TESIS

ANALISIS KOMPARASI TEKNIK DEZCRETIZATION DENGAN K-MEANS CLUSTERING PADA TRANSFORMASI DATA UNTUK PREDIKSI ORGANISME PENGGANGGU TUMBUHAN HAMA PUTIH PALSU MENGGUNAKAN NAÏVE BAYES

Disusun oleh:

Nama : **Fathoni Dwiamtoko**
NIM : **19.52.1272**
Konsentrasi : **Business Intelligence**

PROGRAM STUDI S2 TEKNIK INFORMATIKA

PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

TESIS

ANALISIS KOMPARASI TEKNIK DEZCRETIZATION DENGAN K-MEANS CLUSTERING PADA TRANSFORMASI DATA UNTUK PREDIKSI ORGANISME PENGGANGU TUMBUHAN HAMA PUTIH PALSU MENGGUNAKAN NAÏVE BAYES

COMPARATIVE ANALYSIS OF DEZCRETIZATION TECHNIQUES WITH K-MEANS CLUSTERING ON DATA TRANSFORMATION FOR PREDICTION OF PEST CNAPHALOCROSIS MEDINALIS USING NAÏVE BAYES

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister

Disusun oleh:

Nama	**: Fathoni Dwiatmoko**
NIM	**: 19.52.1272**
Konsentrasi	**: Business Intelligence**

PROGRAM STUDI S2 TEKNIK INFORMATIKA

PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

HALAMAN PENGESAHAN

ANALISIS KOMPARASI TEKNIK DEZCRETIZATION DENGAN K-MEANS CLUSTERING PADA TRANSFORMASI DATA UNTUK PREDIKSI ORGANISME PENGGANGGU TUMBUHAN HAMA PUTIH PALSU MENGGUNAKAN NAÏVE BAYES

COMPARATIVE ANALYSIS OF DEZCRETIZATION TECHNIQUES WITH K-MEANS CLUSTERING ON DATA TRANSFORMATION FOR PREDICTION OF PEST CNAPHALOCROSIS MEDINALIS USING NAÏVE BAYES

Dipersiapkan dan Disusun oleh

Fathoni Dwiatmoko

19.52.1272

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 03 Februari 2022

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 03 Februari 2022
Direktur Program Pascasarjana

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERSETUJUAN

ANALISIS KOMPARASI TEKNIK DEZCRETIZATION DENGAN K-MEANS CLUSTERING PADA TRANSFORMASI DATA UNTUK PREDIKSI ORGANISME PENGGANGGU TUMBUHNA HAMA PUTIH PALSU MENGGUNAKAN NAÏVE BAYES

COMPARATIVE ANALYSIS OF DEZCRETIZATION TECHNIQUES WITH K-MEANS CLUSTERING ON DATA TRANSFORMATION FOR PREDICTION OF PEST CNAPHALOCROSIS MEDINALIS USING NAÏVE BAYES

Dipersiapkan dan Disusun oleh

Fathoni Dwiatmoko

19.52.1272

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 03 Februari 2022

Pembimbing Utama	**Anggota Tim Penguji**
Prof. Dr. Ema Utami, S.Si., M.Kom.	**Dr. Andi Sunyoto, M.Kom.**
NIK. 190302037	**NIK. 190302052**
	Alva Hendi Muhammad, S.T., M.Eng., Ph.D.
Pembimbing Pendamping	**NIK. 190302**
Sudarmawan, M.T.	**Prof. Dr. Ema Utami, S.Si., M.Kom.**
NIK. 190302035	**NIK. 190302037**

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 03 Februari 2022
Direktur Program Pascasarjana

Prof. Dr. Kusrini, M.Kom.
NIK. 19030210

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fathoni Dwiatmoko
NIM : 19.52.1272
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.
Dosen Pembimbing Pendamping : Sudarmawan, M.T.

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 Tim 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, 03 Februari 2022
Yang Menyatakan,

Fathoni Dwiatmoko

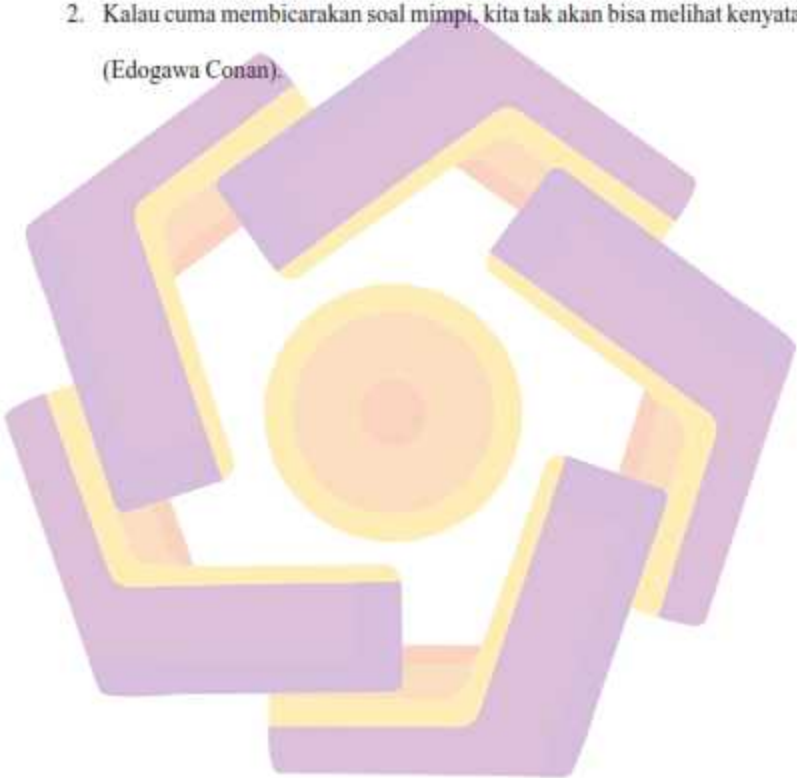
HALAMAN PERSEMBAHAN

Laporan Tesis ini saya persembahkan kepada:

1. Allah SWT, karena rahmat dan kasih sayangNya saya dapat menyelesaikan tesis ini.
2. Orang tua yang selalu mendukung serta mendoakan saya dalam menyelesaikan laporan ini.
3. Kakak dan Adik yang selalu memberikan dukungan untuk saya.
4. Semua saudara dan teman-teman yang telah memberikan dukungan

HALAMAN MOTTO

1. Kehidupan nyata tidaklah sesederhana dengan kehidupan di dalam game (Sawada Hiroki).
2. Kalau cuma membicarakan soal mimpi, kita tak akan bisa melihat kenyataan (Edogawa Conan).



KATA PENGANTAR

Dengan memanjatkan puji dan syukur kehadirat Allah SWT atas segala Rahmat, Taufiq, serta Hidayah-Nya sehingga penulis dapat menyelesaikan Tesis ini dengan judul "Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Penggangu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes" yang merupakan syarat dalam menyelesaikan Program Studi Strata II pada Program Pascasarjana Magister Teknik Informatika, Universitas Amikom Yogyakarta.

Selama penulisan Tesis ini penulis mendapat banyak bantuan dan bimbingan dari berbagai pihak, untuk itu pada kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya. pada:

1. Prof. Dr. M. Suyanto, M. M selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Ema Utami dan Sudarmawan, M. T, selaku Dosen Pembimbing yang telah memberikan bimbingan dan masukan kepada penulis.
3. Seluruh Dosen pengajar Strata dua (S2) Program Pascasarjana Magister Teknik Informatika Universitas Amikom Yogyakarta yang telah mendidik dan memberikan pengetahuan yang tak ternilai kepada penulis selama mengikuti perkuliahan.
4. Orang tua dan keluarga tercinta yang telah memberikan dukungan moril, doa dan kasih sayang
5. Teman-teman Universitas Amikom Yogyakarta angkatan 2019.
6. Semua pihak yang telah membantu penulis dalam menyelesaikan Tesis.

Penulis sadar bahwa tentunya dalam penulisan Tesis ini masih terdapat banyak kekurangan, untuk itu saran dan kritik dari pembaca yang sifatnya membangun sangat diharapkan, demi pengembangan kemampuan penulis ke depan.

Yogyakarta, 03 Februari 2022

Penulis

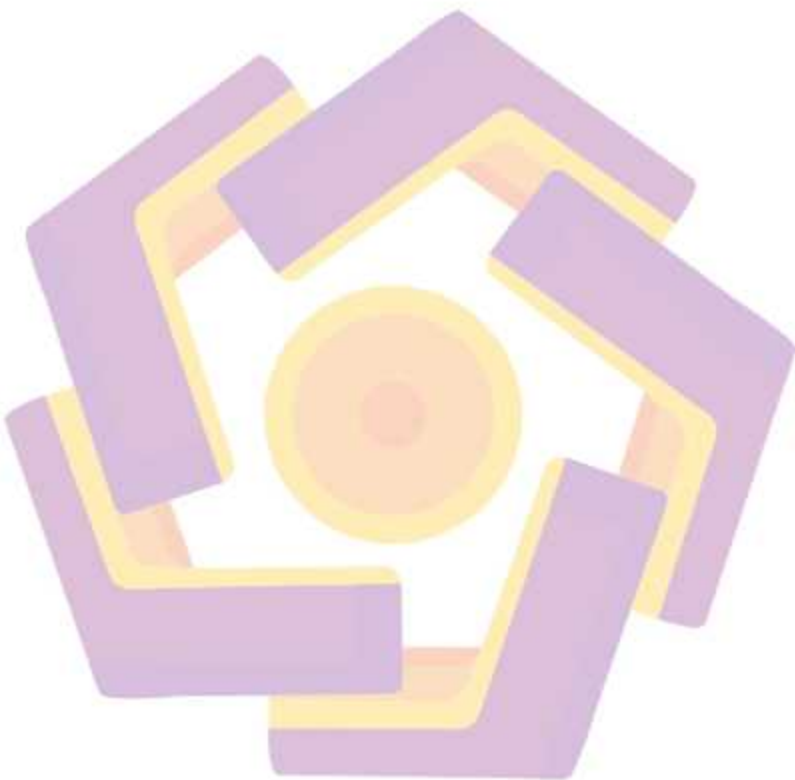
DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR ISTILAH

(jika ada)



INTISARI

Perkembangan teknologi informasi telah menjadi hal yang tidak dapat dipisahkan dari kehidupan manusia. Banyaknya jumlah data di dunia menyebabkan pengolahan data menjadi pesat karena jumlah data yang semakin bertambah. Salah satu bidang yang terdapat pertumbuhan data yang pesat adalah bidang pertanian. Data pertanian dapat di proses dengan data mining untuk dapat di jadikan bahan pembelajaran maupun pengambilan keputusan. Pada data mining terdapat salah satu proses yaitu transformasi data, maka diperlukan metode clustering yang digunakan untuk mengkategorikan data. Pada penelitian ini bertujuan mengetahui perbandingan transformasi data menggunakan discretization dengan k-means clustering pada algoritma naïve bayes dengan data uji data serangan hama hama putih palsu (HPP).

Hasil dari pengujian menggunakan confusion matrix didapatkan teknik discretization dengan nilai akurasi terbaik 75% dengan data training 70% dan 30% data testing. Pengujian menggunakan k-means clustering didapatkan nilai akurasi terbaik 80,56% dengan data training 70% dan 30% data testing. Berdasarkan pengujian menggunakan confusion matrix didapatkan transformasi data terbaik menggunakan k-means clustering dengan nilai akurasi 80,56%.

Kata kunci: k-means, clustering, discretization, transformasi.

ABSTRACT

The development of information technology has become an inseparable part of human life. The large amount of data in the world causes data processing to be fast because the amount of data is increasing. One of the areas where data is growing rapidly is agriculture. Agricultural data can be processed with data mining to be used as learning materials and decision making. In data mining there is one process, namely data transformation, it is necessary to use a clustering method to categorize data. This study aims to compare the data transformation using manual discretization with k-means clustering on the nave Bayes algorithm with test data for white midrib blight (HPP) attack data.

The results of testing using a confusion matrix obtained discretization technique with the best accuracy value of 75% with 70% training data and 30% testing data. Testing using k-means clustering obtained the best accuracy value of 80.56% with 70% training data and 30% testing data. Based on testing using the confusion matrix, the best data transformation was obtained using k-means clustering with an accuracy value of 80.56%.

Keyword: k-means, clustering, discretization, transformation.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi informasi telah menjadi hal yang tidak dapat dipisahkan dari kehidupan manusia. Banyaknya jumlah data di dunia menyebabkan pengolahan data menjadi pesat karena jumlah data yang semakin bertambah. Salah satu bidang yang memiliki pertumbuhan data yang besar adalah bidang pertanian. Bayaknya data dibidang pertanian seperti data serangan hama, curah hujan, luas tanam, waktu tanam, hasil panen dapat dimanfaatkan untuk melihat masa depan dengan menggunakan data mining. Data mining adalah proses yang dilakukan dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai sumber data yang besar (Turban, dkk.2005). Pada penelitian ini peneliti akan melakukan disalah satu sector yaitu sector pertanian. Berdasarkan data BPS Lampung Timur terjadi penurunan hasil produksi padi dari tahun 2016 hingga tahun 2019.

Teknik klasifikasi dalam data mining dapat dijadikan untuk membantu dalam pengambilan keputusan. Proses klasifikasi memerlukan beberapa tahapan salah satunya adalah preprocesing. Preprocessing terdapat beberapa tahapan seperti cleansing dan transformasi data. Proses transformasi data yang perlu dilakukan salah satunya adalah mengelompokan data menjadi beberapa kelompok. Tahapan transformasi data memiliki tingkat kesulitan yang tinggi karena merubah data

numerik menjadi data interval. Untuk itu diperlukan otomatisasi transformasi data numerik menjadi data interval

Pada penelitian Tutus Praningki dan Indra Budi dengan judul Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN, melakukan penelitian tentang pengujian performa metode prediksi dengan tujuan untuk mendukung keputusan dan melakan klasifikasi penyakit kangker serivks. Pada penelitian tersebut pada proses transformasi data dilakukan menggunakan discretization manual sehingga perlu dilakukan komparasi menggunakan Teknik yang lain. Pada processing data dilakukan transformasi data dengan menggunakan Teknik Discretization yaitu dengan melakukan transformasi data dengan merubah data numerik menjadi kedalam data interval. Data yang dirubah yaitu tinggi badan manusia ditransformasi kedalam data interval yaitu tinggi, sedang, dan pendek. Hasil pengujian algoritma Naive Bayes memiliki akurasi terbaik sebesar 94,44%, sedangkan tingkat akurasi yang dihasilkan algoritma CART dan k-NN adalah 88,89%, 85,04% (Praningki & Budi, 2018).

Pada penelitian yang dilakukan oleh Adi Supriyatna dan Wida Prima Mustika tentang komparasi algoritma naïve bayes dengan algoritma SVM untuk memprediksi imunoterapi pada penyakit kutil guna memprediksi keberhasilan pengobatan imunoterapi untuk penyakit kutil, pada penelitan tersebut didapatkan nilai akurasi pada algoritma naïve bayes yaitu 1, sedangkan pada algoritma SVM (Support Vectore Machine) menghasilkan nilai akurasi 0,8. Dari hasil pengujian tersebut menunjukan algoritma naïve bayes lebih unggul daripada algoritma SVM (Supriyatna & Mustika, 2018). Berdasarkan penilitian tersebut maka pada

penelitian ini menggunakan metode naïve bayes untuk melakukan pengujian transformasi data.

Pada penelitian yang dilakukan oleh Mochamad Farid Rifai, Yudhi S. Purwanto, Hendra Jatnika, Sely Karmila dengan judul Pengaruh Kondisi Cuaca Terhadap Serangan Hama Penggerek Batang Pada Tanaman Padi Di Desa Ciaruteun Ilir, Kec. Bungbulang, Kab. Bogor, melakukan penelitian tentang prediksi serangan hama penggerek batang menggunakan metode naïve bayes dan berhasil mendapatkan presentasi keakuratan sebesar 96,76%. Metode preprocessing transformasi yang digunakan yaitu mengklasifikasikan data dengan beberapa kategori yaitu rendah, sedang dan tinggi dengan cara membuat range angka seperti suhu < 24° masuk kategori rendah, suhu 24°-32° masuk kategori sedang, dan suhu > 32° masuk kategori tinggi (Rifai et al., 2020). Pada penelitian tersebut menggunakan metode transformasi data dengan cara manual dan perbedaan dengan penilitian yang akan dilakukan dengan mengkomparasi dezcretization manual dengan metode k-means clustering pada transformasi data.

K-means merupakan salah satu metode non hirarki yang memulai dengan asumsi nilai K untuk menentukan jumlah kelompok data, sehingga data dapat dikelompokkan sesuai kebutuhan (Rachmatin & Sawitri, 2016).

Pada penelitian yang dilakukan oleh Insanul Kamila, Ulya Khairunnisa dan mustakim tentang perbandingan algoritma K-Means dan K-Medoids untuk pengelempokan data transaksi bongkar muat di Provinsi Riau, menghasilkan k-means memiliki nilai DBI sebesar 0,112 dan k-medoids memiliki nilai DBI sebesar 0,119 yang artinya k-means memiliki kemiripan data pada setiap cluster sebesar

88,8% dan k-medoids memiliki kemiripan data pada setiap cluster sebesar 88,1%. (Kamila et al., 2019).

Pada penelitian yang dilakukan oleh Hardian Artanto, istiadi, Fitri Marisa, dan Dwi Purnomo, tentang komparasi algoritma fuzzy c-means dan k-means untuk mengelompokan siswa berdasarkan nilai akademi dan perilaku siswa. Pada penelitian tersebut menghasilkan algoritma k-means memiliki tingkat akurasi 91% dan membutuhkan waktu 4.4105 detik dan pada algoritma fuzzy C-Means memiliki tingkat akurasi 68% dan membutuhkan waktu 5.5416 detik. Dan dapat disimpulkan k-means lebih baik daripada fuzzy c-means (Artanto et al., 2019). Pada penelitian oleh Insanul Kamila, Ulya Khairunnisa dan mustakim dan penelitian yang dilakukan oleh Hardian Artanto, istiadi, Fitri Marisa, dan Dwi Purnomo menghasilakan metode k-means clustering lebih baik sehingga pada penelitian ini menggunakan metode k-means clustering untuk transformasi data.

Pada penelitian (Junaedi et al., 2011) mengenai proses transformasi data pada data mining terdapat metode *smoothing* yaitu metode yang dilakukan apabila data mengandung nilai tidak valid. Dalam proses smoothing terdapat beberapa Teknik yaitu binning atau diskritisasi, clustering, dan regression. Teknik clustering pada penelitian tersebut menggunkan k-means clustering.

Berdasarkan refrensi penelitian diatas maka penelitian ini akan menggunakan ilmu data mining untuk memprediksi serangan OPT menggunakan metode naïve bayes dengan melakukan komparasi pada transformasi data menggunakan k-means clustering dan teknik discretization manual dengan equal-width distance dan jenis atribut ordinal, sehingga menghasilkan prediksi dengan akurasi yang lebih baik dan

memiliki cluster data yang memiliki tingkat kemiripan yang lebih baik sehingga menghasilkan penelitian dengan judul penelitian Analisis Komparasi Teknik Dezcretization Manual dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Oragasme Pengganggu Tumbuhan Hama putih palsu Menggunakan Naïve Bayes.

1.2. Rumusan Masalah

Berikut adalah rumusan masalah yang akan diselesaikan dalam penelitian yaitu membandingkan transformasi data pada metode naïve bayes:

a. Berapa tingkat akurasi yang dihasilkan metode naïve bayes menggunakan transformasi data dengan Teknik discretization manual?
b. Berapa tingkat akurasi yang dihasilkan metode naïve bayes menggunakan transformasi data dengan metode k-means clustering?

1.3. Batasan Masalah

Berikut adalah batasan-batasan dalam penelitian:

a. Lokasi penelitian ini dilakukan di kecamatan Batanghari Nuban, Lampung Timur, Lampung
b. Data yang digunakan data curah hujan, data tanam, dan data serangan OPT (organisme pengganggu tumbuhan) dari tahun 2016 sampai dengan 2019, data yang diperoleh berupa file excel dengan jumlah 21 file. Data didapatkan dari hasil wawancara dengan petugas POPT Kecamatan Batanghari Nuban Kabupaten Lampung Timur, Lampung.

c. Organisme Pengganggu Tumbuhan () yang diteliti adalah OPT tanaman pangan padi.
d. Transformasi data menggunakan Teknik discretization dan menggunakan metode k-means clustering.
e. Menggunakan tools Rapidminer studio 9.9 untuk melakukan penelitian metode k-means dan naïve bayes.
f. Memprediksi intesitas serangan OPT Hama Putih Palsu.
g. Diskritisasi yang digunakan yaitu equal-width (distance) dengan nilai atribut ordinal.
h. Sekenario pengujian prediksi naïve bayes dengan membagi dua sekenario pembagian data yaitu sekenario pertama dengan data training 80% dan data testing 20%. Sekenario kedua dengan pembagian data 70% data training dan 30% data testing.

1.4. Tujuan Penelitian

Berikut adalah tujuan dari penelitian yang akan dilakukan yaitu:

a. Mengetahui tingkat akurasi yang dihasilkan metode naïve bayes menggunakan transformasi data dengan Teknik discretization manual.
b. Mengetahui tingkat akurasi yang dihasilkan metode naïve bayes menggunakan transformasi data dengan metode k-means clustering.
c. Mengetahui hasil perbandingan antara metode transformasi discretization manual dengan metode k-means clustering.

1.5. Manfaat Penelitian

Berikut adalah manfaat dari peneletian yang akan dilakukan yaitu:

a. Dapat membantu mengenai pemilihan metode transformasi data antara discretization manual dengan k-means clustering.

b. Dapat dijadikan rekomendasi untuk penelitian selanjutnya dalam melakukan transformasi data.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

(Rifai et al., 2020) melakukan penilitian tentang prediksi serangan hama padi penggerek batang menggunakan metode naïve bayes. Pada proses transformasi data menggunakan cara memeberikan range seperti data suhu diberikan range < 24 kategori rendah, 24-32 kategori sedang dan >33 kategori tinggi. Berdasarkan data training dapat diklasifikasikan yaitu 138 data training, dan diujikan menggunkan metode naïve bayes didapatkan persentase keakuratan sebesar 96.76%

(Mustafa et al., 2018) melakukan penelitian tentang implementasi data mining guna mengevaluasi kinerja akademik mahasiswa. Pada penlitian tersebut peneliti membuat system guna memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat. Pengujian dilakukan pada data mahasiswa angkatan 2005- 2009 mining NBC menghasilkan nilai precision, recall, dan accuracy masing-masing 83%, 50%, dan 70%. 2. Penentuan data training dapat mempengaruhi hasil pengujian, karena pola data training tersebut akan dijadikan sebagai rule untuk menentukan kelas pada data testing. Sehingga besar atau kecilnya prosentase tingkat precision, recall, dan accuracy dipengaruhi juga oleh penentuan data training. 3. Hasil mining NBC dapat digunakan untuk mengklasifikasikan kinerja akademik mahasiswa tahun ke-2 yang dalam penelitian ini dijadikan data target. Metode naïve bayes menghasilakn tingkat validasi 57,63% dan neural network memiliki tingkat validasi 72,58%.

(Surya Nagari & Inayati, 2019) melakukan peneltian tentang pengelompokan status gizi anak usia dibawah 60 bulan yang dilakukan dengan metode C-Means Clustering. Pada penelitian tersebut berhasil melakukan klastering menggunakan metode K-Means melalui parameter berat badan menurut umur pada klaster 4 yaitu klaster 1 dengan status gizi buruk 23 balita, klaster 2 dengan 17 balita dalam status gizi buruk. status gizi, cluster 3 dengan 7 balita status gizi baik dan cluster 4 dengan 10 balita status gizi lebih.

(Darmansah, 2020) melakukan penelitian tentang kerusakan tanaman cabai, dan mengambil tindakan mengantisipasi penyebab kerusakan pada tanaman cabai dengan menggunakan k-means clustering dan dengan menggunakan data uji 77 maka didapatkan kategori-kategori kerusakan cabai yaitu kerusakan berat dengan jumlah data 25 kerusakan ringan dengan jumlah data 3 mdan kerusakan sedang dengan jumlah data 49.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Pengaruh Kondisi Cuaca Terhadap Serangan Hama Penggerek Batang Pada Tanaman Padi Di Desa Ciaruteun Ilir, Kec. Bungbulang, Kab. Bogor	Mochamad Farid Rifai, Yudhi S. Purwanto, Hendra Jatnika, Sely Karmila 2020	prediksi serangan hama padi penggerek batang menggunakan metode Naïve	Berdasarkan data training dapat mengklasifikasikan 138 data training didapatkan persentase keakuratan sebesar 96.76%	1. Proses transformasi data menggunakan cara memeberikan range seperti data suhu diberikan range < 24 kategori rendah, 24-32 kategori sedang dan >33 kategori tinggi. 2. Pada penelitian tersbut pengujian dilakukan hanya dengan satu data testing menghasilkan akurasi 96.76%, sehingga diperlukan penambahan data testing agar lebih akurat.	Perbedaan antara penelitian sebelumnya dengan penelitian yang akan dilakukan komparasi discretization manual dengan metode k-means clustering, dan dilakukan pengujian dengan membagi antara data training dan data testing sebesar 80% data training dan 20% data testing.

Tabel 2.1. Matriks literatur review dan posisi penelitian
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier	Mustafa, M. Syukri Ramadhan, Muh Rizky Thenata, Angelina P. 2018	memberikan rekomendasi solusi untuk memandu mahasiswa lulus dalam waktu yang paling tepat	Hasil pengujiaan naive bayes didapakan nilai akuasi 70% dengan precision 83% dan recall 50% dari data mahasiswa pada tahun 2005 hingga 2009. Pada penelitian tersebut target data yang digunakan yaitu data akademik mahasiswa pada tahun kedua dengan menggunakan metode naive bayes clasifier	Pada penelitian tersebut dihasikan nilai precision 83%, recall 50%, dan accuracy 70%.	Pada penelitian tersebut nilai akurasi yang dihasilkan masih relative rendah sehingga perlu dilakukan penelitian lanjutan dengan melakukan komparasi transformasi data menggunkan metode discretization manual dengan meotde k-means clustering untuk meningkatkan akurasi naïve bayes.
3	Komparasi Data Mining Naive Bayes Dan Neural Network Memprediksi Masa Studi Mahasiswa S1	Azahari, Yulindawati, Dewi Rosita, dan Syamsuddin Mallala 2020	Menguji metode yang memiliki tingkat akurasi yang tinggi dari dua metode yaitu naïve bayes dan neural network	Metode naïve bayes menghasilakn tingkat validasi 57,63% dan neural network memiliki tingkat validasi 72,58%	Penelitian kedepannya dapat menambahkan atribut yang lebih diskriminan, seperti penghasilan orang tua, jalur masuk kuliah, jurusan pada saat sekolah, ataupun IPK di setiap semester. Penggunaan algoritma	Pada peneltian tersebut dilakukan konversi data untuk dilakaukan prediksi dengan mengkonversi data text menjadi angka bersarkan kategori, seperti pengkonversian asal sekokah terdat ibu kota = 1, kota =2, dan Kabupaten = 3 perbedaan pada penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
					lain untuk memprediksi seperti C4.5, SVM, atau Expectation Maximisation (EM Algorithm) diharapkan lebih meningkatkan nilai kevalidan model.	ini yaitu proses preprosesing datanya menggunakan k-means clustering untuk melakukan pengelompokan data.
4	Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means	Darmansah 2020	mengetahui penyebab kerusakan tanaman cabai, dan mengambil tindakan mengantisipasi penyebab penyebab kerusakan pada tanaman cabai	Setelah melakukan pengujian dengan jumlah data sebanyak 77 maka di dapat hasil untuk C0 menentukan kerusakan cabai berjenis berat yang terdiri dari 25 data, C1 untuk menentukan kerusakan cabai berjenis ringan sebanyak 3 data, sedangkan C2 untuk menentukan kerusakan cabai yang berjenis sedang terdapat 49 data.	Sebaiknya dilakuakn pengembangan lebih lanjut terhadap penelitian penyebab kerusakan tanaman cabai dan membandingkan hasil metode ini dengan metode metode lainnya seperti metode Fuzzy C-Means, SelfOrganizing Map, K-Modes dan lain-lain	Pada penelitian tersebut k-means clustering digunakan untuk menentukan jenis kerusakan cabai dengan 3 tingkatan yaitu berat, sedang dan ringan, sedangkan pada penelitian yang akan dilakukan yaitu melakukan clustering yang bertujuan untuk mengkategorikan tingkat serangan hama terhadap tumbuhan padi.

Tabel 2.1. Matriks literatur review dan posisi penelitian
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS	Stefanny Surya Nagari, Lilik Inayati 2020	mengelompokan status gizi anak usia dibawah 60 bulan yang dilakukan dengan metode C-Means Clustering	Status gizi balita di Ponkesdes Mayangrejo dapat diklaster menggunakan metode K-Means melalui parameter berat badan menurut umur pada klaster 4 yaitu klaster 1 dengan status gizi buruk 23 balita, klaster 2 dengan 17 balita dalam status gizi buruk. status gizi, cluster 3 dengan 7 balita status gizi baik dan cluster 4 dengan 10 balita status gizi lebih.	Pada penelitian tersebut hanya melakukan kluster data status gizi anak usia dibawah 60 bulan.	Pada penelitian sebelumnya telah berhasil menggunkan metode k-means clustering untuk mengelompokan data sehingga pada penelitian ini menggunkan metode k-means clustering untuk transformasi data yang kemudian hasil transformasi data akan dilakukan prediksi menggunkan algoritma naïve bayes.
6	Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN	Tutus Praningki, Indra Budi	mendukung keputusan klinis pada pasien baru	Algoritma Naive Bayes dengan teknik probabilistik mampu dengan baik melakukan klasifikasi terhadap kasus positif dan negatif kanker.	Pada penelitian tersebut pada proses transformasi data dilakukan menggunakan discretization manual sehingga perlu dilakukan komparasi	Pada proses data preprocesing pada penelitian tersebut dilkaukan transformasi data berupa data numerik menjadi interval menggunakan range, pada penelitian ini akan

Tabel 2.1. Matriks literatur review dan posisi penelitian
Analisis Komparasi Teknik Dezcretization Dengan K-Means Clustering Pada Transformasi Data Untuk Prediksi Organisme Pengganggu Tumbuhan Hama Putih Palsu Menggunakan Naïve Bayes (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				serviks, serta menghasilkan tingkat akurasi yang tinggi	menggunakan Teknik yang lain	diusulkan proses tranformasi data menggunakan metode K-Means Clustering

2.3. Landasan Teori

a. Naïve bayess

Naive bayes adalah teknik klasifikasi statistik. teknik klasifikasi dengan naive bayes bekerja degan cara memprediksi probabilitas pada kelompok data tertentu. kinerja naive bayes menunjukan performa dengan tingkat akurasi tinggi jika digunakan untuk klasifikasi dengan data yang besar. naive bayes berguna untuk penyelesaian masalah machine learning karena ketepatan akurasi yang tinggi dan memiliki kesederhanaan dalam proses perhitungannya.. Berikut merupaka aturan dasar naïve bayes berdsarkan Teorema bayes:

$$P(H|X)=\frac{P(x|H)P(H)}{P(x)}$$

X = Data kelas tidak diketahui

H = Hipotesis data pada kelas X

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

(Handayani & Pribadi, 2015).

Teorema Naive Bayes dalam proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema bayes di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{p(F_1 \dots F_n)}$$

Variabel C merupakan kelas, dan variabel F1 ... Fn merupakan atribut yang dibutuhkan untuk proses klasifikasi. Rumus diatas dapat diserdehankaan dengan rumus berikut:

$$Posterior = \frac{Prior\ x\ likelihood}{evidence}$$

Peluang atribut tertentu masuk dalam kelas C Posterior, dikali dengan peluang kemunculan karakteristik sampel pada kelas C disebut juga likelihood, dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global disebut juga evidence (Bustami, 2014).

Pada umumnya, perhitungan naïve bayes lebih mudah pada fitur bertipe kategoris seperti pada status klasifikasi mahasiswa dengan fitur "nilai dengan nilai {lulus, tidak lulus}. Pada fitur dengan tipe numerik (kontinu) terdapat tahapan-tahapan khusus sebelum diklasifikasikan dengan Naïve Bayes. Dengan cara sebagai berikut:

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasikan fitur kontinu ke dalam fitur ordinal.

2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan (Husaini, 2016).

b. K-Means Clustering

K-means clustering adalah algoritma untuk pengelompokan data yang mampu membagi data menjadi kelompok data baru. cara kerja k-means clustering yaitu dengan membagi data berdasrkan karakteristik data yang sama kedalam satu kelompok. pengelompkan data dilakukan dengan tujuan memaksimalkan variasi kelompok-kelompok data dan meminimalisir variasi data dalam suatu kelompok data (Gustientiedina et al., 2019). Secara umum algoritma dasar K-Means Clustering adalah sebagai berikut:

1. Menentukan jumlah kelompok yang akan dibagi.
2. Alokasikan data ke dalam kelompok secara *random*.
3. Hitung centroid/rata-rata dari data yang ada di masing-masing kelompok
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, jika terdapat data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan. Penentuan centroid secara acak ditentukan dengan rumus persamaan yang dituliskan persamaan sebagai berikut:

$$C_i = \min + \frac{(i-1) * (max - min)}{n} + \frac{(max - min)}{2 * n}$$

Keterangan:

C_i: Centroid dari kelas ke-i

Min: nilai terkecil dari data kelas kontinyu

Max: nilai terbesar dari data kelas kontinyu

N: jumlah kelas diskret

Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu Euclidean distance space. Euclidean distance space sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan, dan dapat dituiliskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^{p} \{x_{ik} - x_{jk}\}}$$

Keterangan:

d_{ij}: jarak objek antara obyek i dan j

P: dimensi data (Ahmad et al., 2017).

c. Data Mining

Data mining adalah mengidentifikasi informasi dari berbagai database yang menghasilan informasi dan pengetahuan baru yang dalam prosesnya mengggunan teknik statistik, matematika, kecerdasaan buatan, dan *machine learning*.

Data mining adalah serangkaian proses untuk menggali nilai tambah yang didapatkan dari kumpulan data-data yang akan menghasilkan pola hubungan dari data-data yang berbeda-beda.

Kemajuan luar biasa yang terus berlanjut dalam bidang data minig didorong oleh beberapa faktor, antara lain:

1. Pertumbuhan yang cepat dalam kumpulan data.
2. Penyimpanan data dalam data *warehouse,* sehingga seluruh perusahaan memiliki akses kedalam database yang handal.
3. Adanya penigkatan akses data melalui navigasi web dan intranet.
4. Tekanann kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
5. Perkembangan teknologi perangkat lunak untuk data mining.
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Data mining dibagi menjadi kelompok-kelompok berdasarkan tugas yang dapat dilakukan:

1. Deskripsi

   Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan dalam data.

2. Estimasi

   Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi, variable estimasi mengarah ke data numerik dari pada data kategori.

3. Prediksi

   Prediksi hampir sama dengan klasifikasi dan estimasi dan hasil akan ada pada masa mendatang, beberapa metode dan teknik klasifikasi dan estimasi dapat digunakan untuk prediksi.

4. Klasifikasi

   Klasifikasi merupakan target variabel kategori yang memisahkan data dalam kategori yang ditentukan.

5. Pengklusteran

   Pengklusteran merupakaan pengelompokan data yang memiliki kemiripan dengan data-data dalam cluster lain. Didalam pengklusteran tidak terdapat variabel traget dalam pengklusteran. Algoritma pengklusteran melakaukan pembaian terhadap data menjadi kelompok-kelompok yang memiliki kemiripan dengan data dan akan bernilai maksimal jika dalam satu kelompok, dan akan bernilai minimal jika kemiripan data antar kelompok lain.

6. Asosiasi

   Asosiasi adalah menemukan atribut yang muncul dalam satu waktu, dan biasanya dalam dunia bisnis disebut analisis kerangjang belanja, seperti contohnya menemukan pola penjualan barang dalam supermarket, dimana pembeli akan selalu melakukan pembelian dua atau lebih barang yang berbeda dalam satu transaksi (kusrini, 2009).

7. Diskritisasi

Diskritisasi atau yang sering disebut dengan binning, adalah proses transformasi data numerik menjadi data kategori. Metode ini diterapkan untuk metode data mining yang tidak dapat menangani data numerik, dan membantu dalam mengurangi jumlah nilai noisi pada data numerik, diskritisasi dimunkinkan untuk mengabaikan perbedaan nilai yang tidak relevan (Zaki & Meira, 2013). Terdapat 3 jenis atribut diskritisasi:

a. Nominal yaitu nilai dari set yang tidak berurutan, mis., Warna, profesi

b. Ordinal yaitu nilai dari set yang diperintahkan, mis., Pangkat militer atau akademik

c. Numerik yaitu bilangan real, mis., Bilangan bulat atau bilangan real

Proses binning dimulai dari melihat nilai-nilai yang berada disekitarnya. adapunlangkah-langkah metode binning sebagai berikut:

a. Mengurtukan data terkecil hingga data terbesar.

b. Mempartisai data yang dapat dilakukan dengan dua acara yaitu; equal-width (distance) partitioning dan equaldepth (frequency) partitioning (Junaedi et al., 2011).

8. Symmetrical Uncertainty

Metode pemilihan fitur yang paling sering digunakan adalah berbasis filter seperti informasi timbal balik, korelasi Pearson, uji chi-kuadrat, perolehan informasi, rasio gain, dan relief. metode filter berbasis korelasi cepat (FCBF) untuk menghilangkan fitur yang tidak relevan dan berlebihan.

$$IG(X|Y) = E(X)E(X|Y)$$

$$SU(X,Y) = 2\,x\,\frac{IG(X|Y)}{E(X)+E(Y)}$$

Keterangan:

E(X) dan E(Y) adalah entropi fitur

X dan Y, dan IG(X|Y) adalah perolehan informasi X setelah mengamati Y.

Korelasi C dan korelasi F didefinisikan berdasarkan SU untuk mengukur korelasi antar fitur.

C-korelasi: SU antara fitur Fi dan kelas C, dilambangkan dengan SUi,c .

F-korelasi: SU antara setiap pasangan fitur Fi dan Fj (i 6= j), dilambangkan dengan SUi,j .

Penggunaan ketidakpastian simetris telah terbukti berguna dalam pengurangan dimensi (Piao et al., 2019).

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

a. Jenis

Jenis penelitian yang dilakukan oleh peneliti adalah experiment. Penelitian ini dilakukan untuk memprediksi serangan OPT menggunakan metode K-means clustering dan naïve bayes.

b. Sifat

Sifat penelitian ini yaitu deskriptif, pada penelitian ini akan dilakukan transformasi data menggunakan discretizaztion manual dan k-means clustering untuk membandingkan tingkat akurasi naïve bayes.

c. Pendekatan penelitian

Pendekatan penelitian ini menggunkan pendekatan kuantitatif yaitu hasil dari penerapan metode kalisfikasi dan prediksi berupa text dan angka untuk menunjukan tingkat akurasi dari prediksi serangan OPT.

3.2. Metode Pengumpulan Data

Proses pengumpulan data dengan wawancara dengan Bapak Moh. Arwahni selaku pengamat hama pada wilayah setempat. Hasil dari wawancara tersebut yaitu berupa data set yang berbentuk excel. Berikut dataset yang digunkan untuk penelitian:

a. Data HPP (Hama putih palsu)

Pada data set hpp digunakan sebagai data yang akan diprediksi menggunakan methode naïve bayes.

Table 3.1 Data HPP (Hama Putih Palsu)

No	Bulan		Intensitas Serangan (%)				
			2016	*2017*	*2018*	*2019*	*2020*
1	Desember	TB I	0,0	7,4	10,2	0,0	0,0
		TB II	0,0	7,6	7,6	8,0	7,1
2	Januari	TB I	0,0	7,6	0,0	9,3	11,1
		TB II	3,5	0,0	0,0	4,4	11,1
3	Pebruari	TB I	4,3	0,0	0,0	0,0	11,1
		TB II	11,4	0,0	2,1	0,0	0,0
4	Maret	TB I	0,0	0,0	3,3	0,0	0,0
		TB II	0,0	0,0	0,7	0,0	0,0
5	April	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	0,0	0,0	0,0	0,0	0,0
6	Mei	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	6,7	0,0	0,0	0,0	0,0
7	Juni	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	0,0	0,0	0,0	0,0	0,0
8	Juli	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	0,0	0,0	0,0	0,0	0,0
9	Agustus	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	0,0	0,0	0,0	0,0	6,4
10	September	TB I	0,0	0,0	0,0	0,0	12,8
		TB II	0,0	0,0	0,0	0,0	25,0
11	Oktober	TB I	0,0	1,1	0,0	0,0	0,0
		TB II	0,0	1,6	0,0	0,0	0,0
12	Nopember	TB I	0,0	0,0	0,0	0,0	0,0
		TB II	0,0	0,0	0,0	0,0	0,0
Jumlah Kumulatif			**25,8**	**25,3**	**23,9**	**21,7**	**84,6**

Pada tabel 3.1 total intensitas serangan hama hpp pada tahun 2016 hingga tahun 2020 yaitu 118,5 serangan. Pada tahun 2020 terdapat 84,6 intesitas serangan, tahun 2019 terdapat 21,7 intesitas serangan, tahun 2018 terdapat 23,9 intesitas

serangan, tahun 2017 terdapat 25,3 serangan, dan tahun 2016 terdapat 25,8 intesitas serangan.

b. Laba-laba

Dataset laba-laba merupakan musuh alami yang membantu petani dalam mengurangi serngan hama.

Tabel 3.2 Laba-laba

No	BULAN		Kepadatan Populasi Ekor Per Rumpun				
			2016	2017	2018	2019	2020
1	Desember	TB I	1,18	3,48	17,30	14,20	11,50
		TB II	0,82	2,71	13,00	25,50	36,30
2	Januari	TB I	0,42	2,21	17,00	24,30	22,77
		TB II	1,84	0,00	6,00	25,03	34,50
3	Pebruari	TB I	2,45	0,00	8,00	23,10	29,85
		TB II	1,90	0,00	22,70	21,50	0,00
4	Maret	TB I	1,95	0,64	13,40	0,00	0,00
		TB II	0,00	1,46	28,20	0,00	0,00
5	April	TB I	0,00	1,79	14,10	16,50	0,00
		TB II	0,72	1,81	16,11	27,50	0,00
6	Mei	TB I	1,42	2,10	13,20	23,20	0,00
		TB II	1,90	1,50	21,30	25,26	0,00
7	Juni	TB I	11,25	0,00	0,00	36,40	0,00
		TB II	2,48	0,00	0,00	35,80	0,00
8	Juli	TB I	2,85	0,00	0,00	0,00	0,00
		TB II	1,24	0,00	0,00	0,00	17,50
9	Agustus	TB I	0,00	0,00	5,70	0,00	11,30
		TB II	0,00	0,00	14,80	0,00	16,12
10	September	TB I	0,00	0,00	17,80	0,00	20,60
		TB II	1,02	0,00	13,20	0,00	0,00
11	Oktober	TB I	2,00	0,21	19,70	0,00	0,00
		TB II	0,64	1,43	12,30	0,00	0,00
12	Nopember	TB I	0,86	1,10	1,50	5,30	0,00
		TB II	1,97	0,00	5,50	6,30	6,30
JUMLAH KUMULATIF			**38,91**	**20,44**	**280,81**	**309,89**	**206,74**

Pada tabel 3.2 total populasi laba-laba yaitu 856,79. Pada tahun 2016 terdapat 38.91 populasi, tahun 2017 terdapat 20,44 populasi, tahun 2018 terdapat 280,81 populasi, tahun 2019 terdapat 309,89 populasi dan tahun 206,74 populasi.

c. Paederus SP

Dataset paedarus SP merupakan musuh alami yang membantu petani dalam mengurangi serangan hama.

Table 3.3 Paedarus SP

No	Bulan		Kepadatan Populasi Ekor Per Rumpun				
			2016	*2017*	*2018*	*2019*	*2020*
1	Desember	TB I	0	1,3	3,3	5,01	6,30
		TB II	0	0,57	4	6,2	3,70
2	Januari	TB I	0	0,5	5	6,7	6,67
		TB II	0,79	0	0	3,8	6,44
3	Pebruari	TB I	0,6	0	0	2,4	7,85
		TB II	0,82	0	3,2	2,1	0,00
4	Maret	TB I	0,7	0	3,3	0	0,00
		TB II	0	0,16	3,6	0	0,00
5	April	TB I	0	0,14	2,7	5,3	0,00
		TB II	0	0,48	3,6	9,13	0,00
6	Mei	TB I	0,00	0,60	6,30	6,02	0,00
		TB II	0,28	0,10	6,50	8,14	0,00
7	Juni	TB I	0	0	0	11,4	0,00
		TB II	0	0	0	7,4	0,00
8	Juli	TB I	0	0	0	0	0,00

Table 3.3 Paedarus SP (Lanjutan)

No	Bulan		Kepadatan Populasi Ekor Per Rumpun				
			2016	*2017*	*2018*	*2019*	*2020*
		TB II	0	0	0	0	0,00
9	Agustus	TB I	0,00	0,00	0,00	0,00	0,00
		TB II	0	0	0	0	0,00
10	September	TB I	0,00	0,00	0,00	0,00	5,90
		TB II	0,00	0,00	0,00	0,00	5,88
11	Oktober	TB I	0,8	0,2	5,5	0	0,00
		TB II	0	0,12	6	0	0,00
12	Nopember	TB I	0	0,2	0	0	0,00
		TB II	0		3,2	0	0,00
JUMLAH KUMULATIF			**3,99**	**4,37**	**56,2**	**73,6**	**32,74**

Pada tabel 3.3 total populasi paedarus SP yaitu 170,9. Pada tahun 2016 terdapat 3.99 populasi, tahun 2017 terdapat 4,37 populasi, tahun 2018 terdapat 56,2 populasi, tahun 2019 terdapat 73,6 populasi dan tahun 2020 terdapat 32,74 populasi.

d. Coccinelidae

Dataset coccinelidae merupakan musuh alami yang membantu petani dalam mengurangi serangan hama pengganggu tumbuhan.

Table 3.4 Coccinelidae

No	Bulan		Kepadatan Populasi Ekor Per Rumpun				
			2016	2017	2018	2019	2020
1	Desember	TB I	0,00	1,23	0,00	0,00	0,00
		TB II	0,00	0,76	0,00	0,00	4,50

Table 3.4 Coccinelidae (Lanjutan)

No	Bulan		Kepadatan Populasi Ekor Per Rumpun				
			2016	2017	2018	2019	2020
2	Januari	TB I	0,70	0,60	0,00	0,52	6,99
		TB II	0,40	0,00	0,00	4,70	3,29
3	Pebruari	TB I	0,53	0,00	0,00	3,50	5,45
		TB II	0,30	0,00	0,00	2,60	0,00
4	Maret	TB I	0,80	0,00	1,50	0,00	0,00
		TB II	0,00	0,20	1,14	0,00	0,00
5	April	TB I	0,00	0,35	1,00	0,00	0,00
		TB II	0,00	0,48	2,50	0,00	0,00
6	Mei	TB I	0,00	0,82	0,00	0,00	0,00
		TB II	0,27	0,10	0,00	0,00	0,00
7	Juni	TB I	0,27	0,00	0,00	3,20	0,00
		TB II	0,00	0,00	0,00	5,80	0,00
8	Juli	TB I	0,66	0,00	0,00	0,00	0,00
		TB II	0,00	0,00	0,00	0,00	0,00
9	Agustus	TB I	0,00	0,00	0,00	0,00	0,00
		TB II	0,00	0,00	0,00	0,00	0,00
10	September	TB I	0,00	0,00	0,00	0,00	0,00
		TB II	0,04	0,00	0,00	0,00	0,00
11	Oktober	TB I	0,60	0,00	3,00	0,00	0,00
		TB II	0,00	0,00	0,00	0,00	0,00
12	Nopember	TB I	0,00	0,00	0,00	0,00	0,00
		TB II	0,67	0,00	0,00	0,00	0,00
	Jumlah Kumulatif		**5,24**	**2,55**	**9,14**	**20,32**	**20,23**

Pada table 3.3 total populasi dari tahun 2016 hingga 2020 yaitu 57,48 populasi. Pada tahun 2016 terdapat 5,24 populasi, pada tahun 2017 terdapat 2,55 populasi, pada tahun 2018 terdapat 9,14 populasi, pada tahun 2019 terdapat 20,32 populasi dan pada tahun 2020 terdapat 20,23 populasi.

e. Capung

Dataset capung merupakan musuh alami yang membantu petani dalam mengurangi serangan hama pengganggu tumbuhan.

Table 3.4 Capung

No	Bulan		Kepadatan Populasi Ekor Per Rumpun				
			2016	*2017*	*2018*	*2019*	*2020*
1	Desember	TB I	0,00	0,75	0	0	0
		TB II	0,00	1,68	0	0	0
2	Januari	TB I	0,08	0,37	0	0	0
		TB II	0,00	0,00	0	0	0
3	Pebruari	TB I	0,25	0,00	0	0	0
		TB II	0,25	0,00	0	0	0
4	Maret	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
5	April	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
6	Mei	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
7	Juni	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
8	Juli	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
9	Agustus	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
10	September	TB I	0,00	0,00	0	0	0
		TB II	0,40	0,00	0	0	0
11	Oktober	TB I	0,00	0,00	0	0	0
		TB II	0,00	0,00	0	0	0
12	Nopember	TB I	0,00	0,00	0	0	0
		TB II	0,36	0,00	0	0	0
	Jumlah Kumulatif		**1,34**	**2,80**	0	0	0

Pada tabel 3.4 total populasi dari tahun 2016 dan 2017 yaitu 4,14 populasi. Pada tahun 2016 terdapat 1,34 populasi, pada tahun 2017 terdapat 2,80 populasi, dan pada tahun 2018 hingga 2020 tidak ditemukan populasi capung.

f. Opheonea. Sp

Dataset Opheonea. Sp merupakan musuh alami yang membantu petani dalam mengurangi serangan hama pengganggu tumbuhan.

NO	BULAN		KEPADATAN POPULASI EKOR PER RUMPUN				
			2016	*2017*	*2018*	*2019*	*2020*
1	DESEMBER	TB I	0	0	1,20	2,83	0
		TB II	0	0	12,00	4,70	4,5
2	JANUARI	TB I	0	0	5,00	3,70	4,57
		TB II	0	0	0,00	5,60	4,13
3	PEBRUARI	TB I	0	0	0,00	4,80	6,65
		TB II	0	0	2,30	3,20	0
4	MARET	TB I	0	0	1,90	0,00	0
		TB II	0	0	5,98	0,00	0
5	APRIL	TB I	0	0	0,70	0,00	0
		TB II	0	0	2,04	0,00	0
6	MEI	TB I	0	0	4,90	4,03	0,00
		TB II	0	0	5,00	6,30	0,00
7	JUNI	TB I	0	0	0,00	6,03	0
		TB II	0	0	0,00	5,90	0
8	JULI	TB I	0	0	0,00	0,00	0
		TB II	0	0	0,00	0,00	0
9	AGUSTUS	TB I	0	0	0,00	0,00	0,00
		TB II	0	0	0,00	0,00	0
10	SEPTEMBER	TB I	0	0	0,00	0,00	0,00
		TB II	0	0	0,00	0,00	0,00
11	OKTOBER	TB I	0	0	0,00	0,00	0
		TB II	0	0	7,00	0,00	0
12	NOPEMBER	TB I	0	0	0,00	0,00	0
		TB II	0	0	1,20	0,00	0
JUMLAH KUMULATIF			0	0	**36,02**	**39,56**	**19,85**

Pada tabel 3.5 total populasi dari tahun 2018 hingga 2020 yaitu 95,43 populasi. Pada tahun 2016 dan 2017 tidak ditemukan populasi Opheonea. Sp pada tahun 2018 terdapat 36,02 populasi, pada tahun 2019 terdapat 39,56 populasi, dan pada tahun 2020 terdapat 19,85 populasi.

g. Luas lahan

Dataset luas lahan merupakan jumlah luas lahan yang ditanami padi,

Tabel 3.6 Luas lahan

No	Bulan		Luas Tanam (Ha)				
			2016	*2017*	*2018*	*2019*	*2020*
1	Desember	TB I	620	2.490	1.270	1.250	444
		TB II	2.136	2.815	1.260	1.879	1.385
2	Januari	TB I	2.932	1.915	1.165	2.079	1.528
		TB II	3.182	800	1.070	3.758	1.691
3	Pebruari	TB I	3.082	400	884	3.207	2.193
		TB II	2.732	220	1.200	2.217	2.657
4	Maret	TB I	2.432	690	1.530	1.645	1.808
		TB II	3.233	956	1.730	1.945	1.816
5	April	TB I	2.633	900	2.295	2.579	1.673
		TB II	2.712	900	1.875	1.043	1.468
6	Mei	TB I	2.712	900	1.500	1.043	818
		TB II	2.562	600	809	829	345
7	Juni	TB I	2.442	400	609	829	200
		TB II	1.772	350	510	229	725
8	Juli	TB I	981	75	352	0	850
		TB II	651	0	265	0	880
9	Agustus	TB I	500	0	215	0	879
		TB II	250	0	215	0	879

Tabel 3.6 Luas lahan (Lanjutan)

No	Bulan		Luas Tanam (Ha)				
			2016	*2017*	*2018*	*2019*	*2020*
10	September	TB I	200	0	215	0	534
		TB II	1.290	20	215	0	312
11	Oktober	TB I	1.640	330	160	0	155
		TB II	2.040	590	50	42	155
12	Nopember	TB I	2.090	845	990	82	15
		TB II	2.340	0	1.500	194	85
Jumlah Total			**47.164**	**16.196**	**21.884**	**24.850**	**23.495**

Pada table 3.6 merupakan data dari tahun 2016 hingga 2020 dengan jumlah total 133.589 Ha. Pada tahun 2016 terdapat 47.164 Ha, pada tahun 2017 terdapat 16.196 Ha, pada tahun 2018 terdapat 21.884 Ha, pada tahun 2019 terdapat 24.850 Ha dan pada tahun 2020 23.495 Ha.

h. Curah hujan (mm)

Dataset curah hujan (mm) merupakan data air hujan dalam satuan mm.

Table 3.7 Curah hujan (mm)

No	Bulan		Curah Hujan (mm)				
			2016	2017	2018	2019	2020
1	Desember	TB I	188	83	118	58	224
		TB II	268	50	148	165	192
2	Januari	TB I	224	290	238	238	438
		TB II	163	234	138	138	423
3	Pebruari	TB I	138	218	347	282	142
		TB II	58	309	468	100	210
4	Maret	TB I	180	249	347	439	37

Table 3.7 Curah hujan (mm) (Lanjutan)

No	Bulan		Curah Hujan (mm)				
			2016	2017	2018	2019	2020
		TB II	363	170	160	280	293
5	April	TB I	58	45	127	167	407
		TB II	242	61	123	298	290
6	Mei	TB I	234	38	35	24	325
		TB II	172	9	169	88	534
7	Juni	TB I	49	48	9	4	515
		TB II	45	45	186	30	265
8	Juli	TB I	60	55	1	79	113
		TB II	38	82	9	0	87
9	Agustus	TB I	45	7	0	0	51
		TB II	47	52	0	13	93
10	September	TB I	23	11	36	0	84
		TB II	147	99	96	0	155
11	Oktober	TB I	217	53	0	0	149
		TB II	163	45	89	84	33
12	Nopember	TB I	58	52	0	114	84
		TB II	258	188	0	24	108
Jumlah Total			**3434**	**2493**	**2838**	**2625**	**5252**

Pada tabel 3.7 total jumlah air hujan pada tahun 2016 hingga 2020 terdapat 16.642 mm. pada tahun 2016 terdapat 3.434 mm, pada tahun 2017 teredapat 2.493 mm, pada tahun 2018 terdapat 2.838 mm, pada tahun 2019 terdapat 2.625 mm dan pada tahun 2020 terdapat 5.252 mm.

i. Curah Hujan (Hari)

Dataset curah hujan (hari) merupakan data jumlah hari hujan pada tahun 2016 hingga 2020.

Tabel 3.8 Curah Hujan (hari)

No	Bulan		Jumlah Hari Hujan				
			2016	2017	2018	2019	2020
1	Desember	TBI	13	12	12,5	3	11
		TBII	13	14	14,5	6	16
2	Januari	TBI	12	12	11	15	12
		TBII	11	13	9	11	16
3	Pebruari	TBI	9	9	14,5	13	9
		TBII	6	10	11,5	12	8
4	Maret	TBI	9,5	11	14	14	8
		TBII	13,5	14	11	12	13
5	April	TBI	5,5	8	9	10	12
		TBII	13,5	2	10	12	12
6	Mei	TBI	12	9	3	4	12
		TBII	12	7	8	6	14
7	Juni	TBI	5	8	3	3	12
		TBII	5	9	2	4	9
8	Juli	TBI	4,5	7	0,5	5	11
		TBII	5,5	8	1,5	1	10
9	Agustus	TBI	2,5	2	0	0	6

Tabel 3.8 Curah Hujan (hari)

No	Bulan		Jumlah Hari Hujan				
			2016	2017	2018	2019	2020
		TB II	3,5	6	2	3	8
10	September	TB I	5	2	2,5	0	8
		TB II	8	4	2,5	1	7
11	Oktober	TB I	9,5	8	0	0	11
		TB II	9,5	5	2	2	5
12	Nopember	TB I	6,5	6	0	3	10
		TB II	8,5	12	0	4	11
Jumlah Total			**203**	**198**	**144**	**144**	**251**

Pada gambar 3.8 total jumlah hari hujan pada tahun 2016 hingga 2020 yaitu terdapat 940 hari. pada tahun 2016 terdapat 203 hari, pada tahun 2017 teredapat 198 hari, pada tahun 2018 terdapat 144 hari, pada tahun 2019 terdapat 144 hari dan pada tahun 2020 terdapat 251 hari.

3.3. Cleansing data

Pada tahap ini peneliti melakukan cleansing data dengan deteksi dan memperbaiki (atau menghapus) data set, tabel, dan data yang rusak. Pada proses ini mengacu pada identifikasi data yang kurang lengkap, kurang tepat, kurang benar, dan kurang relevan, yang kemudian dirty data tersebut akan diganti. Ada beberapa fokus area didalam data cleasing, yaitu data kosong, dan data kurang tepat (Riezka et al., 2011).

3.4. Transformasi Data

Pada tahap ini peneliti menentukan jumlah pembagian kategori berdasarkan literatur dan berdasarakan ahli dibidang pertanian.

a. Pembagian kategori intesitas serangan OPT

Tabel 3.9 intensitas serangan OPT

Intensitas Serangan (%)	Tingkat Kerusakan
0,0-1,0	Sehat
1,1-25,0	Ringan
25,0-50,0	Sedang
50,1-75,0	Berat
75,1-100	Sangat Berat

Pada tabel 3.1 menjelaskan pembagian intesitas serangan OPT dengan tingkat serang dalam bentuk persen (%) dan dikateorikan menjadi 5 kategori instesitas serangan OPT (Zeni et al., 2021).

b. Pembagian kategori curah hujan (mm)

BMKG membagi curah hujan bulanan menjadi empat kategori yaitu rendah (0-100 mm), sedang (100-300 mm), tinggi (300-500) dan sangat tinggi (> 500).(Supriyati et al., 2018)

c. Berdasarkan wawancara dengan petugas POPT Kecamatan Batanghari Nuban, didapatkan pembagian kategori populasi musuh alami perumpun dengan kategori sedikit (0-3 ekor), sedang (>3-10 ekor), banyak (>10-20 ekor) dan sangat banyak (>20-40 ekor). Pembagian luas lahan tanam dengan pembagian 4 kategori yaitu luasan sempit (1-600 ha), luasan

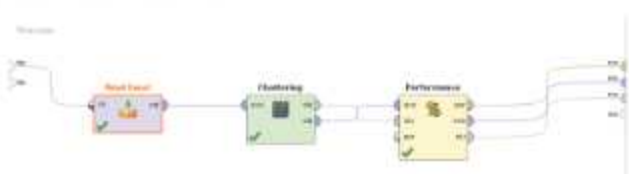
sedang (>600-1300 ha), lausan luas (>1300-2000 ha) dan luasan sangat luas (>2000-4000 ha). Pembagian jumlah hujan (hari) dengan pembagian 3 kategori yaitu sedikit (1-10 hari), sedang (11 – 20 hari) dan banyak (21-31 hari)

d. Transformas data menggunakan teknik discretization

Pada tahap ini peneliti melakukan transformasi data dengan cara manual yaitu dengan membagi kategori menggunakan Microsoft excel dengan menggunakan rumus excel. Berikut rumus yang digunakan untuk membagi kategori IF(AND("range awal","range akhir");"nama kategori".

e. Transformasi data menggunkan algoritma k-means clustering

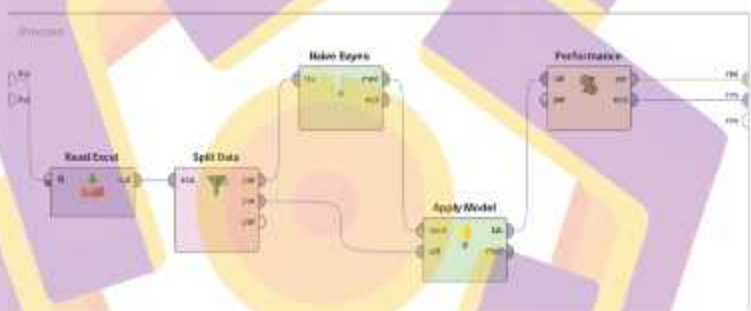
Pada tahap ini peneliti melakukan transformasi data menggunakan algoritma k-means clustering dengan menggunakan rapidminer studio 9.9. pada input data menggunakan data access read excel, pada clustering menggunakan k-measn clustering dengan k= sesuai dengan pengkategorian yang sudah ditentukan, max run =10, measure types=NumericalMeasure, numerical measure=EuclideanDistance, max optimization steps = 100. Pada performance menggunakan performance cluster distance performance, main criterion=Davies Bouldin dan mengaktifkan normalize dan maximize, dengan sekema yang ditunjukan pada gambar 3.8.



Gambar 3.8 k-means pada rapidminer

3.5. Prediksi Menggunakan Naïve Bayes

Pada tahap ini peneliti menggunkana algoritma naïve bayes untuk melakukan prediksi serangan hama menggunakan rapidminer studio 9.9 dengan menggunakan data hasil transfortmasi menggunakan metode discretization manual dan discretization menggunakan algoritma k-means clustering, dan membandingkan hasil prediksi menggunakan data hasil discretization manual dengan data hasil discretization menggunakan k-means clustering, dengan sekema yang ditunjukan pada gambar 3.9.

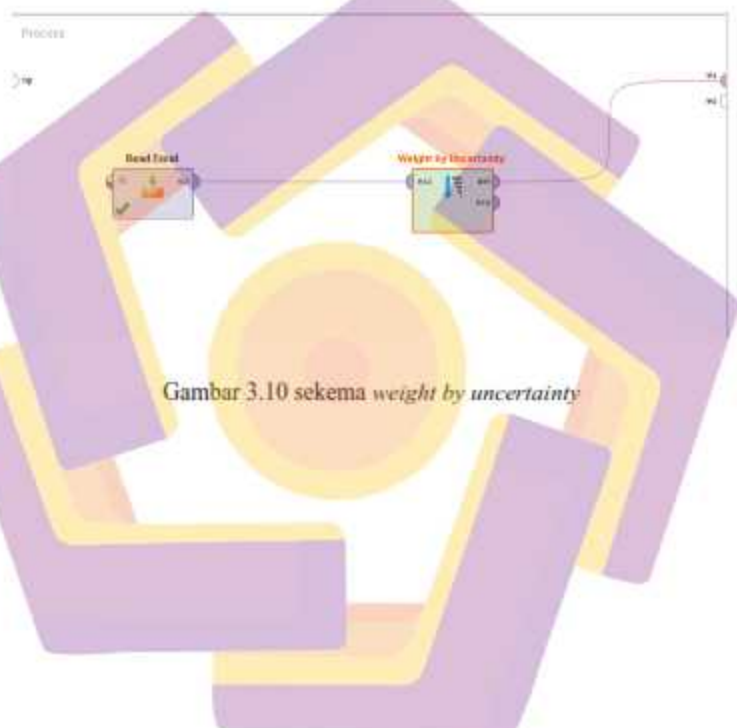


Gambar 3.9 Naïve bayes

Pada algoritma naïve bayes data dibagi menjadi 2 data yaitu data training dan data testing, pembagian data training dan data testing dibagi dengan dua sekenario yaitu sekenario pertama membagi data menjadi 80% data training dan 20% data testing dan sekenario kedua membagi data dengan 70% data training dan 30% data testing. Kelas yang diprediksi yaitu intesitas serangan dengan variabel hpp (Hama Putih Palsu) dengan kelas prediksi sehat, ringan, sedang, berat, sangat berat.

3.6. Pengujian relevansi fitur

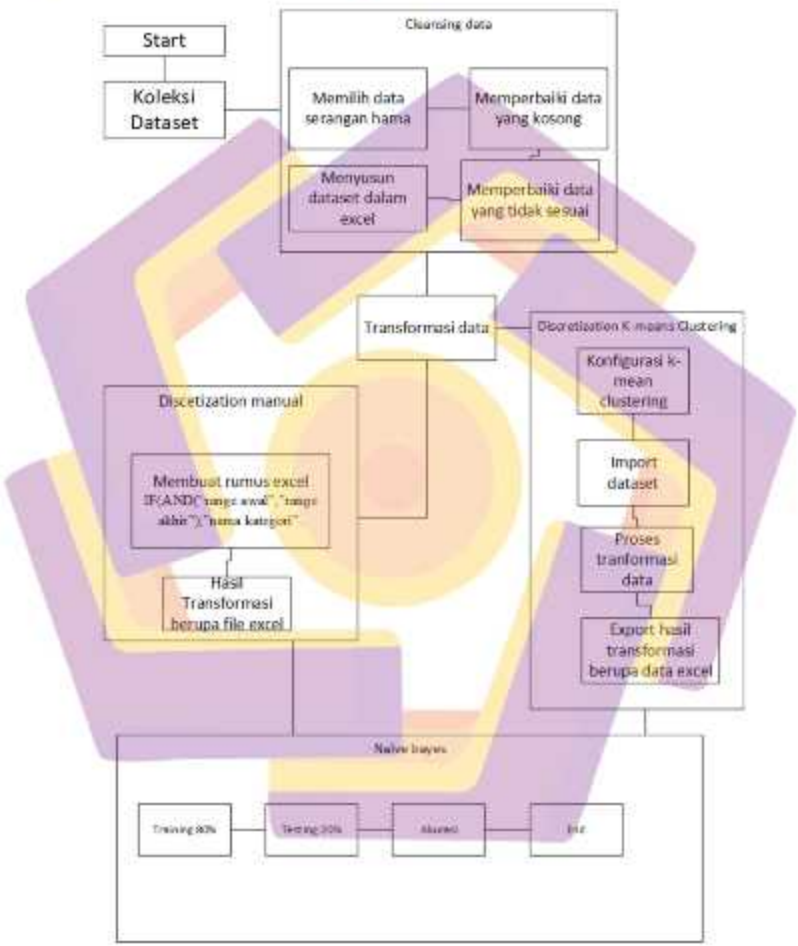
Pada tahap ini peneliti melakukan pengujian relevansi fitur untuk mengetahui fitur-fitur yang mempengaruhi hasil akurasi naïve bayes, pengujian dilakukan menggunkan rapid miner dengan menggunakan metode *Weight by Uncertainty*, dengan sekema pengujian sebagai berikut:



Gambar 3.10 sekema *weight by uncertainty*

3.7. Alur Penelitian

Alur penelitian dapat dilihat pada gambar 3.10 dimulai dari pengumpulan data, *preprocessing* data hingga akurasi prediksi.



Gambar 3.10 Alur penelitian

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Cleansing Data

Pada tahap cleansing data peniliti melakukan perbaikan pada data yang kurang tepat sehingga data menjadi lebih baik. Berikut data yang kurang tepat ditunjukan pada gambar 4.1.

NO	BULAN		Insiden serangan (%)
			2016
1	DESEMBER	TB I	0,0
		TB II	0,0
2	JANUARI	TB I	9,3
		TB II	4,6
3	PEBRUARI	TB I	5,7
		TB II	11,4
4	MARET	TB I	11,4
		TB II	-
5	APRIL	TB I	-
		TB II	-
6	MEI	TB I	0,0
		TB II	0,0
7	JUNI	TB I	5,1
		TB II	11,4
8	JULI	TB I	5,7
		TB II	11,4
9	AGUSTUS	TB I	0,0
		TB II	0,0

Gambar 4.1 data kurang tepat

Pada gambar 4.1 terdapat data yang berisikan dengan tanda (-) pada bulan maret TB II dan pada bulan april TB I dan TB II, sehingga data tersebut kurang tepat maka data akan dirubah menjadi angka nol (0), seperti yang ditunjukan pada gambar 4.2

NO	BULAN		Intensitas serangan (%)
			2016
1	DESEMBER	TB I	0,0
		TB II	0,0
2	JANUARI	TB I	9,3
		TB II	4,6
3	PEBRUARI	TB I	5,7
		TB II	11,4
4	MARET	TB I	11,4
		TB II	0,0
5	APRIL	TB I	0,0
		TB II	0,0
6	MEI	TB I	0,0
		TB II	0,0
7	JUNI	TB I	5,1
		TB II	11,4
8	JULI	TB I	5,7
		TB II	11,4
9	AGUSTUS	TB I	0,0
		TB II	0,0

Gambar 4.2 perbaikan data

Pada gambar 4.2 pada bulan maret dan April sudah diperbaiki dengan mengganti symbol (-) menjadi angka nol (0). Peniliti juga nemukan data yang kosong seeprti yang ditunjukan pada gambar 4.3.

NO	BULAN		Intensitas serangan (%)
			Revisi 2017
1	DESEMBER	TB I	7,4
		TB II	7,6
2	JANUARI	TB I	7,6
		TB II	0,0
3	PEBRUARI	TB I	0,0
		TB II	0,0
4	MARET	TB I	0,0
		TB II	0,0
5	APRIL	TB I	0,0
		TB II	0,0
6	MEI	TB I	0,0
		TB II	0,0
7	JUNI	TB I	0,0
		TB II	0,0
8	JULI	TB I	0,0
		TB II	0,0
9	AGUSTUS	TB I	0,0
		TB II	0,0
10	SEPTEMBER	TB I	0,0
		TB II	0,0
11	OKTOBER	TB I	1,1
		TB II	1,6
12	NOPEMBER	TB I	0,0
		TB II	
	JUMLAH KUMULATIF		25,3

Gambar 4.3 data kosong

Pada gambar 4.3 ditemukan data yang kosong yaitu pada bulan November TB II, sehingga perlu dilakukan perbaikan dengan diberikan nilai 0 sehingga tidak terdapat data yang kosong, dan hasil dari perbaikan ditunjukan pada gambar 4.4.

NO	BULAN		Intensitas serangan (%)
			Rerata 2017
1	DESEMBER	TB I	7,4
		TB II	7,6
2	JANUARI	TB I	7,5
		TB II	0,0
3	PEBRUARI	TB I	0,0
		TB II	0,0
4	MARET	TB I	0,0
		TB II	0,0
5	APRIL	TB I	0,0
		TB II	0,0
6	MEI	TB I	0,0
		TB II	0,0
7	JUNI	TB I	0,0
		TB II	0,0
8	JULI	TB I	0,0
		TB II	0,0
9	AGUSTUS	TB I	0,0
		TB II	0,0
10	SEPTEMBER	TB I	0,0
		TB II	0,0
11	OKTOBER	TB I	1,1
		TB II	1,6
12	NOPEMBER	TB I	0,0
		TB II	0,0
	JUMLAH KUMULATIF		**25,3**

Gambar 4.4 perbaikan nilai yang kosong

4.2. Menyusun Dataset Dalam Excel

Pada tahap ini peneliti melakukan penggabungan data pada masing-masing dataset pada tahun 2016 hingga tahun 2020, berikut ditunjukan hasil penggabungan data dalam bentuk excel pada gambar 4.5.

Tahun	BULAN		Perumpun
2016	DESEMBER	TB I	0,0
		TB II	0,0
	JANUARI	TB I	0,0
		TB II	3,5
	PEBRUARI	TB I	4,3
		TB II	11,4
	MARET	TB I	0,0
		TB II	0,0
2020	AGUSTUS	TB I	0,0
		TB II	0,0
	SEPTEMBER	TB I	0,0
		TB II	0,0
	OKTOBER	TB I	0,0
		TB II	0,0
	NOPEMBER	TB I	0,0
		TB II	0,0

Gambar 4.5 Penggabungan dataset

4.3. Hasil Transformasi Data

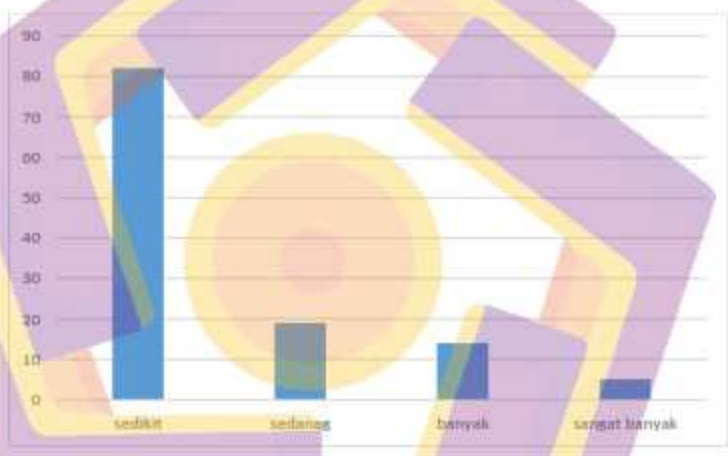
Pada transformasi data yang dilakukan pada preprosesing naïve bayes dataset dirubah dari data numerik menjadi data interval dengan menggunakan discretization manual dan menggunakan k-means clustering.

4.3.1. K-means Clustering

Pada penelitian ini penliti menggunakan rapidminer studio 9.9 untuk melakukan klasifikasi menggunkan k-means clustering, berikut hasil dari penglompokan data menggunakan rapidminer studio 9.9.

a. Laba-laba

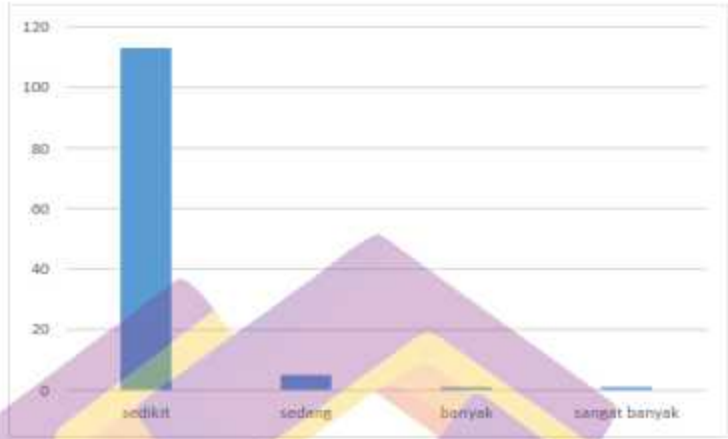
Pada dataset laba-laba dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 1,003 dengan jumlah data 82 item, nilai centroid kategori sedang yaitu 14,136 dengan jumlah data 19 item, nilai centroid kategori banyak yaitu 23,619 dengan jumlah data 14 item, dan nilai centroid kategori sanagat banyak yaitu 34,57 dengan jumlah data 5 item dengan ditunjukan pada gambar 4.6.



Gambar 4.6 Grafik k-means clustering laba-laba

b. Capung

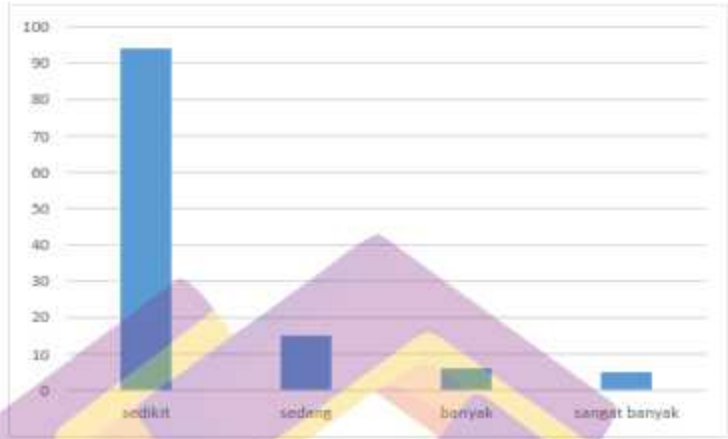
Pada dataset capung dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 0,001 dengan jumlah data 113 item, nilai centroid kategori sedang yaitu 0,326 dengan jumlah data 5 item, nilai centroid kategori banyak yaitu 0,75 dengan jumlah data 1 item, dan nilai centroid kategori sanagat banyak yaitu 1,68 dengan jumlah data 1 item dengan ditunjukan pada gambar 4.7.



Gambar 4.7 Grafik k-means clustering capung

c. Coccinellidae

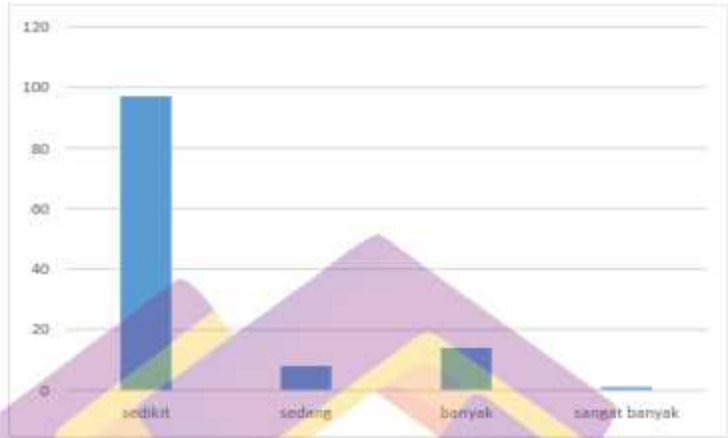
Pada dataset Coccinelidae dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 0,021 dengan jumlah data 94 item, nilai centroid kategori sedang yaitu 0,801 dengan jumlah data 15 item, nilai centroid kategori banyak yaitu 3,015 dengan jumlah data 6 item, dan nilai centroid kategori sanagat banyak yaitu 5,488 dengan jumlah data 5 item dengan ditunjukan pada gambar 4.8.



Gambar 4.8 Grafik k-means clustering coccinelidae

d. Opheonea. Sp

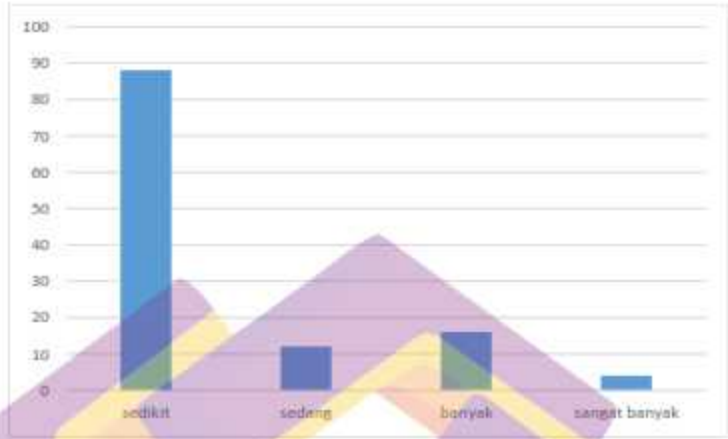
Pada dataset Opheonea. Sp dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 0,032 dengan jumlah data 97 item, nilai centroid kategori sedang yaitu 3,016 dengan jumlah data 8 item, nilai centroid kategori banyak yaitu 5,495 dengan jumlah data 14 item, dan nilai centroid kategori sanagat banyak yaitu 12 dengan jumlah data 1 item dengan ditunjukan pada gambar 4.9.



Gambar 4.9 Grafik k-means clustering Opheonea. Sp

e. Paederus. SP

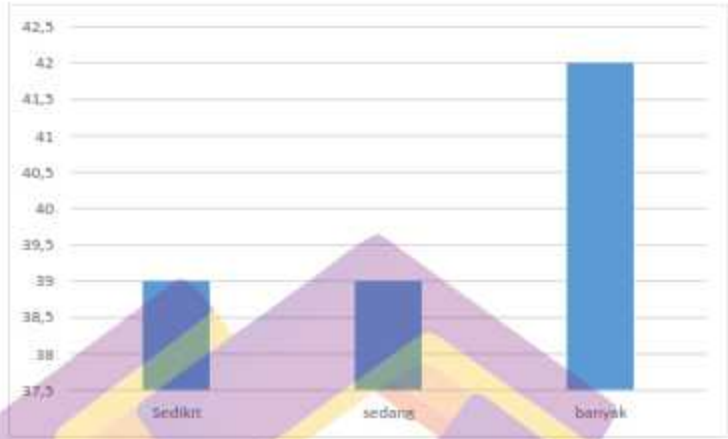
Pada dataset Paederus. SP dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 0,095 dengan jumlah data 88 item, nilai centroid kategori sedang yaitu 3,242 dengan jumlah data 12 item, nilai centroid kategori banyak yaitu 6,07 dengan jumlah data 16 item, dan nilai centroid kategori sanagat banyak yaitu 9,13 dengan jumlah data 4 item dengan ditunjukan pada gambar 4.10.



Gambar 4.10 Grafik k-means clustering Paederus. SP

f. Curah hujan (hari)

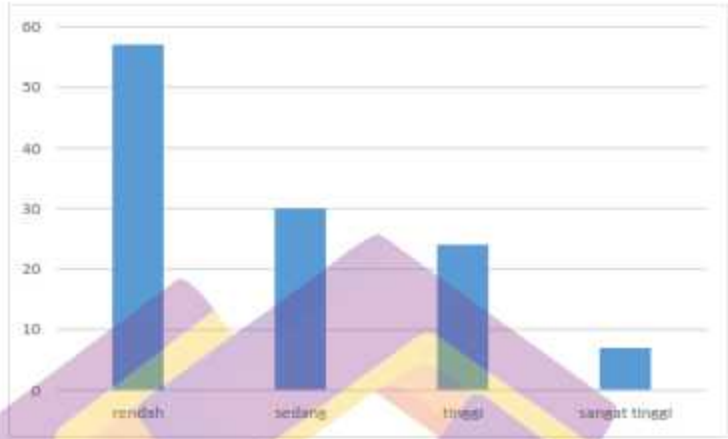
Pada dataset curah hujan (hari) dilakukan pembagian 3 kategori yaitu sedikit, sedang, dan banyak, dengan nilai centroid pada kategori sedikit yaitu 2,5 dengan jumlah data 39 item, nilai centroid kategori sedang yaitu 8,038 dengan jumlah data 39 item, dan nilai centroid kategori banyak yaitu 12,595 dengan jumlah data 42 item, dengan ditunjukan pada gambar 4.11.



Gambar 4.11 Grafik k-means clustering curah hujan (hari)

g. Curah hujan (mm)

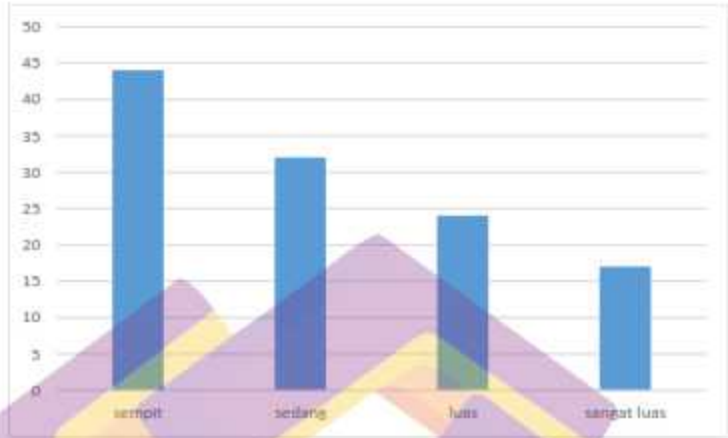
Pada dataset curah hujan (mm) dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, dengan nilai centroid pada kategori sedikit yaitu 38,132 dengan jumlah data 57 item, nilai centroid kategori sedang yaitu 145,323 dengan jumlah data 31 item, nilai centroid kategori banyak yaitu 269,6 dengan jumlah data 25 item, dan nilai centroid kategori sanagat banyak yaitu 460,5 dengan jumlah data 7 item dengan ditunjukan pada gambar 4.12.



Gambar 4.12 Grafik k-means clustering curah hujan (mm)

h. Luas lahan

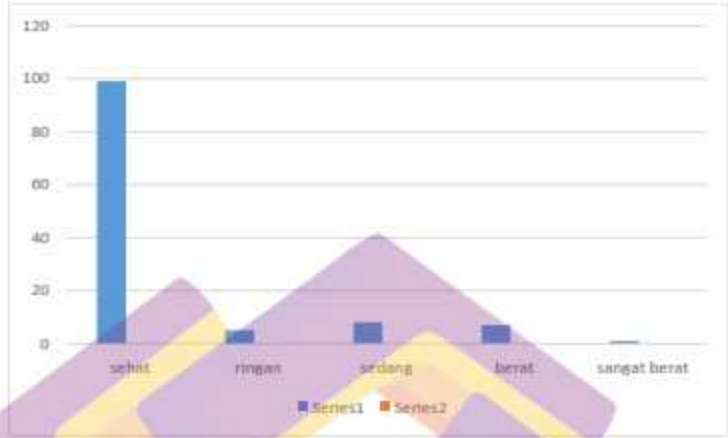
Pada dataset luas lahan dilakukan pembagian 4 kategori yaitu sempit, menengah, luas dan sangat luas, dengan nilai centroid pada kategori sempit yaitu 175,773 dengan jumlah data 44 item, nilai centroid kategori menengah yaitu 909,242 dengan jumlah data 33 item, nilai centroid kategori luas yaitu 1814 dengan jumlah data 25 item, dan nilai centroid kategori sanagat luas yaitu 2805.556 dengan jumlah data 18 item dengan ditunjukan pada gambar 4.13.



Gambar 4.13 Grafik k-means clustering luas lahan

l. HPP

Pada dataset HPP dilakukan pembagian 5 kategori yaitu sehat, ringan, sedang, berat dan sangat berat, nilai centroid pada kategori sehat yaitu 0,034 dengan jumlah data 99 item, nilai centroid pada kategori ringan yaitu 3,51 dengan jumlah data 5 item, nilai centroid kategori sedang yaitu 7,303 dengan jumlah data 8 item, nilai centroid kategori berat yaitu 11,004 dengan jumlah data 7 item, dan nilai centroid kategori sanagat berat yaitu 25 dengan jumlah data 1 item dengan ditunjukan pada gambar 4.14.



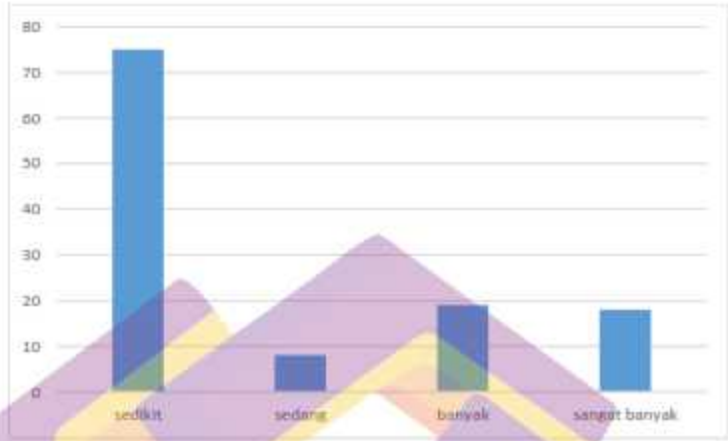
Gambar 4.14 Grafik k-means clustering HPP

4.3.2. Teknik Discretization

Pada penelitian ini dilakukan discretization manual menggunakan Microsoft Office Excel, berikut hasil berikut hasil dari pengelompokan data menggunkan Microsoft Office excel:

a. Laba-laba

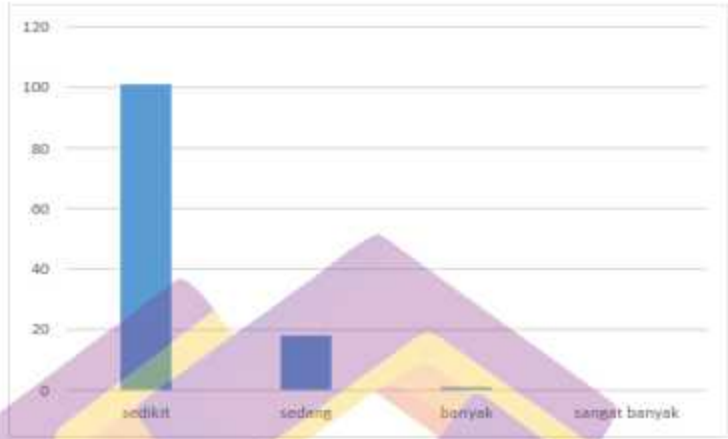
Pada dataset laba-laba dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, nilai range pada kategori sedikit yaitu 0-3 dengan jumlah data 75 item, nilai range kategori sedang yaitu >3-10 dengan jumlah data 9 item, nilai range kategori banyak yaitu >10-20 dengan jumlah data 19 item, dan nilai range kategori sangat banyak yaitu >20-40 dengan jumlah data 18 item dengan ditunjukan pada gambar 4.15.



Gambar 4.15 Grafik discretization manual laba-laba

b. Ophanea SP

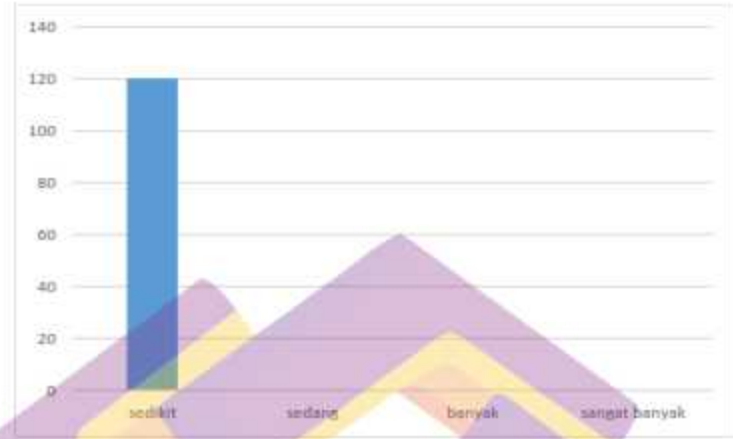
Pada dataset ophanea SP dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, nilai range pada kategori sedikit yaitu 0-3 dengan jumlah data 101 item, nilai range kategori sedang yaitu >3-10 dengan jumlah data 18 item. nilai range kategori banyak yaitu >10-20 dengan jumlah data 1 item, dan nilai range kategori sangat banyak yaitu >20-40 dengan jumlah data 0 item dengan ditunjukan pada gambar 4.16.



Gambar 4.16 Grafik discretization manual ophanea SP

c. Capung

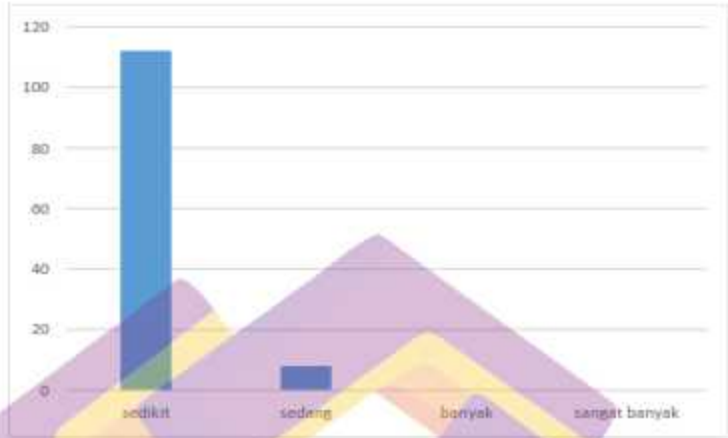
Pada dataset capung dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, nilai range pada kategori sedikit yaitu 0-3 dengan jumlah data 120 item, nilai range kategori sedang yaitu >3-10 dengan jumlah data 0 item, nilai range kategori banyak yaitu >10-20 dengan jumlah data 0 item, dan nilai range kategori sangat banyak yaitu >20-40 dengan jumlah data 0 item dengan ditunjukan pada gambar 4.17.



Gambar 4.17 Grafik discretization manual capung

d. Coccinellidae

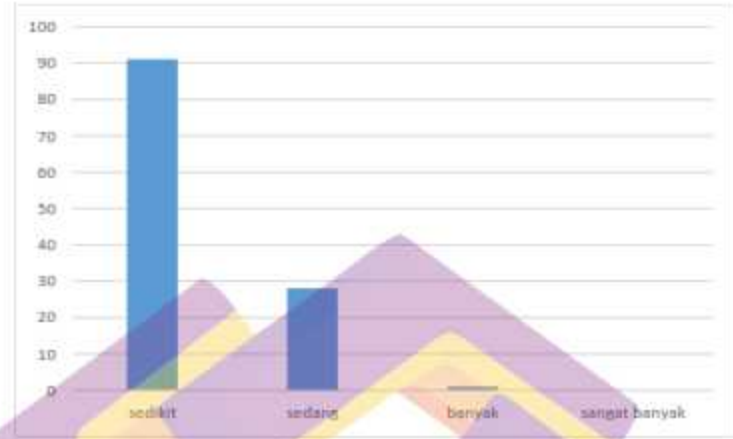
Pada dataset coccinelidae dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, nilai range pada kategori sedikit yaitu 0-3 dengan jumlah data 112 item, nilai range kategori sedang yaitu >3-10 dengan jumlah data 8 item, nilai range kategori banyak yaitu >10-20 dengan jumlah data 0 item, dan nilai range kategori sangat banyak yaitu >20-40 dengan jumlah data 0 item dengan ditunjukan pada gambar 4.18.



Gambar 4.18 Grafik discretization manual capung

e. Paedarus SP

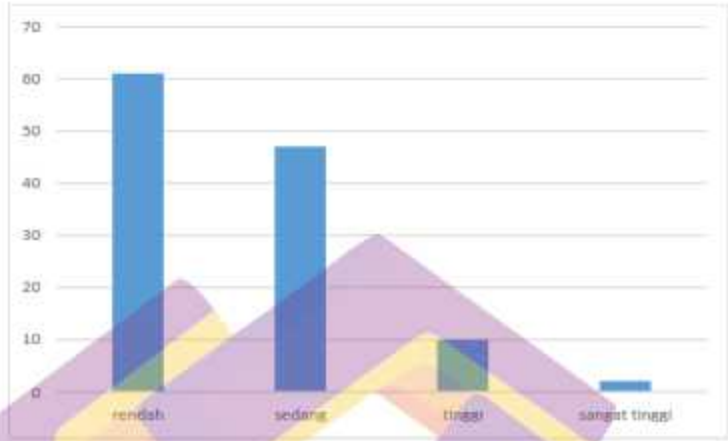
Pada dataset paedarus sp dilakukan pembagian 4 kategori yaitu sedikit, sedang, banyak dan sangat banyak, nilai range pada kategori sedikit yaitu 0-3 dengan jumlah data 91 item, nilai range kategori sedang yaitu >3-10 dengan jumlah data 28 item, nilai range kategori banyak yaitu >10-20 dengan jumlah data 1 item, dan nilai range kategori sangat banyak yaitu >20-40 dengan jumlah data 0 item dengan ditunjukan pada gambar 4.19.



Gambar 4.19 Grafik discretization manual paedarus SP

f. Curah Hujan (mm) dengan 4 kategori

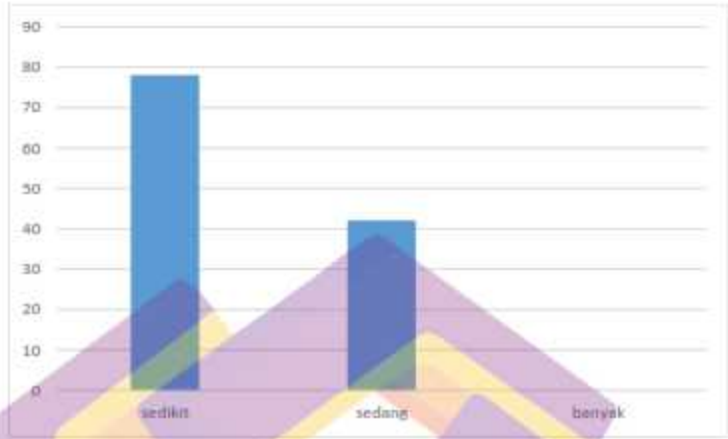
Pada dataset curah hujan (mm) dilakukan pembagian 4 kategori yaitu rendah, sedang, tinggi dan sangat tinggi, dengan nilai range pada kategori rendah yaitu 0-100 dengan jumlah data 61 item, nilai range kategori menengah yaitu 100-300 dengan jumlah data 47 item, nilai range kategori tinggi yaitu 300-500 dengan jumlah data 10 item dan nilai range kategori sangat tinggi yaitu >500 dengan jumlah data 2 item, dengan ditunjukan pada gambar 4.20.



4.20 Grafik discretization manual curah hujan (hari)

g. Curah Hujan (hari) dengan 4 kategori

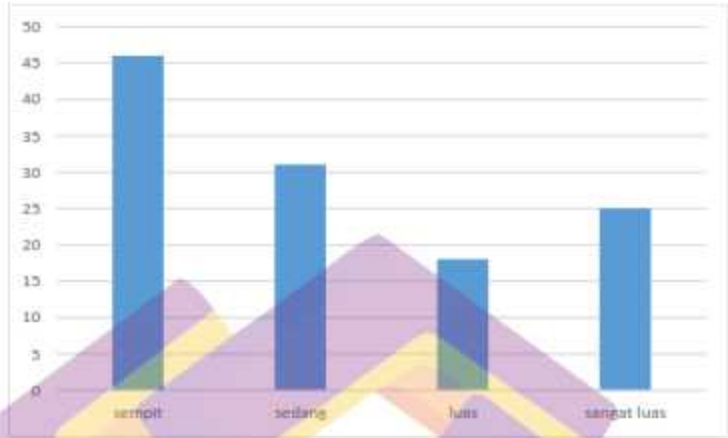
Pada dataset curah hujan (hari) dilakukan pembagian 4 kategori yaitu sedikit, sedang, dan banyak, dengan nilai range pada kategori sedikit yaitu 0-10 dengan jumlah data 78 item, nilai range kategori sedang yaitu 11-20 dengan jumlah data 42 item, dan nilai range kategori banyak yaitu 21-31 dengan jumlah data 0 item, dengan ditunjukan pada gambar 4.21.



Gambar 4.21 Grafik discretization manual curah hujan (hari)

h. Luas Tanam

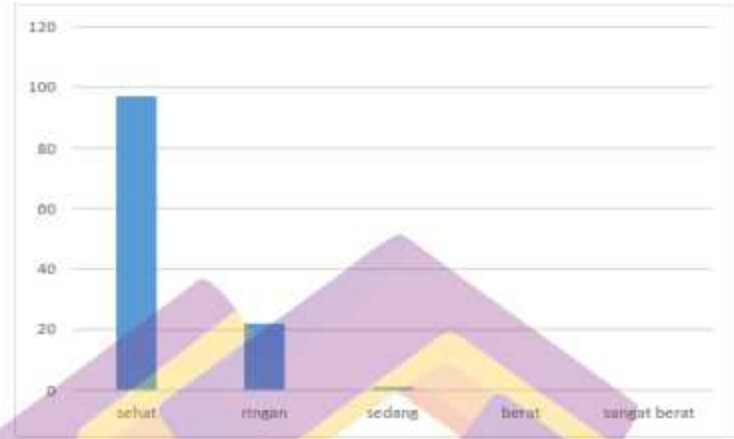
Pada dataset luas lahan dilakukan pembagian 4 kategori yaitu sempit, sedang luas, dan sangat luas dengan nilai range pada kategori sempit yaitu 0-600 dengan jumlah data 46 item, nilai range kategori menengah yaitu >600-1300 dengan jumlah data 31 item, nilai range kategori luas yaitu >1300-2000 dengan jumlah data 18 item, dan nilai range kategori sangat luas yaitu >300-<4000 dengan jumlah data 25 item, dengan ditunjukan pada gambar 4.22.



Gambar 4.22 Grafik discretization manual luas lahan

I. HPP

Pada dataset HPP dilakukan pembagian 4 kategori yaitu sehat, ringan, sedang, berat dan sangat berat, dengan nilai range pada kategori sehat yaitu 0-1 dengan jumlah data 97 item, dan nilai range kategori ringan yaitu >1-25 dengan jumlah data 22 item, nilai range kategori sedang yaitu >25-50 dengan jumlah data 1, nilai range kategori berat yaitu >50-75 dengan jumlah data 0, dan nilai range kategori sangat berat yaitu >75-100 dengan jumlah data 0 dengan ditunjukan pada gambar 4.23.



Gambar 4.23 Grafik discretization manual HPP

4.4. Perbandingan Hasil Transformasi Data

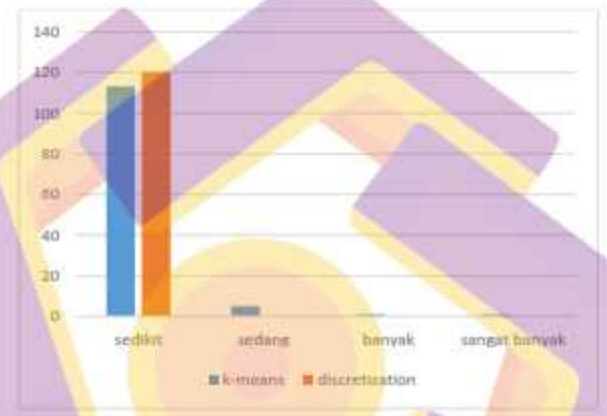
a. Laba-laba

Pada dataset laba-laba dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sedikit, sedang, banyak dan sangat banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.24.

Gambar 4.24 Perbandingan laba-laba

b. Capung

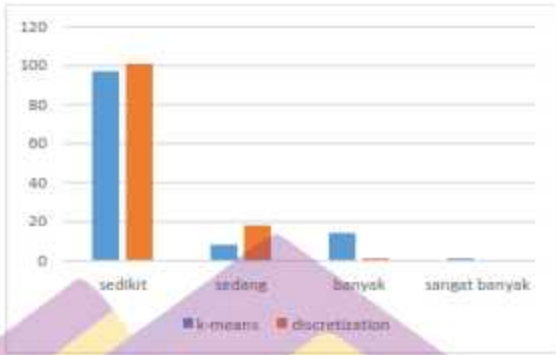
Pada dataset capung dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sedikit, sedang, banyak dan sangat banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.25.



Gambar 4.25 Perbandingan capung

c. Ophanea. SP

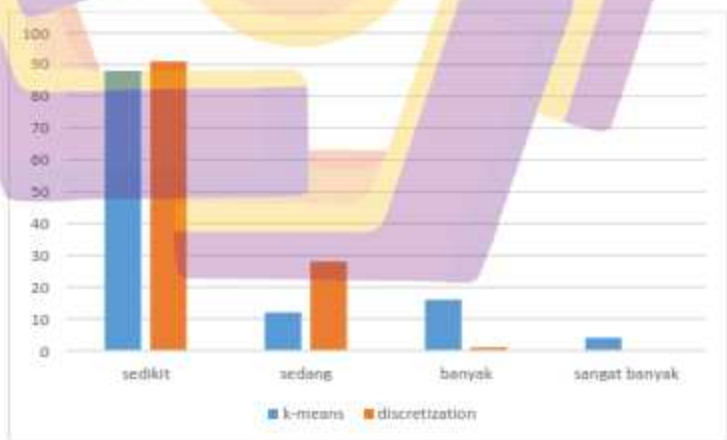
Pada dataset Ophanea. SP dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sedikit, sedang, banyak dan sangat banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.26.



Gambar 4.26 Perbaningan Ophanea. SP

d. Paedarus. SP

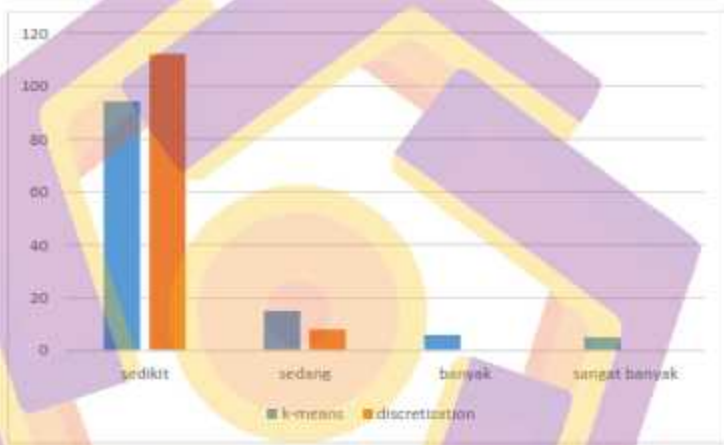
Pada dataset paedarus SP dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sedikit, sedang, banyak dan sangat banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.27.



Gambar 4.27 Perbandingan Paedarus. SP

e. Coccinelidae

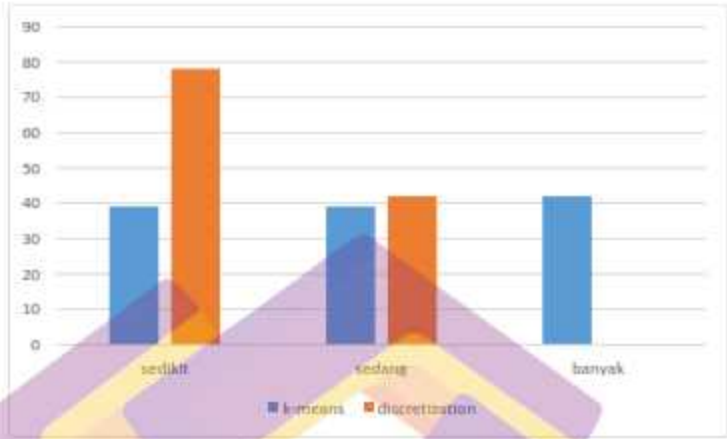
Pada dataset coccinelidae dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sedikit, sedang, banyak dan sangat banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.28.



Gambar 4.28 Perbandingan Coccinelidae

f. Curah Hujan (hari)

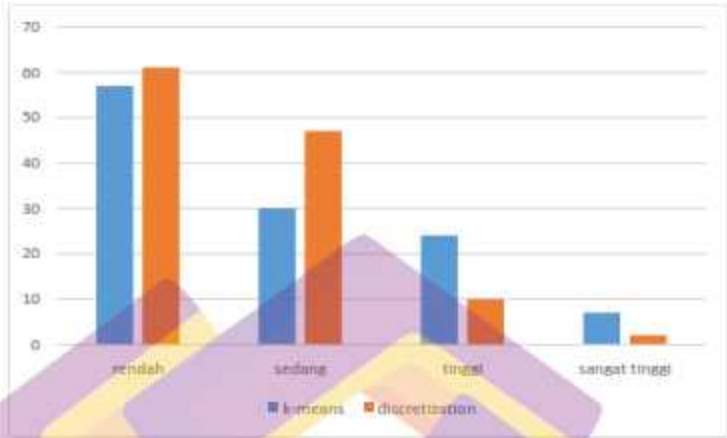
Pada dataset curah hujan (hari) dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 3 yaitu sedikit, sedang, dan banyak. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.29.



Gambar 4.29 Perbandingan curah hujan (hari)

g. Curah Hujan (mm)

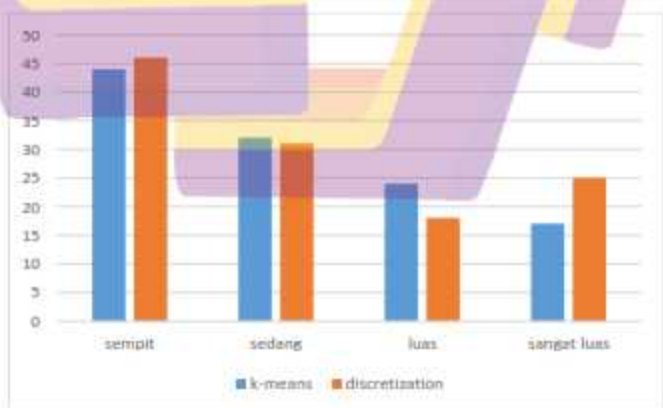
Pada dataset curah hujan (mm) dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu rendah, sedang, tinggi dan sangat tinggi. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.30.



Gambar 4.30 Perbandingan curah hujan (mm)

h. Luas Lahan

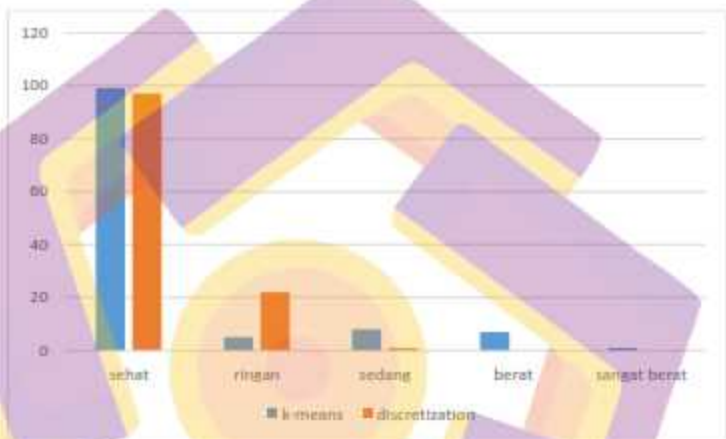
Pada dataset luas lahan dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 4 yaitu sempit, sedang, luas dan sangat luas. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.31.



Gambar 4.31 Perbandingan luas lahan

l. HPP

Pada dataset HPP (Hama Putih Palsu) dilakukan transformasi menggunakan k-means dan discretization dengan pembagian kategori 5 yaitu sehat, ringan, sedang, berat dan sangat berat. Berikut hasil perbandingan jumlah data pada setiap kategori pada gambar 4.25.



Gambar 4.32 Perbandingan HPP

4.5. Hasil Pengujian

Berikut adalah hasil pengujian transformasi data menggunakan metode discretization manual dan k-means clustering pada algoritma naïve bayes.

a. Hasil pengujian discretization dengan sekenario pertama

Pada tahap ini didapatkan hasil pengujian algoritma naïve bayes dengan metode transformasi data menggunakan discretization, dengan jumlah data training 80% data, dan jumlah data testing 20% data, pengujian performance dilakukan dengan pengujian didalam rapidminer studio 9.9 menggunakaan confusion matrix yang ditunjukan pada table 4.1.

Table 4.1 confusion matrix discretization sekenario pertama

	true sehat	true ringan	true sedang
pred. sehat	14	6	1
pred. ringan	3	0	0
pred. sedang	0	0	0

Akurasi = TP/jumlah data

TP=14+0+0 =14

Jumlah data = 24

Akurasi = 14/24 =0,5833

Tabel 4.2 tabel precision discretization sekenario pertama

	Sehat	Ringan	Sedang
TP	14	0	0
FP	7	3	0
Presision	1	0	0

Precision=TP/(TP+FP)

Semua precision = presision sehat+ringan+sedang/jumlah kelas

= (1+0+0)/3= 0,3333

Tabel 4.3 tabel recall discretization sekenario kedua

	Sehat	Ringan	Sedang
TP	14	0	0
FN	3	6	1
recall	0	0	0

Recall=TP/(TP+FN)

Semua recall = recall sehat+ringan+sedang / jumlah kelas

= 2+0+0/3=0,823

Berdasarkan perhitungan mengguankan confusion matrix didapatkan nilai accuracy sebesar 58,33%., precision 0,3333, dan recall 0,823

b. Hasil pengujian discretization dengan sekenario kedua

Pada tahap ini didapatkan hasil pengujian algoritma naïve bayes dengan metode transformasi data menggunakan discretization, dengan jumlah data training 70% data, dan jumlah data testing 30% data, pengujian performance dilakukan dengan pengujian didalam rapidminer studio 9.9 menggunakaan confusion matrix yang ditunjukan pada table 4.4.

Table 4.4 confusion matrix discretization sekenario kedua

	true sehat	true ringan	true sedang
pred. sehat	27	6	1
pred. ringan	2	0	0
pred. sedang	0	0	0

Akurasi = TP/jumlah data

TP=27+0+0 =27

Jumlah data = 36

Akurasi = 27/36 =0,75

Tabel 4.5 tabel precision discretization sekeanrio kedua

	Sehat	Ringan	Sedang
TP	27	0	0
FP	7	2	0
Precision	0,7941	0	0

Precision = TP/(TP+FP)

Jumlah kelas = 3

Semua precision = precision sehat+ringan+sedang/jumlah kelas

= 0,7941+0+0/3=0.2647

Tabel 4.6 tabel recall discretization sekenario kedua

	Sehat	Ringan	Sedang
TP	27	0	0
FN	2	6	1
Recall	0,931	0	0

Recall = TP/(TP+FN)

Jumlah kelas=3

Semua recall = recall sehat+ringan+sedang/jumlah kelas

= 0,931+0+0 /3 = 0,31

Berdasarkan perhitungan menggunakan confusion matrix didapatkan nilai akurasi sebesar 75.00%, precision 0,267 dan recall 0,31.

c. Hasil pengujian k-means dengan sekenario ketiga

Pada tahap ini didapatkan hasil pengujian algoritma naïve bayes dengan metode transformasi data menggunkan k-means clustering, dengan jumlah data training 80% data, dan jumlah data testing 20% data testing, pengujian performance dilakukan dengan pengujian didalam rapid miner menggunkaan confusion matrix yang ditunjukan pada table 4.7

Table 4.7 confusion matrix k-means clustering sekenario ketiga

	true sehat	true sedang	true berat	true sangat berat	true ringan
pred. sehat	17	2	4	1	0
pred. sedang	0	0	0	0	0

Table 4.7 confusion matrix k-means clustering sekenario ketiga (Lanjutan)

	true sehat	true sedang	true berat	true sangat berat	true ringan
pred. berat	0	0	0	0	0
pred. sangat berat	0	0	0	0	0
pred. ringan	0	0	0	0	0

Akurasi: (TP+TN)/jumlah data = 17/24 =0,7083

Tabel 4.8 tabel precision k-means sekenario ketiga

	Sehat	Ringan	Sedang	berat	Sangat berat
TP	17	0	0	0	0
FP	7	0	0	4	1
Precision	0,708	0	0	0	0

Precision = TP/(TP+FP)

Jumlah kelas=5

Semua precision = precision sehat+ringan+sedang+berat+sangat berat/jumlah keas

= 0,708/5=0,1416

Tabel 4.9 tabel recall k-mean clustering sekanario ketiga

	Sehat	Ringan	Sedang	berat	Sangat berat
TP	17	0	0	0	0
FN	0	0	2	4	1
Recall	17/(17+0)=1	0	0/(0+2)=0	0/(0+4)=0	0/(0+1)=0

Recall = TP/(TP+FN)

Jumlah kelas 5

Semua recall = recal sehat+ringan+sedang+berat+sangat berat/jumlah kelas

= 1/5 = 0,2

Berdasarkan perhitungan menggunakan confusion matrix didapatkan nilai akurasi sebesar 70,83%, precision 0,1416, dan recall 0,21

d. Hasil pengujian k-means dengan sekenario keempat

Pada tahap ini didapatkan hasil pengujian algoritma naïve bayes dengan metode transformasi data menggunkan k-means clustering, dengan jumlah data training 70% data, dan jumlah data testing 30% data testing, pengujian performance dilakukan dengan pengujian didalam rapid miner menggunkaan confusion matrix yang ditunjukan pada table 4.10

Table 4.10 confusion matrix k-means clustering sekenario keempat

	true sehat	true sedang	true berat	true sangat berat	true ringan
pred. sehat	29	2	3	1	0
pred. sedang	0	0	0	0	0
pred. berat	0	0	0	0	0
pred. sangat berat	0	0	0	0	0
pred. ringan	0	0	1	0	0

TP=29

Jumlah data = 36

Akurasi =TP/jumlah data

= 29/36=0,8056

Tabel 4.11 tabel precision k-means sekenario keempat

	Sehat	Ringan	Sedang	berat	Sangat berat
TP	29	0	0	0	0
FP	6	1	0	0	1
Precision	0,828	0	0	0	0

Precision = TP/(TP+FP)

Jumlah kelas=5

Semua precision = precision sehat+ringan+sedang+berat+sangat berat/jumlah kelas

= 0,828/5=0,166

Tabel 4.12 tabel 4.12 recall sekenario keempat

	Sehat	Ringan	Sedang	berat	Sangat berat
TP	29	0	0	0	0
FN	0	0	2	3+1	1
Recall	29/(29+0)=1	0	0/(0+2)=0	0/(0+4)=0	0/(0+1)=0

Recall=TP/(TP+FN)

Jumlah kelas=5

Semua recall = recall sehat+ringan+sedang+berat+sangat berat/jumlah kelas

= 1/5=0,2

Berdasarkan perhitungan menggunakan confusion matrix didapatkan nilai akurasi sebesar 80,56%, precision 0,166 dan recall 0,2.

e. Pengujian fitur menggunakan symmetrical uncertainty

Pengujian dilakukan menggunakan rapid miner dengan menggunakan metode *uncertainty* untuk mengetahui relevansi fitur dengan kelas yang diprediksi. Hasil pengujian dengan relevansi tertinggi merupakan fitur yang paling relevan. Berikut merupakan tabel 4.13 yang menampilkan relevansi fitur.

Tabel 4.13 relevansi fitur

Fitur	Relevansi
capung	0,15
laba-laba	0,12
curah hujan(hari)	0,10
coccinelidae	0,10
ophanea. SP	0,10
curah hujan (mm)	0,09
luas lahan	0,07
paedarus	0,07

Pada tabel 4.14 didapatkan relevansi tertinggi yaitu pada fitur capung dengan nilai relevansi sebesar 0,15 dan relevansi terendah didapatkan pada fitur luas lahan dan paedarus dengan nilai 0,09.

Berdasarkan nilai relevansi dilakukan pengujian menggunakan metode naïve bayes dengan menghilangkan masing-masing fitur untuk mengetahui pengaruhnya

terhadap akurasi. Pengujian dilakukan dengan sekenario keempat dan menggunakan transformasi k-means clustering. Berikut merupakan hasil pengujian menggunakan naïve bayes pada tabel 4.14.

Tabel 4.14 relevansi fitur

Tanpa Fitur	Akurasi
capung	80,56%
laba-laba	77,78%
curah hujan(hari)	77,78%
coccinelidae	77,78%
ophanea. SP	80,56%
curah hujan (mm)	80,56%
luas lahan	80,56%
paedarus	80,56%

Pada tabel 4.14 didapatkan nilai akurasi dengan menghilangkan masing-masing fitur untuk mengetahui pengaruhnya terhadap akurasi dan didapatkan hasil laba-laba, curah hujan (hari), dan cocineledae memiliki pengaruh saat dihilangkan,dan terdapat perbedaan pada fitur capung yang memiliki nilai relevansi tertinggi tidak mempengaruhi hasil akurasi saat dihilangkan fitur capung, dan untuk fitur yang lain tidak terdapat pengaruhnya saat setiap fitur dihilangkan.

Berdasarkan pengujian akurasi yang telah dilakukan menggunakan naïve bayes dengan metode transformasi data menggunakan Teknik discretization dan k-means clustering dengan sekenario pembagian data 80% data training dan 20% data

testing, dan sekenario pengujian menggunakan data 70% data training dan 30% data testing, berikut hasil pengujian akurasi pada tabel 4.15.

Tabel 4.15 Hasil akurasi

Transformasi	Sekenario pembagian data					
	80% dan 20%			70% dan 30%		
	akurasi	precision	recall	akurasi	precisio n	recall
k-means	58,33	0,3333	0,823	75	0,267	0,31
Teknis Discretization	70,83	0,1416	0,21	80,56	0,166	0,2

Pada tabel 4.15 didapatkan pembagian data training dan testing mempengaruhi hasil akurasi dan didapatkan transfromasi menggunakan k-means memiliki akurasi yang lebih baik dari pada menggunakan teknik discretization dengan nilai akurasi 80,56% pada sekenario pembagian data 70% training dan 30% data testing.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan peneltian yang telah dilakukan berikut ini kesimpulanya:

1. Hasil transformasi data menggunakan discretization manual degan sekeneario kedua pembagian data training 70% dan data testing 30% didapatkan akurasi terbaik sebesar 75%
2. Hasil transformasi data menggunakan k-means clustering dengan sekenario keempat dengan pembagian data training 70% dan data testing 30% didapatkan akurasi terbaik 80.56%
3. Berdasarkan pengujian transformasi antara discretization manual dan k-means clustering didapatkan akurasi terbaik yaitu transformasi data menggunakan k-means clustering dengan akurasi 80,56%.

5.2. Saran

Untuk penelitan selanjutnya diharapkan dapat melakukan komparasi metode transformasi data dengan metode clustering yang sejenis dan metode discretization sejenis

DAFTAR PUSTAKA

Ahmad, E. D., Kusrini, & Sudarmawan. (2017). Algoritma K-Means untuk Diskretisasi Numerik Kontinyu Pada Klasifikasi Intrusion Detection System Menggunakan Naive Bayes. *Konferensi Nasional Sistem & Informatika*, 61–66.

Artanto, H., Istiadi, Marisa, F., & Purnomo, D. (2019). Implementasi Dan Komparasi Algoritma Fuzzy C-Means Dan K-Means Untuk Mengelompokkan Siswa Berdasarkan Nilai Akademik Dan Perilaku Siswa (Data Survey). *Conference on Innovation and Application of Science and Technology (CIASTECH 2019), Ciastech 2019*, 287–292. http://publishing-widyagama.ac.id/ejournal-v2/index.php/ciastech/article/view/1118

Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika Ahmad Dahlan*, *8*(1), 102632. https://doi.org/10.26555/jifo.v8i1.a2086

Darmansah, N. W. W. (2020). Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 126–134. https://doi.org/10.35957/jatisi.v7i2.309

Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *5*(1), 17–24. https://doi.org/10.25077/teknosi.v5i1.2019.17-24

Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, *7*(1), 19–24. https://doi.org/10.15294/jte.v7i1.8585

Husaini, F. (2016). Algoritma Klasifikasi Naive Bayes Untuk Menilai Kelayakan Kredit (Studi Kasus : Bank Mandiri Kredit Mikro). *Program Studi Teknik Informatika*, *1*(3), 2–12.

Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data Transformation pada Data Mining. *Prosiding Konferensi Nasional Inovasi Dalam Desain Dan Teknologi-IDeaTech*, *7*, 93–99. https://ideatech.stts.edu/proceeding2011/12-000113_INF Hartarto p93-99.pdf

Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *5*(1), 119. https://doi.org/10.24014/rmsi.v5i1.7381

Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data

Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, *4*(2), 151. https://doi.org/10.24076/citec.2017v4i2.106

Piao, M., Piao, Y., & Lee, J. Y. (2019). Symmetrical uncertainty-based feature subset generation and ensemble learning for electricity customer classification. *Symmetry*, *11*(4), 1–11. https://doi.org/10.3390/sym11040498

Praningki, T., & Budi, I. (2018). Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Creative Information Technology Journal*, *4*(2), 83. https://doi.org/10.24076/citec.2017v4i2.100

Rachmatin, D., & Sawitri, K. (2016). *Perbandingan Antara Metode Agglomeratif, Metode Divisif dan Metode K-Means Dalam Analisis Klaster*. *1*, 9–17.

Riezka, A., Atastina, I., & Maulana, K. (2011). Analisis Dan Implementasi Data Cleaning Dengan Menggunakan Metode Multi Pass Neighborhood (Mpn). *Telkom University*.

Rifai, M. F., Purwanto, ; Yudhi S., Jatnika, ; Hendra, & Karmila, ; Sely. (2020). *Pengaruh Kondisi Cuaca Terhadap Serangan Hama Penggerek Batang*. *13*(2), 201–211.

Supriyati, S., Tjahjono, B., & Effendy, S. (2018). Analisis Pola Hujan Untuk Mitigasi Aliran Lahar Hujan Gunungapi Sinabung. *Jurnal Ilmu Tanah Dan Lingkungan*, *20*(2), 95–100. https://doi.org/10.29244/jitl.20.2.95-100

Supriyatna, A., & Mustika, W. P. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, *2*(2), 152. https://doi.org/10.30645/j-sakti.v2i2.78

Surya Nagari, S., & Inayati, L. (2019). IMPLEMENTATION OF CLUSTERING USING K-MEANS METHOD TO DETERMINE NUTRITIONAL STATUS. *Jurnal Biometrika Dan Kependudukan*, *9*(1), 62–68. https://doi.org/10.20473/jbk.v9i1.2020.62-68

Zaki, M. J., & Meira, M. J. (2013). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. https://books.google.com.tr/books?id=Gh9GAwAAQBAJ&lpg=PR9&dq=Data Mining and Analysis: Foundations and Algorithms&hl=tr&pg=PR9#v=onepage&q=Data Mining and Analysis: Foundations and Algorithms&f=false

Zeni, S. A., Rachmawati, N., & Fitrian, A. (2021). Frekuensi Dan Intensitas Serangan Hama Penyakit Pada Bp2Lhk Banjarbaru Kalimantan Selatan. *Jurnal Sylva Scienteae*, *04*(2), 339–345.

LAMPIRAN