

**ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN
ALGORITMA AGGLOMERATIVE HIERARCHICAL
CLUSTERING PADA PLATFORM STEAM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIZQI BIO JANHEFI

19.11.2899

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN
ALGORITMA AGGLOMERATIVE HIERARCHICAL
CLUSTERING PADA PLATFORM STEAM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

RIZQI BIO JANHEFI

19.11.2899

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA
PLATFORM STEAM**

yang disusun dan diajukan oleh

Rizqi Bio Janhefi

19.11.2899

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 18 Januari 2023

Dosen Pembimbing,

Agit Amrullah, S.Kom., M.Kom

NIK. 190302356

HALAMAN PENGESAHAN

SKRIPSI

**ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN ALGORITMA
AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA
PLATFORM STEAM**

yang disusun dan diajukan oleh

Rizqi Bio Janhefi

19.11.2899

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Januari 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mei P Kurniawan, M.Kom.
NIK. 190302187

Anggit Ferdita Nugraha, S.T., M.Eng.
NIK. 190302480

Agit Amrullah, S.Kom., M.Kom
NIK. 190302356

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Januari 2023

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rizqi Bio Janhefi
NIM : 19.11.2899

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA PLATFORM STEAM

Dosen Pembimbing: Agit Amrullah, S.Kom., M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Januari 2023

Yang Menyatakan,



Rizqi Bio Janhefi

HALAMAN PERSEMBAHAN

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi ini. Tanpa bantuan dan doa yang tucurahkan kepada-Nya sehingga penulis dapat menyelesaikan skripsi ini. Maka dari itu penulis mempersembahkan karya ini kepada:

1. Kedua orang tua tercinta yang selalu mendoakan, memberikan semangat dan menjadi motivasi bagi penulis.
2. Dosen wali dan dosen pembimbing yang telah membimbing dan memberikan arahan kepada penulis.
3. Teman-teman kelas Informatika 05 angkatan 2019 yang menjadi rekan seperjuangan kuliah penulis.
4. Teman-teman kost seroja dan sahabat-sahabat penulis yang setiap harinya menghibur dan memberikan semangat kepada penulis.
5. Serta kepada semua pihak yang tidak bisa saya sebutkan satu persatu yang telah berperan secara langsung maupun tidak langsung selama ini.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi ini dengan judul **“ANALISA PERBANDINGAN CLUSTER MENGGUNAKAN ALGORITMA AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA PLATFORM STEAM”**. Dibuatnya skripsi ini guna sebagai syarat untuk mendapatkan gelar Program Sarjana (S1) Jurusan Informatika Universitas Amikom Yogyakarta.

Dengan selesainya skripsi ini, penulis menyadari bahwa skripsi ini tidak mungkin terselesaikan tanpa adanya bantuan, bimbingan dan nasehat dari berbagai pihak selama proses penyusunan skripsi ini, oleh karena itu penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto., MM selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, M. Kom selaku dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradya Duhita, M.Kom selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Agit Amrullah, S.Kom., M. Kom selaku Dosen Pembimbing yang telah memberikan bimbingan selama pembuatan skripsi ini.
5. Bapak / Ibu dosen, staff dan karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu dan bantuan yang bermanfaat.
6. Kedua orang tua, yang telah mendoakan, memberikan semangat dan menjadi motivasi bagi penulis.

Yogyakarta, 18 Januari 2023

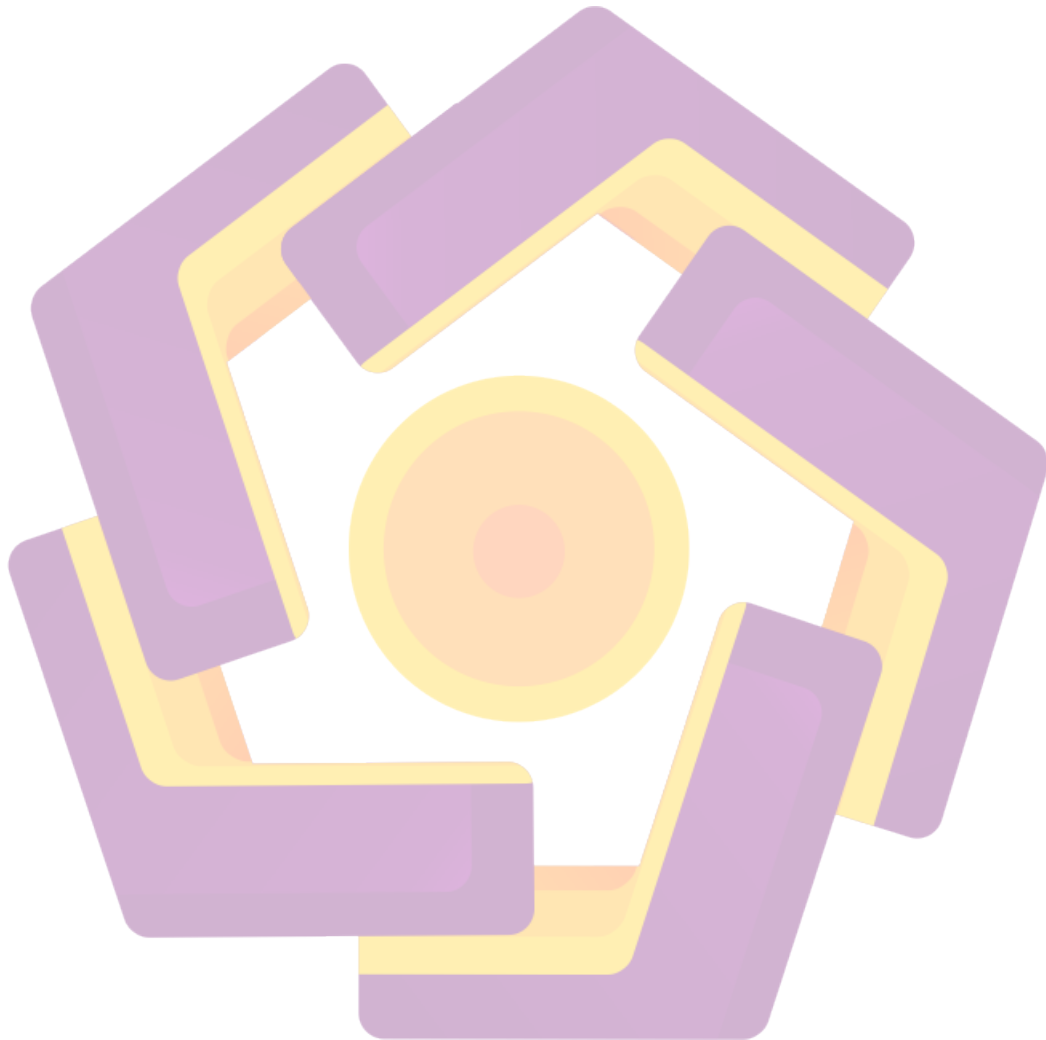
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xvi
ABSTRACT.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.5.1 Manfaat teoritis	4
1.5.2 Manfaat praktis	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 <i>Video Game</i>	10
2.2.1 Platform steam	10
2.3 <i>Dataset</i>	12
2.3.1 Kaggle	12
2.4 <i>Knowledge Discovery in Database</i>	13

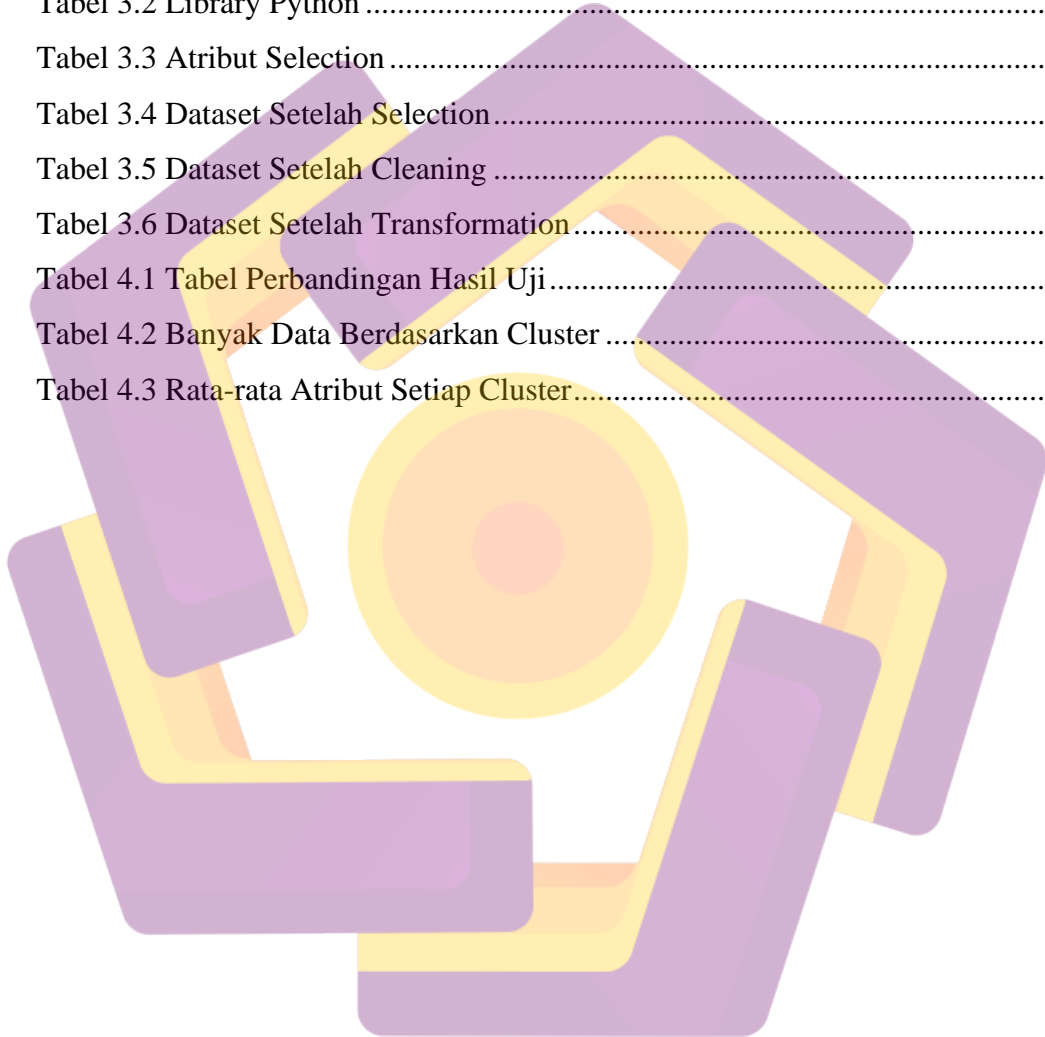
2.4.1	Data Selection	13
2.4.2	<i>Preprocessing/cleaning</i>	13
2.4.3	<i>Transformation</i>	14
2.4.4	<i>Data Mining</i>	14
2.4.5	<i>Interpretation/evaluation</i>	14
2.5	Normalisasi <i>Min-Max</i>	14
2.6	Jarak Euclidean	15
2.7	<i>Agglomerative Hierarchical Clustering</i>	15
2.6.1	<i>Single linkage</i>	15
2.6.2	<i>Complete linkage</i>	16
2.6.3	<i>Average linkage</i>	16
2.6.4	<i>Ward</i>	17
2.8	<i>Koefisien Korelasi Cophenetic</i>	17
BAB III METODE PENELITIAN		19
3.1	Alur Penelitian	19
3.2	Pendahuluan.....	20
3.2.1	Perumusan masalah.....	20
3.2.2	Studi pustaka.....	20
3.3	Pengumpulan Data.....	20
3.4	Analisa	22
3.4.1	Analisa kebutuhan.....	23
3.4.2	Tahapan knowledge discovery in database (KDD).....	24
3.5	Clustering.....	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	Analisa Cluster.....	28
4.1.1	Jarak Euclidean	28
4.1.2	Single Linkage	29
4.1.3	Complete Linkage.....	29
4.1.4	Average Linkage	30
4.1.5	Ward.....	31
4.2	Uji Validitas.....	31

4.3 Metode Cluster Terbaik	32
BAB V PENUTUP	35
5.1 Kesimpulan	35
5.2 Saran	35
DAFTAR PUSTAKA	36



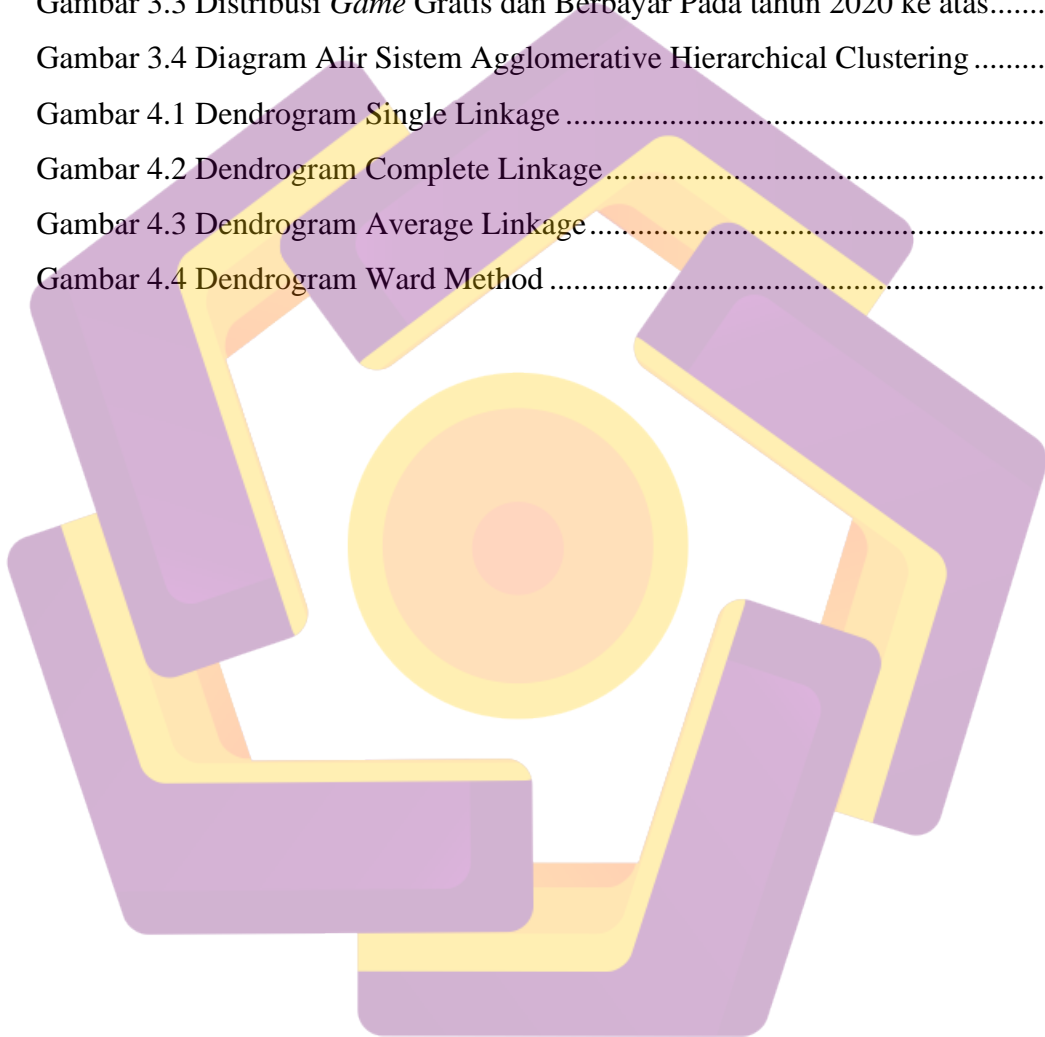
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	7
Tabel 2.2 Klasifikasi Video <i>Game Steam</i>	11
Tabel 3.1 Atribut Dataset Steam	22
Tabel 3.2 Library Python	24
Tabel 3.3 Atribut Selection	25
Tabel 3.4 Dataset Setelah Selection.....	25
Tabel 3.5 Dataset Setelah Cleaning	26
Tabel 3.6 Dataset Setelah Transformation.....	26
Tabel 4.1 Tabel Perbandingan Hasil Uji.....	32
Tabel 4.2 Banyak Data Berdasarkan Cluster	33
Tabel 4.3 Rata-rata Atribut Setiap Cluster.....	33

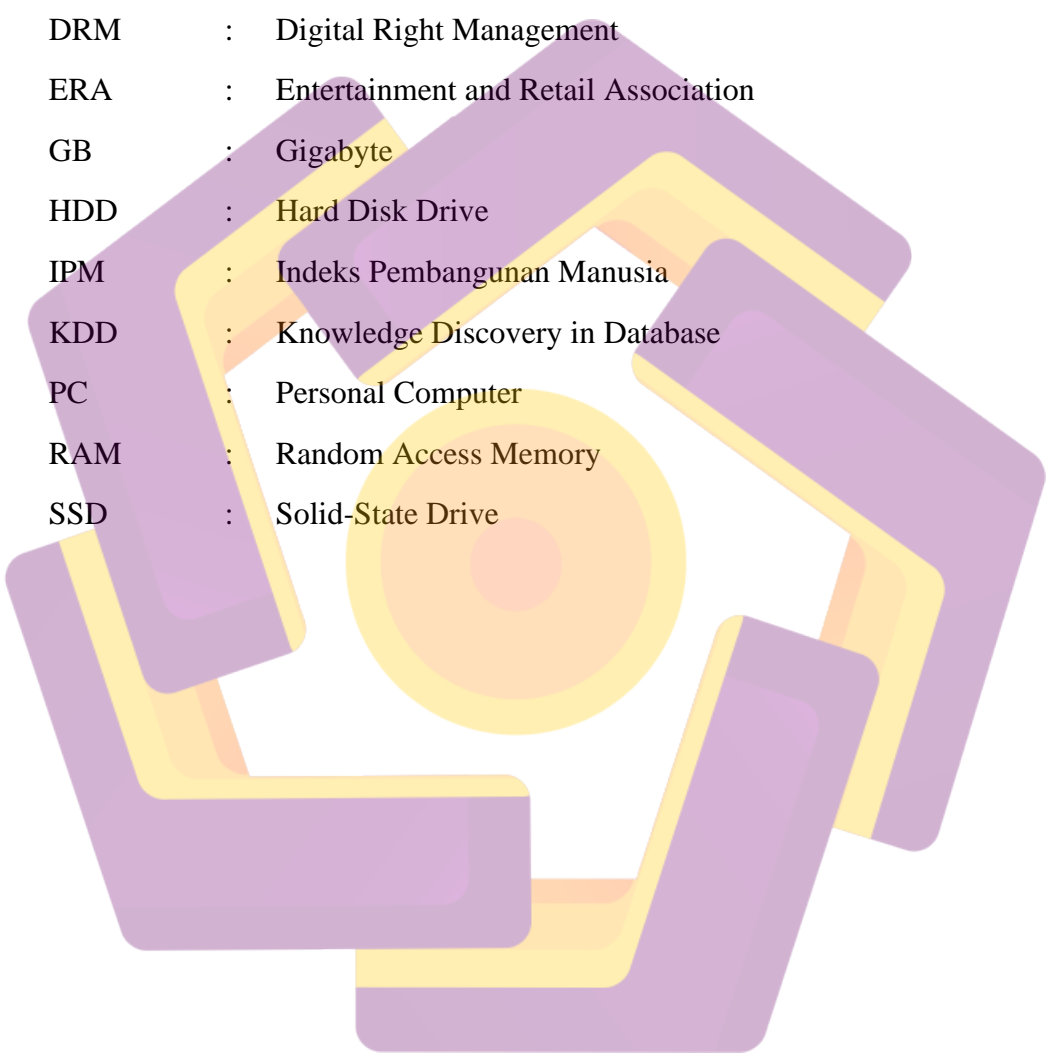


DAFTAR GAMBAR

Gambar 2.1 Proses KDD.....	13
Gambar 3.1 Alur Penelitian	19
Gambar 3.2 Distribusi Jumlah <i>Game</i> Berdasarkan Tahun Rilis	21
Gambar 3.3 Distribusi <i>Game</i> Gratis dan Berbayar Pada tahun 2020 ke atas.....	21
Gambar 3.4 Diagram Alir Sistem Agglomerative Hierarchical Clustering	27
Gambar 4.1 Dendrogram Single Linkage	29
Gambar 4.2 Dendrogram Complete Linkage	29
Gambar 4.3 Dendrogram Average Linkage.....	30
Gambar 4.4 Dendrogram Ward Method	31



DAFTAR LAMBANG DAN SINGKATAN

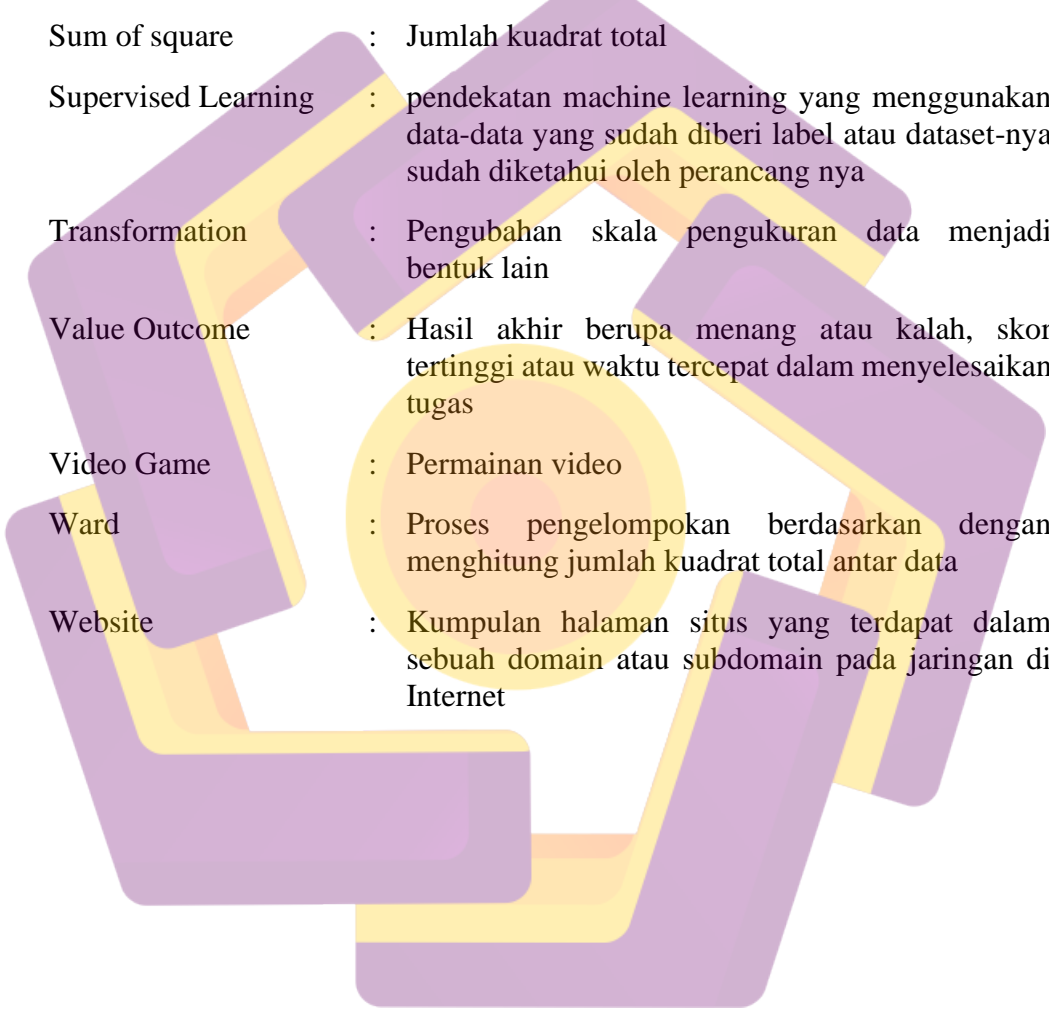


API	:	Application Programming Interface
CD	:	Compact Disc
DRM	:	Digital Right Management
ERA	:	Entertainment and Retail Association
GB	:	Gigabyte
HDD	:	Hard Disk Drive
IPM	:	Indeks Pembangunan Manusia
KDD	:	Knowledge Discovery in Database
PC	:	Personal Computer
RAM	:	Random Access Memory
SSD	:	Solid-State Drive

DAFTAR ISTILAH

Agglomerative hierarchical clustering	: Pengelompokan hirarki yang bertujuan untuk mengelompokkan objek-objek dengan metode bottom-up
Array	: Larik yang berisi kumpulan data bertipe serupa
Average linkage	: Proses pengelompokan berdasarkan jarak rata-rata antar data
Barbershop	: Pangkas rambut
Bottom-up	: Pengelompokan dimana n buah kelompok digabungkan sehingga menjadi cluster tunggal
Cleaning	: Proses pembersihan data
Cluster	: Pengelompokan
Clustering	: Proses pengelompokan objek-objek berdasarkan karakteristik yang dimilikinya
Coding	: Proses kegiatan pengolahan kode
Complete linkage	: Proses pengelompokan berdasarkan nilai tertinggi antar data
Conflict	: Terhadap lawan atau keadaan
Data mining	: proses pengumpulan dan pengolahan data yang bertujuan untuk memperoleh informasi penting pada data
Data scientist	: Orang yang memiliki kemampuan untuk melakukan pengolahan data
Data selection	: Proses pemilihan data
Database	: Kumpulan data yang disimpan dan diakses secara elektronik yang tersusun secara sistematis
Dataset	: Kumpulan data
Dendrogram	: Grafik atau diagram yang disajikan dalam bentuk hirarki
Developer	: Pengembang
Discovery	: Penemuan baru

Dissimilarity distance	:	Jarak ketidakmiripan
Downloadable	:	Dapat diunduh secara online
Euclidean distance	:	Jarak Euclidean
Euclidean space	:	Berkaitan dengan ruang geometri pada jarak Euclidean
Game	:	Permainan
Gamer	:	Orang yang bermain game
Hardware	:	Perangkat keras
Indie video game	:	Permainan video game yang dibuat oleh tim kecil atau bahkan perseorangan.
K-Means	:	Pengelompokan n pengamatan menjadi k kelompok
Knowledge Discovery in Database	:	Kegiatan untuk mengidentifikasi beragam jenis data
Koefisien Korelasi Cophenetic	:	Korelasi antara elemen matriks ketidakmiripan dengan elemen yang dihasilkan oleh cophenetic
Latent Dirichlet Allocation	:	Model statistik generatif yang menjelaskan serangkaian pengamatan melalui kelompok
Library Python	:	Kumpulan modul yang berisi kode yang dapat digunakan pada program yang berbeda
matriks cophenetic	:	Matrik ketidakmiripan
Missing Value	:	Data yang hilang
Multi-Player	:	Lebih dari satu pemain
Player ability	:	Kemampuan yang dimiliki pemain berupa keterampilan, strategi dan keberuntungan
Preprocessing	:	Tahapan sebelum dilakukannya sebuah proses
Private Dataset	:	Dataset yang aksesnya dibatasi hanya untuk kepentingan tertentu
Public Dataset	:	Dataset yang bisa diakses oleh banyak orang
Rating	:	Penilaian dengan standar tertentu
Rules	:	Aturan



Silhouette Coefficient	: Validasi konsistensi suatu data
Single Linkage	: Proses pengelompokan berdasarkan nilai terendah antar data
Social Media	: Sosial media
Software	: Perangkat lunak
Streaming video	: Kegiatan menonton video secara real time
Sum of square	: Jumlah kuadrat total
Supervised Learning	: pendekatan machine learning yang menggunakan data-data yang sudah diberi label atau dataset-nya sudah diketahui oleh perancang nya
Transformation	: Pengubahan skala pengukuran data menjadi bentuk lain
Value Outcome	: Hasil akhir berupa menang atau kalah, skor tertinggi atau waktu tercepat dalam menyelesaikan tugas
Video Game	: Permainan video
Ward	: Proses pengelompokan berdasarkan dengan menghitung jumlah kuadrat total antar data
Website	: Kumpulan halaman situs yang terdapat dalam sebuah domain atau subdomain pada jaringan di Internet

INTISARI

Pada saat ini perkembangan industri *game* digital sangatlah pesat. Berdasarkan laporan dari analisis pasar Newzoo, jumlah *gamer* dunia pada tahun 2022 mencapai 3,2 miliar yang mengalami peningkatan sebesar 2,1% dibandingkan tahun sebelumnya. Bagi para *gamer*, Steam merupakan salah satu layanan digital *video game* terpopuler yang menyediakan ribuan *game* di dalamnya. Steam juga menyajikan ulasan dan *rating* dari para user terhadap *video game*.

Pada penelitian ini, nantinya akan digunakan algoritma *Agglomerative hierarchical clustering* untuk melihat ciri-ciri dari *video game* berdasarkan atribut yang ada. Algoritma *Agglomerative hierarchical clustering* terbagi menjadi beberapa metode, di antaranya metode *single linkage*, *complete linkage*, *average linkage*, dan *ward*. Penelitian ini membandingkan keempat metode dalam *Agglomerative hierarchical clustering* dengan tujuan untuk mendapatkan solusi cluster terbaik dalam kasus *clustering video game* pada platform steam berdasarkan *rating*, banyak *reviews*, harga dan banyaknya jumlah unduhan dari *video game* tersebut.

Hasil dari penelitian diperoleh *complete linkage* sebagai metode terbaik yang memiliki nilai *koefisien korelasi cophenetic* tertinggi jika dibandingkan dengan metode *single linkage*, *average linkage* dan *ward*. Dan banyak hasil *cluster* yang diperoleh adalah sebanyak 4 cluster.

Kata kunci: *Data Mining*, Pengelompokan, Permainan video, Steam, *Agglomerative hierarchical clustering*

ABSTRACT

At this time the development of the digital game industry is very rapid. Based on a report from market analyst Newzoo, the number of world gamers in 2022 will reach 3.2 billion, an increase of 2.1% compared to the previous year. For gamers, Steam is one of the most popular digital video game services that provides thousands of games in it. Steam also presents user reviews and ratings of video games.

In this study, an Agglomerative hierarchical clustering algorithm will be used to see the characteristics of video games based on existing attributes. Agglomerative hierarchical clustering algorithm is divided into several methods, including single linkage, complete linkage, average linkage, and ward methods. This study compares the four methods in Agglomerative hierarchical clustering with the aim of getting the best cluster solution in the case of clustering video games on the Steam platform based on ratings, lots of reviews, price and the number of downloads of the video games.

The results of the study obtained complete linkage as the best method which has the highest cophenetic correlation coefficient when compared to the single linkage, average linkage and ward methods. And the number of cluster results obtained is as many as 4 clusters.

Keyword: *Data Mining, Clustering, Video game, Steam, Agglomerative hierarchical clustering*