roc Curve as our dataset is imbalanced, the best model for predicting long-term deposits is Random Forest.