PROJECT REPORT

CHI - SQUARE AND INFORMATION GAIN FEATURE SELECTION FOR HOTEL REVIEW SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE

NATHANAEL KARUNIA
18.K1.0027

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir:	:	CHI - SQUARE AND INFORMATION GAIN FEATURE SELECTION FOR HOTEL REVIEW SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE
Diajukan oleh	:	Nathanael Karunia Wibowo
NIM	:	18.K1.0027
Tanggal disetujui	:	24 Mei 2022
Telah setujui oleh		
Pembimbing	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 1	:	Yonathan Purbo Santosa S.Kom., M.Sc
Penguji 2	:	Yulianto Tejo Putranto S.T., M.T.
Penguji 3	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
Penguji 4	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 5	:	Rosita Herawati S.T., M.I.T.
Penguji 6	:	Hironimus Leong S.Kom., M.Kom.
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.K1.0027

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : Nathanel Karunia

ID : 18.K1.0027

declare that this work, titled "CHI – SQUARE AND INFORMATION GAIN FEATURE SELECTION FOR HOTEL REVIEW SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE ", 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, April, 29, 2022

10000 METERAI TEMPEL 8A496AJX870885436

NATHANAEL KARUNIA
18.K1.0027

STATEMENT PAGE OF SCIENTIFIC PUBLICATIONS FOR ACADEMIC INTEREST

I, the undersigned:

Name : NATHANAEL KARUNIA

Study Program : Informatics Engineering

Faculty : Computer Science

Type of Work : Thesis

Approve to grant Soegijapranata Catholic University Semarang Non-exclusive Royalty-Free Rights for the scientific works entitled “CHI – SQUARE AND INFORMATION GAIN FEATURE SELECTION FOR HOTEL REVIEW SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE” along with existing tools (if needed). With these rights, database form, maintain, and publish this final project as long as it includes my name as the writer/creator and as the copyright owner.

This statement was made truthfully.

Semarang, Juni, 4, 2022

NATHANAEL KARUNIA

18.K1.0027

ACKNOWLEDGMENT

All praise and gratitude to God Almighty who has given His grace and grace to the author, so that the author can complete this Final Project Report. The writing of the Final Project Report with the title "Chi Square And Information Gain Feature Selection For Hotel Review Sentiment Analysis Using Support Vector Machine" is intended to achieve a Bachelor's Degree in Computer Science at the Informatics Engineering Study Program, Faculty of Computer Science, Soegijapranata Catholic University, Semarang.

The author realizes that in the preparation of this final report, not only from his efforts but from the guidance of various parties. Therefore, the authors would like to thank all parties who helped in the process of writing this Final Project.

1. My Parents and my family who give the best support during lecture.
2. My Mentor Mr. Yonathan that very kindness and helpful on my thesis.
3. All Lectures of Soegijapranata Catholic University IT
4. My friends who push me to do this thesis.

The author realizes that there may still be shortcomings in this Final Project Report. Therefore, criticism and suggestions from readers are very useful for the author. Hopefully, this report can be useful for all those who read it.

ABSTRACT

In the current era, it has become a trend for people to order tickets online through online booking sites and applications, both in terms of transportation such as planes, vacations such as tours, and also lodging such as hotels. To get a good hotel, you need a review from people who have booked it. With the reviews written by visitors to the site or mobile application, they will then be analyzed so that an output can be produced that can be useful. One of the analytical models that can be done is sentiment analysis. The purpose of this study is to find the best method in analyzing sentiment based on the preprocessing of the data and hopefully it can produce knowledge in the form of sentiment analysis classification methods in order to determine a good method devoted to the data preprocessing section.

The algorithm used to make this sentiment classification analysis is the Support Vector Machine using 3 feature selection methods, namely not using the selection feature, using the chi square selection feature, and using the information gain selection feature. The process consists of five steps in this study, which include several activities. namely data collection, preprocessing, feature extraction, feature selection, classification, and calculating accuracy. In the process of calculating accuracy, I used the Confusion Matrix method to find the best method of the three based on the accuracy results obtained.

The results of the 3 uses of the feature selection method that were carried out were using the chi square feature selection method, the highest results were obtained, namely with an average accuracy of 86.68% which was followed by the use of the information gain selection feature which obtained an average accuracy of 85.78% and the last one was followed by the method not using the selection feature which got an average accuracy of 85.24%. From the results of the three methods, it can be concluded that the use of the chi square feature selection method in the case of sentiment analysis on hotel reviews is the best compared to the other two.

Keyword: feature selection, sentiment analysis, hotel review

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE