



Maastricht University

School of Business and Economics

SHOULDICE CASE REPORT

Bernardo Oliveira (I6115981)
Lilian Do Khac (I6114941)
Alex De Vidal De St. Germain (I6115568)

Supervisor: Drs. Marianne Peeters

Issue due date: 10. December 2015

Contents

List of Figures	iii
1 Management Summary	1
2 Introduction	2
2.1 Credit Risk General Overview	2
2.2 Current Status at German Credit	2
3 Analysis Summary Outlook	3
4 Appendix	4
4.1 Review the predictor variables and guess what their role in a credit decision might be. Are there any surprises in the data?	5
4.2 Devide the data into training and validation partitions, and develop classification models using the following data mining techniques: logistic regression, classification trees, and k-nearest neighbor	5
4.3 Chooese one model from each technique and report the confusion matrix and the cost/gain matrix for the validation data. Which technique has the most net profit?	5
4.4 Let us try and improve our performance. Rather than accept the initial classification of all applicants credit status, use the predicted probability of sucess in logistic regression ..where success means 1.. as a basis for selecting the best credit risk first, followed by poorer risk applicants. . . .	5
4.4.1 Sort the validation on predicted probability of success.	5
4.4.2 For each case, calculate the net profit of extending credit.	5
4.4.3 Add another column for cumulative net profit.	5
4.4.4 How far into the validation data do you go to get the maximum net profit?(Often, this is specified as a percentile or rounded to deciles.)	5

Bibliography

6

List of Figures

1 Management Summary

- Point 1
- Point 2
- Point 3
- Point 4

2 Introduction

2.1 Credit Risk General Overview

2.2 Current Status at German Credit

3 Analysis Summary Outlook

4 Appendix

- 4.1 Review the predictor variables and guess what their role in a credit decision might be. Are there any surprises in the data?
- 4.2 Devide the data into training and validation partitions, and develop classification models using the following data mining techniques: logistic regression, classification trees, and k-nearest neighbor
- 4.3 Chooese one model from each technique and report the confusion matrix and the cost/gain matrix for the validation data. Which technique has the most net profit?
- 4.4 Let us try and improve our performance. Rather than accept the initial classification of all applicants credit status, use the predicted probability of sucess in logistic regression ..where success means 1.. as a basis for selecting the best credit risk first, followed by poorer risk applicants.
 - 4.4.1 Sort the validation on predicted probability of success.
 - 4.4.2 For each case, calculate the net profit of extending credit.
 - 4.4.3 Add another column for cumulative net profit.
 - 4.4.4 How far into the validation data do you go to get the maximum net profit?(Often, this is specified as a percentile or rounded to deciles.)
 - 4.4.5 If this logistic regression model is scored to future applicants, what 'probability of success' cutoff should be used in extending credit?

Bibliography

Deloitte. (2015). *2015 global health care outlook* (Report).

Ragsdale, C. T. (2003). *Spreadsheet modeling and decision analysis* [Book]. South-Western.

The World Bank. (n.d.). (Vol. 2015) [Web Page]. The World Bank. Retrieved from <http://data.worldbank.org/data-catalog/world-development-indicators>