PROJECT REPORT

Comparing Random Forest Algorithm and Support Vector Machine for Predicting the Level of Satisfaction with Flights

WAHONO SOESANTIO
18.K1.0045

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir:	:	Comparing Random Forest Algorithm and Support Vector Machine for Predicting the Level of Satisfaction with Flights
Diajukan oleh	:	Wahono Soesantio
NIM	:	18.K1.0045
Tanggal disetujui	:	23 Mei 2022
Telah 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	:	Yulianto Tejo Putranto S.T., M.T.
Penguji 5	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 6	:	Y.b. Dwi Setianto S.T., M.Cs.
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.K1.0045

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : Wahono Soesantio

ID : 18.K1.0045

Declare that this work, titled "Comparing Random Forest Algorithm and Support Vector Machine in predicting the level of satisfaction with flights", 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, 23, May, 2022

Wahono Soesantio

18.K1.0045

SCIENTIFIC PUBLICATIONS STATEMENT PAGE

FOR ACADEMIC INTEREST

I, the undersigned:

Name : WAHONO SOESANTIO

Nim : 18.K1.0045

Progdi : Informatics Engineering

Faculty : Computer Science

Agree to give Soegijapranata Catholic University Semarang the right to non-exclusive royalty free for a scientific work entitled "Comparing Random Forest Algorithm and Support Vector Machine for Predicting the Level of Satisfaction with Flights" along with existing devices. With this Non-exclusive Royalty Free Right, Soegijapranata Catholic University has the right to save, transfer / format media, manage in the form of a database, maintain, and publish this final project as long as I still include my name as the author / creator and as the copyright owner.

This statement I made in truth

Semarang, 23, May, 2022

Wahono Soesantio

18.K1.0045

ACKNOWLEDGMENT

First of all, I thank God, Jesus Christ, who was has allowed me to write a thesis proposal so far, because of his grace I can live healthily until now even in the Covid-19 pandemic situation

I have received a myriad of support, advice, and assistance throughout this document writing. I would like to thank my supervisors Hironimus Leong S. Kom., M. Kom. for formulating this topic. I would also like to thank my friend for guiding me with advice to finish this document.

I would like to thank my family and friends for giving me ceaseless love, support, and advice throughout my study at Soegijapranata Catholic University. You gave me a great escape to rest my mind from my thesis.

For all 2018 Informatics Engineering friends who have accompanied and supported during the lecture period until the writing of this thesis was completed.

Finally, I would like to thank everybody important to the successful realization of this undergraduate thesis. This undergraduate thesis is far from perfect, but it is expected that it will be useful not only for the researcher but also for the readers. For this reason, Constructive thoughtful suggestions and critics are welcomed.

ABSTRACT

This study aims to find out which method is better and more effective so that it can be seen the difference between the two algorithms in predicting the level of passenger satisfaction with public services on the plane which will be displayed using a graph.

For the completion process, what I did was try each algorithm, the first thing I did was to try the random forest algorithm then when I tried it, I kept looking for the TP, TN, FP, FN values to generate prediction scores, accuracy, recall, f-1 scores. then compared with the support vector machine algorithm with more or less the same steps by looking for a true positive, true negative, false negative, false positive then looking for predictive value, accuracy, recall, the f-1 score then displayed on a graph to see which one is better. Both algorithms using a cross-validation split for split data which aims to divide the data into training and testing data and make it easier to calculate accuracy. For the use of split data used with a comparison of 70% training data and 30% testing data from the whole data.

The final result can be seen from the graph of each algorithm that has shown the prediction value, recall, precision, f-1 score and from there it can be concluded which method is better and the differences from each existing algorithm. To get the accuracy results obtained from the determination of the confusion matrix in the form of TP, TN, FP, and FN.

TABLE OF CONTENTS

COVER i

APPROVAL AND RATIFICATION PAGE ii

DECLARATION OF AUTHORSHIP iii

ACKNOWLEDGMENT iv

ABSTRACT v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

CHAPTER 1 1

INTRODUCTION 1

1.1. Background 1

1.2. Problem Formulation 3

1.3. Scope 3

1.4. Objective 3

CHAPTER 2 LITERATURE STUDY 4

CHAPTER 3 RESEARCH METHODOLOGY 8

1. Identification and Literature Study 8
2. Data and Variables 9
3. Perform data processing 9
4. Making a data classification with SVM (Support Vector Machine) 10
5. Making a data classification with Random Forest Classification 11

Algorithm. 11

6. Result 11
7. Report Writing 11

CHAPTER 4 12

ANALYSIS AND DESIGN 12

4.1. Collecting Data 12

4.2.1 SVM Algorithm Steps 15

1. Split Data 15
2. Implementation to Formula 17

3. Finding the Value of Each Attribute 18
4. Conclusion Using Data Testing 21
4.2.2 Random Forest Algorithm Steps 22
1. Split Dataset based on Attribute Value 22
2. Calculation of Satisfaction Amount Based on Age Attribute 23
3. Performing Gini Index Calculations 24
4. Performing a Gini Split Calculation 25
5. Create a Decision Tree 27
6. Random sampling 28
4.3 Accuracy Measurement Support Vector Machine and Random Forest 28
CHAPTER 5 IMPLEMENTATION AND RESULTS 35
5.1 Implementation 35
5.2 Results 41
CHAPTER 6 CONCLUSION 46
REFERENCES 47
APPENDIX 50

LIST OF FIGURES

Figure 1: Decision Tree 27
Figure 2: Confusion Matrix 29
Figure 3:Flowchart Random Forest Algorithm 33

LIST OF TABLES

Table 1: Data Sample Table 13
Table 2: Formula 17
Table 3: Conclusion Table Using Data Testing 21
Table 4: Split Dataset Based on Attribute 22
Table 5: Calculation of the Amount of Satisfaction Based on the Value of Arrivals 23
Table 6: Gini Index Calculations 24
Table 7: Gini Split Calculation 25
Table 8: Gini Splitting Index 26
Table 9: Age ≤ 46 27
Table 10: Age > 46 27
Table 11: Sample Confusion Matrix 29