



PROJECT REPORT
COMPARISON BETWEEN STOCHASTIC GRADIENT
DESCENT AND SUPPORT VECTOR MACHINE ON
ECOMMERCE DATA USING RANDOMIZEDSEARCHCV

KEVIN VICTOR ARAUNA
18.K1.0011

Faculty of Computer Science
Soegijapranata Catholic University
2021

HALAMAN PENGESAHAN



Judul Tugas Akhir: : Comparison Between Stochastic Gradient Descent and Support Vector
Machine on Ecommerce Data Using RandomizedSearchCV

Diajukan oleh : KEVIN VICTOR ARAUNA

NIM : 18.K1.0011

Tanggal disetujui : 30 Januari 2023

Telah setujui oleh

Pembimbing : Rosita Herawati S.T., M.I.T.

Penguji 1 : Yonathan Purbo Santosa S.Kom., M.Sc

Penguji 2 : Hironimus Leong S.Kom., M.Kom.

Penguji 3 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Penguji 4 : Rosita Herawati S.T., M.I.T.

Penguji 5 : Y.b. Dwi Setianto S.T., M.Cs.

Penguji 6 : Yulianto Tejo Putranto S.T., M.T.

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.K1.0011

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : Kevin Victor Arauna

ID : 18.K1.0011

declare that this work, titled "Comparison Between Stochastic Gradient Descent and Support Vector Machine on Ecommerce Data Using RandomizedSearchCV", 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, January, 30, 2023



Kevin Victor Arauna

18.K1.0011

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan dibawah ini:

Nama : Kevin Victor Arauna
Program Studi : Teknik Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneklusif atas karya ilmiah yang berjudul “Comparison Between Stochastic Gradient Descent and Support Vector Machine on Ecommerce Data Using RandomizedSearchCV” beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneklusif ini Universitas Katolik Soegijapranata berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir ini selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 30 Januari 2023

Yang menyatakan



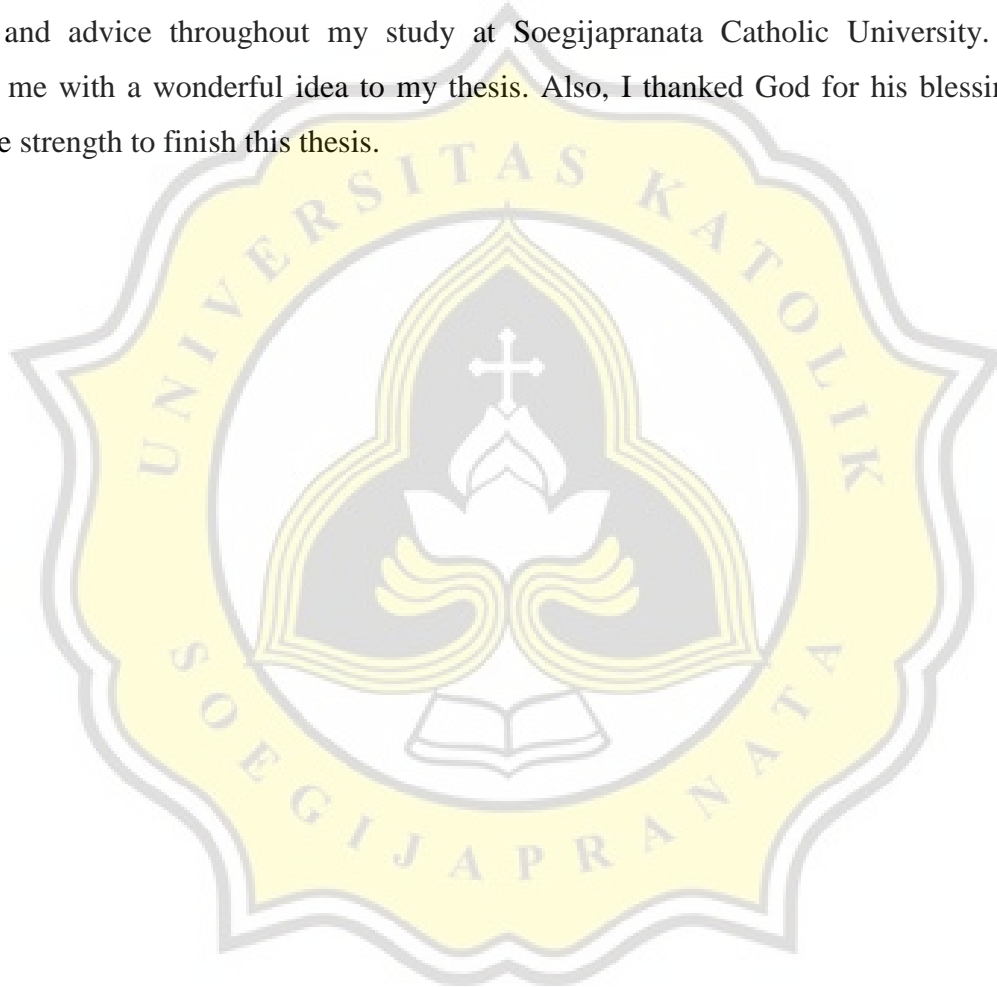
Kevin Victor Arauna

18.K1.0011

ACKNOWLEDGMENT

Throughout the writing of this document, I received a plethora of encouragement, advice, and assistance. I'd like to thank my supervisor, Rosita Herawati S.T., M.I.T. for coming up with the idea and guiding me through the process of creating this topic. I'd also like to thank my friend for providing guidance and advice that enabled me to complete this document.

I would like to thank my family and friends for their patience, ideas, ceaseless love, support, and advice throughout my study at Soegijapranata Catholic University. Everyone provided me with a wonderful idea to my thesis. Also, I thanked God for his blessing and for giving me strength to finish this thesis.



ABSTRACT

In recent years, the use of e-commerce or online shops has increased considerably. Various online stores have sprung up on the internet, both small and large-scale. This has a very important effect on the effective use of time and the level of sales figures. With so many e-commerce emerging and even with almost the same type or product, it will be quite difficult to reach potential buyers for them. The intentions of consumers will also be different for each online store they visit. So the problem with the number of e-commerce that has sprung up is that shop owners must have data that is effective enough to see the intentions of consumers in their stores whether to buy their products or not, regardless of the marketing strategy they apply.

Every e-commerce site has visitor history data especially nowadays that everything is online, it will make it easier for store owners to track consumers who visit their online store. This data includes all information regarding the traces of consumers who visit online stores, both new and repeated. With this data, processing will be carried out to make predictions about consumers' shopping intentions. The data contains attributes that will later be used for processing with the algorithm to be compared. Several algorithms will be used to process this data in order to find which one is the best in predicting consumers' intention to make a purchase or not. The comparison of algorithms will be carried out using the value of the Precision, Recall, F1-Score, Accuracy. Also some values that are considered to support the results of the comparison.

The algorithm are Stochastic Gradient Descent (SGD) and Support Vector Machine (SVM). Comparison of the two algorithms is supported using RandomizedSearchCV as the hyperparameter tuning. The result of which algorithm is better with hyperparameter tuning is SGD. SGD obtained an accuracy value of 0.89 with 60% test data and 40% training data.

Keyword: Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Online Store, Prediction, RandomizedSearchCV

TABLE OF CONTENTS

COVER	i
HALAMAN PENGESAHAN	ii
DECLARATION OF AUTHORSHIP	iii
HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
ACKNOWLEDGMENT	v
TABLE OF CONTENTS	vii
LIST OF FIGURE	ix
LIST OF TABLE	xi
CHAPTER 1 INTRODUCTION	12
1.1. Background.....	12
1.2. Problem Formulation	13
1.3. Scope.....	13
1.4. Objective	13
CHAPTER 2 LITERATURE STUDY	14
CHAPTER 3 RESEARCH METHODOLOGY	19
3.1. Dataset.....	19
3.2. Comparison.....	20
CHAPTER 4 ANALYSIS AND DESIGN	23
4.1. Analysis.....	23
4.2. Algorithm.....	25
4.2.1. Stochastic Gradient Descent (SGD).....	26
4.2.2. Support Vector Machine (SVM).....	28
4.2.3. Hyperparameter Tuning with RandomizedSearchCV	32
4.2.4. Data Balancing with Adaptive Synthetic (ADASYN)	32

4.2.5. Dimensionality Reduction with Principal Component Analysis (PCA) 33

CHAPTER 5 IMPLEMENTATION AND RESULTS 35

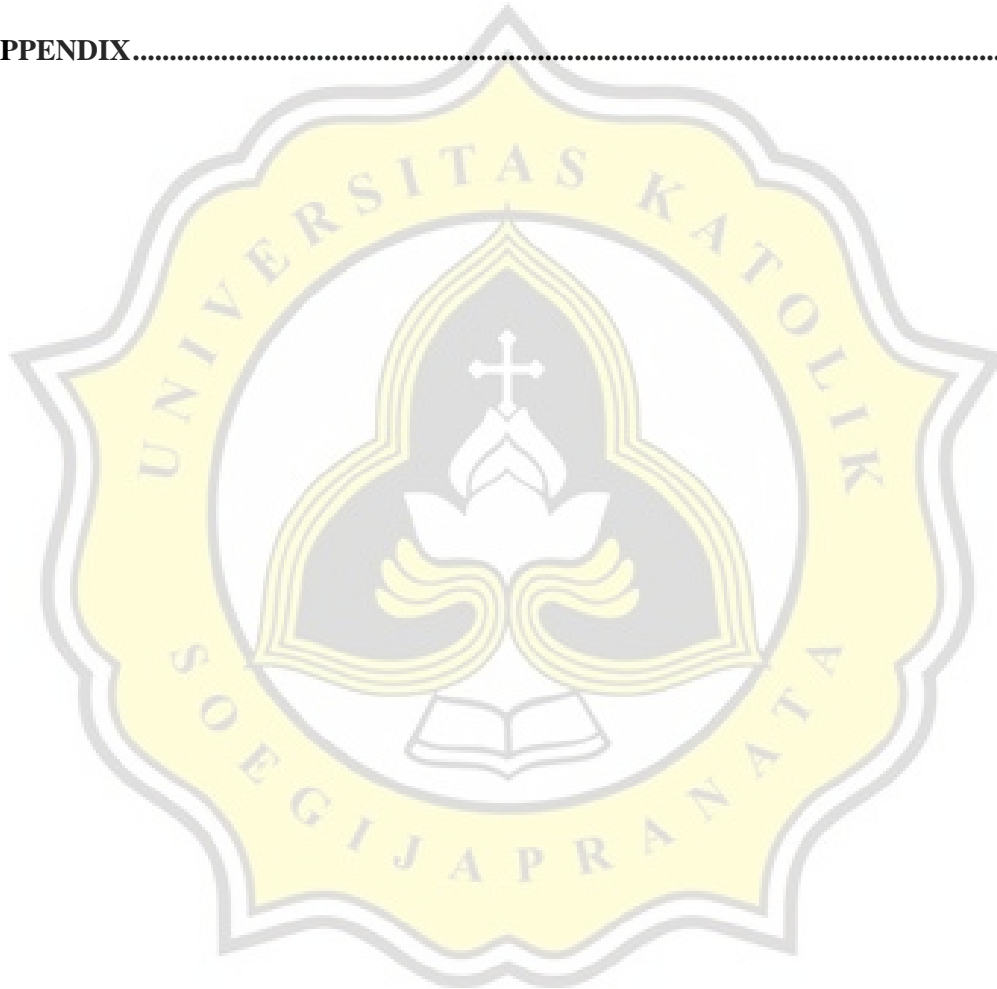
5.1. Implementation 35

5.2. Results..... 40

CHAPTER 6 CONCLUSION 46

REFERENCES..... 47

APPENDIX..... a



LIST OF FIGURE

Figure 3.1 Flowchart	19
Figure 3.1 SVM Process	20
Figure 3.2 SGD Process	21
Figure 4.1 SGD Classifier	27
Figure 4.2 Learning Rate for Classification	27
Figure 4.3 SGD Example	27
Figure 4.1 SVM Example	28
Figure 4.2 RBF Kernel	29
Figure 4.3 C Value	30
Figure 4.4 Gamma Value	31
Figure 4.5 SVM Hyperplane I	31
Figure 4.6 SVM Hyperplane II	31
Figure 4.7 SVM Classifier	31
Figure 5.1 SVM Result	41
Figure 5.2 SGD Result	42
Figure 5.3 Best Parameter Comparison Between SVM and SGD (Label = 0)	43
Figure 5.4 Best Parameter Comparison Between SVM and SGD (Label = 1)	43
Figure 5.5 Accuracy Comparison Between SVM and SGD	44
Figure 5.6 RandomizedSearchCV Comparison on SVM	44
Figure 5.7 RandomizedSearchCV Comparison on SGD	45
Figure 6.1 RandomizedSearchCV Result in Ranged Parameters Value for SVM	h
Figure 6.2 RandomizedSearchCV Result in Specific Parameters Value for SVM	h
Figure 6.3 GridSearchCV Result in Specific Parameters Value for SVM	h

Figure 6.4 GridSearchCV Result in Specific Parameters Value for SVM (failed).....	i
Figure 6.5 GridSearchCV Result in Specific Parameters Value for SGD (failed)	i
Figure 6.6 GridSearchCV Result in Ranged Parameters Value for SGD (failed)	i
Figure 6.7 RandomizedSearchCV Result in Specific Parameters Value for SGD	j
Figure 6.8 RandomizedSearchCV Result in Ranged Parameters Value for SGD	j
Figure 6.9 Classification Report for SVM in Hyperparameter Tuning with RandomizedSearchCV	k
Figure 6.10 Best Parameters Result for SVM in RandomizedSearchCV	k
Figure 6.11 Classification Report for SGD in Hyperparameter Tuning with RandomizedSearchCV	l
Figure 6.12 Best Parameters Result for SGD in RandomizedSearchCV	l
Figure 6.13 Before Data Balancing with ADASYN on SVM	m
Figure 6.14 After Data Balancing with ADASYN on SVM.....	m
Figure 6.15 Before Data Balancing with ADASYN on SGD.....	n
Figure 6.16 After Data Balancing with ADASYN on SGD	n

LIST OF TABLE

Table 4.1. Dataset Columns Details 23

