Machine Learning and the Analysis of Consumer Lending

Maisa Aniceto

University of Brasilia
Brazil

May, 2017

Introduction

One of the main problems in Finance involves the prediction of default (Lin, 2012).

- Individuals $\implies$ loans to meet their consumption needs
- Lending $\implies$ risk
- Lending $\implies$ real and immediate loss

Machine Learning

Aim

- To evaluate the performance and compare machine learning techniques with respect to default risk on consumer loans.
- To contribute to a literature, still under development, on the adequacy of machine learning techniques for the phenomena related to the classification of observations, more particularly for credit risk analysis.

Data

Database

- More than 100,000 consumers.
- Line of credit to individuals.
- A pre-approved limit.
- Fixed interest rate.
- 21 variables
  (income, past loans, savings amount, marital status, type of job, number of dependents, etc).
- High level of credit risk.

Techniques

Decision
Trees

C5.0 algorithm

Support
Vector
Machine

e1071 package

Random
Forest

randomForest
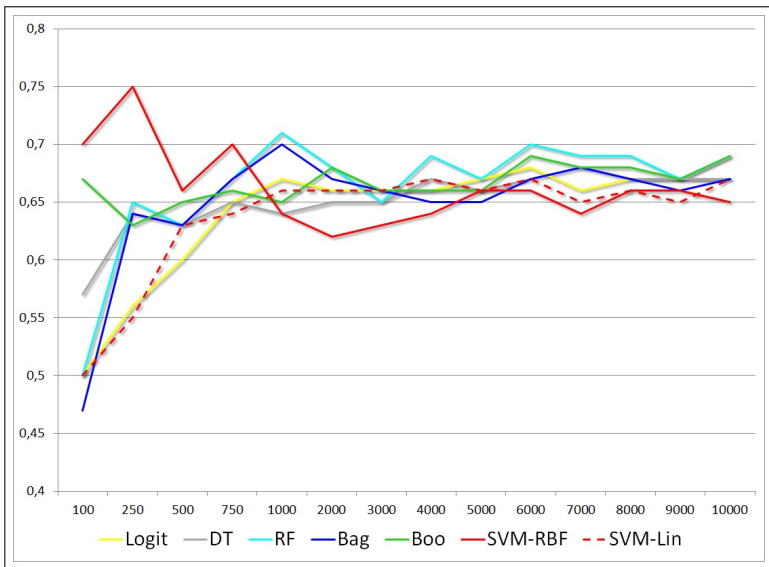algorithm

Bagging

ipred package

AdaBoost

C50 package

Accuracy

Decision Tree	63,37
Random Forest	63,88
AdaBoost	64,41
Bagging	62,47
SVM Linear	63,86
SVM Radial	63,87
Logit	54,44

Accuracy

Conclusion

- Machine learning techniques have a greater precision than the traditional technique.
- Ensemble methods, Random Forest and AdaBoost, have done better than the others.

Next step :

- To use validation measures to compare the results (BRIER score, Kolmogorov-Smirnov statistic, CIER measure, among others).
- To analyze different costs of misclassification.

Thereby our goal is to help financial institutions not to face so much loss on loans.

Thank you !

maisa.c.aniceto@gmail.com
University of Brasilia
Brazil