

Developing a Tool to Manage the Credit Risk using Data Mining

Shamli Rajapakshe

139175M

Dissertation submitted to the Faculty of Information Technology,
University of Moratuwa, Sri Lanka
for the partial fulfillment of the requirements of the
Master of Science in Information Technology.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
March 2016
www.lib.mrt.ac.lk

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Shamli Rajapakshe

Date:

Signature of Student

Supervised by

Mr. Saminda Premaratne  University of Moratuwa, Sri Lanka.
Senior Lecturer www.lib.mrt.ac.lk
Electronic Theses & Dissertations

Faculty of Information Technology

University of Moratuwa

Date:

Signature of Supervisor

Acknowledgement

First and foremost, I would like to express my sincere gratitude towards my supervisor, Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his proper guidance, supervision, advices and sparing valuable time throughout the one year research project.

I would like to express my profound sense of gratitude and respect to all those who helped me in various ways throughout the duration of this research project.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Abstract

Ceasing vehicles are affecting to the credit liquidity of the leasing company. This research has been conducted to develop a tool to manage credit risk in leasing companies using data mining. This tool will predict the ability of recoverability of the loan and determine the most suitable plan for the customer. It is hypothesis that, using data mining technology, the credit risk of leasing companies can be managed. Past dataset from the leasing company has been used to create the data mining model. When a customer comes to lease a vehicle, decision maker will get the information from the customer and enter to the system as inputs then the system will predict the tendency of recoverability of the loan and will give the suitable plans for the customer after evaluating with the previously generated model. This system generated details will support the decision maker to take his decision. The overall design includes frontend software and it is connected to the WEKA API which issued under the GNU General Public License.

The data model that is used in this tool to manage credit risk in leasing companies has been tested by considering a data collected from the medium scale leasing company in Sri Lanka.



University of Moratuwa, Sri Lanka.
Electronic Theses & Dissertations
www.lib.mrt.ac.lk

Table of Content

| | |
|---|-----|
| Declaration..... | i |
| Acknowledgement | ii |
| Abstract..... | iii |
| Table of Content | iv |
| List of Figures | vii |
| List of Tables | vii |
| Chapter 1..... | 1 |
| Introduction..... | 1 |
| 1.1 Prolegomena..... | 1 |
| 1.2 Background and Motivation..... | 2 |
| 1.3 Problem Definition..... | 2 |
| 1.4 Objectives..... | 2 |
| 1.5 Summary..... | 2 |
| Chapter 2..... | 3 |
| Literature Review..... | 3 |
| 2.1 Introduction..... | 3 |
| 2.2 Credit Risk..... | 3 |
| 2.3 Credit Risk Management..... | 3 |
| 2.4 Comparative Study of Data Mining Model for Credit Card Application Scoring in Bank | 4 |
| 2.5 Feature Selection in Credit Scoring Model for Credit Card Applicants | 4 |
| 2.6 Predict the usage of laptops among students in rural areas using WEKA tool.... | 4 |
| 2.7 Data Mining: A prediction Technique for the workers in the PR Department of Orissa..... | 5 |
| 2.8 comparative study of classification algorithms for credit card approval using Waikato Environment of Knowledge Analysis (WEKA)..... | 5 |
| 2.9 Summary | 6 |
| Chapter 3..... | 7 |
| Technology Adopted..... | 7 |
| 3.1 Introduction | 7 |
| 3.2 Data Mining..... | 7 |

| | |
|--|----|
| 3.3 Java..... | 7 |
| 3.4 WEKA / WEKA API | 8 |
| 3.5 SPSS | 8 |
| 3.6 Summary | 8 |
| Chapter 04..... | 9 |
| Data mining approach to manage credit risk | 9 |
| 4.1 Introduction | 9 |
| 4.2 Hypothesis..... | 9 |
| 4.3 Inputs..... | 9 |
| 4.4 Outputs | 9 |
| 4.5 Process and features | 9 |
| 4.7 Users..... | 10 |
| 4.8 Summary | 10 |
| Chapter 5..... | 11 |
| Design of the Prediction Tool..... | 11 |
| 5.1 Introduction | 11 |
| 5.2 Data Input Procedure..... | 11 |
| 5.3 Risk Assessment Process..... | 12 |
| 5.5 Top level Design | 13 |
| 5.5.1 Gather data in the relevant field and Pre preparation | 13 |
| 5.5.2 Load data to the WEKA | 13 |
| 5.5.3 Model building by using suitable classification technology | 13 |
| 5.5.4 Requirement gathering for the application | 13 |
| 5.5.5 Design the application | 14 |
| 5.5.6 Implement a front-end application by using JAVA..... | 15 |
| 5.5.7 Integration..... | 15 |
| 5.5.8 Testing | 15 |
| 5.6 User Interfaces..... | 16 |
| 5.7 Summary | 18 |
| Chapter 6..... | 19 |
| Implementation of the Tool | 19 |
| 6.1 Introduction | 19 |
| 6.2 Implementation of CRMT (Credit Risk Management Tool) | 19 |

| | |
|--|-----|
| 6.2.1 Pre preparation..... | 19 |
| 6.2.2 Best Model Selection..... | 21 |
| 6.2.3 Risk Assessment Process..... | 23 |
| 6.2.4. Plan Prediction Process | 25 |
| 6.2.5 Requirement gathering for the application | 26 |
| 6.2.7 Implement a front-end application by using JAVA..... | 26 |
| 6.2.8 Integration..... | 26 |
| 6.2.9 Testing | 26 |
| 6.2.10 Building an Evaluation Model..... | 26 |
| 6.3 Summary | 26 |
| Chapter 7..... | 27 |
| Evaluation | 27 |
| 7.1 Introduction | 27 |
| 7.2 Data Model Testing..... | 27 |
| 7.3 Test the system for randomly selected values | 277 |
| 7.4 Summary | 30 |
| Chapter 8..... | 31 |
| Conclusion and further work | 31 |
| 8.1 Introduction | 31 |
| 8.2 Conclusion of CRMT (Credit Risk Management Tool)..... | 31 |
| 8.3 Limitations of CRMT (Credit Risk Management Tool) | 31 |
| 8.4 Further work..... | 31 |
| 8.5 Summary | 31 |
| References..... | 32 |
| Appendix A: Categorization | 34 |
| Appendix B: Variable Selection Procedure (Technology: Chi – Squared) | 36 |
| Appendix C: Best Model Selection..... | 44 |



List of Figures

| | |
|---|-----|
| Figure 5.1: Data Input Procedure..... | 11 |
| Figure 5.2: Risk Assessment Procedure..... | 12 |
| Figure 5.3: Plan Prediction Procedure | 12 |
| Figure 5.4: Top level design | 13 |
| Figure 5.5: Use case..... | 14 |
| Figure 5.6: Frontend application with other connected components..... | 15 |
| Figure 5.7: Data Input Interface..... | 16 |
| Figure 5.8: Risk Analysis Interface | 16 |
| Figure 5.9: Model Prediction | 17 |
| Figure 5.10: Requested amount Prediction..... | 17 |
| Table 6.1: Variable Selection Summary table | 20 |
| Figure 6.1: Accuracy levels of algorithms..... | 23 |
| Figure 6.2: Error rates of algorithms..... | 23 |
| Figure 6.3: Building the Classifier..... | 23 |
| Figure 6.4: Using Classifier for Classification | 24 |
| Figure 7.1: Test the selected model with testing data set | 217 |
| Figure 7.2: Model evaluation with the test data set | 218 |
| Figure 7.3: Customer Details..... | 28 |
| Figure 7.4: Most Suitable vehicle model for particular customer..... | 29 |
| Figure 7.5: Most suitable loan ranges for particular user | 29 |

List of Tables

| | |
|---|----|
| Table 6.1: Variable Selection Summary | 21 |
| Table 6.2: Algorithm Accuracy Summary..... | 21 |