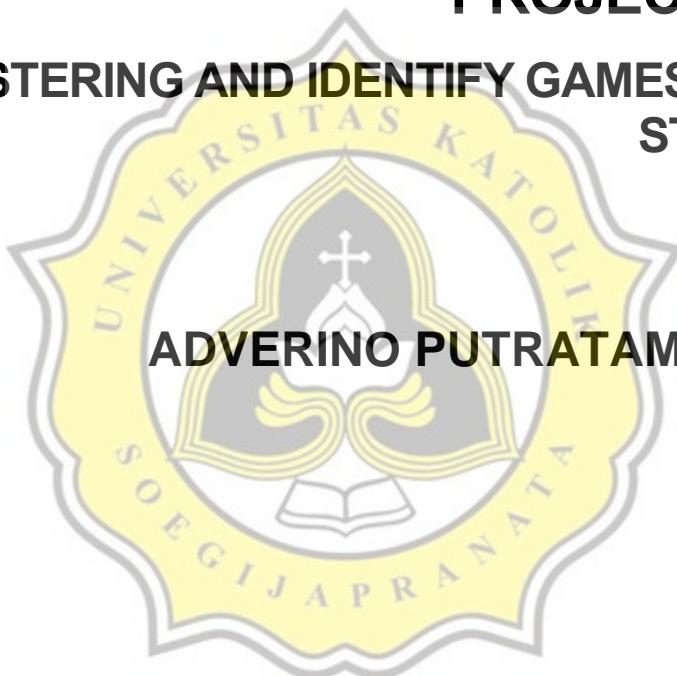




PROJECT REPORT

CLUSTERING AND IDENTIFY GAMES REVIEW ON STEAM STORE

ADVERINO PUTRATAMA PRAYUDA
17.K1.0058



**Faculty of Computer Science
Soegijapranata Catholic University
2021**



HALAMAN PENGESAHAN

Judul Tugas Akhir: : Clustering and Identify Games Review on Steam Store

Diajukan oleh : Adverino Putratama Prayuda

NIM : 17.K1.0058

Tanggal disetujui : 15 Juli 2021

Telah setujui oleh

Pembimbing : Hironimus Leong S.Kom., M.Kom.

Penguji 1 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

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

Penguji 3 : Rosita Herawati S.T., M.I.T.

Penguji 4 : Y.b. Dwi Setianto

Penguji 5 : Yonathan Purbo Santosa S.Kom., M.Sc

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

Ketua Program Studi : Rosita Herawati S.T., M.I.T.

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

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

sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0058

HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : ADVERINO PUTRATAMA PRAYUDA

Program Studi : Teknik Informatika

Fakultas : Ilmu Komputer

Jenis Karya : Penelitian

Menyetujui untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneksklusif atas karya ilmiah yang berjudul “CLUSTERING AND IDENTIFY GAMES REVIEW ON STEAM STORE” beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif 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, 15 Juli 2021

Yang menyatakan



ADVERINO PUTRATAMA PRAYUDA

DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : ADVERINO PUTRATAMA PRAYUDA

ID : 17.K1.0058

declare that this work, titled "CLUSTERING AND IDENTIFY GAMES REVIEW ON STEAM STORE", 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, July 15th 2021



ADVERINO PUTRATAMA PRAYUDA

17.K1.0058

ACKNOWLEDGMENT

First of all, I want to praise and give thanks to the Lord Jesus, with His grace I was able to finish this thesis. 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 guiding me to make this topic. I would also like to thank my friends for support and guiding with advice to finish this document.

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



Semarang, 15 Juli 2021

A handwritten signature in black ink, appearing to read "ADVERINO PUTRATAMA PRAYYUDA".

ADVERINO PUTRATAMA PRAYYUDA

ABSTRACT

Modern entertainment can already be achieved very easily, one example is a game that can be purchased online and downloaded immediately. As a means, online stores provide a review feature as a means to convince buyers through feedback from other buyers. However, this feature is often misused for one-sided advantage, including the Steam platform which has a large number of users.

This research will discuss whether Steam has fake reviews and, whether each review is a helpful review. These questions will be solved by using a data mining algorithm. The use of K-Means will be focused on determining the majority of reviews are Recommended or Not Recommended reviews.

The first result is big upvotes on each review mostly come from 2 factors, which is Steam Curators and big developers. The second result is long review and total play hour is not a determinant for the number of helpful votes. The third result is the accuracy from clustering Review and Recommendation using K-Means algorithm show a high percentage, with 500 sample data, this algorithm can have 100% accuracy, but with different amount of sample data the accuracy score can changed, in example if I use 100 sample data, the algorithm can only achieve 84% accuracy.

Keyword: *K-Means, TF-IDF, Steam, Review, Clustering*

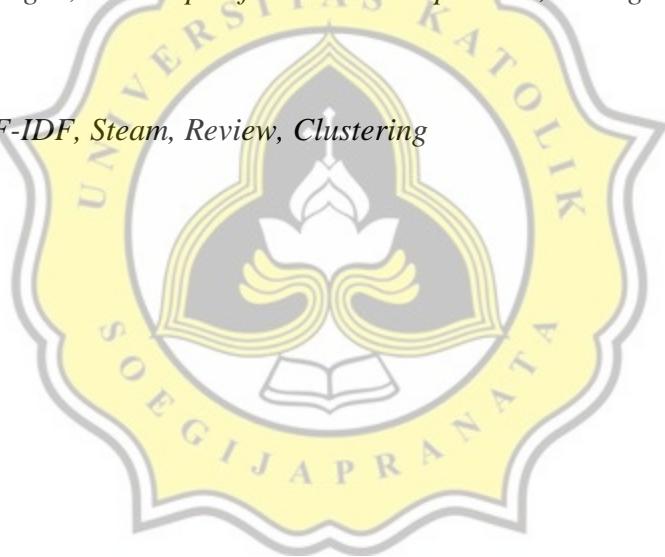


TABLE OF CONTENTS

COVER.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iii
DECLARATION OF AUTHORSHIP.....	iv
ACKNOWLEDGMENT	v
TABLE OF CONTENTS	vii
LIST OF FIGURES	x
LIST OF TABLES.....	xi
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	5
3.1 Journal and References	5
3.2 Data.....	5
3.3 Design.....	5
3.4 Code Program	6
3.5 Implementation and Analysis.....	6
3.6 Conclusion	6
CHAPTER 4 ANALYSIS AND DESIGN	7
4.1 Analysis	7
4.1.1 Data Collecting	7

4.1.2 Preprocessing Data.....	11
4.1.2.1 Data Lowering	11
4.1.2.2 Removing Unescaped Character	12
4.1.2.3 Stop Words Removal	12
4.1.2.4 Punctuation Removal	13
4.1.2.5 Emoji and Emoticon Removal	13
4.1.2.6 URL Removal.....	13
4.1.2.7 Word Lemmatization	13
4.1.3 Term Frequency-Inverse Document Frequency	14
4.1.4 Processing Data with K-Means Algorithm.....	18
4.2 Design.....	23
CHAPTER 5 IMPLEMENTATION AND TESTING	24
5.1 Implementation	24
5.2 Preprocessing	24
5.3 TF-IDF.....	26
5.4 K-Means Implementation.....	28
5.5 Analysis	33
5.5.1 Data Spread on hour_play and helpful	34
5.5.2 Correlation between Hour of Played and Helpfulness.....	35
5.5.3 Correlation between Recommendation and Review.....	36
5.5.4 Accuracy	39
5.5.5 Evaluation	41
CHAPTER 6 CONCLUSION	42
6.1 Conclusion	42
6.2 Suggestions	42
REFERENCES	43

APPENDIX.....a



LIST OF FIGURES

Figure 4. 1 FlowChart	23
Figure 5. 1 Output from line 134 - 140.....	28
Figure 5. 2 Plot Graph for Clustered Data	33
Figure 5. 3 Helpful and Weight Plot.....	34
Figure 5. 4 Hour Played and Weight Plot	34
Figure 5. 5 Hours of Played and Helpful Plot.....	35
Figure 5. 6 Weight and Recommendation DataFrame.....	36
Figure 5. 7 Plot of Unclusterd Weight and Recommendation data	37
Figure 5. 8 Plot of Clustered Weight and Reccomendation data.....	37
Figure 5. 9 Correlation between Recommendation and Helpful	38
Figure 5. 10 Plot of Clustered Data between Weight and Helpful	41
Figure 5. 11 Plot of Clustered Data between Weight and Hour of Played	41

LIST OF TABLES

Table 4. 1 Example Data.....	7
Table 4. 2 Example Training Data.....	10
Table 4. 3 Example of Data Lowering.....	11
Table 4. 4 Unescape Table.....	12
Table 4. 5 Example of Stop Words.....	12
Table 4. 6 Example of Punctuation Removal	13
Table 4. 7 Example of Word Lemmatization.....	13
Table 4. 8 Table of TF and IDF	14
Table 4. 9 Table of Weight each word(tf x idf)	16
Table 4. 10 List of All Document's Weights	17
Table 4. 11 Table of first Centroids.....	18
Table 4. 12 First Iteration.....	19
Table 4. 13 Centroid Value from First Iteration	20
Table 4. 14 Second Iteration.....	20
Table 4. 15 Centroid Value from Second Iteration.....	21
Table 4. 16 Third Iteration	21
Table 4. 17 Centroid value from third iteration	22
Table 5. 1 Table of Figure 5. 8 contents	38
Table 5. 2 Accuracy Table for 100 Sample Data.....	39
Table 5. 3 Accuracy Table for 500 Sample Data.....	39