

skripsi

by Jasin Y

Submission date: 17-Aug-2022 11:31PM (UTC+1000)

Submission ID: 1883553980

File name: skripsi.docx (1.9M)

Word count: 9065

Character count: 57730

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

SKRIPSI

Diajukan Untuk Skripsi Guna Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Program Studi Sistem Informasi Fakultas Teknik
Universitas Nusantara PGRI Kediri



OLEH :

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA

UN PGRI KEDIRI

2022

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

SKRIPSI

Diajukan Untuk Skripsi Guna Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Program Studi Sistem Informasi Fakultas Teknik
Universitas Nusantara PGRI Kediri



OLEH :

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI**

2022

ii

Skripsi oleh:

MUHAMMAD ZAINAL ARIFFIN

NPM : 18.1.03.03.0023

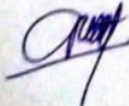
Judul:

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

Telah Disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program
Studi Sistem Informasi UN PGRI Kediri

Tanggal: 22 juli 2022

Pembimbing I



Aldina Ristyawan, M.Kom

NIDN. 0721018801

Pembimbing II



M.Nailiyyah Muzaki, M.Cs

NIDN. 0706098902

Skripsi oleh:

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

Judul:

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**




Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi Program

Studi Sistem Informasi FT UN PGRI Kediri

Pada tanggal: 22 juli 2022

Dan Dinyatakan telah Memenuhi Persyaratan

Panitian Penguji:

- | | | |
|---------------|-----------------------------|--|
| 1. Ketua | : Aidina Ristyawan, M.Kom |  |
| 2. Penguji | : Anita Sari Wardani, M.Kom |  |
| 3. Penguji II | : M.Najibulloh Muzaki, M.Cs |  |

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd
NIP. 19640202 199103 1002

PERNYATAAN

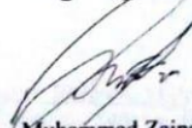
Yang bertanda tangan di bawah ini saya,

Nama : Muhammad Zainal Ariifin
Jenis Kelamin : Laki-Laki
Tempat/tgl.lahir : Nganjuk/ 23 Juli 1999
NPM : 18.1.03.03.0023
Fak./Jur./Prodi. : Teknik / Sistem Informasi

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 22 juli 2022

Yang Menyatakan



Muhammad Zainal Ariifin
NPM: 18.1.03.03.0023

MOTTO

Kertas kosong menjadi bernilai saat tergores oleh tinta

Jiwa yang kosong menjadi berarti saat hati memikirkan makna kehidupan

Abstrak

Muhammad Zainal Arifin Data Mining Perbandingan Performa Cluster Dalam Penentuan Jenis Bunga Iris, Skripsi, Fakultas Teknik, Universitas Nusantara PGRI Kediri, 2022.

Penelitian ini dilatar belakangi dari Sudah banyak algoritma *clustering* yang dipakai dalam penelitian namun ada beberapa algoritma *clustering* yang populer digunakan dalam analisis komparasi seperti algoritma *K-Means*, *K-Medoids*, *C-Means*, *Hard C-Means*, dan *X-Means*. Namun banyak penelitian yang hanya masih menggunakan satu algoritma *clustering* dalam penentuan jenis bunga *iris*.

Penelitian ini menggunakan 3 algoritma clustering dalam analisis komparasi algoritma yaitu *K-Means*, *X-Means*, dan *K-Medoids*. Dataset menggunakan data *iris* dari *UCI Machine Learning Repository*. Penelitian ini menggunakan nilai k yang digunakan adalah tiga kluster untuk diuji menggunakan nilai *Davies Bouldin Index* (DBI) serta menggunakan Validasi *K-Fold Cross Validation* untuk mendapatkan nilai interval dengan membagi data training dan testing dan metode perbandingan uji beda T-test untuk mengetahui ada perbedaan signifikan atau tidak dari perbandingan nilai *Davies Bouldin Index*, sehingga diperoleh algoritma yang memiliki sifat dominan pada tahapan preprosesing data dengan menggunakan data tranformation yaitu berupa metode *Z-score normalization*.

Kesimpulan hasil dari penelitian ini adalah nilai *Davies Bouldin Index* (DBI) yang didapat adalah sebesar 0,839 untuk algoritma *K-Means*, *X-Means* sebesar 0,841, dan *K-Medoids* sebesar 1,498, sedangkan dari hasil uji beda t-test yang dilakukan bahwa algoritma *K-Means* dan *X-Means* merupakan algoritma dengan hasil yang paling dominan dari ketiga algoritma yang diterapkan pada dataset *iris*. Sedangkan untuk algoritma *K-Medoids* merupakan algoritma dengan hasil yang kurang baik dengan nilai *Davies Bouldin Index* (DBI) sebesar 1.498.

Kata Kunci – *Clustering, Algoritma K-Means, Algoritma X-Means, Algoritma K-Medoids*

KATA PENGANTAR

Puji Syukur Kami panjat kan kehadiran Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenan- Nya tugas penyusunan skripsi ini dapat diselesaikan.

Skripsi dengan judul "**Perbandingan Performa Cluster Dalam Penentuan Jenis Bunga Iris**" ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi UN PGRI Kediri.

Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Dr. Zaenal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Bapak Dr. Suryo Widodo, M.Pd. selaku Dekan Fakultas Teknik Universitas Nusantara PGRI Kediri.
3. Ibu Rina Firliana, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.
4. Bapak Aidina Ristyawan, M.Kom selaku Dosen pembimbing I yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
5. Bapak M.Najibulloh Muzaki, M.Cs selaku Dosen pembimbing II yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
6. Ibu Anita Sari Wardani, M.Kom selaku Penguji yang telah memberikan saran dan masukan untuk perbaikan dan penyelesaian skripsi ini.
7. Orangtua tercinta, Ayah Ibunda serta saudara-saudaraku atas ketulusan doa, motivasi dan nasihat-nasehat sepanjang waktu sehingga dapat menyelesaikan skripsi ini.

Kediri, 22 Juli 2022



MUHAMMAD ZAINAL ARIFIN
NPM : 18.1.03.03.0023

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PESETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO.....	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Batasan Masalah.....	4
1.3. Rumusan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	7
2.1. Kajian Teori.....	7
2.1.1. Data Mining.....	7
2.1.2. Cluster.....	10
2.1.3. K-Means.....	11
2.1.4. X-Means.....	12
2.1.5. K-Medoids.....	14
2.1.6. Perhitungan Jarak (<i>Measures Distance</i>).....	15
2.1.7. <i>Euclidean Distance</i>	16
2.1.8. <i>Davies Buldin index</i>	16
2.1.9. <i>K-Fold Cross Validation</i>	18
2.1.10. Parametrik uji beda T-test.....	19
2.1.11. Dataset.....	20
2.1.12. Rapidminer.....	21
2.2. Kajian Penelitian terdahulu.....	21
BAB III METODE PENELITIAN.....	24

3.1. Metode Penelitian	24
3.2. Metode Pengumpulan Data	24
3.3. Metode Analisis Data	25
3.3.1. Jenis Analisis	25
3.4. Alur Penelitian	25
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Pengumpulan Data	29
4.2. <i>Preprocessing</i> Data	30
4.2.1. Transformasi Data	31
4.3. Implementasi Algoritma Clustering	33
4.3.1. Pemodelan <i>K-Means</i>	33
4.3.2. Pemodelan <i>X-Means</i>	34
4.3.3. Pemodelan K-Medoids	36
4.4. Evaluasi	37
4.5. Metode Perbandingan	39
BAB V PENUTUP	41
4.6. Kesimpulan	41
4.7. Saran	41
DAFTAR PUSTAKA	43
LAMPIRAN	45

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	22
Tabel 4.1 <i>Data sets</i> yang digunakan	29
Tabel 4.2 Hasil Normalisasi Data	32
Tabel 4.3 Perbandingan Performance tiga Algoritma.....	38
Tabel 4.4 Hasil uji beda T-test	39

DAFTAR GAMBAR

Gambar 1.1 Bunga Iris	1
Gambar 2.1 Bidang Ilmu <i>Data Mining</i>	8
Gambar 2.2 Tahapan KDD	9
Gambar 2.3 Skenario Pengujian <i>10-Fold Cross Validation</i>	19
Gambar 3.1 Alur Penelitian.....	26
Gambar 4.1 visualisasi data iris	30
Gambar 4.2 Rancangan Model <i>K-Means</i>	33
Gambar 4.3 Desain Bagian Cross Validation	34
Gambar 4.4 Hasil <i>description Performance Vector</i>	34
Gambar 4.5 Rancangan Model <i>X-Means</i>	35
Gambar 4.6 Desain Bagian Cross Validation	35
Gambar 4.7 Hasil <i>description Performance Vector</i>	36
Gambar 4.8 Rancangan Model <i>K-Medoids</i>	36
Gambar 4.9 Desain Bagian Cross Validation	37
Gambar 4.10 Hasil <i>description Performance Vector</i>	37
Gambar 4.11 Desain Perbandingan uji beda T-test.....	39

DAFTAR LAMPIRAN

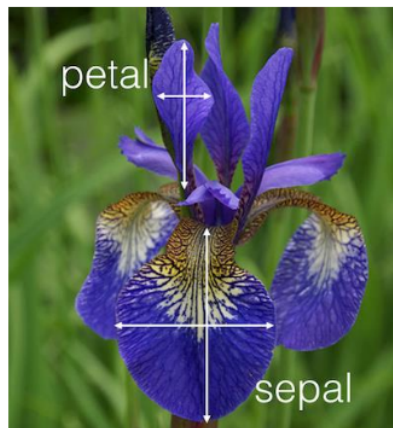
Lampiran 1 <i>Data Iris</i>	1
Lampiran 2 <i>Data Iris Z-Score Normalization</i>	8

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Data iris merupakan data dari 150 bunga, Gambar 1.1 menunjukkan bahwa bunga iris diidentifikasi berdasarkan *Sepal Length*, *Sepal Width*, *Petal Length*, *Petal Width*. Dari 150 data tersebut pada umumnya peneliti-peneliti sebelumnya mengelompokkan menjadi tiga kelompok bunga, yaitu *iris setosa*, *iris virginica* dan *iris versi color*. Dalam pengujian metode cluster banyak dari peneliti sebelumnya yang menggunakan *data iris*, karena data iris merupakan data yang gampang untuk didapatkan dan sederhana (Candrasari Hermanto, 2017).



Gambar 1.1 Bunga iris (Taha Chicho et al, 2021)

Data mining adalah ekstraksi informasi dari sejumlah data yang besar untuk melihat pengetahuan yang tersembunyi dan memfasilitasi pemakainya secara *real time*. *Data mining* memiliki algoritma yang berbagai macam untuk analisis data.

Beberapa algoritma *data mining* sering digunakan untuk analisis data adalah *Clustering*, *Association*, *Classification* dan lain-lain. *Cluster* adalah teknik untuk analisis data eksplorasi secara efektif (Herviany et al, 2021).

Analisis *cluster* adalah teknik multivariat yang memiliki tujuan primer buat mengelompokkan objek-objek menurut karakteristik yang dimiliki (Awalludin & Taufik, 2017). Analisis cluster mengklasifikasi objek sehingga objek-objek yang paling dekat kesamaanya menggunakan objek lain berada pada cluster yang sama (Sari & Sukestiyarno, 2021).

Di antara banyak algoritma *clustering*, algoritma *clustering K-means* banyak digunakan karena algoritmanya yang sederhana dan konvergensi yang cepat. Namun, nilai K dari pengelompokan perlu diberikan terlebih dahulu dan pilihan nilai K secara langsung mempengaruhi hasil konvergensi (Yuan & Yang, 2019).

K-Medoids adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (*medoids*) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster*. Metode *K-medoids* lebih cocok untuk mengelompokkan data dibandingkan metode *K-Means* (Marlina et al, 2018).

Dalam Adhitama et al. (2020), Pelleg dan Moore melakukan penelitian pada tahun 2000 untuk melakukan ekstensi kepada metode *K-Means* untuk menghasilkan proses komputasi yang lebih cepat dan melakukan estimasi jumlah klaster yang paling efisien dalam pemrosesan data. Hasil dari penelitian tersebut adalah metode *X-Means* (extended *K-Means*) dapat digunakan pada dataset yang kecil dan proses komputasi yang lebih cepat dibandingkan dengan *K-Means* .

Sudah banyak algoritma ⁵ *clustering* yang dipakai dalam penelitian namun ada beberapa algoritma *clustering* yang populer digunakan dalam analisis komparasi seperti algoritma *K-Means*, *K-Medoids*, *C-Means*, *Hard C-Means*, dan *X-Means*. Namun banyak penelitian yang hanya masih menggunakan satu algoritma *clustering* dalam penentuan jenis bunga *iris* (Candrasari Hermanto, 2017).

Data yang digunakan didalam penelitian ini adalah *data set* iris, data tersebut dipublikasikan oleh UCI Repository yang mana data tersebut sudah diakui oleh penelitian yang bergerak dalam data science, Metode yang dipakai dalam penelitian ini adalah Algoritma *X-Means*, *K-Means*, *K-Medoids*. Penelitian ini membandingkan performa dari ketiga algoritma Dengan menghitung nilai *Davies Bouldin Index* (DBI) serta menggunakan Validasi *K-Fold Cross Validation* untuk mendapatkan nilai interval dengan membagi data ⁵ *training* dan *testing* dan metode perbandingan uji beda T-test untuk mengetahui ada perbedaan signifikan atau tidak dari perbandingan nilai *davies bouldin index*, sehingga diperoleh algoritma yang optimal dari ketiga algoritma dengan *data set iris*. Pada tahapan preprosesing data dengan menggunakan *data transformation* yaitu berupa metode *Z-score normalization*.

Clustering dipilih dalam pemrosesan data karena mampu menentukan K terbaik pada pengelompokan dalam *data set*. Penganalisaan k suatu cara efektif karena hasil *cluster* bergantung pada jumlah k yang digunakan.

Berdasarkan latar belakang yang diuraikan maka dilakukan komparasi algoritma yang paling optimal dalam clustering bunga iris.

1.2. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut :

1. *Dataset* yang digunakan dalam penelitian ini berasal dari data sekuler yang diperoleh dari *UCI Repository* berupa data iris.
2. Algoritma *X-Means*, *K-Means* dan *K-Medoids* menggunakan Metode *Validasi K-Fold Cross Validation* dan Evaluasi DBI.
3. Untuk metode perbandingan menggunakan metode uji beda T-test

1.3. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana menetapkan performa *clustering* yang optimal dari komparasi Algoritma *X-Means*, *K-Means*, dan *K-Medoids*?
2. Bagaimana membandingkan hasil *cluster* dari 3 algoritma menggunakan metode Evaluasi DBI dan uji beda T-test?

1.4. Tujuan Penelitian

Sesuai dengan rumusan masalah yang telah dibahas diatas maka tujuan dari penelitian ini adalah:

1. Membandingkan performa dari ketiga algoritma dengan menghitung nilai *Davies Bouldin Index* (DBI) yang menggunakan metode DBI.
2. Mengetahui algoritma *Cluster* yang optimal untuk penelitian data iris.

1.5. Manfaat Penelitian

Dalam penelitian ini diharapkan dapat memberikan manfaat berupa, menjadi perbandingan peneliti lain dalam menganalisis menggunakan metode *clustering*.

1.6. Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri dari lima bagian, yaitu sebagai berikut:

Bab 1: Pendahuluan

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

Bab 2: Landasan Teori

Bab ini berisi tentang teori-teori yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini yaitu teori-teori yang berhubungan dengan *data mining*, *clustering*, *K-Means*, *K-Medoids*, *X-Means*, dan beberapa penelitian terdahulu.

Bab 3: Metodologi Penelitian

Bab ini berisi tentang tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja.

⁴ Bab 4: Hasil Dan Pembahasan

Bab ini berisi tentang proses yang terdapat pada langkah penelitian yang telah direncanakan.

Bab 5: Penutup

Bab ini berisi tentang kesimpulan dari semua uraian pada bab-bab sebelumnya dan saran-saran yang diharapkan berguna untuk pengembangan penelitian selanjutnya.

BAB II

KAJIAN TEORI

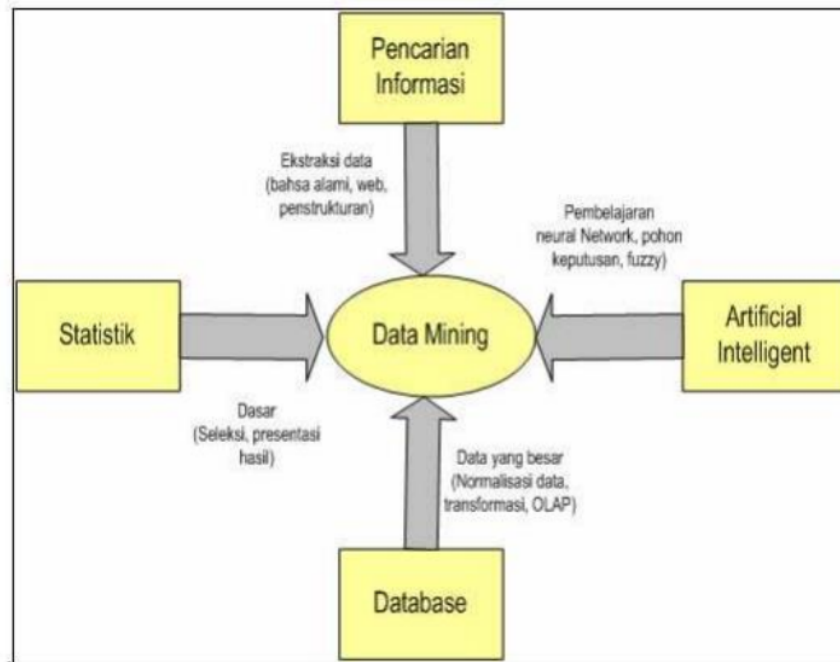
2.1. Kajian Teori

2.1.1. Data Mining

Data Mining adalah sub bidang interdisipliner ilmu komputer dan statistik dengan tujuan keseluruhan mengekstrak informasi (*intelligent methods*) dari kumpulan data dan mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut (Hastie, Tibshirani, & Friedman, 2009).

Data Mining sebagai proses untuk mendapatkan informasi yang berguna dari Gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai penggalian informasi baru dari sejumlah besar data yang berguna untuk pengambilan keputusan. Istilah *data mining* terkadang disebut sebagai penemuan pengetahuan (Prasetyo, 2012).

Peran utama dari *data mining* adalah estimasi, prediksi, klasifikasi, clustering, dan asosiasi. Dari semua peranan *data mining* tersebut terbagi menjadi 2 berdasarkan metode pembelajarannya yaitu *Supervised Learning*, *Unsupervised Learning* (Santosa, 2007). Perbedaan dari kedua metode pembelajaran pada algoritma *data mining* tersebut adalah jika dalam supervised learning harus memiliki data sampel atau sering disebut juga dengan data training. Sedangkan dalam unsupervised learning tidak membutuhkan data training. Salah satu contoh peran *data mining* dengan metode unsupervised learning adalah *clustering*.

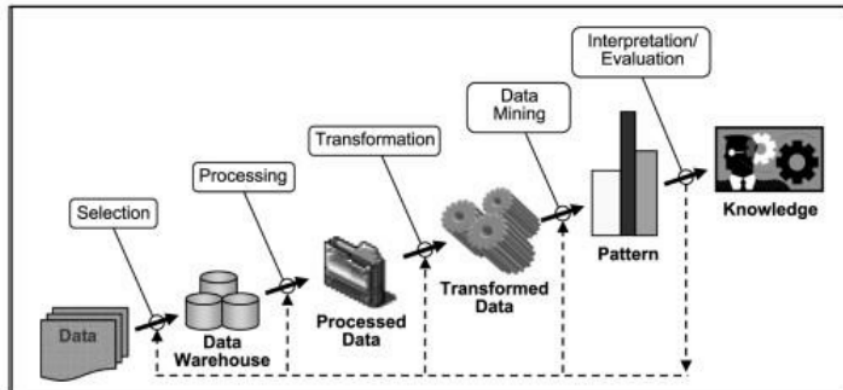


Gambar 2.1 Bidang Ilmu *data mining* (Pramudiono, 2006)

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan *data mining* adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dulu. Gambar 2.1 menunjukkan bahwa *data mining* memiliki akar yang panjang dari bidang ilmu seperti *artificial intelligent*, *machine learning*, statistik, database dan juga information retrieval (Pramudiono, 2006).

Data mining adalah bagian dari *Knowledge Discovery in Database* (KDD), *Knowledge Discovery in Database* (KDD) adalah seluruh proses menemukan dan mengidentifikasi pola atau informasi data tentang di mana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Secara garis besar proses

KDD terdiri atas beberapa tahapan (Fayyad et al, 1996).



Gambar 2.2 Tahapan KDD (Fayyad et al., 1996)

1. Data Selection

Pemilihan (seleksi) data dilakukan dari suatu kumpulan data operasional. Sebelum tahap penggalian informasi dalam KDD dimulai proses ini perlu dilakukan. Data hasil seleksi disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Data Cleaning

Proses pembersihan secara khusus mencakup penghapusan data duplikat, pengecekan data yang tidak konsisten, dan perbaikan kesalahan data seperti salah ketik. Juga menjalankan proses *enrichment* yang “memperkaya” data yang ada dengan data atau informasi lain yang dibutuhkan terkait KDD, Data dan informasi eksternal dll.

3. Data Transformation

Proses transformasi data yang dipilih adalah pengkodean, yang membuatnya cocok untuk proses penambangan data. Proses coding dalam

KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Proses Mining

Data mining merupakan proses untuk mencari suatu pola atau informasi yang menarik dalam data yang terpilih dengan teknik atau metode tertentu. Ada berbagai metode atau algoritma untuk *data mining*. Pilihan metode dan algoritme yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation/Evaluation

Interpretasi adalah proses menampilkan pola informasi yang dihasilkan oleh pemangku kepentingan dari proses *data mining*. Tahap ini meliputi pemeriksaan terhadap pola atau informasi yang ditemukan agar tidak bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.1.2. Cluster

Dalam analisis *cluster*, data (objek) dikelompokkan hanya berdasarkan informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya. Tujuan adalah agar objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam yang lain. Lebih besar kemiripannya (homogenitas) dalam kelompok dan lebih besar perbedaannya di antara kelompok yang lain, konsep inilah yang akan dibahas dalam pengelompokan (Prasetyo, 2012).

Analisis *cluster* ⁴ dimana objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain dan berbeda dalam kelompok

dan lebih besar perbedaannya diantara kelompok lain. Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data clustering yaitu *hierarchical* (hirarki) data clustering dan *non-hierarchical* (non hirarki) data *clustering*.

2.1.3. *K-Means*

K-Means merupakan salah satu metode pengelompokan secara *partitioning* yang memisahkan data ke dalam *cluster* yang berbeda. Dengan *partitioning* secara iterative, *K-Means Clustering* mampu memindahkan rata-rata jarak setiap data ke klasternya (James et al, 1967).

K-Means Clustering merupakan algoritma pengklasteran iteratif yang melakukan partisi set data ke dalam sejumlah *K Cluster* yang sudah ditetapkan di awal. Pengelompokan *K-Means* mudah diimplementasikan dan dieksekusi, relatif cepat, mudah beradaptasi, dan bahkan digunakan secara luas. Secara historis, *K-Means Clustering* menjadi salah satu algoritma yang paling penting dalam bidang *data mining* (Prasetyo, 2012).

Adapun Langkah-langkah untuk *K-Means Clustering* adalah sebagai berikut (Prasetyo, 2012):

1. Menentukan *K* sebagai jumlah *cluster* yang ingin dibentuk
2. Menentukan pusat *cluster* (*centroid*) secara acak dari objek yang ada pada masing-masing *cluster* dengan persamaan:

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{nj}}{n} \dots\dots\dots (2.1)$$

Dimana:

C_{kj} = pusat *cluster* ke-k pada variabel ke j ($j = 1, 2, \dots, p$)

n = banyak data pada *cluster* ke-k

- Menentukan jarak setiap objek dengan setiap *centroid* dengan perhitungan jarak setiap objek dengan setiap *centroid* menggunakan jarak *Euclidean distance*

$$d(X_i, X_g) = \sqrt{\sum_{j=1}^p (X_{ij} - X_{gj})^2} \dots \dots \dots (2.2)$$

- Mengalokasikan masing-masing data ke *centroid*/rata-rata terdekat yang dirumuskan sebagai berikut:

$$a_{ij} = \begin{cases} 1, & s = \min\{d(x_i, C_{kj})\} \\ 0, & \text{lainnya} \end{cases} \dots \dots \dots (2.3)$$

a_{ij} adalah nilai keanggotaan titik x_i ke pusat *cluster* C_{kj} , s adalah dari data x_i ke pusat *cluster* C_{kj} setelah dibandingkan.

- Lakukan iterasi, Kemudian tentukan posisi *centroid* baru menggunakan persamaan (2.2)
- Mengulangi kembali langkah 3-6 sampai tidak ada lagi perpindahan objek atau tidak ada perubahan pada fungsi objektifnya.

2.1.4. X-Means

X -means clustering adalah variasi pengelompokan K -means yang memperlakukan alokasi *cluster* dengan mencoba partisi berulang dan

mempertahankan pemisahan hasil yang optimal, sampai beberapa kriteria tercapai. Tujuan dari *X-mean cluster* adalah melakukan pengelompokan intrinsik dalam kumpulan data yang tidak berlabel. Menyediakan cara yang cepat dan efisien untuk mengklasifikasikan data tidak terstruktur, penggunaan konkurensi dengan mempercepat proses konstruksi dan penggunaan model (Zizwan et al, 2020).

Dalam Adhitama et al. (2020), Pelleg dan Moore Mendefinisikan algoritma *X-Means* yang disesuaikan dengan *K-Means*. *X-Means* bekerja setelah setiap iterasi *K-Means* dan menentukan bagian mana dari pusat aliran yang akan dipisahkan untuk mendapatkan data yang lebih baik. Pembagian keputusan dilakukan dengan menghitung kriteria BIC. Langkah-langkah algoritma *X-Means* adalah sebagai berikut :

1. Inialisasi $K = K_{min}$
2. Jalankan algoritma *K-Means*
3. Untuk $k=1, \dots, k$: Ganti setiap centroid μ_k dengan dua centroid μ_1 dan μ_2

Dua centroid baru buat inialisasi masing-masing prosedur pemecahan *K-means* diperoleh dengan mempengaruhi centroid orisinal pada dua arah berlawanan sepanjang vektor yang dipilih secara random menggunakan jumlah yang proporsional dengan ukuran kluster yang bersangkutan.

4. Jalankan algoritma *K-means* dengan $K=2$ di atas k -cluster
5. Ganti atau pertahankan setiap centroid berdasarkan kriteria pemilihan model. Algoritma melakukan tes pemilihan model BIC untuk menentukan apakah dua cluster baru adalah model yang lebih baik daripada kluster tunggal asli di masing-masing kasus. $BIC_{\{k\}} = -2 * \log L + K \log N$ yang N

adalah jumlah observasi dan k adalah jumlah cluster dan $\log L$ adalah log-likelihood

6. Jika Kondisi konvergensi tidak puas, ulangi ke langkah 2

2.1.5. *K-Medoids*

K-Medoids Clustering, juga dikenal sebagai Partitioning Around Medoids (PAM), adalah varian dari metode *K-Means*. Hal ini didasarkan pada penggunaan medoids bukan dari pengamatan mean yang dimiliki oleh setiap klaster, dengan tujuan mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai yang ada dalam dataset (Vercellis, 2009).

K-Medoids Clustering hadir untuk mengatasi kelemahan *K-Means Clustering* yang sensitive terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data. Algoritma ini menggunakan objek dalam kumpulan objek untuk mewakili sebuah cluster. Objek yang dipilih untuk membentuk sebuah cluster yang disebut medoid. Adapun tahapan *K-Medoids Clustering* diantaranya (Fitriyadi & ana kurniawati, 2021) :

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)^T (x_i - x_j)} \dots \dots \dots (2.4)$$

Dimana $i=1, \dots, n; j=1, \dots, n$ dan p adalah banyak variable, serta V adalah matrik varian kovarian.

3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai *medoid*.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster anggota cluster masing-masing

2.1.6. Perhitungan Jarak (*Measures Distance*)

Perhitungan jarak sering digunakan untuk menentukan apakah dua vektor serupa. Sehingga metode ini banyak digunakan untuk melakukan pengenalan pola (Wurdianarto, Novianto, & Rosyidah, 2014). Metode jarak yang tersedia meliputi *Euclidean Distance*, *Chebyshev*, *Angular Separation*, *Canberra Distance*, *Haming Distance* dan *Sorrensen Distance*. Pada algoritma *K-Means*, *X-Means* dan *K-Medoids*, proses *clustering* menggunakan metode jarak *Euclidean Distance*.

Perhitungan jarak berdasarkan pada transformasi yang hanya mempertahankan perilaku spektral sinyal ucapan yang telah diterapkan dalam pengakuan, tugas identifikasi dan verifikasi, dan dalam vocoding laju bingkai variabel. Dalam tes identifikasi pembicara, Pfeifer menunjukkan bahwa tanpa matriks bobot, hasil yang buruk diperoleh dengan menggunakan ukuran jarak *Euclidean* langsung pada koefisien filter inversi prediksi linier atau pada koefisien refleksi, sementara skor

tinggi diperoleh menggunakan *root mean square (rms) Euclidean* mengukur jarak antara tes dan referensi log spektra dari analisis prediksi linier. Pendekatan terakhir akan disebut sebagai ukuran spektral log rms.

2.1.7. *Euclidean Distance*

Euclidean Distance adalah metric yang paling sering digunakan untuk menghitung kesamaan dua vector. Dengan asumsi dua objek, A dengan koordinat (x1, y1) dan B dengan koordinat (x2, y2), jarak antara dua objek dapat diukur dengan:

$$\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \dots\dots\dots(2.5)$$

Jarak Euclidean Distance adalah jarak antar titik dalam garis lurus. Metode jarak ini menggunakan teorema *Pythagoras*. Dan merupakan perhitungan jarak yang paling sering digunakan dalam proses *machine learning*. Rumus dari *Euclidean Distance* dapat dilihat sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \dots\dots\dots(2.6)$$

Keterangan:

d_{ij} = jarak untuk menghitung kemiripan

n = jumlah vektor.

X_{ik} = Nilai atau data dari objek ke-I pada variabel ke-k

X_{jk} = Nilai atau data dari objek ke-j pada variabel ke-k

2.1.8. *Davies Buldin index*

Davies bouldin index (DBI) merupakan salah satu metode yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin. *Davies Bouldin Index* digunakan untuk mengevaluasi *cluster* secara umum berdasarkan perbandingan rasio *cluster*

ke-I dan cluster ke-j. semakin kecil nilai *Davies Bouldin Index* maka semakin baik cluster yang dihasilkan. Perhitungan nilai DBI disajikan pada persamaan (Adhitama et al., 2020).

Pengukuran dengan *Davies-Bouldin Index* ini memaksimalkan jarak inter-cluster antara cluster C_i dan C_j dan pada waktu yang sama mencoba untuk meminimalkan jarak antar titik dalam sebuah cluster. Jika jarak inter-cluster maksimal, berarti kesamaan karakteristik antar masing-masing cluster sedikit sehingga perbedaan antar-cluster terlihat lebih jelas. Jarak minimum dalam sebuah cluster berarti bahwa karakteristik (referensi) dari setiap objek dalam cluster sangat mirip. Untuk menghitung *Davies-Bouldin Index*:

4 1. *Sum of square within cluster (SSW)*

Untuk mengetahui kohesi dalam sebuah cluster ke-i adalah dengan menghitung nilai dari *Sum of square within cluster (SSW)*. Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat cluster dari sebuah cluster yang diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of square within cluster* adalah sebagai berikut :

$$SSW_i = \frac{1}{m_i} = \sum_{j=1}^{m_i} d(x_j, c_j) \dots\dots\dots(2.7)$$

4
Dari persamaan tersebut, m_i merupakan jumlah data dalam cluster ke-i, c_i adalah *centroid* cluster ke-i, dan $d()$ merupakan jarak setiap ke *centroid* yang dihitung menggunakan jarak *euclidean*.

2. *Sum of Square Between-cluster (SSB)*

Perhitungan *Sum of Square Between-cluster (SSB)* bertujuan untuk mengetahui separasi antar cluster. Persamaan yang digunakan untuk menghitung

nilai *Sum of Square Between-cluster* adalah sebagai berikut :

$$SSB_{ij} = d(c_i, c_j) \dots\dots\dots(2.8)$$

3. *Ratio* (Rasio)

Ratio bertujuan untuk untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j. *Cluster* yang baik adalah *cluster* dengan koehsi terkecil dan separasi yang sebesar mungkin. Nilai rasio dihitung dengan menggunakan rumus berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \dots\dots\dots(2.9)$$

Nilai rasio yang diperoleh tersebut digunakan untuk mencari *davies-bouldin index* (DBI) dari persamaan berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{ij} (R_{ij}) \dots\dots\dots(2.10)$$

Dari rumus ini, k adalah jumlah cluster yang digunakan. Semakin kecil nilai DBI yang dihasilkan (non-negatif) ≥ 0 , semakin baik cluster yang diperoleh dari pengelompokan.

2.1.9. *K-Fold Cross Validation*

K-fold cross validation merupakan metode pengujian data terhadap algoritma dengan cara membagi sejumlah data n-fold yang diinginkan kemudian data tersebut akan dibagi dalam n sebuah partisi dengan jumlah yang sama A,B,C,D, selanjutnya dilakukan proses *testing* sebanyak n kali. Dalam iterasi ke-n partisi A akan dijadikan data *testing* lalu sisanya akan menjadi data *training*. Langkah selanjutnya adalah mengacak nilai atribut input secara iteratif untuk mendapatkan hasil pengujian untuk beberapa atribut input (Pitria, 2019).

Pengujian ini dirancang untuk mengetahui seberapa akurat metode *K-Means*, *X-Means*, dan *K-Medoids*. menurut (Garcia, Luengo, & Herrera, 2015) Dalam bukunya yang berjudul “*Data Preprocessing in Data Mining*” dianjurkan menggunakan *10-folds cross validation* dalam membuat model. Pada Gambar 2.3 Dijelaskan ilustrasi evaluasi menggunakan *10-fold cross validation* (Ren, Li, & Han, 2019).

n-data										
	Percobaan A	Percobaan B	Percobaan C	Percobaan D	Percobaan E	Percobaan F	Percobaan G	Percobaan H	Percobaan I	Percobaan J
Percobaan 1	A	B	C	D	E	F	G	H	I	J
Percobaan 2	A	B	C	D	E	F	G	H	I	J
Percobaan 3	A	B	C	D	E	F	G	H	I	J
Percobaan 4	A	B	C	D	E	F	G	H	I	J
Percobaan 5	A	B	C	D	E	F	G	H	I	J
Percobaan 6	A	B	C	D	E	F	G	H	I	J
Percobaan 7	A	B	C	D	E	F	G	H	I	J
Percobaan 8	A	B	C	D	E	F	G	H	I	J
Percobaan 9	A	B	C	D	E	F	G	H	I	J
Percobaan 10	A	B	C	D	E	F	G	H	I	J
Keterangan :										
	Data Training									
	Data Testing									

Gambar 2.3 Skenario pengujian *10-fold cross validation*

2.1.10. Parametrik uji beda T-test

T-test adalah metode pengujian hipotesis yang menggunakan satu individu (objek penelitian) dalam dua proses yang berbeda. Walaupun dengan objek yang sama digunakan, tetapi sampel tetap terbagi menjadi dua. Yaitu data dari proses pertama dan data dari proses kedua. Kinerja dapat ditentukan dengan membandingkan keadaan objek pada penelitian pertama dengan keadaan objek pada penelitian kedua (Hasan, Hikmah, & Utami, 2018).

Uji beda T-test (uji T) merupakan salah satu uji yang digunakan untuk menguji bahwa tidak ada perbedaan yang signifikan antara dua mean sampel yang diambil secara acak dari populasi yang sama. Pengambilan keputusan ditentukan berdasarkan nilai signifikansi. Dasar pengujian hasil regresi biasanya ditetapkan

dengan nilai alpa sebesar 0,05. Jenis-jenis dari t-test antara lain:

1. *One Sample T-test*

One sample t-test adalah teknik analisis buat membandingkan satu variabel bebas. Teknik ini dipakai buat menguji apakah nilai eksklusif tidak selaras secara signifikan atau tidak menggunakan homogen-homogen sebuah sampel.

2. *Paired Sample T-test*

Analisis Paired sample t-test adalah mekanisme yang dipakai buat membandingkan homogen-homogen dua variabel pada satu grup. Artinya analisis ini bermanfaat buat melakukan pengujian terhadap satu sampel yang menerima suatu treatment lalu akan dibandingkan homogen-homogen menurut sampel tadi antara sebelum & sehabis treatment.

3. *Independent Sample T-test*

Independent sample t-test merupakan uji yang digunakan untuk memilih apakah dua sampel yang tidak berhubungan mempunyai homogen-homogen yang tidak selaras. Jadi tujuan metode ini merupakan membandingkan homogen-homogen dua kelompok yang tidak mempunyai interaksi satu sama lain.

2.1.11. Dataset

Set Data (*data set*) dapat dipandang sebagai kumpulan objek data. Nama lain yang sering digunakan adalah *record*, *point*, *vector*, *pattern*, *even*, *observation*, *case*, atau bahkan data. Sementara objek data digambarkan dengan sejumlah atribut yang menangkap (*capture*) karakter dasar objek data, contohnya tinggi badan yang memberikan nilai kuantitatif tinggi badan seseorang, waktu yang menangkap saat sebuah peristiwa terjadi. Atribut terkadang juga disebut variable, karakteristik,

medan (*field*), fitur, atau dimensi (Prasetyo, 2012).

2.1.12. Rapidminer

Rapidminer merupakan perangkat lunak yang bersifat terbuka (*open source*) yang dibuat oleh Dr. Markus Hofmann dari Institute of Technology Blanchardstown dan Ralf Klinkenberg dari rapid-i.com. Rapidminer adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, text mining dan analisis prediksi. Rapidminer menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapidminer memiliki kurang lebih 500 operator *data mining*, termasuk operator untuk input, output, data preprocessing dan visualisasi. Rapidminer merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. Rapidminer ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi (Fitriyadi & ana kurniawati, 2021).

2.2. Kajian Penelitian terdahulu

Guna mempermudah dalam melihat dukungan hasil penelitian terdahulu yang berkaitan dengan landasan perumusan hipotesis disajikan pada table 2.1 sebagai berikut:

Tabel 2.1 Penelitian terdahulu

Peneliti	Hasil
(Adhitama et al., 2020), ”Penentuan Jumlah Cluster Ideal SMK Di Jawa Tengah Dengan Metode <i>X-Means Clustering</i> dan <i>K-Means Clustering</i> ”	Hasil dari penelitian tersebut yang telah dilakukan menunjukkan nilai DBI yang konsisten antara <i>X-Means</i> dan <i>K-Means</i> clustering untuk mengukur hasil cluster persebaran data pokok SMK dengan enam buah
(Wahidin & Sensuse, 2021). ” Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air”	Hasil dari penelitian tersebut yang telah dilakukan pengukuran kinerja dengan melakukan perbandingan nilai DBI dari algoritma <i>K-Means</i> , <i>X-Means</i> Dan <i>K-Medoids</i> yang telah dilakukan dengan menggunakan 100 data, Algoritma <i>K-Means</i> menghasilkan nilai DBI lebih kecil yaitu 0,792 dibandingkan dengan nilai DBI algoritma <i>X-Means</i> sebesar 0.812 dan algoritma <i>K-Medoids</i> sebesar 1,700, sehingga <i>K-Means</i> menjadi algoritma terbaik untuk klasterisasi awak kabin lion air.
(Muhamad, Prasajo, Sugianto, Surtiningsih, & Cholissodin, 2017), ”Optimasi <i>Naïve Bayes Classifier</i> Dengan Menggunakan <i>Particle Swarm Optimization</i> Pada Data Iris ”	Hasil dari penelitian tersebut Pada optimasi <i>Naïve Bayes Classifier</i> dengan menggunakan <i>Particle Swarm Optimization</i> pada data iris. Klasifikasi dilakukan dengan menentukan bobot atribut optimum dengan menggunakan <i>Particle Swarm Optimization</i> . Hasil klasifikasi diperoleh dari <i>fitness</i> tertinggi.
(Candrasari Hermanto, 2017), ”Analisis Algoritma Clustering Dalam Kasus	Berdasarkan hasil pengujian penelitian tersebut membandingkan dua algoritma clustering yaitu <i>K-Means</i> dan <i>K-Medoids</i> , dengan

Peneliti	Hasil
Penentuan ⁵ Jenis Bunga Iris”	membandingkan nilai Davies Bouldin Index dan nilai dari number of cluster dengan uji beda parametrik t-test dihasilkan algoritma <i>K-Medoids</i> memiliki nilai <i>Davies Bouldin index</i> sebesar 0.291 tetapi nilai number of cluster paling sedikit 2.000,sifatnya tidak dominan dengan lainnya. Sedangkan algoritma <i>K-Means</i> Memiliki nilai <i>Davies Bouldin Index</i> Paling rendah yaitu 0.167, tetapi memiliki sifat paling dominan diantara ke dua algoritma lainnya
(G.Soni & Patel, 2017), ”Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data”	Hasil Penelitian tersebut menunjukkan keunggulan K-medoids dari K-means dalam hal waktu eksekusi, kualitas kelas clustered dan juga jumlah record
(Muzaki, 2019), “Penerapan Triangular Kernel Nearest Neighbor Sebagai Metode Clustering Dasar Pada Metode Bagging”	Hasil Penelitian tersebut Multi-clustering bagging dengan kombinasi TKNN dan Ward’s linkage dapat memberikan kualitas hasil cluster yang lebih baik dari metode single clustering yang dihasilkan, terutama untuk <i>dataset iris, flame, pathbased1, aggregation, R15</i> dan <i>D31</i> .

BAB III

METODOLOGI PENELITIAN

3.1. Metode Penelitian

Dalam penelitian ini jenis penelitian yang digunakan ialah kausal komparatif atau perbandingan. Penelitian ini berfokus pada hubungan yang mempengaruhi variabel-variabel yang diselidiki, serta memberikan alasan dan penjelasannya (Hasibuan, 2007). Peneliti memilih penelitian komparatif karena komparatif merupakan penelitian yang menggunakan teknik membandingkan suatu objek dengan objek lain. Pada penelitian ini peneliti berusaha mencari tahu tentang Algoritma yang terbaik untuk pengelompokan data iris dan apakah ada perbedaan signifikan terhadap algoritma *X-Means*, *K-Means* dan *K-Medoids*. Menurut Emzir (2010) Metode penelitian kausal komparatif memiliki 5 alur penelitian yaitu Permasalahan penelitian, Menentukan kelompok data, pemilihan kelompok pembanding, pengumpulan data dan analisa data. Hasil penelitian ini kemudian diharapkan dapat dijadikan pedoman perbandingan peneliti lain dalam menganalisis menggunakan metode clustering.

3.2. Metode Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data sekuler, data diambil dari *Machine Learning Repository UCI* (Universities California, Invence) dengan alamat web <https://archive.ics.uci.edu/> yang mana data tersebut berupa data Iris yang bersifat data publik.

3.3. Metode Analisis Data

Pemrosesan *data mining* dalam penelitian ini akan digunakan metode clustering. Metode clustering termasuk dalam *unsevised learning*. Algoritma pada metode clustering yang akan dipakai dalam penelitian ini adalah algoritma *K-Means*, *X-Means* dan *K-Medoids*. Ketiga algoritma tersebut merupakan metode clustering non hierarki, yang mana algoritma *X-Means* dan *K-Medoids* merupakan variasi dari *K-Means*.

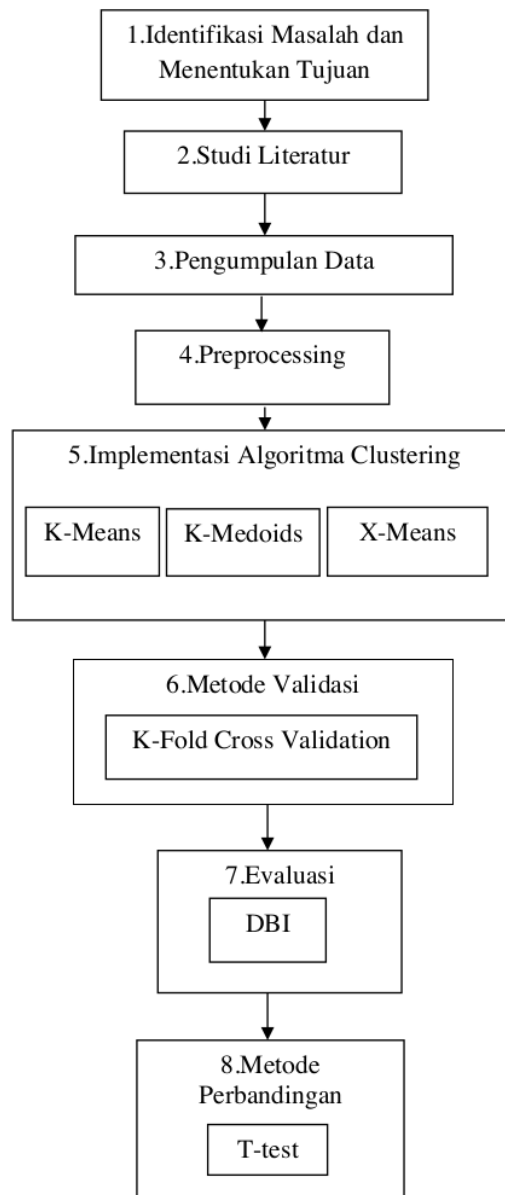
Alat yang digunakan dalam penelitian ini menggunakan tools Rapid Miner. Dimana proses mining yang dimulai dari pemrosesan awal (*preprocessing*), pemodelan dari masing-masing algoritma serta melakukan validasi, perbandingan evaluasi, dan pengujian uji beda.

3.3.1. Jenis Analisis

Jenis Analisis yang dilakukan dalam penelitian menggunakan uji beda t-test yang tergolong dalam uji perbandingan (komparatif) yang bertujuan untuk membandingkan apakah masing-masing algoritma yang diuji perbeda secara signifikan atau tidak.

3.4. Alur Penelitian

Adapun alur penelitian dalam penelitian ini akan digunakan pada bagan Gambar 3.1:



Gambar 3.1 Alur penelitian

Keterangan:

1. Identifikasi masalah dan Menentukan Tujuan

Tahapan pertama dalam penelitian ini adalah identifikasi masalah yang mana ⁵ banyak penelitian yang hanya masih menggunakan satu algoritma *clustering* dalam penentuan jenis bunga *iris*. Terdapat permasalahan dalam bagaimana menetapkan performa clustering yang signifikan dari komparasi Algoritma clustering.

2. Studi Literatur

Pada tahap ini, peneliti melakukan studi literatur dengan melakukan pemahaman lebih lanjut terhadap teknik *data mining*. Dalam pemahaman serta penjelasan konsep ini, penulis mendapatkan dari buku-buku referensi, jurnal penelitian, paper, artikel yang didapat dari internet, maupun literatur lainnya yang berkaitan dengan metode yang digunakan serta objek yang diteliti.

3. Pengumpulan data

Tahapan ini merupakan tahapan untuk mempersiapkan data yang akan digunakan pada penelitian ini. Data yang digunakan yaitu *Data Set Iris*

4. *Preprocessing* Data

Sebelum melakukan proses *data mining*, perlu Preprosesing data dilakukan ⁶ dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem. Proses ini dilakukan agar dapat memperoleh hasil yang benar-benar akurat. Preprosesing data ⁶

dapat berupa data cleaning, data integration, data reduction, dan data transformation.

5. Implementasi Algoritma *Clustering*

Tahap selanjutnya yakni Proses penentuan komparasi clustering dengan *data set iris* dilakukan dengan pelatihan model algoritma *K-Means*, *X-Means* dan *K-Medoids*.

6. Metode Validasi

Kemudian untuk validasi, penelitian ini menggunakan *10-fold cross validation*. *10-fold cross validation* akan mengulang pengujian sebanyak 10 kali.

7. Metode Evaluasi

Pada tahap ini peneliti melihat hasil evaluasi dari Nilai *Davies Bouldin Index* (DBI), kemudian hasil tersebut akan digunakan oleh peneliti untuk perbandingan performance dari masing-masing algoritma.

8. Metode Perbandingan

Peneliti menggunakan metode perbandingan uji beda parametrik t-test untuk membandingkan nilai *Davies Bouldin Index* (DBI) algoritma clustering. Nilai performa yang diperoleh dibandingkan menggunakan t-test untuk memastikan apakah ada perbedaan signifikan pada algoritma clustering.

BAB IV

4 HASIL DAN PEMBAHASAN

Bab ini merupakan penjelasan tentang pemodelan dan pengujian metode Algoritma X-Means, K-Means dan K-Medoids dengan ditetapkannya Hasil Evaluasi nilai DBI, serta Hasil uji beda menggunakan T-test yang didukung oleh tools RapidMiner.

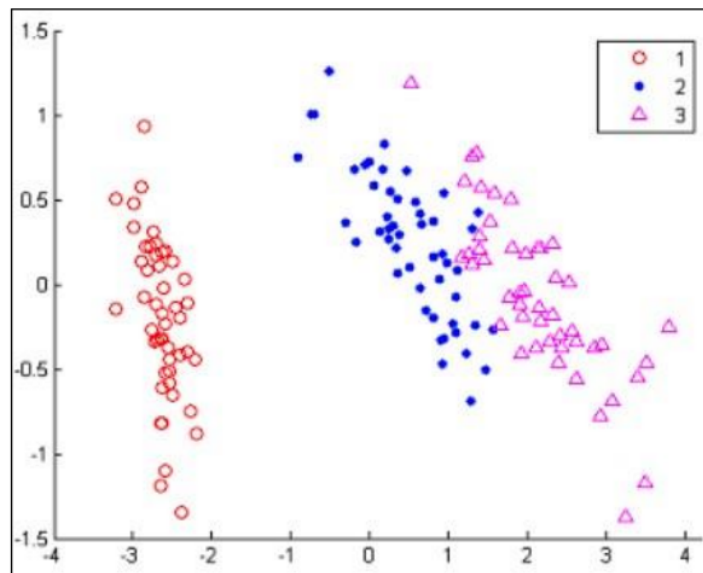
4.1. Pengumpulan Data

Adapun data yang digunakan dalam penelitian ini merupakan data sekunder yang dipublikasi oleh UCI *Repository*. *Data set* iris ini mempunyai jumlah data sebanyak 150 dengan 3 kategori yaitu 50 Iris virginica, 50 Iris setosa, dan 50 Iris Versicolor. Jumlah Atribut pada *data set* iris ini ada 5 yaitu 4 atribut bertipe Real, dan 1 atribut bertipe Nominal. 5 atribut tersebut adalah *Sepal length* (cm), *Sepal width* (cm), *Petal length* (cm), *Petal width* (cm), Species *Iris-setosa*, *Iris-virginica* dan *Iris-versicolor*. Tabel 4.1 menunjukkan informasi dalam dataset Iris mengenai jumlah data, jumlah atribut, dan jumlah kelas. Visualisasi yang menunjukkan distribusi dan bentuk data ditunjukkan pada Gambar 4.1.

Tabel 4.1 *Data sets* yang digunakan

Id	label	Sepal length	Sepal width	Petal length	Petal width
id_1	Iris-setosa	5.1	3.5	1.4	0.2
id_2	Iris-setosa	4.9	3.0	1.4	0.2
id_3	Iris-setosa	4.7	3.2	1.3	0.2

Id	label	Sepal length	Sepal width	Petal length	Petal width
id_4	Iris-setosa	4.6	3.1	1.5	0.2
id_5	Iris-setosa	5.0	3.6	1.4	0.2
id_6	Iris-setosa	5.4	3.9	1.7	0.4
id_7	Iris-setosa	4.6	3.4	1.4	0.3
id_8	Iris-setosa	5.0	3.4	1.5	0.2
id_9	Iris-setosa	4.4	2.9	1.4	0.2
...
id_150	Iris-virginica	5.9	3	5.1	1.8



Gambar 4.1 Visualisasi data iris (Muzaki, 2019)

4.2. Preprocessing Data

Preprocessing data merupakan sebuah tahap awal yang harus dilakukan pada

data mining. Tujuan praproses dalam *data mining* adalah untuk mempersiapkan data mentah sebelum dilakukan proses lain. *Preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem. *Preproses* juga dilakukan untuk mendapatkan hasil yang lebih akurat, pengurangan waktu perhitungan untuk large scale problem, dan membuat nilai data menjadi lebih kecil tanpa merubah informasi yang didalamnya. *Preprocessing* data dapat berupa data *cleaning*, data *integration*, data *reduction*, dan data *transformation*.

4.2.1. Transformasi Data

Tahap *preprocessing* dalam penelitian ini hanya menggunakan data *Transformation* agar data dapat diolah pada tahap *data mining*. Pada beberapa dataset iris terdapat rentang nilai yang berbeda disetiap atribut. Perbedaan rentang nilai pada setiap atribut menyebabkan tidak berfungsinya atribut yang memiliki nilai jauh lebih kecil dibandingkan dengan atribut-atribut lainnya. Oleh karena itu, diperlukan adanya transformasi data dengan normalisasi untuk menyamakan rentang nilai pada setiap atribut dengan skala tertentu. Agar dapat menghasilkan *data mining* yang lebih baik. Peneliti melakukan Transformation data dengan menggunakan metode *Z-score normalization*. *Z-score normalization* merupakan metode normalisasi berdasarkan *mean* (nilai rata-rata) dan *standard deviation* (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual minimum dan maksimum dari data. Proses *Z-score normalisasi* ditunjukkan dalam persamaan 4.1 Dan hasil normalisasi terdapat pada Tabel 4.2

$$v^1 = \frac{v - \mu_a}{\sigma_a} \dots\dots\dots(4.1)$$

Keterangan:

v^1 = Data ternormalisasi (data baru)

v = Data awal

μ_a = *mean*

σ_a = *standard deviation*

Tabel 4.2 Hasil normalisasi data

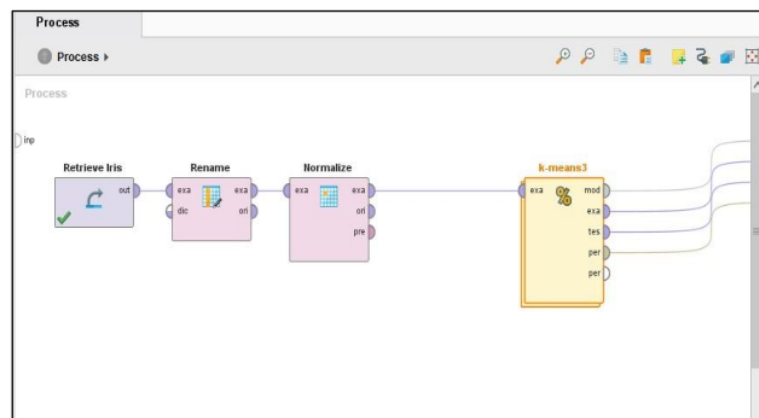
Id	label	Sepal length	Sepal width	Petal length	Petal width
id_1	Iris-setosa	-0.898	1.029	-1.337	-1.309
id_2	Iris-setosa	-1.139	-0.125	-1.337	-1.309
id_3	Iris-setosa	-1.381	0.337	-1.393	-1.309
id_4	Iris-setosa	-1.501	0.106	-1.280	-1.309
id_5	Iris-setosa	-1.018	1.259	-1.337	-1.309
id_6	Iris-setosa	-0.535	1.951	-1.167	-1.047
id_7	Iris-setosa	-1.501	0.798	-1.337	-1.178
id_8	Iris-setosa	-1.018	0.798	-1.280	-1.309
id_9	Iris-setosa	-1.743	-0.355	-1.337	-1.309
id_10	Iris-setosa	-1.139	0.106	-1.280	-1.440
...
Id_150	Iris-virginica	0.068	-0.125	0.760	0.788

4.3. Implementasi Algoritma Clustering

Setelah melalui proses *preprocessing* kemudian hasil *preprocessing* digunakan untuk proses mining menggunakan algoritma *K-Means*, *X-Means* dan *K-Medoids*. Dataset yang digunakan berisi 150 record yang telah dinormalisasi. nilai k yang digunakan adalah tiga kluster dan dilakukan dengan bantuan *framework data mining* yaitu Rapidminer. Dalam Implementasi menggunakan *Davies Bouldin Index* (DBI) juga menggunakan validasi *X-fold cross validation* untuk membagi data menjadi data training dan testing tujuannya mencari nilai interval yang nantinya akan diolah dalam proses uji beda T-test.

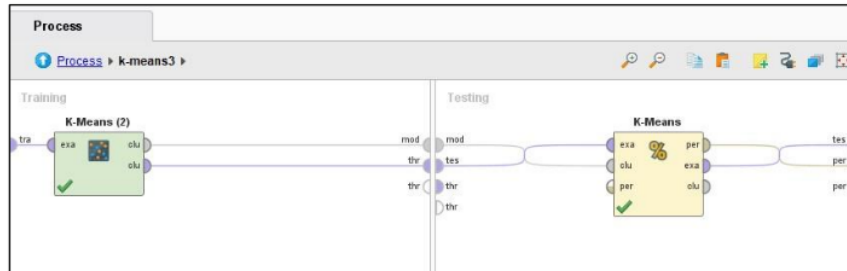
4.3.1. Pemodelan *K-Means*

Pada dataset yang telah ditentukan model pertama dari metode yang akan digunakan adalah *K-Means* dengan grafik dari software RapidMiner yang digunakan dapat dilihat pada gambar 4.2 .



Gambar 4.2 Rancangan model *k-means*

Dari pemodelan *K-Means* pada Gambar 4.2, terdapat desain baru di dalam Cross Validation. Desain bagian dari Cross Validation yang ditunjukkan pada Gambar 4.3



Gambar 4.3 Desain bagian cross validation

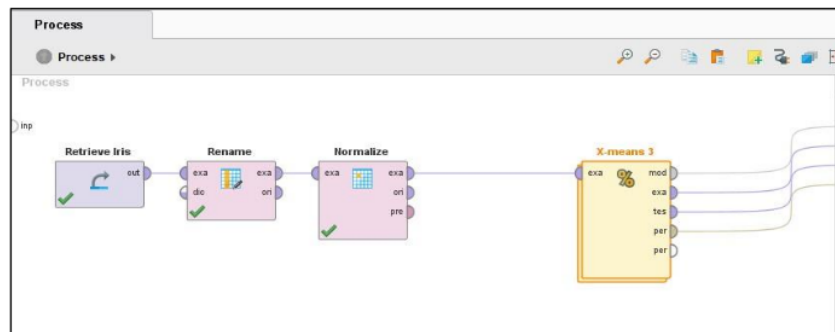
Dari pemodelan memperoleh hasil antar *centroid* untuk nilai *micro average cluster 0* sebesar -0,675, nilai *micro average cluster 1* sebesar -1,238, dan nilai *micro average cluster 2* sebesar -0,943 selain mendapatkan hasil mengenai jarak rata-rata antar *centroid* proses *cluster* juga mendapatkan dan nilai dari *Davies Bouldin Index (DBI)* untuk nilai *DBI* mendapatkan nilai *micro average* sebesar -0,838 untuk lebih jelasnya lagi dapat dilihat pada Gambar 4.4 :

PerformanceVector	
Performance	PerformanceVector:
	Avg. within centroid distance: -0.975 +/- 0.275 (micro average: -0.975)
	Avg. within centroid distance_cluster_0: -0.675 +/- 0.207 (micro average: -0.675)
	Avg. within centroid distance_cluster_1: -1.238 +/- 0.645 (micro average: -1.238)
	Avg. within centroid distance_cluster_2: -0.943 +/- 0.364 (micro average: -0.943)
Description	Davies Bouldin: -0.838 +/- 0.107 (micro average: -0.838)
Annotations	

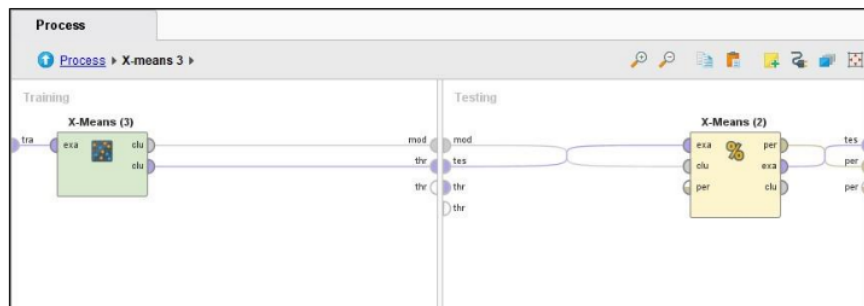
Gambar 4.4 Hasil *description performance vector*

4.3.2. Pemodelan *X-Means*

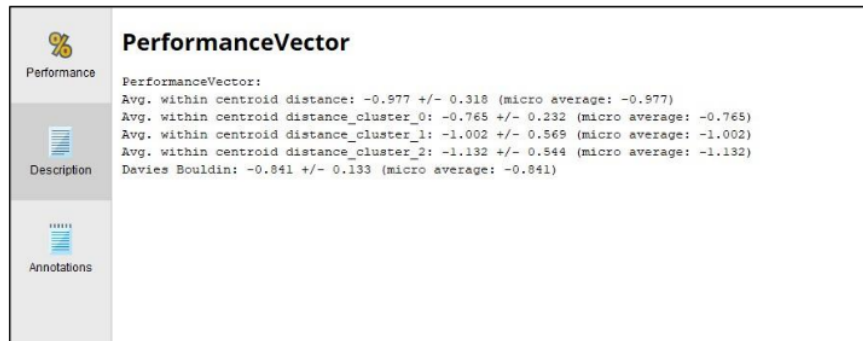
Pada dataset yang telah ditentukan model kedua dari metode yang akan digunakan adalah *X-Means* dengan grafik dari *software* RapidMiner yang digunakan dapat dilihat Gambar 4.5 :

Gambar 4.5 Rancangan model *x-means*

Dari pemodelan *X-Means* pada Gambar 4.5, terdapat desain baru di dalam *Cross Validation*. Desain bagian dari *Cross Validation* yang ditunjukkan pada Gambar 4.6

Gambar 4.6 Desain bagian *cross validation*

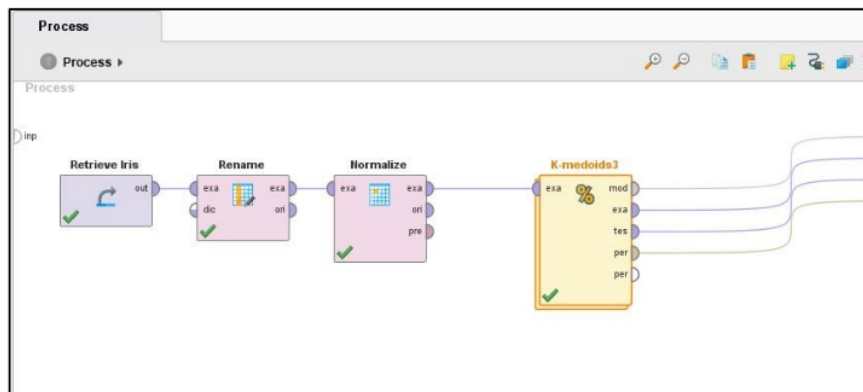
Hasil dari pemodelan memperoleh hasil antar *centroid* untuk nilai *micro average cluster* 0 sebesar -0,765, nilai *micro average cluster* 1 sebesar -1,002, dan nilai *micro average cluster* 2 sebesar -1,132 selain mendapatkan hasil mengenai jarak rata-rata antar *centroid* proses cluster juga mendapatkan dan nilai dari *Davies Bouldin Index* (DBI) untuk nilai DBI mendapatkan nilai *micro average* sebesar -0,841 untuk lebih jelasnya lagi dapat dilihat pada Gambar 4.7 .



Gambar 4.7 Hasil *description performance vector*

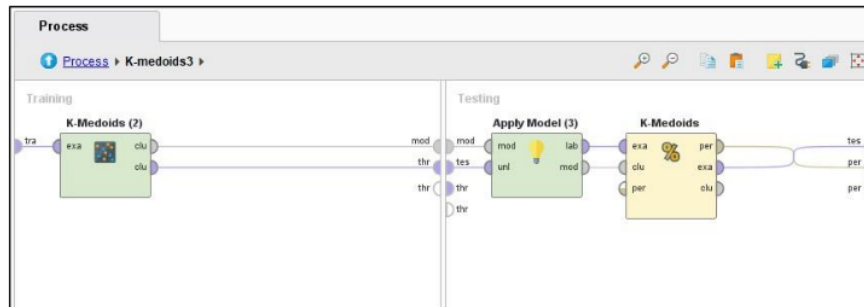
4.3.3. Pemodelan K-Medoids

Pada dataset yang telah ditentukan model kedua dari metode yang akan digunakan adalah *K-Medoids* dengan grafik dari software RapidMiner yang digunakan dapat dilihat pada Gambar 4.8:



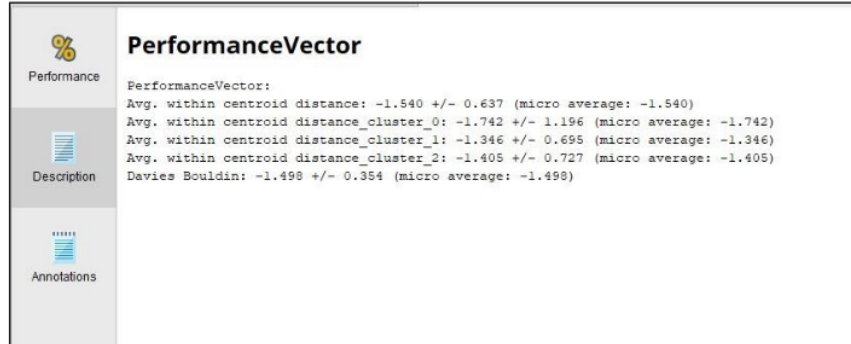
Gambar 4.8 Rancangan model *k-medoids*

Dari pemodelan *K-Medoids* pada Gambar 4.8, terdapat desain baru di dalam *Cross Validation*. Desain bagian dari *Cross Validation* yang ditunjukkan pada Gambar 4.9 .



Gambar 4.9 Desain bagian cross validation

Hasi dari pemodelan memperoleh hasil antar *centroid* untuk nilai *micro average cluster* 0 sebesar -1,742, nilai *micro average cluster* 1 sebesar -1,346, dan nilai *micro average cluster* 2 sebesar -1,405 selain mendapatkan hasil mengenai jarak rata-rata antar *centroid* proses *cluster* juga mendapatkan dan nilai dari *Davies Bouldin Index* (DBI) untuk nilai DBI mendapatkan nilai *micro average* sebesar -1,498 untuk lebih jelasnya lagi dapat dilihat pada Gambar 4.10 .



Gambar 4.10 Hasil description performance vector

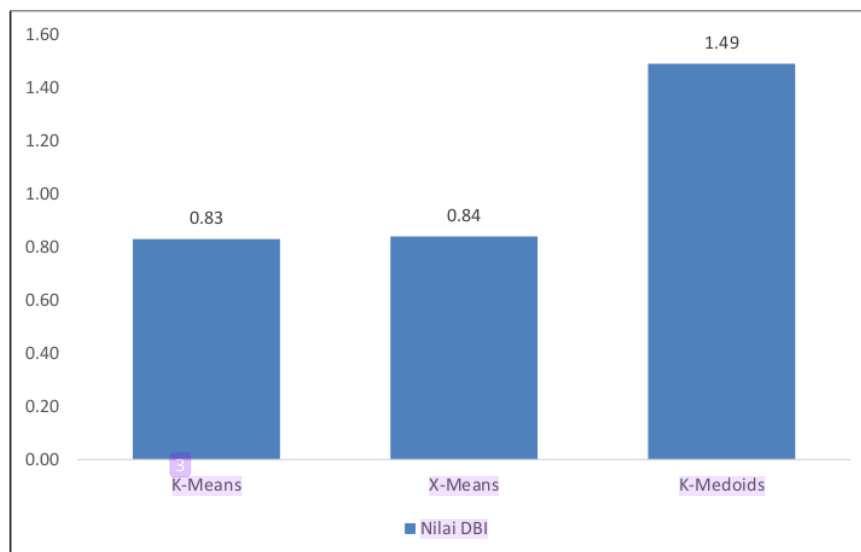
4.4. Evaluasi

Hasil clustering telah diperoleh bahwa dari masing-masing pemodelan algoritma menghasilkan nilai *Davies Bouldin Index* (DBI), maka tahap selanjutnya adalah membandingkan nilai *Davies Bouldin Index* (DBI). Untuk kemudian dicari hasil

nilai terkecil dari seluruh percobaan. Hasil nilai *Davies Bouldin Index* (DBI) dari jumlah Cluster dapat ditunjukkan pada tabel 4.3

Tabel 4.3 Perbandingan performance tiga algoritma

Algoritma	Davies Bouldin Index
<i>K-Means</i>	0,838
<i>X-Means</i>	0,841
<i>K-Medoids</i>	1,498



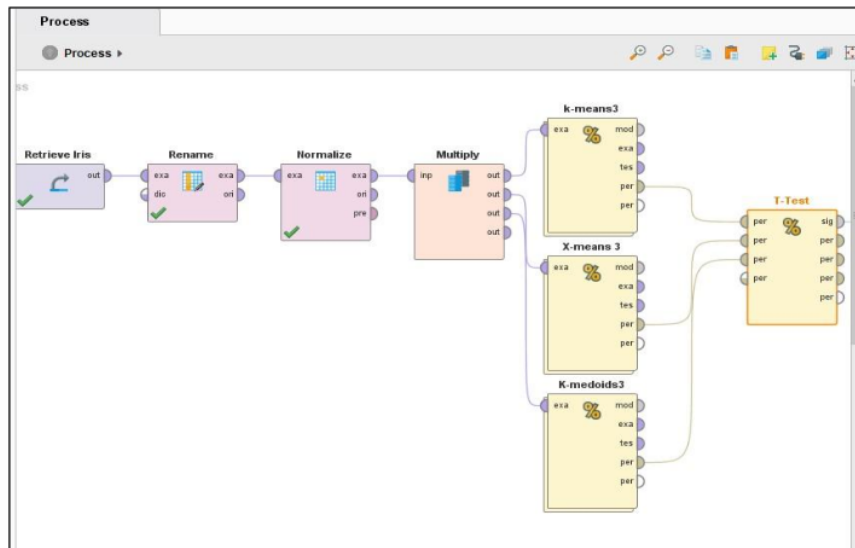
Gambar 4.11 Visualisasi komparasi nilai DBI

Kemudian pada Tabel 4.3 Tersebut Dari hasil pengujian nilai performa algoritma *K-Means*, *X-Means* dan *K-Medoids* didapatkan nilai *Davies Bouldin Index* (DBI) untuk Algoritma *K-Means* sebesar 0,838. Kemudian untuk Algoritma *X-Means* sebesar 0,841. Sedangkan untuk Algoritma *K-Medoids* mendapatkan nilai sebesar 1,498. Visualisasi yang menunjukkan komparasi nilai DBI ditunjukkan pada Gambar 4.11 jadi dapat ditarik kesimpulan bahwa nilai *K-Means* memperoleh

nilai terkecil untuk pengujian *Davies Bouldin Index (DBI)*.

4.5. Metode Perbandingan

Setelah mendapatkan nilai DBI dari 3 algoritma, penelitian selanjutnya adalah melakukan pengujian dengan uji beda t-test. Desain dari Perbandingan uji beda T-test ditunjukkan pada Gambar 4.12.



Gambar 4.12 Desain perbandingan uji beda t-test

Pada pengujian ini, uji beda t-test memiliki nilai *signifikan* yang berbeda pada setiap algoritmanya. Berikut ini Tabel 4.4 dari hasil uji beda T-test dari keseluruhan algoritma *cluster*.

Tabel 4.4 Hasil uji beda t-test

	K-Means	X-Means	K-Medoids
K-Means	-	0,953	0,000
X-Means	0,957	-	0,000
K-Medoids	0,000	0,000	-

Dari hasil uji beda t-test diatas, diketahui bahwa untuk hasil variabel antara algoritma *K-Means* dengan *X-Means* diperoleh nilai signifikansi yang lebih kecil dari 0,05 menunjukkan adanya perbedaan yang signifikan yaitu: algoritma *K-Means* dengan *K-Medoids*, algoritma *X-Means* dengan *K-Medoids*. Sedangkan untuk algoritma *K-Means* dengan *X-Means* menunjukkan hasil nilai signifikansi lebih besar dari 0,05 sehingga tidak ada perbedaan yang signifikan.

BAB V

PENUTUP

4.6. Kesimpulan

Dari hasil penelitian yang telah dilakukan pada data bunga iris maka dapat disimpulkan bahwa metode *clustering data mining* Dari hasil pengujian nilai performa algoritma *K-Means*, *X-Means* dan *K-Medoids* yang ditunjukkan pada Tabel 4.2 didapatkan nilai *Davies Bouldin Index* (DBI). hasil terbaik berdasarkan nilai *Davies-Bouldin Index* terdapat pada algoritma *K-Means* dengan nilai DBI sebesar 0,838, nilai tersebut dikatakan baik karena hasil tersebut merupakan nilai terkecil yang diperoleh, karena semakin kecil nilai DBI atau semakin mendekati nilai nol maka semakin optimum kluster yang dihasilkan. Untuk pengujian uji beda T-test dari masing-masing algoritma *K-Means*, *X-Means* dan *K-Medoids* yang ditunjukkan pada Tabel 4.3 didapatkan kesimpulan bahwa ada perbedaan yang signifikan antara Algoritma *K-Means* dengan *K-Medoids*, Algoritma *X-Means* dengan *K-Medoids*. Sedangkan untuk Algoritma *K-Means* dengan *X-Means* menunjukkan hasil lebih dari nilai alpha sehingga tidak ada perbedaan yang signifikan. Dengan demikian Performa yang paling optimal dari ketiga algoritma dalam pemrosesan *data set* iris yaitu Algoritma *K-Means* dan *X-Means*.

4.7. Saran

Agar penelitian ini bisa ditingkatkan berikut ini adalah saran-saran untuk mendapatkan hasil yang lebih baik:

1. Penelitian ini dapat dikembangkan lebih lanjut dengan melakukan uji beda

T-Test dengan membandingkan tiga algoritma untuk melihat algoritma mana yang lebih dominan atau signifikan.

2. Penelitian ini dapat dikembangkan lagi dengan membandingkan dengan metode lainnya seperti *Self-Organizing Map* (SOM), *Fuzzy C-Means* (FCM), dan lain-lain.
3. Tidak semua kasus atau permasalahan harus diselesaikan dengan satu algoritma pada *data mining*. Karena belum tentu algoritma yang digunakan merupakan algoritma yang paling akurat. Oleh karena itu untuk menentukan algoritma yang paling akurat ini perlu dilakukan komparasi beberapa algoritma.


DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Ananda, R. (2020). Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 1–5. <https://doi.org/10.33387/jiko.v3i1.1635>
- Awalludin, A. S., & Taufik, I. (2017). *Analisis Cluster Data Longitudinal pada Pengelompokan Daerah Berdasarkan Indikator IPM di Jawa Barat*. 978, 187–194.
- Candrasari Hermanto, D. M. (2017). ANALISIS ALGORITMA CLUSTERING DALAM KASUS PENENTUAN JENIS BUNGA IRIS. *Jurnal Pendidikan Teknologi Dan Kejuruan*.
- Emzir. (2010). *Metodologi Penelitian Pendidikan Kuantitatif dan Kualitatif*. Jakarta: Raja Grafindo Persada.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*.
- Fitriyadi, akhmad upi, & ana kurniawati. (2021). Algoritma K-Means dan K-Medoids Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional. *Kilat*, 10(1), 157–168. <https://doi.org/10.33322/kilat.v10i1.1174>
- G.Soni, K., & Patel, D. A. (2017). Comparative analysis of FCM and HCM algorithm on Iris data set. *International Journal of Computer Applications*, 5(2), 33–37. <https://doi.org/10.5120/888-1261>
- Garcia, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining*.
- Hasan, F. N., Hikmah, N., & Utami, D. Y. (2018). Perbandingan Algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung jawab BSI Entrepreneur Center. *Jurnal Pilar Nusa Mandiri*, 14(2), 169. <https://doi.org/10.33480/pilar.v14i2.908>
- Hasibuan, Z. A. (2007). Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi. *Konsep, Teknik, Dan Aplikasi*, (Universitas Indonesia), 194.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*.
- Herviany, M., Delima, S. P., Nurhidayah, T., & Kasini. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 34–40.
- James and others, M. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.
- Marlina, D., Lina, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle

- Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 180. <https://doi.org/10.25126/jtiik.201743251>
- Muzaki, M. (2019). Penerapan Triangular Kernel Nearest Neighbor Sebagai Metode Clustering Dasar Pada Metode Bagging. *Jurnal Teknologi Dan Riset Terapan (JATRA)*, 1(1 SE-Research Articles). Retrieved from <https://jurnal.polibatam.ac.id/index.php/JATRA/article/view/1313>
- Pitria, P. (2019). Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes. *Undergraduate Theses from JBPTUNIKOMPP*.
- Pramudiono, Iko. (2006). *Apa itu data mining?*
- Prasetyo, eko. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*.
- Ren, Q., Li, M., & Han, S. (2019). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives. *Big Earth Data*, 3(1), 8–25. <https://doi.org/10.1080/20964471.2019.1572452>
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 602–610. Retrieved from <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Taha Chicho, B., Mohsin Abdulazeez, A., Qader Zeebaree, D., & Assad Zebari, D. (2021). Machine Learning Classifiers Based Classification For IRIS Recognition. *Qubahan Academic Journal*, 1(2), 106–118. <https://doi.org/10.48161/qaj.v1n2a48>
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Milan: WILEY.
- Wahidin, A. J., & Senses, D. I. (2021). Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *Jurnal ICT: Information Communication & Technology*, 20(2), 298–302. <https://doi.org/10.36054/jict-ikmi.v20i2.387>
- Wurdianarto, S., Novianto, S., & Rosyidah, U. (2014). Perbandingan Euclidean Distance Dengan Canberra Distance Pada Face Recognition. *Techno.Com*, 13(1), 31–37. Retrieved from <https://publikasi.dinus.ac.id/index.php/technoc/article/view/539>
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, 2(2), 226–235. <https://doi.org/10.3390/j2020016>
- Zizwan, P. A., Zarlis, M., & Nababan, E. B. (2020). *Bigdata Clustering using X-means method with Euclidean Distance*. <https://doi.org/10.1088/1742-6596/1566/1/012103>

LAMPIRAN-LAMPIRAN

Lampiran 1 Berita Acara



PERSETUJUAN BAHU

BERITA ACARA KEMAJUAN PEMBIMBINGAN PENULISAN KARYA TULIS ILMIAH

1. NAMA MAHASISWA : Muhammad Zainal Arifin
 NPM : 191.03.03.0023
 Fak/Jur/Prodi : Teknik / SISTEM INFORMASI
 Alamat Rumah : Ds. Pandan, kec. Durenan, kab. Trenggalek
 Alamat email : ~~muhammad~~ Muhammad Zainal 213@gmail.com
 No. Telp. / HP : 0856115 922 522

2. DOSEN PEMBIMBING I : AIDINA RIST YAWAN, M.Kom
 Alamat Rumah : Jl. Kloti Mangrove RT.43/Rw.03 Dsa Sragi, Ds. Dika
 Alamat email : aidinaristy@unpkediri.ac.id
 No. Telp. / HP : 0812326244817

3. DOSEN PEMBIMBING II : M. NASIBULLAH MUZAM, M.Cs
 Alamat Rumah : Ds. Karangsejo, kec. Ngasem, kab. Kediri
 Alamat email : muhammadnasibullah@gmail.com
 No. Telp. / HP : 085 245 503 242

4. JUDUL KTI : Peningkatan Performa Cluster Pulau Pemantren Jenis Burayud 10's

Catatan :

1. Periode Bimbingan (Sesuai SK Rektor) : _____
2. Jadwal Bimbingan : _____

	Hari	Pukul	Tempat / Ruang
Pembimbing I	Selasa	13.00	Pradi 51
	Kamis	13.00	Pradi 51
Pembimbing II	Senin	10.00	Pradi 51

3. Kemajuan Bimbingan

Pembimbing I

NO.	TANGGAL	MATERI	MASALAH	TT. DOSEN
1.	10-06-2022	BAB I	LATAR belakang masalah diaraniki	
2.	14-06-2022	BAB I	DATASAH MASALAH konsistensi mata	
3.	17-06-2022	BAB II	KAJIAN TEORI Hafidz Lengkari	
4.	24-06-2022	BAB II	Paragraf dan Indole Pendidikan Terpadu	
5.	28-06-2022	BAB III	Penumbuhan Alur Penelitian	
6.	01-07-2022	BAB IV	Perbaikan Hasil dan Pembahasan	
7.	08-07-2022	BAB V	Penumbuhan hasil Riset dan Survei	
8.	08-07-2022	BAB V		
			ACE Sidang	

Pembimbing II

NO.	TANGGAL	MATERI	MASALAH	TT. DOSEN
1.	02-06-2022	BAB I	Penulisan Latar belakang	
2.	09-06-2022	BAB I	meningkatkan gambar dan gambar iris	
3.	16-06-2022	BAB I	Perbaikan Rumusan masalah	
4.	23-06-2022	BAB II	Perbaikan Penulisan kajian Teori	
5.	30-06-2022	BAB II	meningkatkan ketepatan kajian Teori	
6.	04-07-2022	BAB III	Perbaikan Penulisan Metodologi Penelitian	
7.	07-07-2022	BAB III	Perbaikan Penulisan hasil Riset dan Survei	
			ACE Sidang	

Mengetahui,
Kaprod

Rita F. S. S. M. M. M.
NIDN 9731087703

Kediri, 14 Juli - 2022
Mahasiswa Yhs

Muhammad Zuhairi
NPM 19102030022

Lampiran 2 Berita Acara



YAYASAN PEMBINA LEMBAGA PENDIDIKAN PERGURUAN TINGGI PGRI KEDIRI
UNIVERSITAS NUSANTARA PGRI KEDIRI

Status "Terakreditasi Baik Sekali"

SK BAN PT Nomor 671/SK/BAN-PT/Akred/PT/VII/2021, Tanggal 21 Juli 2021

Alamat : Jalan K.H. Achmad Dahlan Nomor 76, Telepon& Faksimile (0354) 771576, 771503 Kediri

website : <http://www.unpkediri.ac.id/> email : admin@unpkediri.ac.id

FAK ILMU PEND. & KEGURUAN:	6 PGSD	FAK TEKNIK:	FAK ILMU KEH. & SAINS:	6 D3 Kelemban
1. SI Biologi dan Kesehatan	7. PG PAUD	1. D3 Teknik Mesin	1. D3 Keperawatan	
2. SI Pend. Sejarah	FAK EKONOMI:	2. D3 Teknik Industri	2. S1 Perikanan	PASCASARJANA:
3. SI Pend. PPKn	1. SI Manajemen	3. D3 Teknik Elektro	3. S1 Pend. Matematika	1. S2 Keguruan Olahraga
4. SI Pend. Etno- dan Sntro Indonesia	2. SI Akuntansi	4. S1 Teknik Informatika	4. S1 Pend. Biologi	2. S2 Pendidikan Ekonomi
5. SI Pend. Etno. Jajngn	3. SI Pend. Ekonomi	5. S1 Sistem Informatika	5. S1 Prejurisdi	

BERITA ACARA UJIAN SKRIPSI

Pada hari ini **Jum'at** , Tanggal **22 Juli 2022** Pukul **09.00-09.30 WIB** bertempat di Ruang **M5**

Kampus Universitas Nusantara PGRI Kediri, telah dilaksanakan Ujian Skripsi Mahasiswa

Universitas Nusantara PGRI Kediri :

NAMA : **Muhammad Zainal Arifin**
 NPM : **18.1.03.03.0023**
 FAK - PRODI : **FT-Sistem Informasi**
 JUDUL : **Perbandingan Performa Cluster Dalam Penentuan Jenis Bunga Iris**
 Dengan Hasil : **89 (A)**

Ketua Penguji : **Aidina Ristyawan, M. Kom**

Penguji 1 : **Anita Sari Wardani, M. Kom**

Penguji 2 : **M Najibulloh Muzaki, M. Cs**

Kediri, 25 Juli 2022

Mengetahui,

Ketua Program Studi

Sistem Informasi



Anita Farhana, M. Kom

Lampiran 3 *Data set iris*

Id	Label	A1	A2	A3	A4
id_1	Iris-setosa	5.1	3.5	1.4	0.2
id_2	Iris-setosa	4.9	3.0	1.4	0.2
id_3	Iris-setosa	4.7	3.2	1.3	0.2
id_4	Iris-setosa	4.6	3.1	1.5	0.2
id_5	Iris-setosa	5.0	3.6	1.4	0.2
id_6	Iris-setosa	5.4	3.9	1.7	0.4
id_7	Iris-setosa	4.6	3.4	1.4	0.3
id_8	Iris-setosa	5.0	3.4	1.5	0.2
id_9	Iris-setosa	4.4	2.9	1.4	0.2
id_10	Iris-setosa	4.9	3.1	1.5	0.1
id_11	Iris-setosa	5.4	3.7	1.5	0.2
id_12	Iris-setosa	4.8	3.4	1.6	0.2
id_13	Iris-setosa	4.8	3.0	1.4	0.1
id_14	Iris-setosa	4.3	3.0	1.1	0.1
id_15	Iris-setosa	5.8	4.0	1.2	0.2
id_16	Iris-setosa	5.7	4.4	1.5	0.4
id_17	Iris-setosa	5.4	3.9	1.3	0.4
id_18	Iris-setosa	5.1	3.5	1.4	0.3
id_19	Iris-setosa	5.7	3.8	1.7	0.3
id_20	Iris-setosa	5.1	3.8	1.5	0.3

Id	Label	A1	A2	A3	A4
	1 Iris-setosa				
id_21	Iris-setosa	5.4	3.4	1.7	0.2
id_22	Iris-setosa	5.1	3.7	1.5	0.4
id_23	Iris-setosa	4.6	3.6	1.0	0.2
id_24	Iris-setosa	5.1	3.3	1.7	0.5
id_25	Iris-setosa	4.8	3.4	1.9	0.2
id_26	Iris-setosa	5.0	3.0	1.6	0.2
id_27	Iris-setosa	5.0	3.4	1.6	0.4
id_28	Iris-setosa	5.2	3.5	1.5	0.2
id_29	Iris-setosa	5.2	3.4	1.4	0.2
id_30	Iris-setosa	4.7	3.2	1.6	0.2
id_31	Iris-setosa	4.8	3.1	1.6	0.2
id_32	Iris-setosa	5.4	3.4	1.5	0.4
id_33	Iris-setosa	5.2	4.1	1.5	0.1
id_34	Iris-setosa	5.5	4.2	1.4	0.2
id_35	Iris-setosa	4.9	3.1	1.5	0.1
id_36	Iris-setosa	5.0	3.2	1.2	0.2
id_37	Iris-setosa	5.5	3.5	1.3	0.2
id_38	Iris-setosa	4.9	3.1	1.5	0.1
id_39	Iris-setosa	4.4	3.0	1.3	0.2
id_40	Iris-setosa	5.1	3.4	1.5	0.2
id_41	Iris-setosa	5.0	3.5	1.3	0.3

Id	Label	A1	A2	A3	A4
id_42	Iris-setosa	4.5	2.3	1.3	0.3
id_43	Iris-setosa	4.4	3.2	1.3	0.2
id_44	Iris-setosa	5.0	3.5	1.6	0.6
id_45	Iris-setosa	5.1	3.8	1.9	0.4
id_46	Iris-setosa	4.8	3.0	1.4	0.3
id_47	Iris-setosa	5.1	3.8	1.6	0.2
id_48	Iris-setosa	4.6	3.2	1.4	0.2
id_49	Iris-setosa	5.3	3.7	1.5	0.2
id_50	Iris-setosa	5.0	3.3	1.4	0.2
id_51	Iris-versicolor	7.0	3.2	4.7	1.4
id_52	Iris-versicolor	6.4	3.2	4.5	1.5
id_53	Iris-versicolor	6.9	3.1	4.9	1.5
id_54	Iris-versicolor	5.5	2.3	4.0	1.3
id_55	Iris-versicolor	6.5	2.8	4.6	1.5
id_56	Iris-versicolor	5.7	2.8	4.5	1.3
id_57	Iris-versicolor	6.3	3.3	4.7	1.6
id_58	Iris-versicolor	4.9	2.4	3.3	1.0
id_59	Iris-versicolor	6.6	2.9	4.6	1.3
id_60	Iris-versicolor	5.2	2.7	3.9	1.4
id_61	Iris-versicolor	5.0	2.0	3.5	1.0
id_62	Iris-versicolor	5.9	3.0	4.2	1.5

Id	Label	A1	A2	A3	A4
	1				
id_63	Iris-versicolor	6.0	2.2	4.0	1.0
id_64	Iris-versicolor	6.1	2.9	4.7	1.4
id_65	Iris-versicolor	5.6	2.9	3.6	1.3
id_66	Iris-versicolor	6.7	3.1	4.4	1.4
id_67	Iris-versicolor	5.6	3.0	4.5	1.5
id_68	Iris-versicolor	5.8	2.7	4.1	1.0
id_69	Iris-versicolor	6.2	2.2	4.5	1.5
id_70	Iris-versicolor	5.6	2.5	3.9	1.1
id_71	Iris-versicolor	5.9	3.2	4.8	1.8
id_72	Iris-versicolor	6.1	2.8	4.0	1.3
id_73	Iris-versicolor	6.3	2.5	4.9	1.5
id_74	Iris-versicolor	6.1	2.8	4.7	1.2
id_75	Iris-versicolor	6.4	2.9	4.3	1.3
id_76	Iris-versicolor	6.6	3.0	4.4	1.4
id_77	Iris-versicolor	6.8	2.8	4.8	1.4
id_78	Iris-versicolor	6.7	3.0	5.0	1.7
id_79	Iris-versicolor	6.0	2.9	4.5	1.5
id_80	Iris-versicolor	5.7	2.6	3.5	1.0
id_81	Iris-versicolor	5.5	2.4	3.8	1.1
id_82	Iris-versicolor	5.5	2.4	3.7	1.0
id_83	Iris-versicolor	5.8	2.7	3.9	1.2

Id	Label	A1	A2	A3	A4
id_84	Iris-versicolor	6.0	2.7	5.1	1.6
id_85	Iris-versicolor	5.4	3.0	4.5	1.5
id_86	Iris-versicolor	6.0	3.4	4.5	1.6
id_87	Iris-versicolor	6.7	3.1	4.7	1.5
id_88	Iris-versicolor	6.3	2.3	4.4	1.3
id_89	Iris-versicolor	5.6	3.0	4.1	1.3
id_90	Iris-versicolor	5.5	2.5	4.0	1.3
id_91	Iris-versicolor	5.5	2.6	4.4	1.2
id_92	Iris-versicolor	6.1	3.0	4.6	1.4
id_93	Iris-versicolor	5.8	2.6	4.0	1.2
id_94	Iris-versicolor	5.0	2.3	3.3	1.0
id_95	Iris-versicolor	5.6	2.7	4.2	1.3
id_96	Iris-versicolor	5.7	3.0	4.2	1.2
id_97	Iris-versicolor	5.7	2.9	4.2	1.3
id_98	Iris-versicolor	6.2	2.9	4.3	1.3
id_99	Iris-versicolor	5.1	2.5	3.0	1.1
id_100	Iris-versicolor	5.7	2.8	4.1	1.3
id_101	Iris-virginica	6.3	3.3	6.0	2.5
id_102	Iris-virginica	5.8	2.7	5.1	1.9
id_103	Iris-virginica	7.1	3.0	5.9	2.1
id_104	Iris-virginica	6.3	2.9	5.6	1.8

Id	Label	A1	A2	A3	A4
	1				
id_105	Iris-virginica	6.5	3.0	5.8	2.2
id_106	Iris-virginica	7.6	3.0	6.6	2.1
id_107	Iris-virginica	4.9	2.5	4.5	1.7
id_108	Iris-virginica	7.3	2.9	6.3	1.8
id_109	Iris-virginica	6.7	2.5	5.8	1.8
id_110	Iris-virginica	7.2	3.6	6.1	2.5
id_111	Iris-virginica	6.5	3.2	5.1	2.0
id_112	Iris-virginica	6.4	2.7	5.3	1.9
id_113	Iris-virginica	6.8	3.0	5.5	2.1
id_114	Iris-virginica	5.7	2.5	5.0	2.0
id_115	Iris-virginica	5.8	2.8	5.1	2.4
id_116	Iris-virginica	6.4	3.2	5.3	2.3
id_117	Iris-virginica	6.5	3.0	5.5	1.8
id_118	Iris-virginica	7.7	3.8	6.7	2.2
id_119	Iris-virginica	7.7	2.6	6.9	2.3
id_120	Iris-virginica	6.0	2.2	5.0	1.5
id_121	Iris-virginica	6.9	3.2	5.7	2.3
id_122	Iris-virginica	5.6	2.8	4.9	2.0
id_123	Iris-virginica	7.7	2.8	6.7	2.0
id_124	Iris-virginica	6.3	2.7	4.9	1.8
id_125	Iris-virginica	6.7	3.3	5.7	2.1

Id	Label	A1	A2	A3	A4
	1				
id_126	Iris-virginica	7.2	3.2	6.0	1.8
id_127	Iris-virginica	6.2	2.8	4.8	1.8
id_128	Iris-virginica	6.1	3.0	4.9	1.8
id_129	Iris-virginica	6.4	2.8	5.6	2.1
id_130	Iris-virginica	7.2	3.0	5.8	1.6
id_131	Iris-virginica	7.4	2.8	6.1	1.9
id_132	Iris-virginica	7.9	3.8	6.4	2.0
id_133	Iris-virginica	6.4	2.8	5.6	2.2
id_134	Iris-virginica	6.3	2.8	5.1	1.5
id_135	Iris-virginica	6.1	2.6	5.6	1.4
id_136	Iris-virginica	7.7	3.0	6.1	2.3
id_137	Iris-virginica	6.3	3.4	5.6	2.4
id_138	Iris-virginica	6.4	3.1	5.5	1.8
id_139	Iris-virginica	6.0	3.0	4.8	1.8
id_140	Iris-virginica	6.9	3.1	5.4	2.1
id_141	Iris-virginica	6.7	3.1	5.6	2.4
id_142	Iris-virginica	6.9	3.1	5.1	2.3
id_143	Iris-virginica	5.8	2.7	5.1	1.9
id_144	Iris-virginica	6.8	3.2	5.9	2.3
id_145	Iris-virginica	6.7	3.3	5.7	2.5
id_146	Iris-virginica	6.7	3.0	5.2	2.3

Id	Label	A1	A2	A3	A4
id_147	¹ Iris-virginica	6.3	2.5	5.0	1.9
id_148	Iris-virginica	6.5	3.0	5.2	2.0
id_149	Iris-virginica	6.2	3.4	5.4	2.3
id_150	Iris-virginica	5.9	3.0	5.1	1.8

Lampiran 4 Data *iris* z-score normalization

Id	Label	A1	A2	A3	A4
id_1	Iris-setosa	-0.898	1.029	-1.336	-1.308
id_2	Iris-setosa	-1.139	-0.124	-1.336	-1.308
id_3	Iris-setosa	-1.380	0.336	-1.393	-1.308
id_4	Iris-setosa	-1.501	0.106	-1.280	-1.308
id_5	Iris-setosa	-1.018	1.259	-1.336	-1.308
id_6	Iris-setosa	-0.535	1.951	-1.166	-1.046
id_7	Iris-setosa	-1.501	0.797	-1.336	-1.177
id_8	Iris-setosa	-1.018	0.797	-1.280	-1.308
id_9	Iris-setosa	-1.743	-0.355	-1.336	-1.308
id_10	Iris-setosa	-1.139	0.106	-1.280	-1.439
id_11	Iris-setosa	-0.535	1.489	-1.280	-1.308
id_12	Iris-setosa	-1.259	0.797	-1.223	-1.308
id_13	Iris-setosa	-1.259	-0.124	-1.336	-1.439
id_14	Iris-setosa	-1.863	-0.124	-1.506	-1.439
id_15	Iris-setosa	-0.052	2.181	-1.450	-1.308
id_16	Iris-setosa	-0.173	3.104	-1.280	-1.046
id_17	Iris-setosa	-0.535	1.951	-1.393	-1.046
id_18	Iris-setosa	-0.898	1.028	-1.336	-1.177

Id	Label	A1	A2	A3	A4
id_19	Iris-setosa	-0.173	1.720	-1.166	-1.177
id_20	Iris-setosa	-0.898	1.720	-1.280	-1.177
id_21	Iris-setosa	-0.535	0.798	-1.166	-1.308
id_22	Iris-setosa	-0.898	1.489	-1.280	-1.046
id_23	Iris-setosa	-1.501	1.259	-1.563	-1.308
id_24	Iris-setosa	-0.898	0.567	-1.166	-0.915
id_25	Iris-setosa	-1.259	0.798	-1.053	-1.308
id_26	Iris-setosa	-1.018	-0.125	-1.223	-1.308
id_27	Iris-setosa	-1.018	0.798	-1.223	-1.046
id_28	Iris-setosa	-0.777	1.028	-1.280	-1.308
id_29	Iris-setosa	-0.777	0.798	-1.336	-1.308
id_30	Iris-setosa	-1380	0.337	-1.223	-1.308
id_31	Iris-setosa	-1259	0.106	-1.223	-1.308
id_32	Iris-setosa	-0.535	0.798	-1.280	-1.046
id_33	Iris-setosa	-0.777	2.412	-1.280	-1.439
id_34	Iris-setosa	-0.415	2.643	-1.336	-1.308
id_35	Iris-setosa	-1.139	0.106	-1.280	-1.439
id_36	Iris-setosa	-1.018	0.337	-1.450	-1.308
id_37	Iris-setosa	-0.415	1.028	-1.393	-1.308

Id	Label	A1	A2	A3	A4
id_38	Iris-setosa	-1.139	0.106	-1.280	-1.439
id_39	Iris-setosa	-1.743	-0.125	-1.393	-1.308
id_40	Iris-setosa	-0.898	0.798	-1.280	-1.308
id_41	Iris-setosa	-1.018	1.028	-1.393	-1.177
id_42	Iris-setosa	-1.622	-1.738	-1.393	-1.177
id_43	Iris-setosa	-1.743	0.337	-1.393	-1.308
id_44	Iris-setosa	-1.018	1.028	-1.223	-0.784
id_45	Iris-setosa	-0.898	1.720	-1.053	-1.046
id_46	Iris-setosa	-1.259	-0.125	-1.336	-1.177
id_47	Iris-setosa	-0.898	1.720	-1.223	-1.308
id_48	Iris-setosa	-1.501	0.337	-1.336	-1.308
id_49	Iris-setosa	-0.656	1.489	-1.280	-1.308
id_50	Iris-setosa	-1.018	0.567	-1.336	-1.308
id_51	Iris-versicolor	1.396	0.337	0.534	0.264
id_52	Iris-versicolor	0.672	0.337	0.420	0.395
id_53	Iris-versicolor	1.276	0.106	0.647	0.395
id_54	Iris-versicolor	-0.415	-1.738	0.136	0.132

Id	Label	A1	A2	A3	A4
id_55	Iris-versicolor	0.793	-0.586	0.477	0.395
id_56	Iris-versicolor	-0.173	-0.586	0.420	0.133
id_57	Iris-versicolor	0.551	0.567	0.534	0.526
id_58	Iris-versicolor	-1.139	-1.508	-0.259	-0.260
id_59	Iris-versicolor	0.914	-0.355	0.477	0.133
id_60	Iris-versicolor	-0.777	-0.816	0.080	0.264
id_61	Iris-versicolor	-1.018	-2.430	-0.146	-0.260
id_62	Iris-versicolor	0.068	-0.125	0.250	0.395
id_63	Iris-versicolor	0.189	-1.969	0.137	-0.260
id_64	Iris-versicolor	0.310	-0.355	0.534	0.264
id_65	Iris-versicolor	-0.294	-0.355	-0.090	0.133
id_66	Iris-versicolor	1.034	0.106	0.363	0.264
id_67	Iris-versicolor	-0.294	-0.125	0.420	0.395

Id	Label	A1	A2	A3	A4
id_68	Iris-versicolor	-0.052	-0.816	0.193	-0.260
id_69	Iris-versicolor	0.431	-1.969	0.420	0.395
id_70	Iris-versicolor	-0.294	-1.277	0.080	-0.129
id_71	Iris-versicolor	0.068	0.337	0.590	0.788
id_72	Iris-versicolor	0.310	-0.586	0.137	0.133
id_73	Iris-versicolor	0.551	-1.277	0.647	0.395
id_74	Iris-versicolor	0.310	-0.586	0.534	0.002
id_75	Iris-versicolor	0.672	-0.355	0.307	0.133
id_76	Iris-versicolor	0.914	-0.125	0.363	0.264
id_77	Iris-versicolor	1.155	-0.586	0.590	0.264
id_78	Iris-versicolor	1.034	-0.125	0.704	0.657
id_79	Iris-versicolor	0.189	-0.355	0.420	0.395
id_80	Iris-versicolor	-0.173	-1.047	-0.147	-0.260

Id	Label	A1	A2	A3	A4
id_81	Iris-versicolor	-0.415	-1.508	0.023	-0.129
id_82	Iris-versicolor	-0.415	-1.508	-0.033	-0.260
id_83	Iris-versicolor	-0.052	-0.816	0.080	0.002
id_84	Iris-versicolor	0.189	-0.816	0.760	0.526
id_85	Iris-versicolor	-0.535	-0.125	0.420	0.395
id_86	Iris-versicolor	0.189	0.798	0.420	0.526
id_87	Iris-versicolor	1.034	0.106	0.534	0.395
id_88	Iris-versicolor	0.551	-1.738	0.363	0.133
id_89	Iris-versicolor	-0.294	-0.125	0.193	0.133
id_90	Iris-versicolor	-0.415	-1.277	0.137	0.133
id_91	Iris-versicolor	-0.415	-1.047	0.363	0.002
id_92	Iris-versicolor	0.310	-0.125	0.477	0.264
id_93	Iris-versicolor	-0.052	-1.047	0.137	0.002

Id	Label	A1	A2	A3	A4
id_94	Iris-versicolor	-1.018	-1.738	-0.260	-0.260
id_95	Iris-versicolor	-0.294	-0.816	0.250	0.133
id_96	Iris-versicolor	-0.173	-0.125	0.250	0.002
id_97	Iris-versicolor	-0.173	-0.355	0.250	0.133
id_98	Iris-versicolor	0.431	-0.355	0.307	0.133
id_99	Iris-versicolor	-0.898	-1.277	-0.430	-0.129
id_100	Iris-versicolor	-0.173	-0.586	0.193	0.133
id_101	Iris-virginica	0.551	0.567	1.270	1.705
id_102	Iris-virginica	-0.052	-0.816	0.760	0.919
id_103	Iris-virginica	1.517	-0.125	1.213	1.181
id_104	Iris-virginica	0.551	-0.355	1.043	0.788
id_105	Iris-virginica	0.793	-0.125	1.156	1.312
id_106	Iris-virginica	2.121	-0.125	1.610	1.181
id_107	Iris-virginica	-1.139	-1.277	0.420	0.657
id_108	Iris-virginica	1.759	-0.355	1.440	0.788
id_109	Iris-virginica	1.034	-1.277	1.156	0.788

Id	Label	A1	A2	A3	A4
id_110	Iris-virginica	1.638	1.259	1.326	1.705
id_111	Iris-virginica	0.793	0.337	0.760	1.050
id_112	Iris-virginica	0.672	-0.816	0.874	0.919
id_113	Iris-virginica	1.155	-0.125	0.987	1.181
id_114	Iris-virginica	-0.173	-1.277	0.704	1.050
id_115	Iris-virginica	-0.052	-0.586	0.760	1.574
id_116	Iris-virginica	0.672	0.337	0.874	1.443
id_117	Iris-virginica	0.793	-0.125	0.987	0.788
id_118	Iris-virginica	2.242	1.720	1.667	1.312
id_119	Iris-virginica	2.242	-1.047	1.780	1.443
id_120	Iris-virginica	0.189	-1.969	0.704	0.395
id_121	Iris-virginica	1.276	0.337	1.100	1.443
id_122	Iris-virginica	-0.294	-0.586	0.647	1.050
id_123	Iris-virginica	2.242	-0.586	1.667	1.050
id_124	Iris-virginica	0.551	-0.816	0.647	0.788
id_125	Iris-virginica	1.034	0.567	1.100	1.181
id_126	Iris-virginica	1.638	0.337	1.270	0.788
id_127	Iris-virginica	0.431	-0.586	0.590	0.788
id_128	Iris-virginica	0.310	-0.125	0.647	0.788

Id	Label	A1	A2	A3	A4
id_129	Iris-virginica	0.672	-0.586	1.043	1.181
id_130	Iris-virginica	1.638	-0.125	1.156	0.526
id_131	Iris-virginica	1.879	-0.586	1.326	0.919
id_132	Iris-virginica	2.483	1.720	1.496	1.050
id_133	Iris-virginica	0.672	-0.586	1.043	1.312
id_134	Iris-virginica	0.551	-0.586	0.760	0.395
id_135	Iris-virginica	0.310	-1.047	1.043	0.264
id_136	Iris-virginica	2.242	-0.125	1.326	1.443
id_137	Iris-virginica	0.551	0.798	1.043	1.574
id_138	Iris-virginica	0.672	0.106	0.987	0.788
id_139	Iris-virginica	0.189	-0.125	0.590	0.788
id_140	Iris-virginica	1.276	0.106	0.930	1.181
id_141	Iris-virginica	1.034	0.106	1.043	1.574
id_142	Iris-virginica	1.276	0.106	0.760	1.443
id_143	Iris-virginica	-0.052	-0.816	0.760	0.919
id_144	Iris-virginica	1.155	0.337	1.213	1.443
id_145	Iris-virginica	1.034	0.567	1.100	1.705
id_146	Iris-virginica	1.034	-0.125	0.817	1.443
id_147	Iris-virginica	0.551	-1.277	0.704	0.919

Id	Label	A1	A2	A3	A4
id_148	Iris-virginica	0.793	-0.125	0.817	1.050
id_149	Iris-virginica	0.431	0.798	0.930	1.443
id_150	Iris-virginica	0.068	-0.125	0.760	0.788

skripsi

ORIGINALITY REPORT

26%

SIMILARITY INDEX

26%

INTERNET SOURCES

15%

PUBLICATIONS

16%

STUDENT PAPERS

PRIMARY SOURCES

1	www.87994.com Internet Source	8%
2	forge.greyc.fr Internet Source	6%
3	repository.bsi.ac.id Internet Source	3%
4	123dok.com Internet Source	3%
5	docplayer.info Internet Source	3%
6	jurnal.unimed.ac.id Internet Source	2%

Exclude quotes On

Exclude bibliography On

Exclude matches < 2%

skripsi

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70

PAGE 71

PAGE 72

PAGE 73

PAGE 74

PAGE 75

PAGE 76

PAGE 77
