PROJECT REPORT

LOAN APPROVAL PREDICTION USING LOGISTIC REGRESSION AND EXTREME GRADIENT BOOSTING ALGORITHMS

YESICA SUGIARTO, TING
18.K1.0024

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir	:	Loan Approval Prediction Using Logistic Regression and Extreme GradientBoosting Algorithms
Diajukan oleh	:	Yesica Sugiarto, Ting
NIM	:	18.K1.0024
Tanggal disetujui	:	23 Mei 2022Telah setujui oleh
Pembimbing	:	Hironimus Leong S.Kom., M.Kom.
Penguji 1	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2	:	Hironimus Leong S.Kom., M.Kom.
Penguji 3	:	Rosita Herawati S.T., M.I.T.
Penguji 4	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 5	:	Yulianto Tejo Putranto S.T., M.T.
Penguji 6	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Ketua Program Studi	:	Rosita Herawati S.T., M.I.T.
Dekan	:	Dr. Bernardinus Harnadi S.T., M.T.

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

sintak.unika.ac.id/skripsi/verifikasi/?id=18.K1.0024

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : YESICA SUGIARTO, TING

ID : 18.K1.0024

declare that this work, titled ”LOAN APPROVAL PREDICTION USING LOGISTIC REGRESSION AND EXTREME GRADIENT BOOSTING ALGORITHMS”, and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, Juni, 01, 2022

Yesica Sugiarto

18.K1.0024

DECLARATION OF AUTHORSHIP

I, the undersigned :

Name : YESICA SUGIARTO, TING

Study Program : Informatics Engineering

Faculty : Computer Science

Type of Work : Thesis

Approve to grant Soegijapranata Catholic University Semarang Nom-exclusive Royality-Free Rights for the scientific works entitled "LOAN APPROVAL PREDICTION USING LOGISTIC REGRESSION AND EXTREME GRADIENT BOOSTING" along with existing tools. With these rights, database form, maintain, and publish this final project as long it includes my name as the creator or writer and as the copyright owner.

This statement I made truthfully.

Semarang, Juni, 01, 2022

Yesica Sugiarto

18.K1.0024

ACKNOWLEDGMENT

First of all I would like to thank God Almighty because I had the opportunity to have the experience of working on this project entitled “Loan Approval Prediction Using Logistic Regression and Extreme Gradient Boosting Algorithms”

During the work on this project, I am very grateful and appreciate the support and cooperation from:

1. Supervisor Mr. Marlon who has guided me from the beginning of the process of working on this project with great patience and guidance in the technique of preparing this project.
2. family who support me wholeheartedly and wish the best for the smooth running of this project.
3. Friends who support me by providing motivation, criticism, suggestions and support in completing this project.

With full awareness I realize that this thesis is less than perfect, therefore I always hope for criticism and suggestions from readers which will certainly be very helpful for the addition of my knowledge and progress, and hopefully useful for readers.

Semarang, Juni, 01, 2022

Yesica Sugiarto

18.K1.0024

ABSTRACT

In the savings and loan business or banking, it is very important to determine whether the borrower is able to pay or cannot repay the loan. This is very necessary in order to avoid losses.

In this process, predictions to classify categories of borrowers including those who are able to pay or not, require a computer approach using machine learning with some of the algorithm methods, namely logistic regression and extreme gradient boosting.

Both models can predict the outcome of the decision quite well with the acquisition of logistic regression 69% and extreme gradient boosting 82%. Which when compared to extreme gradient boosting has better results.

Keyword: loan, logistic regression, extreme gradient boosting

TABLE OF CONTENTS

HALAMAN PENGESAHAN ii

DECLARATION OF AUTHORSHIP iii

DECLARATION OF AUTHORSHIP iv

ACKNOWLEDGMENT v

ABSTRACT vi

TABLE OF CONTENTS vii

LIST OF FIGURE ix

LIST OF TABLE x

CHAPTER 1 INTRODUCTION 1

1.1. Background 1

1.2. Problem Formulation 2

1.3. Scope 2

1.4. Objective 2

CHAPTER 2 LITERATURE STUDY 3

CHAPTER 3 RESEARCH METHODOLOGY 6

3.1. Data Collection 6

3.2. Algorithms 6

3.3. Design 6

3.4. Coding 7

3.5. Analysis 7

CHAPTER 4 ANALYSIS AND DESIGN 8

4.1. Data Collection 9

4.2. EDA (Exploratory Data Analysis) 11

4.3. Data Cleaning 15

4.4. Encoding 18
4.5. Features Selection 18
4.6. Splitting Dataset 19
4.7. Building model using Logistic Regression 20
4.8. Building model using Extreme Gradient Boosting 25
CHAPTER 5 IMPLEMENTATION AND RESULTS 31
5.1. Implementation 31
5.2. Result 49
5.3. Analysis 53
CHAPTER 6 56
CONCLUSION 56
REFERENCES 57
APPENDIX 58

LIST OF FIGURE

Figure 4.1 Workflow 8

Figure 4.6.1 Splitting Dataset 19

LIST OF TABLE

Table 4.1.1 Dataset 9

Table 4.1.2 Dataset Information 10

Table 4.2.1 EDA 11

Table 4.3.1 Sum of Missing Value 15

Table 4.3.2 Before Data Cleaning 16

Table 4.3.3 After Data Cleaning 17

Table 4.4.1 Encoding 18

Table 4.5.1 Features Selection 18

Table 4.7.1 Initializes Weight and Bias 20

Table 5.2.1 Logistic Regression Result 60:40 49

Table 5.2.2 Logistic Regression Result 70:30 50

Table 5.2.3 Extreme Gradient Boosting Result 60:40 51

Table 5.2.4 Extreme Gradient Boosting Result 70:30 52

Table 5.3.1 Extreme Gradient Boosting Sample Data 54