PROJECT REPORT

SENTIMENT ANALYSIS ON MOVIE REVIEW USING RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHM

LADY VIONA SUGIANTO
18.K1.0023

Faculty of Computer Science
Soegijapranata Catholic University
2022

HALAMAN PENGESAHAN

Judul Tugas Akhir:	:	Sentiment Analysis on Movie Review Using Random Forest and Logistic Regression Algorithm
Diajukan oleh	:	Lady Viona Sugianto
NIM	:	18.K1.0023
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	:	Y.b. Dwi Setianto S.T., M.Cs.
Penguji 5	:	R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D
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.K1.0023

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : LADY VIONA SUGIANTO

ID : 18.K1.0023

declare that this work, titled "SENTIMENT ANALYSIS ON MOVIE REVIEW USING RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHM", 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, June, 03, 2022

LADY VIONA SUGIANTO

18.K1.0023

STATEMENT PAGE OF SCIENTIFIC PUBLICATIONS FOR ACADEMIC INTEREST

I, the undersigned:

Name : LADY VIONA SUGIANTO

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 "SENTIMENT ANALYSIS ON MOVIE REVIEW USING RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHM" 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, June, 03, 2022

LADY VIONA SUGIANTO

18.K1.0023

ACKNOWLEDGMENT

First of all, I would like to thank God for His blessings and grace, so that I can complete a thesis entitled "Sentiment Analysis on Movie Review Using Random Forest and Logistic Regression Algorithm".

I am also very grateful to Hironimus Leong S. Kom., M. Kom as a supervisor at Soegijapranata University who has assisted in writing this thesis.

I would like to thank my family and friends who always support and motivate me so that this thesis can run well and smoothly.

Finally, I realize that the writing of this thesis is still far from perfect. Therefore, I ask for his criticism and suggestions for the sake of building the perfection of this thesis. Hopefully writing this thesis can be useful and provide more knowledge for readers.

Semarang, June, 03, 2022

LADY VIONA SUGIANTO

18.K1.0023

ABSTRACT

In the analysis of the movie classification of a movie, users tend to see whether a movie is good or not from the movie's rating. Most users are not interested in seeing other user reviews before, even though the rating is always tied to the reviews given by users to find out whether the film is good or not. Due to the large number of reviews left by users on a movie website, users have difficulty knowing the classification of positive reviews and negative reviews, and users also do not know the quality of a movie.

Therefore, in this project, sentiment analysis was made for a film review. By using 2 algorithms, namely Random Forest Classifier and Logistic Regression to determine the sentiment analysis of each film. Both algorithms determine positive and negative sentiments. The data structure used is a database, which will train data and test data for comparison purposes. The dataset is also converted to CSV for easier analysis. To evaluate the model, a comparison of algorithms based on accuracy was carried out.

From implementing the Random Forest algorithm with a data range of 650-2000, the resulting analysis sentiment is 9 positive and 1 negative. While the Logistic Regression algorithm produces 8 positive and 2 negative sentiment analyses. And from the comparison of accuracy, Random Forest is better and more suitable for this research because it has an average accuracy of 72,9261% while the average accuracy of Logistic Regression is 63,516%.

Keyword: sentiment analysis, movie review, logistic regression, random forest

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE