

SKRIPSI

**MODEL ANALISIS DATA UNTUK PENYAKIT PADA DINAS
KESEHATAN KOTA BOGOR**

**Oleh :
Waldemar Naposo
0651 16 330**



**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PAKUAN**

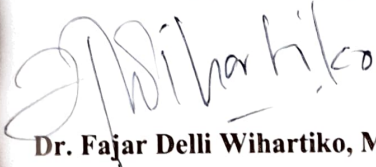
2023

HALAMAN PENGESAHAN

Judul : MODEL ANALISIS DATA UNTUK PENYAKIT PADA DINAS
KESEHATAN KOTA BOGOR
Nama : WALDEMAR NAPOSO
Npm : 0651 16 330

Mengesahkan,

Pembimbing Pendamping
Program Studi Ilmu Komputer
FMIPA-UNPAK



Dr. Fajar Delli Wihartiko, M.M M.Kom.

Pembimbing Utama
Program Studi Ilmu Komputer
FMIPA-UNPAK



Arie Qur'ania, M.Kom.

Mengetahui,

Ketua Jurusan Ilmu Komputer
FMIPA – UNPAK



Arie Qur'ania, M.Kom.

Dekan
FMIPA – UNPAK



Asep Denih, S.Kom., M.Sc., Ph.D.

PERNYATAAN PELIMPAHAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Waldemar Naposo
NPM : 065116330
Judul Skripsi : Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan
Kota Bogor

Dengan ini saya menyatakan bahwa Paten dan Hak Cipta dari produk Skripsi dan Tugas Akhir di atas adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apapun kepada perguruan tinggi manapun.

Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun yang tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan Paten, hak cipta dari karya tulis saya kepada Universitas Pakuan

Bogor, April 2023

Waldemar Naposo
065116330

PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI

Dengan ini saya menyatakan bahwa:

Sejauh yang saya ketahui, karya tulis ini bukan merupakan karya tulis yang pernah di publikasikan atau sudah pernah dipakai untuk mendapatkan gelar sarjana di Universitas lain, kecuali pada bagian-bagian di mana sumber informasinya dicantumkan dengan cara referensi yang semestinya.

Demikian pernyataan ini saya buat dengan sebenar-benarnya. Apabila kelak dikemudian hari terdapat gugatan, penulis bersedia dikenakan sanksi sesuai dengan peraturan yang berlaku.

Bogor, April 2023

Waldemar Naposo
065116330

RIWAYAT HIDUP



Penulisan dilahirkan di Bogor pada tanggal 02 Juni 1997 dari pasangan Bapak Saidi Halim dan Ibu Lenni Rosevi sebagai anak pertama dari empat bersaudara. Penulis memulai pendidikan dasar di SDN Babakan Madang 01, kemudian tahun 2009 masuk MTs Al – Hidayah Citaringgul di kabupaten Bogor. Pada tahun 2015 menjadi alumni dari SMK 2 Triple “J” dan meneruskan pendidikan mengambil Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam di Universitas Pakuan. Selama menjadi mahasiswa penulis aktif dalam organisasi Unit Kegiatan Mahasiswa UKM Sepak Bola & Futsal Universitas Pakuan. Pada tahun 2023, penulis menyelesaikan penelitiannya dengan judul Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor.

RINGKASAN

Waldemar Naposo, 2023. Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor. Dibawah bimbingan Arie Qur'ania, M.Kom dan Fajar Delli W., S. Si., M. Kom.

Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor pengambilan data berasal dari Dinas Kesehatan Kota Bogor dengan rentang waktu dari tahun 2019-2021 yang akan diolah menggunakan metode *K-Medoid* dan *Naïve Bayes*

Dalam Penelitian ini menggunakan kategori sebanyak 4 kategori diantaranya puskesmas, jumlah kasus, tahun, jumlah penduduk dengan Jumlah kasus penyebaran penyakit tuberkulosis (TB), diare. pneumonia, dan demam berdarah dengue (DBD). Pada penerapan metode *K-Medoid* dilakukan sebanyak 3 kali percobaan dengan pemodelan $K=2$ $K=3$ dan $K=4$ didapatkan nilai cluster optimal terdapat pada $K=2$ dengan nilai DBI 0,923 dengan dominan yang bergerombol terdapat pada 2 didominasi oleh tingkat infeksi penyakit sedang sebesar 50% dengan terdiri dari 12 puskesmas. Sedangkan pada metode *Naïve Bayes* menggunakan 22 data training selanjutnya untuk data uji menggunakan sebanyak 3 data. Maka hasil prediksi dapat dihasilkan bahwa puskesmas dengan tingkat infeksi yang tinggi di tahun 2019 berdasarkan kriteria penyakit dan jumlah penduduk adalah puskesmas kayu manis dengan persentase tingkat infeksi sebesar 30,078%, di tahun 2020 puskesmas kayu manis dengan persentase tingkat infeksi sebesar 28,254%, di tahun 2021 puskesmas kayu manis dengan persentase tingkat infeksi sebesar 30,65%, dan pada 2022 prediksi tingkat infeksi tertinggi adalah puskesmas kayu manis dengan persentase tingkat infeksi sebesar 33,046%.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada kehadiran Allah SWT, karena rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi ini yang berjudul “Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor“. Penulisan tugas akhir ini merupakan salah satu syarat memperoleh gelar Sarjana Komputer di Program Studi Ilmu Komputer Matematika Dan Ilmu Pengetahuan Alam Universitas Pakuan Bogor.

Dalam penulisan tugas akhir ini, penulis dengan senang hati ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Arie Qur'ania, M.Kom, Ketua Program Studi Ilmu Komputer sekaligus dosen pembimbing utama yang telah memberikan dorongan moril dan motivasi penulis
2. Fajar Delli W., S. Si., M. Kom, selaku pembimbing pendamping yang telah memberikan masukan dan saran yang sangat membantu dalam pembuatan projek ini.
3. Orang tua dan keluarga yang tiada henti memberikan semangat serta doa'a.
4. Sahabat seperjuangan yang memberikan semangat dan saling menguatkan.

Menyadari keterbatasan waktu dan kemampuan dalam penulisan tugas akhir ini masih jauh dari sempurna. Oleh karena itu segala kritik dan saran yang membangun akan di terima dengan senang hati, mudah-mudahan Allah SWT akan membalas semua kebaikan kepada semua pihak yang membantu. Akhir kata semoga laporan ini dapat bermanfaat bagi kita semua.

Bogor, April 2023

Waldemar Naposo

DAFTAR ISI

HALAMAN PENGESAHAN	i
PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI.....	ii
KATA PENGANTAR	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
1.1. Latar Belakang	1
1.2. Tujuan Penelitian	2
1.3. Ruang Lingkup.....	2
1.4. Manfaat	2
BAB II TINJAUAN PUSTAKA	3
2.1 Tinjauan Pustaka	3
2.1.1 Clustering	3
2.1.2 Prediksi	3
2.1.3 Klasifikasi	3
2.1.4 Penyakit	3
2.1.5 Metode K-Medoids	4
2.1.6 Davies Bouldin Index (DBI)	4
2.1.7 Algoritma Naïve Bayes	5
2.2 Penelitian Terdahulu	6
2.3 Tabel Perbandingan Penelitian.....	8
BAB III METODE PENELITIAN	9
3.1 Metode Penelitian	9
3.1.1 Data Cleaning	9
3.1.2 Data Selection dan Transformation	9
3.1.3 Data Mining	9
3.1.4 Evaluasi Pola (Pattern Evaluation)	10
3.1.5 Presentasi Pengetahuan (Knowledge Presentation)	10
3.2 Waktu dan Tempat Penelitian.....	10
3.3 Alat dan Bahan.....	10
3.3.1 Alat Penelitian	10
3.3.2 Bahan Penelitian	10
BAB IV PERANCANGAN DAN IMPLEMENTASI	11

4.1	Analisis Data	11
4.2	Data Cleaning.....	11
4.3	Seleksi Data dan Transformasi Data	12
4.4	Data Mining	13
4.4.1	Metode K-Medoid	13
4.4.2	Metode Naïve Bayes.....	15
4.5	Implementasi.....	23
4.5.1	Implementasi K-Medoids.....	24
4.5.2	Implementasi Naïve Bayes Klasifikasi	25
4.5.3	Implementasi Naïve Bayes Prediksi	28
4.6	Implementasi Sistem	31
BAB V HASIL DAN PEMBAHASAN		32
5.1	Data Cleaning.....	32
5.2	Data Transformation	32
5.3	Data Mining	33
5.3.1	Hasil <i>Clustering</i> dengan Metode <i>K-Medoids</i>	34
5.3.2	Proses Keseluruhan Klasterisasi <i>K-Medoids</i>	35
5.3.3	Proses Pembagian Data Latih dan Data Uji.....	37
5.3.4	Pembuatan Model <i>Naïve Bayes</i>.....	38
5.3.5	Hasil Presentase Infeksi per tahun.....	38
5.4	Pembahasan.....	39
5.4.1	Uji Akurasi K-Medoids dengan menggunakan Davies Bouldin Index (DBI).....	40
5.4.2	Uji Akurasi <i>Naïve Bayes</i> dengan menggunakan <i>Confusion Matrix</i>	41
BAB VI KESIMPULAN DAN SARAN		42
6.1	Kesimpulan	42
6.2	Saran	42
DAFTAR PUSTAKA		43

DAFTAR TABEL

Tabel 1. Perbandingan Penelitian	8
Tabel 2. Alat Penelitian	10
Tabel 3. Data Sampel Puskesmas Tahun 2019	11
Tabel 4. Data Sampel Puskesmas Tahun 2020	11
Tabel 5. Data Sampel Puskesmas Tahun 2021	12
Tabel 6. Sampel Data Puskesmas Tahun 2019	12
Tabel 7. Sampel Data Puskesmas Tahun 2020	12
Tabel 8. Sampel Data Puskesmas Tahun 2021	13
Tabel 9. Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 1	14
Tabel 10. Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 2	14
Tabel 11. Mean Seluruh Kriteria	16
Tabel 12. Standar Deviasi Seluruh Kriteria	17
Tabel 13. Data Testing	17
Tabel 14. Hasil Distribusi Gaussian	18
Tabel 15. Hasil Probabilitas Akhir	19
Tabel 16. Data Training Tahun 2019.....	19
Tabel 17. Data Training Tahun 2020.....	19
Tabel 18. Data Training Tahun 2021.....	19
Tabel 19. Nilai Distribusi Probabilitas Tahun 2019.....	20
Tabel 20. Nilai Distribusi Probabilitas Tahun 2020.....	20
Tabel 21. Nilai Distribusi Probabilitas Tahun 2021.....	20
Tabel 22. Nilai $P(X C_i)$ Tahun 2019.....	21
Tabel 23. Nilai $P(X C_i)$ Tahun 2020.....	21
Tabel 24. Nilai $P(X C_i)$ Tahun 2021.....	21
Tabel 25. Hasil Persentase $P(X C_i)$ Tahun 2019.....	22
Tabel 26. Hasil Persentase $P(X C_i)$ Tahun 2020.....	22
Tabel 27. Hasil Persentase $P(X C_i)$ Tahun 2021.....	22
Tabel 28. Hasil Persentase Setiap Kriteria.....	23
Tabel 29. Hasil Prediksi	23
Tabel 30. Nilai DBI K-Medoids	41

DAFTAR GAMBAR

Gambar 1. Siklus Hidup Proses Data Mining (Jiawei Han et al., 2022)	9
Gambar 2. Flowchart Algoritma K-Medoid.....	13
Gambar 3. Flowchart Klasifikasi Algoritma Naïve Bayes	15
Gambar 4. Flowchart Prediksi Algoritma Naïve Bayes	16
Gambar 5. Implementasi Menggunakan Python.....	23
Gambar 6. Proses Import Library	24
Gambar 7. Proses Membaca data .csv	24
Gambar 8. Proses Visualisasi data.....	24
Gambar 9. Proses Cek NaN.....	25
Gambar 10. Proses K-Medoid Clustering.....	25
Gambar 11. Proses Davies Bouldin Index	25
Gambar 12. Proses import library	25
Gambar 13. Proses membaca data .csv	26
Gambar 14. Proses pembagian data training dan data uji	26
Gambar 15. Proses menentukan data yang akan digunakan.....	26
Gambar 16. Proses memfilter model	27
Gambar 17. Hasil Confusion matrix	27
Gambar 18. Membaca data testing.....	27
Gambar 19. Hasil klasifikasi prediksi infeksi	28
Gambar 20. Proses import library	28
Gambar 21. Proses membaca data .csv	28
Gambar 22. Proses cek probabilitas dari masing masing kriteria.....	29
Gambar 23. Hasil distribusi probabilitas.....	29
Gambar 24. Hasil distribusi probabilitas tiap kriteria per tahun.....	29
Gambar 25. Proses menentukan hasil presentase tiap kriteria per tahun	30
Gambar 26. Hasil presentase infeksi per tahun.....	30
Gambar 27. Hasil presentase prediksi infeksi tiap puskesmas tahun 2022.....	31
Gambar 28. Total Keseluruhan Data	32
Gambar 29. Data Transformasi.....	32
Gambar 30. Jumlah atribut kelas pada masing masing dataset	33
Gambar 31. Hubungan Clustering Jumlah Penduduk dan Pneumonia.....	34
Gambar 32. Hubungan Clustering Jumlah Penduduk dan DBD	34
Gambar 33. Hubungan Clustering Jumlah Penduduk dan Tuberculosis	34
Gambar 34. Hubungan Clustering Jumlah Penduduk dan Diare	35
Gambar 35. Scatter Plot Hasil Clustering TB dan DBD	36
Gambar 36. Scatter Plot Hasil Clustering Pneumonia dan Diare	36
Gambar 37. Pembagian Data Latih dan Data Uji.....	37
Gambar 39. Pembuatan Model Naïve Bayes.....	38
Gambar 40. Hasil Presentase Infeksi pertahun	38
Gambar 41. Hasil presentasi prediksi infeksi tiap puskesmas tahun 2022	39
Gambar 42. Hasil Keseluruhan Klasifikasi Prediksi	40
Gambar 43. Grafik Nilai DBI Menggunakan Algoritma K-Medoids.....	40
Gambar 44. Hasil Confusion Matrix.....	41

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian.....	45
Lampiran 2. Perhitungan Manual.....	47

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dinas kesehatan kota Bogor merupakan unsur pemerintah yang bertanggung jawab dalam bidang kesehatan meliputi kesehatan masyarakat, pencegahan dan pengendalian penyakit, pelayanan, dan sumber daya kesehatan. Penyakit merupakan suatu kondisi dimana tubuh ataupun pikiran mengalami ketidaknyamanan atau disfungsi terhadap orang yang dipengaruhinya (Juninda et al. 2019). Kondisi kesehatan di Indonesia yang mengkhawatirkan disebabkan karena pola hidup sehat masyarakat yang masih buruk, karena rendahnya kesadaran menjaga lingkungan dan mengatur pola makan. Sedangkan untuk kondisi kesehatan di Kota Bogor sendiri tidak jauh berbeda namun pada Kota Bogor *Universal Health Coverage* (UHC) telah mencapai 95% dari penduduk yang telah terdaftar sebagai peserta program Jaminan Kesehatan Nasional (JKN). Menurut *Unicef* beberapa faktor yang mempengaruhi kesehatan adalah faktor keturunan, faktor lingkungan, faktor pelayanan kesehatan, dan faktor perilaku.

Namun berdasarkan riset kesehatan dasar (Riskesdas) tahun 2019 mencatat terdapat beberapa penyakit dengan prevalensi tertinggi di Indonesia yang menyebar di provinsi provinsi di Indonesia. Penyakit yang berprevalensi tertinggi tersebut adalah hipertensi sebesar 25,8%, penyakit sendi sebesar 24,7%, penyakit hepatitis B sebesar 21,8%, stroke sebesar 12,1%, dan kurang gizi sebesar 19,6%. Berdasarkan data Dinas Kesehatan kota Bogor 2016-2019, jumlah kasus penyebaran penyakit tuberkulosis (TB), pneumonia, dan demam berdarah dengue (DBD) mengalami peningkatan. Dengan wilayah yang luas dan kondisi cuaca yang tidak menentu menyebabkan dinas kesehatan kota Bogor sulit untuk memantau penyebaran penyakit dengan jumlah kasus terbanyak, hal ini menyebabkan dari akibat lambatnya pencegahan maupun penanggulangan penyakit tersebut. Maka dari itu, perlu dilakukan prediksi untuk pencegahan dan penanganan lebih awal agar tidak masuk kedalam tingkatan yang lebih berbahaya. Prediksi penyakit dalam ilmu kedokteran pada waktu yang tepat adalah masalah utama bagi para profesional untuk pencegahan dan rencana perawatan yang efektif. Menganalisis suatu permasalahan diatas maka dibutuhkan metode yang menghasilkan akurasi yang akurat dengan menggunakan metode *K-Medoids* dan *Naïve Bayes* yang merupakan metode yang berbasis *clustering* yang relative dan untuk melakukan prediksi, baik digunakan dalam kasus klasifikasi maupun regresi dan sangat populer belakangan ini *K-Medoids* dan *Naïve Bayes*.

Beberapa penelitian mengenai Implementasi *Algoritma K-Medoids* untuk Perbandingan *Algoritma K-Means* Dan *K-Medoids* Dalam Pengelompokkan Tingkat Kebahagiaan Provinsi Di Indonesia yaitu pada penelitian yang dilakukan oleh penelitian (Citra Fathia Palembang 2022). Ada pula penelitian yang pernah dilakukan oleh Penelitian (Emir Ramon 2022). Tujuan dalam penelitian nya yaitu Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan *Algoritma Support Vector Machine (SVM)*. Ada pula penelitian yang pernah dilakukan oleh penelitian (Andik 2022). Tujuan dalam penelitian nya yaitu Data Mining Klasifikasi Kelulusan Mahasiswa Menggunakan Metode *Naïve Bayes*. Ada pula penelitian yang pernah dilakukan oleh penelitian (Ike Wahyuni Septiani 2022). Tujuan dalam penelitian nya yaitu Implementasi *Algoritma K-Medoids* Dengan Evaluasi *Davies-*

Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. Ada pula penelitian yang pernah dilakukan oleh Penelitian (Muhammad Bhakti Fajri 2022). Tujuan dalam penelitiannya yaitu Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Medoid Clustering Tugas akhir ini akan dilakukan Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor menggunakan metode *K-Medoids* dan metode *Naïve Bayes*.

1.2. Tujuan Penelitian

Tujuan penelitian ini adalah Model Analisis Data Untuk Penyakit Pada Dinas Kesehatan Kota Bogor menggunakan metode *K-Medoids* dan metode *Naïve Bayes*.

1.3. Ruang Lingkup

Pembatasan ruang lingkup pada penelitian ini adalah analisis data menggunakan bahasa pemrograman Python dengan menggunakan metode *Naïve Bayes* untuk klasifikasi dan prediksi dan *K-Medoid* untuk clustering, dengan pengambilan data berasal dari Dinas Kesehatan Kota Bogor dengan rentang waktu dari tahun 2019-2021 dengan kriteria yang digunakan yaitu Tuberkulosis, Diare, Demam Berdarah, Pneumonia, dan Jumlah Penduduk.

1.4. Manfaat

Manfaat dari penelitian yang dilakukan adalah :

1. Meng-klasterisasi data untuk mengetahui klasifikasi dan prediksi jumlah kasus tinggi dan rendahnya suatu penyakit di dinas kesehatan kota Bogor
2. Diharapkan dapat membantu meningkatkan tingkat pertimbangan dalam mengambil kebijakan untuk pencegahan dan penanggulangan penyakit pada dinas kesehatan kota Bogor
3. Menjadi bahan referensi untuk penelitian penelitian selanjutnya yang berkaitan dengan proses *clusterin*, klasifikasi dan prediksi.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

2.1.1 Clustering

Clustering yaitu menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari *cluster* adalah meminimalkan jarak didalam *cluster* dan memaksimalkan jarak antar *cluster*. *Clustering* merupakan proses pembentukan suatu kelompok dari sebuah data yang tidak diketahui kelompok-kelompoknya berdasarkan kesamaan (Aprilawati, 2019).

2.1.2 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar selisih antara sesuatu yang terjadi dengan hasil perkiraan dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Rohmawati, 2017).

2.1.3 Klasifikasi

Klasifikasi dapat digambarkan sebagai berikut. Data input, disebut juga training set, terdiri atas banyak contoh (*record*), yang masing-masing memiliki beberapa atribut. Selanjutnya, tiap contoh diberi sebuah label *class* khusus. Tujuannya untuk menganalisa data *input* dan mengembangkan 12 deskripsi atau model akurat untuk tiap *class* menggunakan fitur-fitur pada data. Deskripsi *class* ini digunakan untuk mengklasifikasikan data pengujian lainnya dengan label *class* tidak diketahui. Deskripsi tersebut juga dapat digunakan untuk memahami tiap *class* dalam data (Rizki. 2020).

2.1.4 Penyakit

Penyakit merupakan suatu keadaan abnormal bagi tubuh ataupun pikiran yang mengalami ketidaknyamanan atau disfungsi terhadap orang yang dipengaruhinya. Salah satu faktornya seperti lingkungan dan pola hidup sehat yang menjadi pengaruh besar terhadap penyakit yang diderita oleh penderita. Berikut beberapa jenis penyakit yaitu :

1. Tuberkulosis (TB)

Tuberkulosis (TB) merupakan penyakit yang menular di sebabkan oleh infeksi bakteri *mycobacterium tuberculosis*. Beban penyakit yang disebabkan oleh tuberkulosis dapat diukur dengan *case notification rate* (CNR) *prevalensi*, dan *mortalitas/kematian* (Dinkes, 2017)

2. Diare

Penyakit diare merupakan penyakit endemis di Indonesia dan merupakan penyakit potensi KLB yang sering disertai dengan kematian, angka kematian (CFR) saat KLB diare diharapkan <1% (Dinkes, 2016)

3. Demam Berdarah Dengue

Penyakit demam berdarah dengue (DBD) merupakan penyakit yang disebabkan oleh virus dengue yang digolongkan sebagai *arthropod-bovirus*, genus flavivirus

dan *family flaviviridae*. DBD di tularkan melalui gigitan nyamuk dari *genus aedes*, terutama *aedes aegypti* atau *aedes albopictus* (Dinkes, 2016)

4. Pneumonia

Pneumonia merupakan penyakit infeksi akut yang mengalami jaringan paru paru (*alveoli*) yang dapat disebabkan oleh berbagai *mikroorganism*e seperti virus, jamur dan bakteri. Gejala penyakit pneumonia yaitu menggigil, demam, sakit kepala, batuk. Mengeluarkan dahak dan sesak napas (Dinkes, 2017)

2.1.5 Metode K-Medoids

K-Medoids Clustering merupakan metode pengklasteran partisi untuk mengklasterkan sekumpulan n objek menjadi sejumlah k klaster. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah klaster. Objek yang mewakili sebuah klaster disebut dengan medoids. Klaster dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek non medoids.

Langkah-langkah *K-Medoids* adalah:

1. Pilih poin k sebagai inisial centroid / nilai tengah (medoids) sebanyak k *cluster*.
2. Cari semua poin yang paling dekat dengan medoids, dengan cara menghitung setiap jarak *Vector* antar dokumen dengan menggunakan *Euclidian Distance*. Rumusnya adalah sebagai berikut :

$$dij = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (1)$$

Dimana $d(i,j)$ = jarak antara objek ke i dengan x_{ij} = nilai objek i pada variabel ke-k

x_{jk} = nilai objek j pada variabel ke-k
 p = banyaknya variable yang di amati

3. Secara acak, pilih poin yang bukan medoids.
4. Hitung total jarak antar medoids.
5. Jika TD baru < TD awal, tukar posisi medoids dengan medoids baru, jadilah medoids yang baru.
6. Ulangi langkah 2-5 sampai medoids tidak berubah lagi

2.1.6 Davies Bouldin Index (DBI)

David L. Davies dan Donald W. Bouldin memperkenalkan sebuah metode yang diberi nama dengan nama mereka berdua, yaitu *Davies Bouldin Index* (DBI) yang digunakan untuk mengevaluasi *cluster* (Nawrin, et al. 2017). Evaluasi menggunakan *Davies Bouldin Index* ini memiliki skema evaluasi internal *cluster*, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster* (Bates & Kalita 2016).

a. Sum of Square Within Cluster (SSW)

Untuk mengetahui kohesi dalam sebuah *cluster* ke-i adalah dengan menghitung nilai dari *Sum of Square Within-cluster* (SSW). Kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang Universitas Sumatera Utara 9 diikuti. Persamaan yang digunakan untuk memperoleh nilai *Sum of Square Within cluster* adalah sebagai berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(X_j, C_i) \quad (2)$$

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

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

$$SSB_{i,j} = d(C_i, C_j) \quad (3)$$

c. *Ratio (Rasio)*

Bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j. Untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*, digunakan persamaan berikut (Nawrin, et al. 2017).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

d. *Davies Bouldin Index*

Nilai rasio yang diperoleh dari persamaan (8) digunakan untuk mencari nilai *Davies-Bouldin Index (DBI)* dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^k \text{Max}_{i \neq j} (R_{i,j}) \quad (5)$$

Dari persamaan tersebut, k adalah jumlah *cluster*. Semakin kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (non-negatif ≥ 0), maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan algoritma *clustering* (Bates & Kalita 2016).

2.1.7 Algoritma Naïve Bayes

Naïve Bayes adalah metode klasifikasi dengan *probabilistic* sederhana yang dihitung berdasarkan frekuensi dan komposisi dari nilai yang ada. Ciri utama dari klasifikasi ini adalah asumsi yang kuat akan independensi masing-masing kondisi atau kejadian. Keuntungan menggunakan *naïve bayes* yaitu jumlah data training yang dibutuhkan tidak terlalu banyak untuk menentukan estimasi parameter dalam proses klasifikasi (Gustientiedinaa dan Desnelitab 2019). *Naïve Bayes* didefinisikan sebagai pengklasifikasian depan metode *probabilitas* dan *statistic* dengan memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Armansyah 2021). Langkah-langkah algoritma *Naïve Bayes* adalah :

1. Baca data training

Untuk rumus *teorema bayes* sebagai berikut :

$$\begin{aligned} \frac{P(H|X)}{P(X)} &= \frac{P(X|H)P(H)}{P(X)} \\ &= \frac{P(X|H)}{P(H)/P(X)} \end{aligned}$$

Keterangan:

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X yang merupakan suatu *class* yang lebih spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi X
(posteriori *probability*)

$P(H)$ = Probabilitas hipotesis H (prior *probability*)

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$ = Probabilitas X

2. Hitung jumlah *class*/label
3. Hitung jumlah kasus yang sama dengan *class* yang sama
4. Kalikan semua nilai hasil sesuai dengan data X yang dicari kelasnya

2.2 Penelitian Terdahulu

1. Peneliti : Tri Septiar Syamfithriani, dkk. (2022)
Judul : Perbandingan Algoritma K-Means dan K-Medoids Untuk Pemetaan Daerah Penanganan Diare Pada Balita di Kabupaten Kuningan
Isi : Dengan metode tersebut diperoleh hasil bahwa pada algoritma K-Means jumlah optimum cluster adalah 3 cluster sedangkan algoritma K-Medoids adalah 2 cluster. Evaluasi cluster terbaik menggunakan metode Davies-Bouldin Index (DBI) dan diperoleh hasil bahwa nilai DBI K-Means selalu lebih kecil daripada algoritma K-Medoids baik pada 2 cluster ataupun 3 cluster, hal ini menunjukkan bahwa algoritma K-Means lebih baik daripada algoritma K-Medoids. Berdasarkan hasil tersebut, maka merekomendasikan pemetaan daerah prioritas penanganan penyakit diare menggunakan algoritma K-Means dengan jumlah 3 cluster, yaitu daerah prioritas sedang terdiri dari 9 daerah, daerah prioritas tinggi terdiri dari 3 daerah dan daerah prioritas rendah terdiri dari 25 daerah. Hasil pemetaan tersebut dapat menjadi masukan bagi Dinas Kesehatan Kabupaten Kuningan untuk membuat strategi penanggulangan dan pencegahan penyakit diare pada balita.
2. Peneliti : Emir Ramon, dkk. (2022)
Judul : Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma *Support Vector Machine (SVM)*
Isi : Dalam penelitian ini, penulis melakukan penelitian tentang klasifikasi status gizi bayi di posyandu kecamatan Bangun Purba, Rokan Hulu. Penelitian ini bertujuan untuk menghitung nilai akurasi dari algoritma *SVM* dalam mengklasifikasi status gizi balita pada kecamatan Bangun Purba dan untuk memberikan informasi serta pengetahuan yang dapat dimanfaatkan oleh berbagai pihak, salah satunya posyandu kecamatan Bangun Purba, Rokan Hulu, Riau.
3. Peneliti : Dyan Dyanmita Putri, dkk. (2018).
Judul : Klasifikasi Penyakit Kulit Pada Manusia Menggunakan

Metode Binary Decision Tree Support Vector Machine (BDTSVM) (Studi Kasus: Puskesmas Dinoyo Kota Malang)

Isi` : Kulit merupakan organ tubuh pada manusia yang sangat penting karena terletak pada bagian luar tubuh yang berfungsi untuk menerima rangsangan seperti sentuhan, rasa sakit dan pengaruh lainnya dari luar. Penyakit kulit salah satu penyakit yang sering dijumpai pada negara beriklim tropis seperti Indonesia. Kurangnya pengetahuan tentang jenis penyakit kulit serta tidak mengetahui cara pencegahannya mengakibatkan seseorang dapat terkena penyakit kulit tingkat akut. Sehingga dengan adanya bantuan teknologi komputer diharapkan penyakit yang menyerang kulit tubuh manusia dapat diketahui secara dini dan hal tersebut dapat memperkecil terjadinya penyakit yang lebih berbahaya. Penelitian ini bertujuan untuk menentukan klasifikasi penyakit kulit pada manusia menggunakan metode Binary Decision Tree Support Vector Machine (BDTSVM). Berdasarkan hasil pengujian didapatkan nilai akurasi terbaik sebesar 97,14% dengan pengujian parameter SVM yaitu nilai λ (lambda) = 0,5, C (complexity) = 1, konstanta γ (gamma) = 0,01, dan itermax = 10.

4. Peneliti : Ike Wahyuni Septiani, dkk. (2022)
Judul : Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru
Isi` : Penanganan lebih dini yang perlu dilakukan untuk mengurangi tingkat kematian pada pasien penderita kanker paru-paru setelah melakukan operasi toraks, dengan cara pengumpulan data-data dari setiap pasien mengenai suatu informasi ini menyebabkan suatu masalah yang baru diantaranya adalah data yang diperoleh meliputi data yang berdimensi tinggi dan mempunyai banyak atribut sehingga bisa menghasilkan informasi yang kurang akurat. Maka diperlukannya perhitungan data mining clustering. Pada umumnya metode dalam melakukan klustering dikelompokkan menjadi empat bagian diantaranya adalah partitioning, hierarchical, grid-based and model-based. Penelitian ini menggunakan algoritma k-medoids karena mampu mengatasi data sensitif terhadap outlier dan memiliki akurasi yang tinggi dan efisien dalam memproses objek dalam jumlah besar. Hasil dari perhitungan K-Medoids dievaluasi menggunakan euclidean distance Davies Bouldin Index yang menghasilkan nilai DBI sebesar 0,93543 menunjukkan bahwa algoritma k-medoids mencapai pengelompokkan yang baik karena hasil akhir

dari perhitungannya kurang dari 0. Dari hasil evaluasi menggunakan DBI menunjukkan bahwa algoritma k-medoid mempunyai akumulasi rata-rata rata-rata pada saat eksekusi cukup cepat dan kualitas cluster yang baik.

5. Peneliti : Muhammad Bhakti Fajri, dkk. (2022)
 Judul : Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering
 Isi : Dengan menggunakan teknik data mining serta memanfaatkan algoritma K-Means clustering untuk mengelompokkan usia pasien berdasarkan kelompok usia yang berpotensi dalam penyebaran penyakit pada Puskesmas Muara Enim. Data tersebut akan dilakukan cleaning, selection dan akan ditentukan jarak usia pasien yang akan di bagi menjadi 4 kelompok yaitu Anak-anak, Remaja, Dewasa dan Lansia, yang selanjutnya akan dilakukan clustering data. Dibantu dengan aplikasi rapidminer, data yang sudah diolah akan dihitung secara manual yang menghasilkan sebanyak total 5015 pasien ditahun 2020 dan 5466 pasien ditahun 2021. Berdasarkan data tersebut dapat disimpulkan bahwa kelompok rentang usia tertinggi yang mengidap penyakit adalah Pasien Dewasa dan Anak – Anak kemudian pasien lansia dan terakhir pasien remaja pada setiap tahunnya. Pada tahun 2020 dan 2021 penyakit Asma, Diare, Penyakit tekanan darah tinggi, Tuberkulosis, Imunisasi dan Demam, batuk pilek merupakan penyakit dengan penderita tertinggi.

2.3 Tabel Perbandingan Penelitian

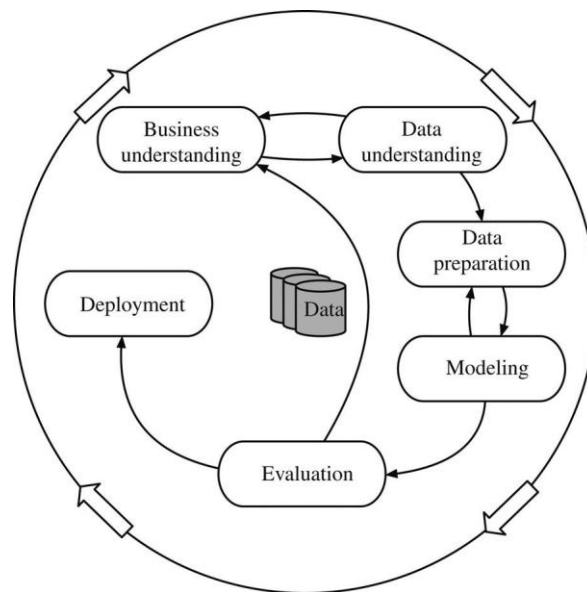
Tabel 1. Perbandingan Penelitian

Penelitian	Software			Metode			
	Anaconda	Rapid Minner	Excel	K-Medoid	Naive bayes	K-Means	SVM
Syamfithriani T.S. (2022)		√	√	√		√	
Ramon E (2022)	√		√				√
Putri D.D. (2022).	√		√				√
Septiani I. W (2022)	√		√	√			
Fajri M. B (2022)		√	√			√	
Naposo W (2022)	√	√	√	√	√		

BAB III METODE PENELITIAN

3.1 Metode Penelitian

Data Mining didefinisikan sebagai proses menemukan pola dalam data. Prosesnya harus otomatis atau (biasanya) semi otomatis. Pola yang ditemukan harus bermakna karena mengarah pada beberapa keuntungan misalnya keuntungan ekonomi. Data selalu hadir dalam jumlah yang substansial. Data sebelum diolah menggunakan algoritma dari Data Mining terlebih dahulu melalui tahapan siklus hidup proses data mining. Proses siklus hidup Data Mining dapat dilihat pada gambar 1.



Gambar 1. Siklus Hidup Proses Data Mining (Jiawei Han et al., 2022)

3.1.1 Data Cleaning

Data Cleaning merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Serta Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru (Tosida et al. 2016). Data yang digunakan pada penelitian ini yaitu berasal dari dinas kesehatan Bogor guna sebagai sumber data.

3.1.2 Data Selection dan Transformation

Seleksi data adalah langkah pertama dimana memilih data yang sesuai untuk dianalisis yang digunakan dari sebuah *database*. Seleksi data juga digunakan untuk mengevaluasi dengan menyesuaikan terhadap kriteria yang ditentukan (Ersoz et al, 2017). Sementara, *Data Transformation* adalah proses dimana data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan (Tosida, 2015).

3.1.3 Data Mining

Proses *Mining* adalah proses yang paling penting dimana akan dilakukan berbagai teknik yang diaplikasikan untuk mengekstrak berbagai pola-pola potensial

untuk mendapatkan data yang berguna (Hoerudin, 2019). Proses mining menggunakan algoritma *K-Medoids* dan metode *Naïve Bayes* sebagai metode *clustering*, klasifikasi dan prediksi untuk data penyakit. Di mana data diambil dari dinas kesehatan kota Bogor.

3.1.4 Evaluasi Pola (Pattern Evaluation)

Evaluasi merupakan mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai (Praja, 2020) Evaluasi dari penelitian ini menggunakan pengolahan data menggunakan *python jupyter anaconda*.

3.1.5 Presentasi Pengetahuan (Knowledge Presentation)

Proses Presentasi Pengetahuan atau yang bisa disebut dengan Visualisasi merupakan penyajian pengetahuan mengenai metode yang telah digunakan memperoleh pengetahuan oleh pengguna. Dalam penelitian ini hasil berupa grafik, perbandingan hasil *cluster* yang lebih optimal dari kedua metode yaitu metode *K-Medoids* dan *Naïve Bayes*.

3.2 Waktu dan Tempat Penelitian

Waktu pelaksanaan penelitian ini dilakukan dari bulan November 2022 sampai Desember 2022 (3 Bulan). Penelitian ini dilaksanakan dinas kesehatan Bogor.

3.3 Alat dan Bahan

3.3.1 Alat Penelitian

Alat yang digunakan dalam penelitian ini terdiri dari perangkat keras (*Hardware*) dan perangkat lunak (*Software*) dapat dilihat dalam Tabel 2.

Tabel 2. Alat Penelitian

No	Perangkat Keras	Perangkat Lunak
1	Intel(R) Core(TM) i5-8300H CPU @2.30GHz	OS Windows 10 Pro
2	RAM 8GB.	Microsoft Office 2019
3	Solid State Drive (SSD) 512 GB	Google Chrome
4	Hardisk 1 TB	Visual Studio Code

3.3.2 Bahan Penelitian

Bahan-bahan yang diperlukan dalam penelitian ini meliputi :

1. Data penyakit didapat dari dinas kesehatan (Dinkes) kota Bogor
2. Jurnal-jurnal terkait dari orisinil studi kasus serta orisini metode yang digunakan.

BAB IV PERANCANGAN DAN IMPLEMENTASI

4.1 Analisis Data

Analisis data merupakan kegiatan setelah data dari sumber data terkumpul (Yusup, 2015). Data yang digunakan yaitu data penyakit dari rentang waktu 2019-2021 dengan jenis penyakit yang akan digunakan dalam penelitian ini yaitu tuberkulosis, diare, demam berdarah dengue (DBD) dan pneumonia yang digunakan dalam analisis data menggunakan 4 atribut data diantaranya diantaranya puskesmas, jumlah kasus, tahun, jumlah penduduk. Data yang sudah diperoleh selanjutnya dianalisis dengan dilakukan tahap *data cleaning*, *transformasi* data dan tahap terakhir dilakukan proses *data mining* dengan menggunakan algoritma *K-Medoids* dan *Naïve Bayes* untuk memperoleh hasil.

4.2 Data Cleaning

Data cleaning atau pembersihan data merupakan proses identifikasi informasi yang tidak lengkap, atau tidak relevan dan kemudian memodifikasi atau menghapus data yang kotor (Saini, 2019). Dalam penelitian ini tidak ada data yang dibersihkan dikarenakan data yang digunakan berasal dari data dari dinas kesehatan Bogor sehingga data sudah siap untuk diproses. Data dapat dilihat pada tabel 3,4, dan 5. Untuk data selengkapnya dapat dilihat pada lampiran 1.

Tabel 3.Data Sampel Puskesmas Tahun 2019

No	Puskesmas	Jenis Penyakit Tahun 2019				Jumlah Penduduk
		Tuberkulosis	DBD	Pneumonia	Diare	
1	Mulyaharja	30	29	20	1330	17921
2	Bogor Timur (Baranang Siang & Katulampa)	42	17	15	1709	36069
3	Bogor Utara	74	6	25	1747	168894
4	Kedung Badak	14	14	13	1036	25949
5	Kayu Manis	17	23	18	1085	12477

Tabel 4. Data Sampel Puskesmas Tahun 2020

No	Puskesmas	Jenis Penyakit Tahun 2020				Jumlah Penduduk
		Tuberkulosis	DBD	Pneumonia	Diare	
1	Mulyaharja	29	31	19	1423	18355
2	Bogor Timur (Baranang Siang & Katulampa)	46	19	15	1793	37127
3	Bogor Utara	87	8	26	1826	170920
4	Kedung Badak	30	16	15	1102	26809
5	Kayu Manis	18	25	20	1126	12894

Tabel 5. Data Sampel Puskesmas Tahun 2021

No	Puskesmas	Jenis Penyakit Tahun 2021				Jumlah Penduduk
		Tuberculosis	DBD	Pneumonia	Diare	
1	Mulyaharja	32	34	23	1487	18863
2	Bogor Timur (Baranang Siang & Katulampa)	50	22	17	1884	39210
3	Bogor Utara	93	7	28	1860	179679
4	Kedung Badak	33	18	17	1166	27438
5	Kayu Manis	20	26	21	1189	13132

4.3 Seleksi Data dan Transformasi Data

Seleksi data dilakukan untuk menyeleksi atau memilih atribut-atribut yang ada dalam *database* yang akan digunakan dalam proses analisis *data mining*. Kemudian, *transformasi* adalah proses untuk merubah data menjadi menjadi format yang ditentukan dan menjadi data yang siap untuk diproses pada tahap *data mining*. Pada data sampel yang ada akan dilakukan seleksi dengan menghilangkan atribut-atribut yang tidak akan digunakan dan hanya menyisakan atribut Puskesmas, Jumlah Penduduk, Tuberculosis (TB), Demam Berdarah Dengue (DBD), Pneumonia, dan Diare. Pada penelitian ini saya menggunakan 15 data dengan 5 atribut. Sampel data dapat dilihat pada Tabel 6, 7, dan 8. Untuk keseluruhan data dapat dilihat pada lampiran 1.

Tabel 6. Sampel Data Puskesmas Tahun 2019

No	Puskesmas	Jenis Penyakit Tahun 2019				Jumlah Penduduk
		Tuberculosis	DBD	Pneumonia	Diare	
1	Mulyaharja	30	29	20	1330	17921
2	Bogor Timur (Baranang Siang & Katulampa)	42	17	15	1709	36069
3	Bogor Utara	74	6	25	1747	168894
4	Kedung Badak	14	14	13	1036	25949
5	Kayu Manis	17	23	18	1085	12477

Tabel 7. Sampel Data Puskesmas Tahun 2020

No	Puskesmas	Jenis Penyakit Tahun 2020				Jumlah Penduduk
		Tuberculosis	DBD	Pneumonia	Diare	
1	Mulyaharja	29	31	19	1423	18355
2	Bogor Timur (Baranang Siang & Katulampa)	46	19	15	1793	37127
3	Bogor Utara	87	8	26	1826	170920
4	Kedung Badak	30	16	15	1102	26809
5	Kayu Manis	18	25	20	1126	12894

Tabel 8. Sampel Data Puskesmas Tahun 2021

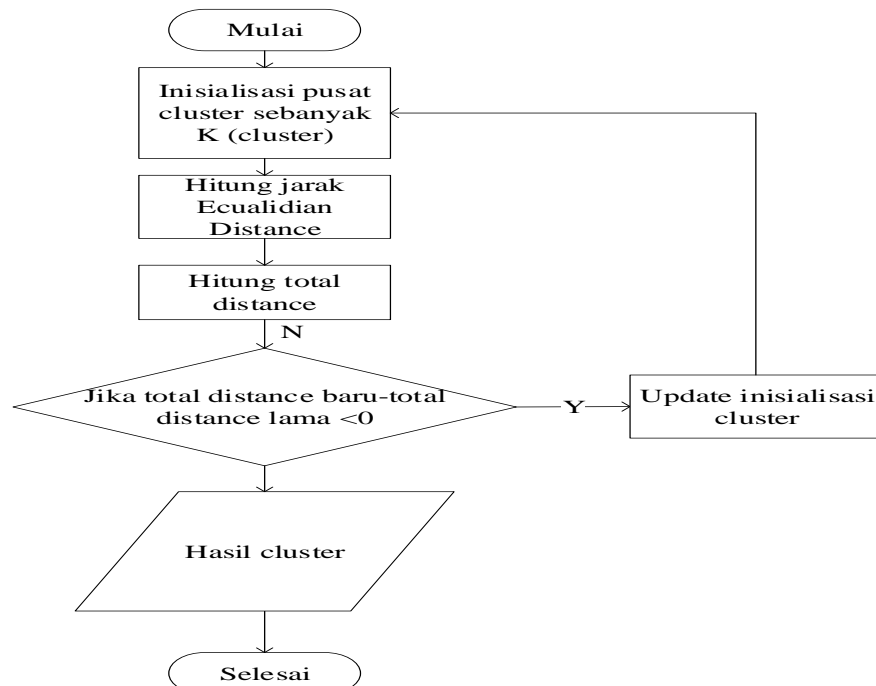
No	Puskesmas	Jenis Penyakit Tahun 2021				Jumlah Penduduk
		Tuberculosis	DBD	Pneumonia	Diare	
1	Mulyaharja	32	34	23	1487	18863
2	Bogor Timur (Baranang Siang & Katulampa)	50	22	17	1884	39210
3	Bogor Utara	93	7	28	1860	179679
4	Kedung Badak	33	18	17	1166	27438
5	Kayu Manis	20	26	21	1189	13132

4.4 Data Mining

Metode yang akan digunakan dalam penelitian ini adalah *K-Medoids* dan *Naïve Bayes*. Untuk melakukan penelitian ini akan menggunakan *Software Anaconda Navigator* dengan pemrograman *python*.

4.4.1 Metode K-Medoid

K-Medoids Clustering merupakan metode pengklasteran partisi untuk mengklasterkan sekumpulan n objek menjadi sejumlah k klaster. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah klaster. Objek yang mewakili sebuah klaster disebut dengan medoids. Klaster dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek non medoids (Setyawati, 2017). Dapat dilihat pada gambar 2.



Gambar 2. Flowchart Algoritma *K-Medoid*

- **Langkah pertama**, kita tentukan jumlah k sebanyak 3.
- **Langkah kedua**, kita pilih data 1 sebagai medoid 1, data 2 sebagai medoid 2 dan data 3 sebagai medoid ke 3.
- **Langkah ketiga** adalah menghitung jarak objek ke masing masing medoid yang telah di Pilih. Kita akan gunakan rumus jarak paling sering dipakai yaitu

Euclidean. Dapat dilihat pada tabel 9.

Tabel 9. Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 1

Data	Jarak objek ke-1 medoid	Jarak objek ke-2 medoid	Jarak objek ke-3 medoid	Terdekat	Klaster yang diikuti
1	0	18.151,966	150.973,584	0	1
2	18.151,966	0	132.825,010	0	2
3	150.973,584	132.825,010	0	0	3
4	8.033,415	10.142,392	142.946,781	8.033,415	1
5	5.449,530	23.600,265	156.418,412	0	1
6	443,860	17.716,319	150.539,357	443,860	1
7	19.211,590	1.061,338	131.767,012	1.061,338	2
8	152.999,816	134.851,059	2.027,582	2.027,582	3
9	8.890,935	9.279,881	142.086,471	8.890,935	1
10	5.031,153	23.182,346	156.001,247	5.031,153	1
11	955,014	17.207,445	150.031,234	955,014	1
12	21.296,218	3.145,886	129.684,076	3.145,886	2
13	161.758,882	143.610,089	10.785,609	10.785,609	3
14	9.518,420	8.648,069	141.457,200	8.648,069	2
15	4.791,087	22.942,907	155.763,010	4.791,087	1
Jumlah	567.505,47	566.364,972	1.753.306,585		
Total	2.887.177,027				

- **Langkah keempat** adalah melakukan iterasi pada medoid yang baru dan akan berhenti jika total simpangannya lebih besar dari 0. Kita pilih responden 13 sebagai medoid 1, responden 14 sebagai medoid 2 dan responden 15 sebagai medoid 3

Tabel 10. Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 2

Data	Jarak objek ke-1 medoid	Jarak objek ke-2 medoid	Jarak objek ke-3 medoid	Terdekat	Klaster yang diikuti
1	161.758,882	9.518,420	4.791,087	4.791,087	3
2	143.610,089	8.648,069	22.942,907	8.648,069	2
3	10.785,609	141.457,200	155.763,010	10.785,609	1
4	153.732,230	1.494,796	12.817,923	1.494,796	2
5	167.203,814	14.961,229	663,225	663,225	3
6	161.324,607	9.086,645	5.228,250	5.228,250	3
7	142.552,024	9.709,275	24.002,617	9.709,275	2
8	8.759,068	143.483,529	157.789,301	8.759,068	1
9	152.871,893	632,261	13.677,285	632,261	2
10	166.786,633	14.544,064	246,209	246,209	3
11	160.816,446	8.581,023	5.738,761	5.738,761	3
12	140.469,010	11.793,889	26.087,277	11.793,889	2
13	0	152.242,594	166.548,369	0	1
14	152.242,594	0	14.306,027	0	2
15	166.548,369	14.306,027	0	0	3
Jumlah	1.889.461,223	540.459,021	610.602,248		
Total	3.040.522,492				

- **Hasil Cluster**

Tahap selanjutnya menghitung total simpangan yang berfungsi untuk mengetahui anggota *cluster* yang terbentuk pada masing masing medoid

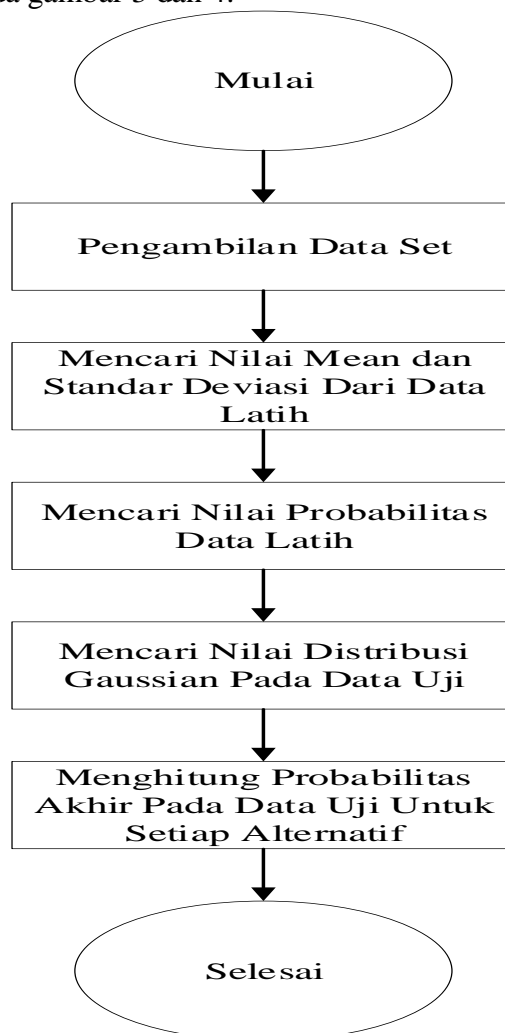
$$S = B - A$$

$$S = 3.040.522,492 - 2.887.177,027$$

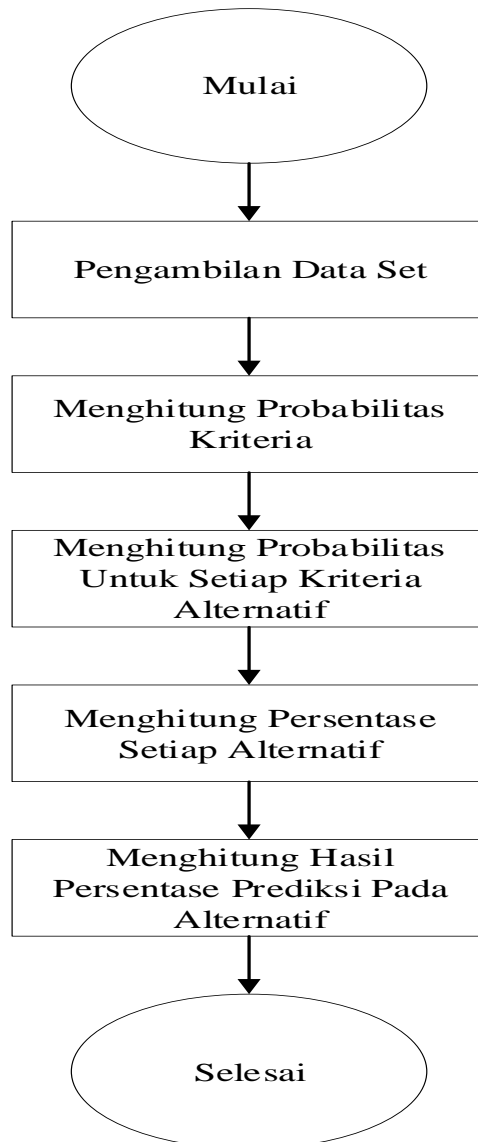
$$= 153.345,465$$

4.4.2 Metode Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas serta statistik sederhana yang berdasar pada teorema Bayes dengan asumsi independensi yang kuat (Patima, 2018). merupakan memprediksi kesempatan di masa depan bersumber pada pengalaman di masa sebelumnya (Lishania, 2019). Dapat dilihat pada gambar 3 dan 4.



Gambar 3. *Flowchart* Klasifikasi Algoritma *Naïve Bayes*



Gambar 4. *Flowchart* Prediksi Algoritma *Naïve Bayes*

1. Klasifikasi

- **Langkah 1** : mencari nilai *Mean* dan Standar Deviasi dari data training, berikut adalah nilai mean untuk seluruh kriteria, dapat dilihat pada tabel 11.

Tabel 11. Mean Seluruh Kriteria

Kelas	Jumlah	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Tinggi	2	24,333	28	20,167	1273,333	15607
Rendah	1	85	7	26,333	1811	173164,333
Sedang	2	35,833	17,667	16,333	1462,833	32100,333

Berikut adalah nilai standar deviasi dari seluruh kriteria, dapat dilihat pada tabel 12.

Tabel 12. Standar Deviasi Seluruh Kriteria

Prediksi	Jumlah	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Tinggi	2	6	3,333	0,5	140	2802,5
Rendah	1	0	0	0	0	0
Sedang	2	10,166	1,666	1	332,5	5473,4

- Langkah 2 : Mencari nilai probabilitas

Setelah menghitung nilai mean dan standar deviasi untuk data numerik, maka selanjutnya menentukan nilai

probabilitas setiap fitur pada setiap kelasnya $P(\text{Tinggi}) = 2/5 = 0,4$

$P(\text{Rendah}) = 1/5 = 0,2$ $P(\text{Sedang}) = 2/5 = 0,4$

- Langkah 3 : Nilai Distribusi *Gaussian*

Menghitung nilai *probabilitas* untuk data testing puskesmas dari tahun 2019-2021.

Berikut adalah data Testing yang digunakan, dapat dilihat pada tabel 13.

Tabel 13. Data Testing

Pusk esmas	Kriteria Tahun 2019-2021					Kelas Infeksi
	Jenis Penyakit					
	Tu berculosis	DBD	Pneumo nia	Diare	Jumlah Penduduk	
Lawang Gintung	71.333	11.333	14	1375	7377.667	?
Pulo Armyan	30	23.333	27.333	458	40301.667	?
Tegal Gundil	43	23	19.667	1549.667	41033	?

Untuk mengetahui nilai probabilitas data testing, maka harus mencari nilai distribusi *gaussian* terlebih dahulu. Hasil keseluruhan dari nilai distribusi *gaussian* untuk seluruh puskesmas dapat dilihat pada Tabel 14.

Tabel 14. Hasil Distribusi *Gaussian*

Puskesmas	Tuberculosis			DBD			PN			Diare			Jumlah Penduduk		
	Tinggi	Re nd a h	Se da ng	Tin ggi	R e n d a h	Se da ng	Tin ggi	R e n d a h	Se da ng	Tin ggi	R e n d a h	Se da ng	Tin ggi	R e n d a h	Se da ng
Lawang	0,000000000315	0	0,000834	0,000448	0	0,00017463	0,00000000000737	0	0,00026248891	0,00002189668	0	0,000158	0,00019103	0	2.707E-09
Pulo Armyan	0,0042575031	0	0,0033295574	0,0044919339	0	0,000739622	1.99E-45	0	1.51E-29	1.23E-10	0	0,000124757	0,000278833	0	1.412E-10
Tegal Gundil	0,000526044	0	0,0030616247	0,00388602	0	0,001430249	0,484062851	0	0,001539244	0,00406349	0	0,00159894	0,0000000019	0	0,0000192

- Langkah 4 : Probabilitas Akhir Setiap Kelas.

Hasil Probabilitas akhir setiap kelas untuk seluruh puskesmas dapat dilihat pada tabel 15.

Tabel 15. Hasil Probabilitas Akhir

No	Puskesmas	Tinggi	Rendah	Sedang	Kelas Infeksi
1	Lawang	4,31939E-63	0	1.27015E-21	Sedang
2	Pulo Armin	1,30487E-65	0	6.56718E-49	Sedang
3	Tegal Gundil	7.63987E-31	0	1.50103E-15	Sedang

1. Prediksi

- Langkah 1 : Baca Data Training

Data set diambil dari data Dinas Kesehatan pada tahun 2019 – 2021 pada tabel 12.

Tabel 16. Data Training Tahun 2019

Puskesmas	Kriteria Tahun 2019				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	30	29	20	1330	17921
Bogor Timur (Baranang Siang & Katulampa)	42	17	15	1709	36069
Bogor Utara	74	6	25	1747	168894
Kedung Badak	14	14	13	1036	25949
Kayu Manis	17	23	18	1085	12477

Tabel 17. Data Training Tahun 2020

Puskesmas	Kriteria Tahun 2020				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	29	31	19	1423	18355
Bogor Timur (Baranang Siang & Katulampa)	46	19	15	1793	37127
Bogor Utara	87	8	26	1826	170920
Kedung Badak	30	16	15	1102	26809
Kayu Manis	18	25	20	1126	12894

Tabel 18. Data Training Tahun 2021

Puskesmas	Kriteria Tahun 2021				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	32	34	23	1487	18863
Bogor Timur (Baranang	50	22	17	1884	39210

Siang & Katulampa)					
Bogor Utara	93	7	28	1860	179679
KedungBadak	33	18	17	1253	27438
Kayu Manis	20	26	21	1189	13132

Untuk memudahkan dalam pembacaan data, nilai distribusi probabilitas yang telah didapatkan dari pencarian di atas selanjutnya akan ditampilkan ke dalam bentuk tabel seperti yang terlihat pada tabel 19, 20, dan 21.

Tabel 19. Nilai Distribusi Probabilitas Tahun 2019

Puskesmas	Kriteria Tahun 2019				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	0,001552	0,001500	0,001034	0,068805	0,927108
Bogor Timur (Baranang Siang & Katulampa)	0,001110	0,000449	0,000396	0,045149	0,952895
Bogor Utara	0,000433	0,000035	0,000146	0,010231	0,989153
Kedung Badak	0,000518	0,000518	0,000481	0,038333	0,960149
Kayu Manis	0,001248	0,001688	0,001321	0,079662	0,916079

Tabel 20. Nilai Distribusi Probabilitas Tahun 2020

Puskesmas	Kriteria Tahun 2020				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	0,001460	0,001516	0,000957	0,071662	0,924359
Bogor Timur (Baranang Siang & Katulampa)	0,001179	0,000487	0,000385	0,045974	0,951974
Bogor Utara	0,000503	0,000046	0,000150	0,010563	0,988737
Kedung Badak	0,001072	0,000572	0,000536	0,039396	0,958422
Kayu Manis	0,001278	0,001775	0,001420	0,079954	0,915572

Tabel 21. Nilai Distribusi Probabilitas Tahun 2021

Puskesmas	Kriteria Tahun 2021				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	0,001566	0,001663	0,001125	0,072753	0,922892
Bogor Timur (Baranang Siang & Katulampa)	0,001214	0,000534	0,000413	0,045747	0,952092
Bogor Utara	0,000512	0,000038	0,000154	0,010238	0,989057
Kedung Badak	0,001147	0,000626	0,000591	0,043569	0,954066
Kayu Manis	0,001390	0,001807	0,001459	0,082638	0,912705

- Langkah 3 : Menghitung $P(X|Ci)$ untuk setiap kriteria dan kelas

Untuk memudahkan dalam pembacaan data, nilai $P(X|Ci)$ Probabilitas yang telah didapatkan dari pencarian di atas selanjutnya akan ditampilkan ke dalam bentuk tabel seperti yang terlihat pada tabel 22, 23, dan 24.

Tabel 22. Nilai $P(X|Ci)$ Tahun 2019

Puskesmas	Kriteria Tahun 2019				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	0,319278	0,357995	0,252326	0,284107	0,195370
Bogor Timur (Baranang Siang & Katulampa)	0,228348	0,107159	0,096632	0,186427	0,200805
Bogor Utara	0,089075	0,008353	0,035625	0,042245	0,208445
Kedung Badak	0,106562	0,123628	0,117374	0,158283	0,202233
Kayu Manis	0,256738	0,402858	0,322357	0,328937	0,193046

Tabel 23. Nilai $P(X|Ci)$ Tahun 2020

Puskesmas	Kriteria Tahun 2020				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	0,265841	0,351499	0,277552	0,289486	0,195051
Bogor Timur (Baranang Siang & Katulampa)	0,214676	0,109659	0,111659	0,185717	0,200878
Bogor Utara	0,091588	0,010358	0,043503	0,042670	0,208635
Kedung Badak	0,195193	0,128799	0,155452	0,159144	0,202239
Kayu Manis	0,232702	0,399684	0,264151	0,322982	0,193197

Tabel 24. Nilai $P(X|Ci)$ Tahun 2021

Puskesmas	Kriteria Tahun 2021				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	0,268657	0,356255	0,300641	0,285367	0,195081
Bogor Timur (Baranang Siang & Katulampa)	0,208267	0,114396	0,110369	0,179439	0,201253
Bogor Utara	0,087837	0,008140	0,041154	0,040158	0,209067
Kedung Badak	0,196775	0,134104	0,157937	0,170896	0,201671
Kayu Manis	0,238463	0,387104	0,389898	0,324140	0,192928

- Langkah 4 : Menghitung Persentase $P(X|Ci)$ setiap kriteria.

Untuk mempermudah dalam membaca hasil persentase, maka dibuat di dalam bentuk tabel 25, 26, dan 27.

Tabel 25. Hasil Persentase P(X|Ci) Tahun 2019

Puskesmas	Kriteria Tahun 2019				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	31,93	35,80	25,23	28,41	19,54
Bogor Timur (Baranang Siang & Katulampa)	22,83	10,72	9,66	18,64	20,08
Bogor Utara	8,91	0,83	3,56	4,22	20,84
Kedung Badak	10,66	12,36	11,74	15,82	20,22
Kayu Manis	25,67	40,29	32,24	32,89	19,30

Tabel 26. Hasil Persentase P(X|Ci) Tahun 2020

Puskesmas	Kriteria Tahun 2020				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	26,58	35,15	27,75	28,95	19,50
Bogor Timur (Baranang Siang & Katulampa)	21,47	10,97	11,17	18,57	20,09
Bogor Utara	9,15	1,04	4,35	4,27	20,86
Kedung Badak	19,52	12,88	15,54	15,91	20,22
Kayu Manis	23,27	39,97	26,41	32,30	19,32

Tabel 27. Hasil Persentase P(X|Ci) Tahun 2021

Puskesmas	Kriteria Tahun 2021				
	Jenis Penyakit				
	Tuberkulosis	DBD	Pneu monia	Diare	Jumlah Penduduk
Mulyaharja	26,87	35,62	30,06	28,54	19,51
Bogor Timur (Baranang Siang & Katulampa)	20,83	11,44	11,04	17,94	20,12
Bogor Utara	8,78	0,81	4,11	4,02	20,91
Kedung Badak	19,68	13,40	15,79	17,09	20,17
Kayu Manis	23,85	38,71	38,99	32,41	19,29

Untuk mempermudah membaca hasil persentase pertahun nya maka dibuat dalam bentuk tabel 28.

Tabel 28. Hasil Persentase Setiap Kriteria

Puskesmas	Persentase Tahun 2019	Persentase Tahun 2020	Persentase Tahun 2021	Tingkat Infeksi
Mulyaharja	28,182 %	27,586	28,12 %	Tinggi
Bogor Timur	16,386 %	16,454 %	16,274 %	Sedang
Bogor Utara	7,672 %	7,934 %	7,726 %	Rendah
Kedung Badak	14,16 %	16,814 %	17,226 %	Sedang
Kayu Manis	30,078 %	28,254 %	30,65 %	Tinggi

Dari hasil perhitungan di atas agar mudah dipahami hasil prediksi perhitungan datanya maka dibuat dalam bentuk Tabel 29.

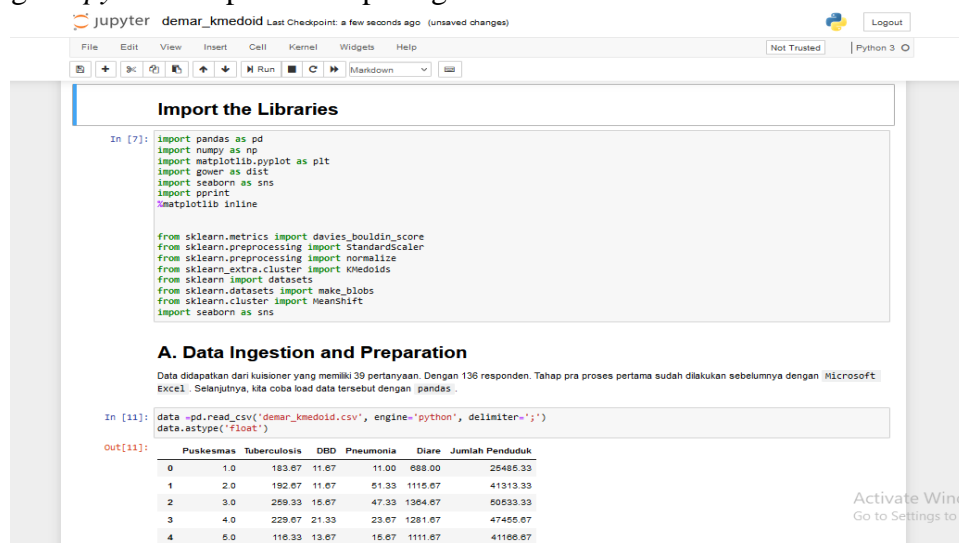
Tabel 29. Hasil Prediksi

Puskesmas	Persentase Akhir Tahun 2022
Mulyaharja	28,654%
Bogor Timur	16,094%
Bogor Utara	7,518%
Kedung Badak	17,638%
Kayu Manis	33,046%

Dari hasil prediksi pada Tabel 11 di atas dapat dihasilkan bahwa puskesmas dengan tingkat infeksi yang tinggi di tahun 2019 berdasarkan kriteria penyakit dan jumlah penduduk adalah puskesmas Kayu Manis dengan persentase tingkat infeksi sebesar 30,078%, di tahun 2020 puskesmas Kayu Manis dengan persentase tingkat infeksi sebesar 28,254%, di tahun 2021 puskesmas Kayu Manis dengan persentase tingkat infeksi sebesar 30,65%, dan pada 2022 prediksi tingkat infeksi tertinggi adalah puskesmas Kayu Manis dengan persentase tingkat infeksi sebesar 33,046%.

4.5 Implementasi

Implementasi merupakan kegiatan memperoleh dan mengintegrasikan sumber daya fisik dan konseptual yang menghasilkan suatu sistem yang bekerja. Pada tahap implementasi ini merupakan penggabungan dari semua elemen dan aktivitas sistem. Proses untuk memodelkan *clustering* pada responden diimplementasikan pada pemrogram *python*. Dapat dilihat pada gambar 5.



Gambar 5. Implementasi Menggunakan *Python*

4.5.1 Implementasi K-Medoids

Setelah data disimpan, selanjutnya dilakukan clustering terhadap data tersebut. Dengan memasukkan fungsi yang akan digunakan untuk menganalisis dengan metode *K-medoids* yang nantinya akan ditampilkan ke dalam grafik. Langkah-langkah untuk implementasi pemrosesan data menggunakan *K-Medoids* dijelaskan sebagai berikut :

1. Persiapkan terlebih dahulu *software Anaconda*. Untuk memulai nya pertama, mengimport *library* untuk membaca file csv, mengolah data *array* dan untuk preprocessing. Dapat dilihat pada Gambar 6.

```
In [60]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import gower as dist
import seaborn as sns
import pprint
%matplotlib inline

from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import normalize
from sklearn_extra.cluster import KMedoids
from sklearn.datasets import make_blobs
from sklearn.cluster import MeanShift
import seaborn as sns
```

Gambar 6. Proses Import Library

2. Kemudian untuk memulai nya pertama memanggil data yang akan digunakan dengan memakai "pd.read_csv" dan melakukan penindexan berbasis lokasi. Dapat dilihat pada Gambar 7.

```
In [11]: data = pd.read_csv('demar_kmedoid.csv', engine='python', delimiter=',')
data.astype('float')
```

Out[11]:

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
0	1.0	183.67	11.67	11.00	688.00	25485.33
1	2.0	192.67	11.67	51.33	1116.67	41313.33
2	3.0	259.33	15.67	47.33	1364.67	50533.33
3	4.0	229.67	21.33	23.67	1281.67	47456.67
4	5.0	116.33	13.67	15.67	1111.67	41166.67
5	6.0	180.67	46.00	90.00	1719.67	63695.67
6	7.0	213.67	15.67	110.00	1207.00	44708.00
7	8.0	211.67	30.67	201.00	1947.00	72104.67
8	9.0	341.00	22.00	173.33	2046.00	75781.33
9	10.0	87.00	21.33	107.67	1467.00	54335.33
10	11.0	106.67	5.33	6.00	253.33	9372.67

Gambar 7. Proses Membaca data .csv

3. Kemudian membuat ringkasan statistik data. Bagian ini, visualisasi untuk memahami bagaimana setiap fitur terkait dengan yang lain. Dapat dilihat pada gambar 8.

```
In [12]: from sklearn.preprocessing import normalize
data_scaled = normalize(data)
data_scaled = pd.DataFrame(data_scaled, columns=data.columns)
data_scaled.head()
```

Out[12]:

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
0	0.000039	0.007204	0.000458	0.000431	0.026985	0.999810
1	0.000048	0.004662	0.000282	0.001242	0.026995	0.999824
2	0.000059	0.005130	0.000310	0.000936	0.026995	0.999822
3	0.000084	0.004638	0.000449	0.000499	0.026998	0.999824
4	0.000121	0.002825	0.000332	0.000381	0.026994	0.999831

Gambar 8. Proses Visualisasi data

4. Kemudian cek NaN pada data apakah data kosong atau tidak. Dapat dilihat pada gambar 9.

```
In [32]: data.isnull().any()
Out[32]: Puskesmas           False
Tuberculosis                 False
DBD                           False
Pneumonia                     False
Diare                         False
Jumlah Penduduk               False
label                         False
dtype: bool
```

Gambar 9. Proses Cek NaN

5. Setelah itu proses akan dilakukan menggunakan *clustering K-Medoids* dengan menampilkan modeling *K-Medoids* menggunakan *cluster* berjumlah 3 menggunakan *matrix Euclidian Distance*. Dapat dilihat pada Gambar 10.

```
In [12]: import warnings
warnings.filterwarnings("ignore")
model = KMedoids(n_clusters=3, random_state=0).fit(x)
labels = model.labels_
predict = model.predict
cluster_center = np.unique(model.cluster_centers_)
inertia = model.inertia_
```

Gambar 10. Proses K-Medoid Clustering

6. Evaluasi dilakukan dengan menggunakan *Davies Bouldin Index (DBI)*. Dapat dilihat pada Gambar 11.

```
In [14]: print(davies_bouldin_score(x, labels))
2.8236420097757016
```

Gambar 11. Proses *Davies Bouldin Index*

4.5.2 Implementasi Naïve Bayes Klasifikasi

Setelah data berhasil disimpan, selanjutnya dilakukan *clustering* terhadap data tersebut. Dengan memasukkan fungsi yang akan digunakan untuk menganalisis dengan metode *naïve bayes* yang nantinya akan ditampilkan ke dalam grafik.

1. Persiapkan terlebih dahulu *software Anaconda*. Untuk memulainya pertama, *import library* untuk membaca file *.csv*, mengolah data *array* dan untuk preprocessing. Data dapat dilihat pada gambar 12.

```
In [2]: import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer

from jcopml.pipeline import num_pipe, cat_pipe
```

Gambar 12. Proses *import library*

2. Kemudian untuk memulai nya pertama memanggil data yang akan digunakan dengan memakai "pd.read_csv" dan melakukan penindexan berbasis lokasi. Dapat dilihat pada gambar 13.

```
In [3]: #baca data excel
data = pd.read_csv('demar_klasifikasi.csv.xls', engine='python', delimiter=';')
data.astype('float')
```

Out[3]:

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Infeksi
0	1.0	183.67	11.67	11.00	688.00	25485.33	2.0
1	2.0	192.67	11.67	51.33	1115.67	41313.33	0.0
2	3.0	259.33	15.67	47.33	1364.67	50533.33	0.0
3	4.0	229.67	21.33	23.67	1281.67	47455.67	0.0
4	5.0	116.33	13.67	15.67	1111.67	41166.67	0.0
5	6.0	180.67	46.00	90.00	1719.67	63695.67	1.0
6	7.0	213.67	15.67	110.00	1207.00	44708.00	0.0
7	8.0	211.67	30.67	201.00	1947.00	72104.67	1.0
8	9.0	341.00	22.00	173.33	2046.00	75781.33	0.0
9	10.0	87.00	21.33	107.67	1467.00	54335.33	1.0
10	11.0	106.67	5.33	6.00	253.33	9372.67	2.0

Gambar 13. Proses membaca data .csv

3. Kemudian untuk melatih model data dan membaginya mejadi 22 data training dan 3 data uji. Dapat dilihat pada gambar 14.

```
In [11]: #melatih data model dan tes menggunakan 18% data dari 25 data yang diatih
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1, random_state=0)
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

Out[11]: ((22, 6), (3, 6), (22,), (3,))

```
In [12]: x_test
```

Out[12]:

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
5	6	180.67	46.00	90.00	1719.67	63695.67
2	3	259.33	15.67	47.33	1364.67	50533.33
19	20	98.00	26.67	125.00	1517.00	56183.67

Gambar 14. Proses pembagian data training dan data uji

4. Kemudian untuk menentukan data apa yang akan di gunakan. Dapat dilihat pada gambar 15.

```
In [13]: preprocessor = ColumnTransformer([
    ('numeric', num_pipe(), ["Tuberculosis", "DBD", "Pneumonia", "Diare", "Jumlah Penduduk"]),
])
```

Gambar 15. Proses menentukan data yang akan digunakan

5. Kemudian tahap memfilter model. Dapat dilihat pada gambar 16.

```
In [15]: pipeline.fit(x_train, y_train) // untuk memfilter model
Out[15]: Pipeline(memory=None,
               steps=[('prep',
                       ColumnTransformer(n_jobs=None, remainder='drop',
                                          sparse_threshold=0.3,
                                          transformer_weights=None,
                                          transformers=[('numeric',
                                                         Pipeline(memory=None,
                                                                steps=[('imputer',
                                                                 SimpleImputer(add_indicator=False,
                                                                 copy=True,
                                                                 fill_value=None,
                                                                 missing_values=nan,
                                                                 strategy='median',
                                                                 verbose=0))],
                                                                verbose=False),
                                                         ['Tuberculosis', 'DBD',
                                                          'Pneumonia', 'Diare',
                                                          'Jumlah Penduduk'])],
                                                                verbose=False)),
                       ('algo', GaussianNB(priors=None, var_smoothing=1e-09))],
               verbose=False)
```

Gambar 16. Proses memfilter model

6. Kemudian tahap untuk mendapatkan hasil *confusion matrix*. Dapat dilihat pada gambar 17.

```
In [18]: #hasil confusion matrix
from jcopml.plot import plot_confusion_matrix
plot_confusion_matrix(x_train, y_train, x_test, y_test, pipeline)
```

Gambar 17. Hasil Confusion matrix

7. Kemudian tahap menginput data testing dengan memanggil data yang akan digunakan dengan memakai "pd.read_csv" dan melakukan penindexan berbasis lokasi. Dapat dilihat pada gambar 18.

```
In [22]: #baca data testing
X_pred = pd.read_csv('data_testingK.csv.xls', engine='python', delimiter=';')
X_pred
```

```
Out[22]:
```

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
0	Lewi Liang	592	193	40	7824	125562
1	Cianpea	540	203	78	8931	170208

Gambar 18. Membaca data testing

- Kemudian tahap untuk menentukan hasil klasifikasi prediksi dari infeksi dapat dilihat gambar 19.

```
In [24]: #Hasil klasifikasi prediksi
X_pred["Infeksi"] = pipeline.predict(X_pred)
X_pred
```

```
Out[24]:
```

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Infeksi
0	Leuwiliang	592	193	40	7824	125552	1
1	Ciamepa	540	203	76	8931	170206	1

Gambar 19. Hasil klasifikasi prediksi infeksi

4.5.3 Implementasi Naïve Bayes Prediksi

Setelah data berhasil disimpan, selanjutnya dilakukan *clustering* terhadap data tersebut. Dengan memasukkan fungsi yang akan digunakan untuk menganalisis dengan metode *naive bayes* yang nantinya akan ditampilkan ke dalam grafik

- Persiapkan terlebih dahulu *software Anaconda*. Untuk memulai nya pertama mengimport *library* untuk membaca file csv, mengolah data *array* dan untuk *preprocessing*. Dapat dilihat pada gambar 20.

```
In [1]: import pandas as pd
import numpy as np

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer

from jcopml.pipeline import num_pipe, cat_pipe
```

Gambar 20. Proses *import library*

- Kemudian Kemudian untuk memulai nya pertama memanggil data yang akan digunakan dengan memakai "pd.read_csv" dan melakukan pen indexan berbasis lokasi. Dapat dilihat pada gambar 21.

```
In [2]: data =pd.read_csv('naivebayesp_demar.csv.xls', engine='python', delimiter=';')
data
```

```
Out[2]:
```

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Tuberculosis.1	DBD.1	Pneumonia.1	Diare.1	Jumlah Penduduk.1	Tuberculosis.2	DBD.2	Pi
0	Bogor Selatan	88	19	19	684	25333	198.0	2	9.0	688.0	25490	265	14	
1	Mulyaharja	70	22	106	1087	40265	205.0	7	37.0	1116.0	41316	303	6	
2	Cipaku	57	22	129	1340	49619	354.0	8	1.0	1365.0	50541	367	17	
3	Bondongan	85	33	29	1291	47797	217.0	14	25.0	1282.0	47467	387	17	
4	Lawang Gintung	33	25	12	1103	40855	190.0	9	3.0	1112.0	41174	126	7	
5	Bogor Timur	76	94	118	1695	62772	203.0	21	85.0	1720.0	63704	263	23	
6	Pulo Armyn	99	19	277	1201	44487	406.0	15	42.0	1207.0	44717	136	13	
7	Bogor Utara	187	41	295	1914	70877	113.0	14	170.0	1947.0	72111	335	37	
8	Warung Jambu	157	41	288	1999	74039	409.0	3	143.0	2046.0	75786	457	22	
9	Tegal Gundil	100	35	270	1466	54284	95.0	9	43.0	1467.0	54349	66	20	
10	Bogor Tengah	28	10	6	258	9541	133.0	2	10.0	253.0	9373	159	4	

Gambar 21. Proses membaca data .csv

3. Kemudian cek *probabilitas* dari masing masing kriteria dari rentan waktu 2019 sampai 202. Dapat dilihat pada gambar 22.

```
In [3]: #menghitung probabilitas masing - masing kriteria tahun 2019
Pc119Tuber = data.iloc[:,25,1].sum()/25
Pc119DBD = data.iloc[:,25,2].sum()/25
Pc119Pneu = data.iloc[:,25,3].sum()/25
Pc119Dia = data.iloc[:,25,4].sum()/25
Pc119P = data.iloc[:,25,5].sum()/25
ProbP119 = data.iloc[0,1:6].sum()
ProbP219 = data.iloc[1,1:6].sum()
ProbP319 = data.iloc[2,1:6].sum()
ProbP419 = data.iloc[3,1:6].sum()
ProbP519 = data.iloc[4,1:6].sum()
ProbP619 = data.iloc[5,1:6].sum()
ProbP719 = data.iloc[6,1:6].sum()
ProbP819 = data.iloc[7,1:6].sum()
ProbP919 = data.iloc[8,1:6].sum()
ProbP1019 = data.iloc[9,1:6].sum()
ProbP1119 = data.iloc[10,1:6].sum()
ProbP1219 = data.iloc[11,1:6].sum()
ProbP1319 = data.iloc[12,1:6].sum()
ProbP1419 = data.iloc[13,1:6].sum()
ProbP1519 = data.iloc[14,1:6].sum()
ProbP1619 = data.iloc[15,1:6].sum()
ProbP1719 = data.iloc[16,1:6].sum()
ProbP1819 = data.iloc[17,1:6].sum()
ProbP1919 = data.iloc[18,1:6].sum()
ProbP2019 = data.iloc[19,1:6].sum()
ProbP2119 = data.iloc[20,1:6].sum()
ProbP2219 = data.iloc[21,1:6].sum()
ProbP2319 = data.iloc[22,1:6].sum()
ProbP2419 = data.iloc[23,1:6].sum()
ProbP2519 = data.iloc[24,1:6].sum()
```

Gambar 22. Proses cek probabilitas dari masing masing kriteria

4. Kemudian proses cek hasil distribusi probabilitas. Dapat dilihat pada gambar 23.

```
In [21]: #Hasil Distribusi Probabilitas
DistribusiP = {'Puskesmas': ['Bogor Selatan', 'Mulyaharja', 'Cipaku', 'Bondongan', 'Lawang Gintung', 'Bogor Timur', 'Pulo Arany', 'Bogor Utara', 'Tegal Quntill', 'Bogor Tengah'],
'Tuberculosis2019': [PTuberP119, PTuberP219, PTuberP319, PTuberP419, PTuberP519, PTuberP619, PTuberP719, PTuberP819, PTuberP919, PTuberP1019, PTuberP1119, PTuberP1219, PTuberP1319, PTuberP1419, PTuberP1519, PTuberP1619, PTuberP1719, PTuberP1819, PTuberP1919, PTuberP2019, PTuberP2119, PTuberP2219, PTuberP2319, PTuberP2419, PTuberP2519],
'DBD2019': [PDBD119, PDBD219, PDBD319, PDBD419, PDBD519, PDBD619, PDBD719, PDBD819, PDBD919, PDBD1019, PDBD1119, PDBD1219, PDBD1319, PDBD1419, PDBD1519, PDBD1619, PDBD1719, PDBD1819, PDBD1919, PDBD2019, PDBD2119, PDBD2219, PDBD2319, PDBD2419, PDBD2519],
'Diare2019': [PDIA119, PDIA219, PDIA319, PDIA419, PDIA519, PDIA619, PDIA719, PDIA819, PDIA919, PDIA1019, PDIA1119, PDIA1219, PDIA1319, PDIA1419, PDIA1519, PDIA1619, PDIA1719, PDIA1819, PDIA1919, PDIA2019, PDIA2119, PDIA2219, PDIA2319, PDIA2419, PDIA2519],
'Pneumonia2019': [PPNP119, PPNP219, PPNP319, PPNP419, PPNP519, PPNP619, PPNP719, PPNP819, PPNP919, PPNP1019, PPNP1119, PPNP1219, PPNP1319, PPNP1419, PPNP1519, PPNP1619, PPNP1719, PPNP1819, PPNP1919, PPNP2019, PPNP2119, PPNP2219, PPNP2319, PPNP2419, PPNP2519],
'Pneumonia2020': [PTuberP120, PTuberP220, PTuberP320, PTuberP420, PTuberP520, PTuberP620, PTuberP720, PTuberP820, PTuberP920, PTuberP1020, PTuberP1120, PTuberP1220, PTuberP1320, PTuberP1420, PTuberP1520, PTuberP1620, PTuberP1720, PTuberP1820, PTuberP1920, PTuberP2020, PTuberP2120, PTuberP2220, PTuberP2320, PTuberP2420, PTuberP2520],
'DBD2020': [PDBD120, PDBD220, PDBD320, PDBD420, PDBD520, PDBD620, PDBD720, PDBD820, PDBD920, PDBD1020, PDBD1120, PDBD1220, PDBD1320, PDBD1420, PDBD1520, PDBD1620, PDBD1720, PDBD1820, PDBD1920, PDBD2020, PDBD2120, PDBD2220, PDBD2320, PDBD2420, PDBD2520],
'Diare2020': [PDIA120, PDIA220, PDIA320, PDIA420, PDIA520, PDIA620, PDIA720, PDIA820, PDIA920, PDIA1020, PDIA1120, PDIA1220, PDIA1320, PDIA1420, PDIA1520, PDIA1620, PDIA1720, PDIA1820, PDIA1920, PDIA2020, PDIA2120, PDIA2220, PDIA2320, PDIA2420, PDIA2520],
'Pneumonia2020': [PPNP120, PPNP220, PPNP320, PPNP420, PPNP520, PPNP620, PPNP720, PPNP820, PPNP920, PPNP1020, PPNP1120, PPNP1220, PPNP1320, PPNP1420, PPNP1520, PPNP1620, PPNP1720, PPNP1820, PPNP1920, PPNP2020, PPNP2120, PPNP2220, PPNP2320, PPNP2420, PPNP2520],
'Jumlah Penduduk2021': [P3PP121, P3PP221, P3PP321, P3PP421, P3PP521, P3PP621, P3PP721, P3PP821, P3PP921, P3PP1021, P3PP1121, P3PP1221, P3PP1321, P3PP1421, P3PP1521, P3PP1621, P3PP1721, P3PP1821, P3PP1921, P3PP2021, P3PP2121, P3PP2221, P3PP2321, P3PP2421, P3PP2521],
'Jumlah Penduduk2021': [P3PP121, P3PP221, P3PP321, P3PP421, P3PP521, P3PP621, P3PP721, P3PP821, P3PP921, P3PP1021, P3PP1121, P3PP1221, P3PP1321, P3PP1421, P3PP1521, P3PP1621, P3PP1721, P3PP1821, P3PP1921, P3PP2021, P3PP2121, P3PP2221, P3PP2321, P3PP2421, P3PP2521]
}
DProbabilitas = pd.DataFrame(DistribusiP)
DProbabilitas
```

	Puskesmas	Tuberculosis2019	DBD2019	Pneumonia2019	Diare2019	Jumlah Penduduk2019	Tuberculosis2020	DBD2020	Pneumonia2020	Diare2020	Per
0	Bogor Selatan	0.003386	0.000727	0.000727	0.026164	0.969017	0.007504	0.000076	0.000341	0.026073	
1	Mulyaharja	0.001685	0.000528	0.002551	0.026101	0.969073	0.004803	0.000164	0.000867	0.026147	
2	Cipaku	0.001114	0.000430	0.002521	0.026189	0.969746	0.006773	0.000153	0.000019	0.026115	
3	Bondongan	0.001726	0.000670	0.000589	0.026221	0.970793	0.004428	0.000286	0.000510	0.026161	
4	Lawang Gintung	0.000785	0.000595	0.000286	0.026244	0.972090	0.004472	0.000212	0.000071	0.026172	
5	Bogor Timur	0.001174	0.001452	0.001822	0.026176	0.969377	0.003088	0.000319	0.001293	0.026166	
6	Pulo Arany	0.002148	0.000412	0.000611	0.026062	0.965307	0.005762	0.000323	0.000905	0.026020	
7	Bogor Utara	0.002551	0.000559	0.004024	0.026107	0.968759	0.001520	0.000188	0.002286	0.026185	

Gambar 23. Hasil distribusi probabilitas

5. Kemudian menentukan hasil distribusi probabilitas tiap kriteria per tahun. Dapat dilihat pada Gambar 24

```
In [37]: #Hasil Distribusi Probabilitas Tiap Kriteria Per Tahun
DistribusiPxCi = {'Puskesmas': ['Bogor Selatan', 'Mulyaharja', 'Cipaku', 'Bondongan', 'Lawang Gintung', 'Bogor Timur', 'Pulo Arany', 'Bogor Utara', 'Tegal Quntill', 'Bogor Tengah'],
'Tuberculosis2019': [PxCITC12019, PxCITC22019, PxCITC32019, PxCITC42019, PxCITC52019, PxCITC62019, PxCITC72019, PxCITC82019, PxCITC92019, PxCITC102019, PxCITC112019, PxCITC122019, PxCITC132019, PxCITC142019, PxCITC152019, PxCITC162019, PxCITC172019, PxCITC182019, PxCITC192019, PxCITC202019, PxCITC212019, PxCITC222019, PxCITC232019, PxCITC242019, PxCITC252019],
'DBD2019': [PxCIDB12019, PxCIDB22019, PxCIDB32019, PxCIDB42019, PxCIDB52019, PxCIDB62019, PxCIDB72019, PxCIDB82019, PxCIDB92019, PxCIDB102019, PxCIDB112019, PxCIDB122019, PxCIDB132019, PxCIDB142019, PxCIDB152019, PxCIDB162019, PxCIDB172019, PxCIDB182019, PxCIDB192019, PxCIDB202019, PxCIDB212019, PxCIDB222019, PxCIDB232019, PxCIDB242019, PxCIDB252019],
'Diare2019': [PxCIDI12019, PxCIDI22019, PxCIDI32019, PxCIDI42019, PxCIDI52019, PxCIDI62019, PxCIDI72019, PxCIDI82019, PxCIDI92019, PxCIDI102019, PxCIDI112019, PxCIDI122019, PxCIDI132019, PxCIDI142019, PxCIDI152019, PxCIDI162019, PxCIDI172019, PxCIDI182019, PxCIDI192019, PxCIDI202019, PxCIDI212019, PxCIDI222019, PxCIDI232019, PxCIDI242019, PxCIDI252019],
'Pneumonia2019': [PxCIPN12019, PxCIPN22019, PxCIPN32019, PxCIPN42019, PxCIPN52019, PxCIPN62019, PxCIPN72019, PxCIPN82019, PxCIPN92019, PxCIPN102019, PxCIPN112019, PxCIPN122019, PxCIPN132019, PxCIPN142019, PxCIPN152019, PxCIPN162019, PxCIPN172019, PxCIPN182019, PxCIPN192019, PxCIPN202019, PxCIPN212019, PxCIPN222019, PxCIPN232019, PxCIPN242019, PxCIPN252019],
'Pneumonia2020': [PxCITC12020, PxCITC22020, PxCITC32020, PxCITC42020, PxCITC52020, PxCITC62020, PxCITC72020, PxCITC82020, PxCITC92020, PxCITC102020, PxCITC112020, PxCITC122020, PxCITC132020, PxCITC142020, PxCITC152020, PxCITC162020, PxCITC172020, PxCITC182020, PxCITC192020, PxCITC202020, PxCITC212020, PxCITC222020, PxCITC232020, PxCITC242020, PxCITC252020],
'DBD2020': [PxCIDB12020, PxCIDB22020, PxCIDB32020, PxCIDB42020, PxCIDB52020, PxCIDB62020, PxCIDB72020, PxCIDB82020, PxCIDB92020, PxCIDB102020, PxCIDB112020, PxCIDB122020, PxCIDB132020, PxCIDB142020, PxCIDB152020, PxCIDB162020, PxCIDB172020, PxCIDB182020, PxCIDB192020, PxCIDB202020, PxCIDB212020, PxCIDB222020, PxCIDB232020, PxCIDB242020, PxCIDB252020],
'Diare2020': [PxCIDI12020, PxCIDI22020, PxCIDI32020, PxCIDI42020, PxCIDI52020, PxCIDI62020, PxCIDI72020, PxCIDI82020, PxCIDI92020, PxCIDI102020, PxCIDI112020, PxCIDI122020, PxCIDI132020, PxCIDI142020, PxCIDI152020, PxCIDI162020, PxCIDI172020, PxCIDI182020, PxCIDI192020, PxCIDI202020, PxCIDI212020, PxCIDI222020, PxCIDI232020, PxCIDI242020, PxCIDI252020],
'Pneumonia2020': [PxCIPN12020, PxCIPN22020, PxCIPN32020, PxCIPN42020, PxCIPN52020, PxCIPN62020, PxCIPN72020, PxCIPN82020, PxCIPN92020, PxCIPN102020, PxCIPN112020, PxCIPN122020, PxCIPN132020, PxCIPN142020, PxCIPN152020, PxCIPN162020, PxCIPN172020, PxCIPN182020, PxCIPN192020, PxCIPN202020, PxCIPN212020, PxCIPN222020, PxCIPN232020, PxCIPN242020, PxCIPN252020],
'Jumlah Penduduk2021': [PxCIDP12021, PxCIDP22021, PxCIDP32021, PxCIDP42021, PxCIDP52021, PxCIDP62021, PxCIDP72021, PxCIDP82021, PxCIDP92021, PxCIDP102021, PxCIDP112021, PxCIDP122021, PxCIDP132021, PxCIDP142021, PxCIDP152021, PxCIDP162021, PxCIDP172021, PxCIDP182021, PxCIDP192021, PxCIDP202021, PxCIDP212021, PxCIDP222021, PxCIDP232021, PxCIDP242021, PxCIDP252021],
'Jumlah Penduduk2021': [PxCIDP12021, PxCIDP22021, PxCIDP32021, PxCIDP42021, PxCIDP52021, PxCIDP62021, PxCIDP72021, PxCIDP82021, PxCIDP92021, PxCIDP102021, PxCIDP112021, PxCIDP122021, PxCIDP132021, PxCIDP142021, PxCIDP152021, PxCIDP162021, PxCIDP172021, PxCIDP182021, PxCIDP192021, PxCIDP202021, PxCIDP212021, PxCIDP222021, PxCIDP232021, PxCIDP242021, PxCIDP252021]
}
PxCi = pd.DataFrame(DistribusiPxCi)
PxCi
```

	Puskesmas	Tuberculosis2019	DBD2019	Pneumonia2019	Diare2019	Jumlah Penduduk2019	Tuberculosis2020	DBD2020	Pneumonia2020	Diare2020	Jumlah Penduduk2020	Tuber
0	Bogor Selatan	0.066553	0.049529	0.010911	0.040018	0.040018	0.057095	0.015174	0.010952	0.039940	0.039951	
1	Mulyaharja	0.033309	0.036054	0.038298	0.040014	0.040021	0.036546	0.032633	0.027837	0.040054	0.040034	
2	Cipaku	0.022025	0.029301	0.037848	0.040056	0.040048	0.051533	0.030640	0.000614	0.040004	0.039990	
3	Bondongan	0.034134	0.045677	0.008842	0.040106	0.040092	0.033693	0.057192	0.016382	0.040074	0.040059	
4	Lawang Gintung	0.015524	0.040538	0.004296	0.040141	0.040145	0.034026	0.042406	0.002267	0.040091	0.040078	
5	Bogor Timur	0.023205	0.098926	0.027356	0.040036	0.040033	0.023495	0.063957	0.041524	0.040083	0.040081	
6	Pulo Arany	0.042475	0.029098	0.090237	0.039862	0.039868	0.066597	0.064736	0.029074	0.039859	0.039868	
7	Bogor Utara	0.050431	0.038111	0.060406	0.039931	0.039925	0.011564	0.037694	0.073417	0.040111	0.040109	
8	Warung Jambu	0.040564	0.036513	0.056499	0.039955	0.039957	0.039701	0.007662	0.056500	0.039903	0.039905	
9	Tegal Quntill	0.035209	0.042475	0.072181	0.039930	0.039922	0.012917	0.032195	0.024673	0.040155	0.040164	
10	Bogor Tengah	0.056243	0.059236	0.009151	0.040091	0.040031	0.035571	0.040977	0.032864	0.039664	0.039673	

Gambar 24. Hasil distribusi probabilitas tiap kriteria per tahun

6. Kemudian menentukan hasil presentase tiap kriteria per tahun. Dapat dilihat pada gambar 25.

```
In [53]: #Hasil Presentase Tiap Kriteria Per Tahun
PercentasePxCI = {'Puskesmas': ["Bogor Selatan","Mulyaharja","Cipaku","Bondongan","Lawang Gintung","Bogor Timur","Pulo Armyrn","Bog
'Tuberculosis2019': [PTCP12019,PTCP22019,PTCP32019,PTCP42019,PTCP52019,PTCP62019,PTCP72019,PTCP82019,PTCP92019,PTCF
'DBD2019': [POBD12019,POBD22019,POBD32019,POBD42019,POBD52019,POBD62019,POBD72019,POBD82019,POBD92019,POBD102019,PC
'Pneumonia2019': [PPneu12019,PPneu22019,PPneu32019,PPneu42019,PPneu52019,PPneu62019,PPneu72019,PPneu82019,PPneu92019,PPne
'Diare2019': [PDia12019,PDia22019,PDia32019,PDia42019,PDia52019,PDia62019,PDia72019,PDia82019,PDia92019,PDia102019,PDia
'Jumlah Penduduk2019': [PJ2019,PJP2019,PJP32019,PJP42019,PJP52019,PJP62019,PJP72019,PJP82019,PJP92019,PJP102019,PC
'Tuberculosis2020': [PTCP12020,PTCP22020,PTCP32020,PTCP42020,PTCP52020,PTCP62020,PTCP72020,PTCP82020,PTCP92020,PTCF
'DBD2020': [POBD12020,POBD22020,POBD32020,POBD42020,POBD52020,POBD62020,POBD72020,POBD82020,POBD92020,POBD102020,PC
'Pneumonia2020': [PPneu12020,PPneu22020,PPneu32020,PPneu42020,PPneu52020,PPneu62020,PPneu72020,PPneu82020,PPneu92020,PPne
'Diare2020': [PDia12020,PDia22020,PDia32020,PDia42020,PDia52020,PDia62020,PDia72020,PDia82020,PDia92020,PDia102020,PC
'Jumlah Penduduk2020': [PJ2020,PJP2020,PJP32020,PJP42020,PJP52020,PJP62020,PJP72020,PJP82020,PJP92020,PJP102020,PC
'Tuberculosis2021': [PTCP12021,PTCP22021,PTCP32021,PTCP42021,PTCP52021,PTCP62021,PTCP72021,PTCP82021,PTCP92021,PTCF
'DBD2021': [POBD12021,POBD22021,POBD32021,POBD42021,POBD52021,POBD62021,POBD72021,POBD82021,POBD92021,POBD102021,PC
'Pneumonia2021': [PPneu12021,PPneu22021,PPneu32021,PPneu42021,PPneu52021,PPneu62021,PPneu72021,PPneu82021,PPneu92021,PPne
'Diare2021': [PDia12021,PDia22021,PDia32021,PDia42021,PDia52021,PDia62021,PDia72021,PDia82021,PDia92021,PDia102021,PC
'Jumlah Penduduk2021': [PJ2021,PJP2021,PJP32021,PJP42021,PJP52021,PJP62021,PJP72021,PJP82021,PJP92021,PJP102021,PC
PercentasePxCI = pd.DataFrame(PercentasePxCI)
PercentasePxCI
```

Out[53]:

	Puskesmas	Tuberculosis2019	DBD2019	Pneumonia2019	Diare2019	Jumlah Penduduk2019	Tuberculosis2020	DBD2020	Pneumonia2020	Diare2020	Jumlah Penduduk2020	Tubu
0	Bogor Selatan	6.655303	4.952850	1.091051	4.001784	4.001830	5.709521	1.517365	1.095244	3.994030	3.995123	
1	Mulyaharja	3.330946	3.608350	3.829850	4.001396	4.002065	3.654633	3.283322	2.783721	4.005369	4.003447	
2	Cipaku	2.202549	2.930149	3.784832	4.005602	4.004843	5.153278	3.064049	0.061435	4.000381	3.998987	
3	Bondongan	3.413387	4.567694	0.884242	4.010563	4.009167	3.369332	5.719231	1.638167	4.007380	4.005915	
4	Lawang Gintung	1.552444	4.053763	0.428637	4.014115	4.014523	3.402607	4.240589	0.226732	4.009142	4.007810	
5	Bogor Timur	2.320496	9.882626	2.735619	4.003588	4.003319	2.349833	6.395667	4.152350	4.008279	4.008055	
6	Pulo Armyrn	4.247516	2.809764	9.023726	3.986165	3.986758	6.659694	6.473396	2.907445	3.985877	3.986821	
7	Bogor Utara	5.043073	3.611132	6.040627	3.993080	3.992509	1.156359	3.769362	7.341711	4.011148	4.010899	
8	Warung Jambu	4.056417	3.651264	5.649912	3.995472	3.995677	3.970121	0.766174	5.858016	3.998292	3.998484	
9	Tegal Gundil	3.520888	4.247532	7.218089	3.992993	3.992187	1.291657	3.219523	2.467323	4.015522	4.016437	
10	Bogor Tengah	5.624335	6.923561	0.915104	4.009085	4.003077	10.357075	4.097708	3.286393	3.966381	3.967254	

Gambar 25. Proses menentukan hasil presentase tiap kriteria per tahun

7. Kemudian proses menentukan hasil presentase infeksi per tahun. Dapat dilihat pada gambar 26.

```
In [57]: #Hasil Presentase Infeksi Per Tahun
PrediksiPuskesmas = {'Puskesmas': ["Bogor Selatan","Mulyaharja","Cipaku","Bondongan","Lawang Gintung","Bogor Timur","Pulo Armyrn","E
'Persentase Prediksi 2019': [PersentaseP12019,PersentaseP22019,PersentaseP32019,PersentaseP42019,PersentaseP52019,Pe
'Persentase Prediksi 2020': [PersentaseP12020,PersentaseP22020,PersentaseP32020,PersentaseP42020,PersentaseP52020,Pe
'Persentase Prediksi 2021': [PersentaseP12021,PersentaseP22021,PersentaseP32021,PersentaseP42021,PersentaseP52021,Pe
}
Prediksi = pd.DataFrame(PrediksiPuskesmas)
Prediksi
```

Out[57]:

	Puskesmas	Persentase Prediksi 2019	Persentase Prediksi 2020	Persentase Prediksi 2021
0	Bogor Selatan	4.140584	3.282257	4.214038
1	Mulyaharja	3.754521	3.546098	3.121254
2	Cipaku	3.385595	3.265828	3.425314
3	Bondongan	3.377011	3.748005	3.750954
4	Lawang Gintung	2.812898	3.177378	3.078844
5	Bogor Timur	4.591130	4.182837	3.803825
6	Pulo Armyrn	4.810788	4.802685	2.788393
7	Bogor Utara	4.578084	4.057898	4.938389
8	Warung Jambu	4.289749	3.718217	4.037374
9	Tegal Gundil	4.594338	3.002092	2.810090
10	Bogor Tengah	4.295032	5.134982	4.999098

Gambar 26. Hasil presentase infeksi per tahun

8. Kemudian proses menentukan hasil persentase prediksi infeksi tiap puskesmas tahun 2022. Dapat dilihat pada gambar 27.

```
In [59]: #Hasil Persentase Prediksi Infeksi Tiap Puskesmas Tahun 2022
PrediksiP = {'Puskesmas' : ["Bogor Selatan","Mulyaharja","Cipaku","Bondongan","Lawang Gintung","Bogor Timur","Pulo Armyrn","Bogor Uta
'Persentase Prediksi 2022' : [PrediksiP12022,PrediksiP22022,PrediksiP32022,PrediksiP42022,PrediksiP52022,PrediksiP62022,
]}
Prediksi2021 = pd.DataFrame (PrediksiP)
Prediksi2021
```

Out[59]:

	Puskesmas	Persentase Prediksi 2022
0	Bogor Selatan	5.185816
1	Mulyaharja	2.696409
2	Cipaku	3.595002
3	Bondongan	3.753902
4	Lawang Gintung	2.976311
5	Bogor Timur	3.424413
6	Pulo Armyrn	0.770100
7	Bogor Utara	5.814842
8	Warung Jambu	4.358531
9	Tegal Gundil	2.218088
10	Bogor Tengah	4.863234

Gambar 27. Hasil *presentase* prediksi infeksi tiap puskesmas tahun 2022

4.6 Implementasi Sistem

Implementasi pembuatan website *clustering*, klasifikasi dan prediksi untuk data penyakit di dinas kesehatan kota Bogor dibuat menggunakan bahasa pemrograman *python, html, javascript, web framework flask* dengan *software visual studio code*.

BAB V HASIL DAN PEMBAHASAN

5.1 Data Cleaning

Data Cleaning bertugas untuk menghilangkan noise dan data yang tidak konsisten. Pada tahap ini dilakukan penghapusan pada data yang tidak memiliki kelengkapan atribut sesuai yang dibutuhkan (Savitri, 2018). Pada penelitian ini proses pembersihan tidak dilakukan karena data yang didapatkan dari dinas kesehatan Bogor tanpa adanya nilai yang hilang, data juga sudah konsisten, dan tidak ada data yang kosong. Dapat dilihat pada gambar 28.

```
Out[3]:
```

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Infeksi
0	1.0	183.67	11.67	11.00	688.00	25485.33	2.0
1	2.0	192.67	11.67	51.33	1115.67	41313.33	0.0
2	3.0	259.33	15.67	47.33	1364.67	50533.33	0.0
3	4.0	229.67	21.33	23.67	1281.67	47455.67	0.0
4	5.0	116.33	13.67	15.67	1111.67	41166.67	0.0
5	6.0	180.67	46.00	90.00	1719.67	63695.67	1.0
6	7.0	213.67	15.67	110.00	1207.00	44708.00	0.0
7	8.0	211.67	30.67	201.00	1947.00	72104.67	1.0
8	9.0	341.00	22.00	173.33	2046.00	75781.33	0.0
9	10.0	87.00	21.33	107.67	1467.00	54335.33	1.0
10	11.0	106.67	5.33	6.00	253.33	9372.67	2.0

Gambar 28. Total Keseluruhan Data

5.2 Data Transformation

Pada tahap transformasi data memiliki fungsi akan mengubah data maupun diinisialisasi dalam bentuk numerik, kategori atau digabungkan ke dalam format yang untuk proses data mining (Geeks, 2020). Proses transformasi data dilakukan dengan menghapus cell yang kosong atau disebut dengan NaN, menghapus data duplikasi, serta atribut yang tidak relevan. Hal ini dilakukan untuk memperkecil volume dokumen yang akan dilakukan. Data yang digunakan dalam penelitian ini adalah sumber data yang diperoleh dari dinas kesehatan kota Bogor. Data yang dikumpulkan adalah data penyakit tuberkulosis (TB), diare, demam berdarah dengue (DBD) dan pneumonia dengan atribut puskesmas, jenis kasus, jumlah penduduk, tahun. Beberapa data yang diambil bersifat kualitatif yaitu data yang bukan angka.

```
In [8]: #Inisiasi Kolom Yang Digunakan
x = data.drop(columns="Infeksi")
y = data.Infeksi
x
```

```
Out[8]:
```

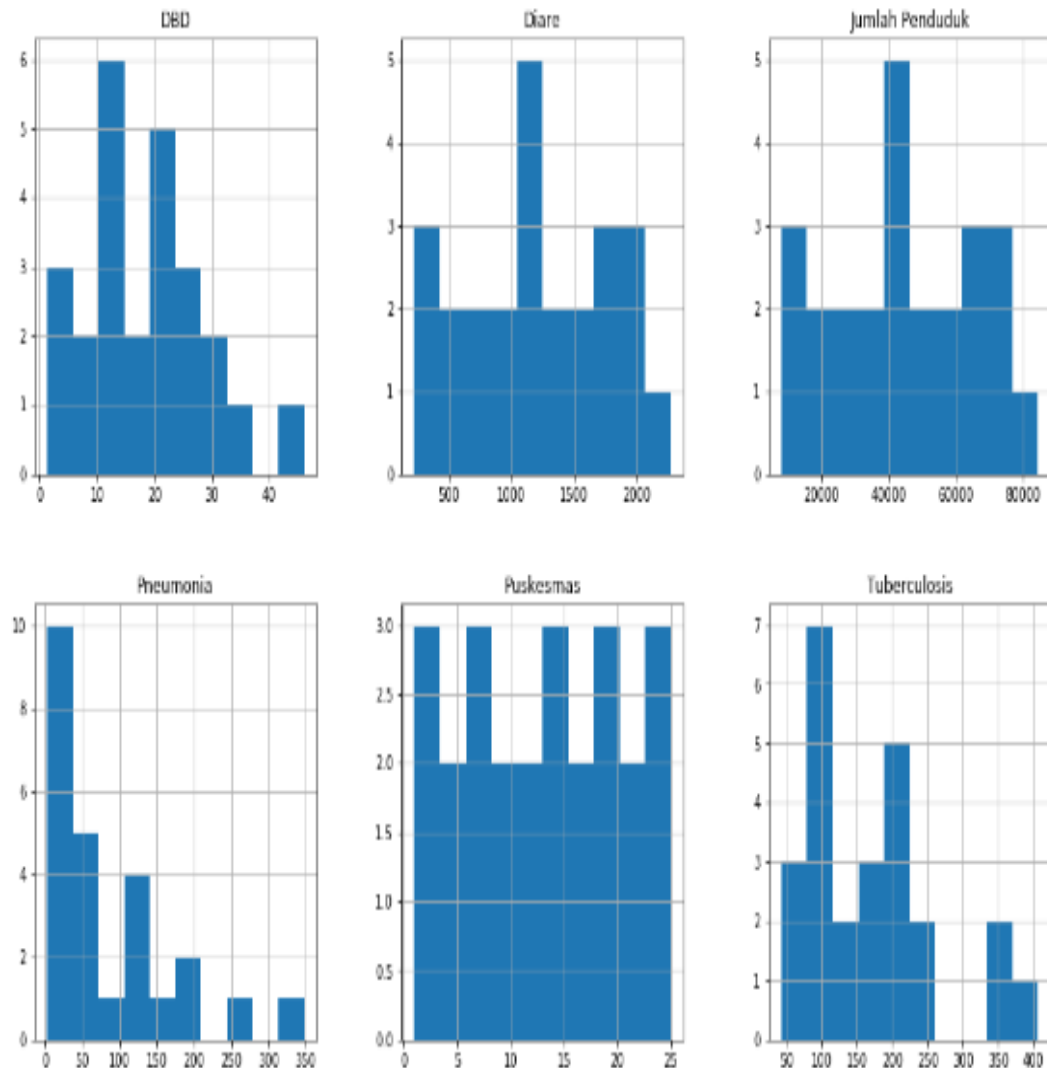
	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
0	1	183.67	11.67	11.00	688.00	25485.33
1	2	192.67	11.67	51.33	1115.67	41313.33
2	3	259.33	15.67	47.33	1364.67	50533.33
3	4	229.67	21.33	23.67	1281.67	47455.67
4	5	116.33	13.67	15.67	1111.67	41166.67
5	6	180.67	46.00	90.00	1719.67	63695.67
6	7	213.67	15.67	110.00	1207.00	44708.00
7	8	211.67	30.67	201.00	1947.00	72104.67
8	9	341.00	22.00	173.33	2046.00	75781.33
9	10	87.00	21.33	107.67	1467.00	54335.33
10	11	106.67	5.33	6.00	253.33	9372.67
11	12	153.67	14.67	4.33	705.00	26110.67
12	13	43.00	7.33	3.00	496.00	18369.33
13	14	79.67	1.33	3.00	246.00	9118.00
14	15	97.67	10.67	52.00	1133.33	41970.67
15	16	108.00	14.00	119.33	986.00	36526.67

Gambar 29. Data Transformasi

5.3 Data Mining

Data Mining adalah suatu pengekstrakan informasi baru yang diambil dari sebuah data besar yang membantu dalam pengambilan keputusan atau bisa dikatakan sebagai proses pengumpulan informasi penting dari suatu data yang besar untuk diubah menjadi sebuah pengetahuan (Sembiring, 2019). Proses mining dilakukan dengan menggunakan *library scikit-learn* dengan bahasa pemrograman *Python*. Data yang digunakan dari rentang waktu 2019-2021 dengan 4 atribut diantaranya puskesmas, jumlah kasus, tahun dan jumlah penduduk dengan jenis penyakit yang akan digunakan yaitu tuberkulosis(TB), diare, demam berdarah dengue (DBD) dan pneumonia. Metode *K-Medoids* dan *Naïve Bayes* akan diterapkan pada penelitian ini untuk *clustering*, klasifikasi dan prediksi dengan 3 target kelas yaitu rendah, sedang dan tinggi yang dapat dilihat pada Gambar 28

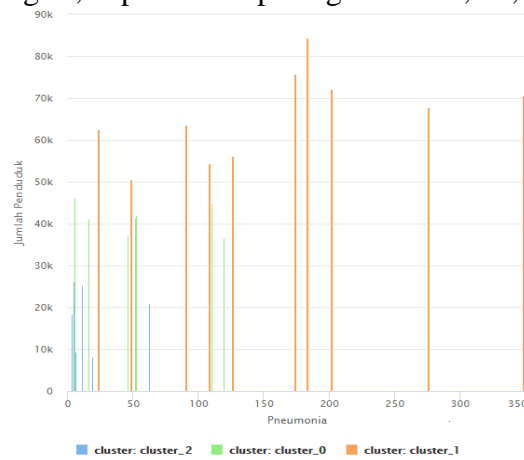
Jumlah atribut kelas yang akan diklaster dibuat menjadi 6 dataset. Dataset pertama merupakan dataset puskesmas. Kemudian dataset kedua merupakan dataset tuberkulosis. Kemudian dataset ketiga merupakan dataset diare. Kemudian dataset keempat merupakan dataset demam berdarah (DBD). Kemudian dataset kelima merupakan dataset pneumonia dan dataset enam merupakan dataset jumlah penduduk.



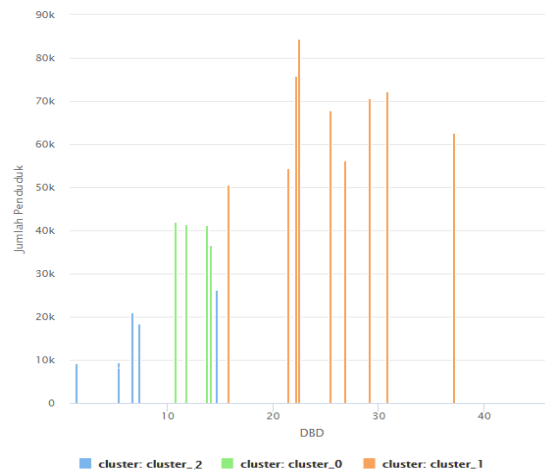
Gambar 30. Jumlah atribut kelas pada masing masing dataset

5.3.1 Hasil Clustering dengan Metode K-Medoids

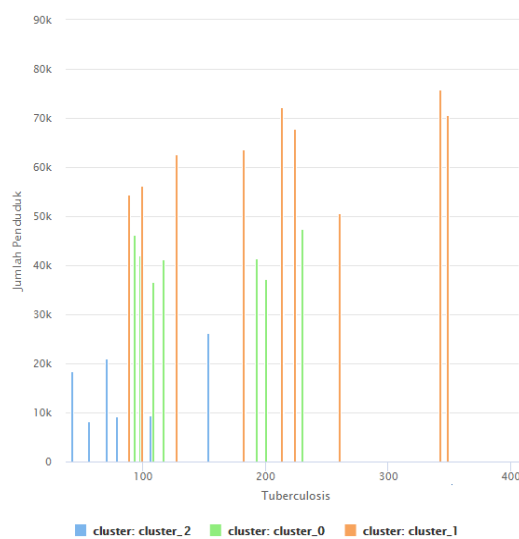
Hasil clustering menggunakan metode K-Medoids pada data penyakit yang terbagi ke dalam 4 kategori, dapat dilihat pada gambar 31, 32, 33, dan 34.



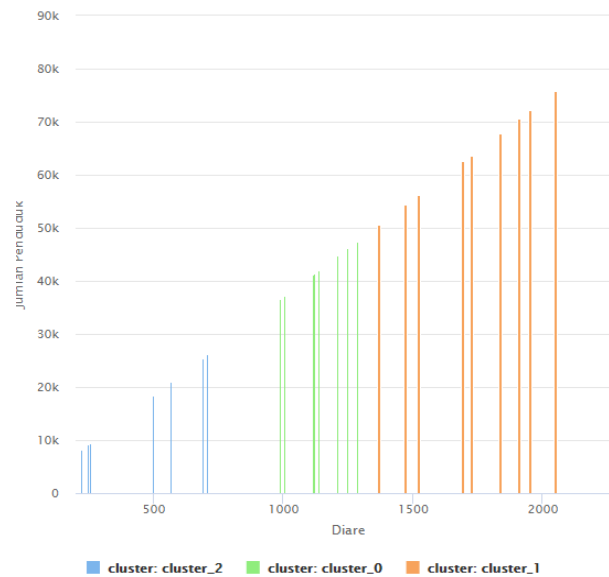
Gambar 31. Hubungan Clustering Jumlah Penduduk dan Pneumonia



Gambar 32. Hubungan Clustering Jumlah Penduduk dan DBD



Gambar 33. Hubungan Clustering Jumlah Penduduk dan Tuberculosis

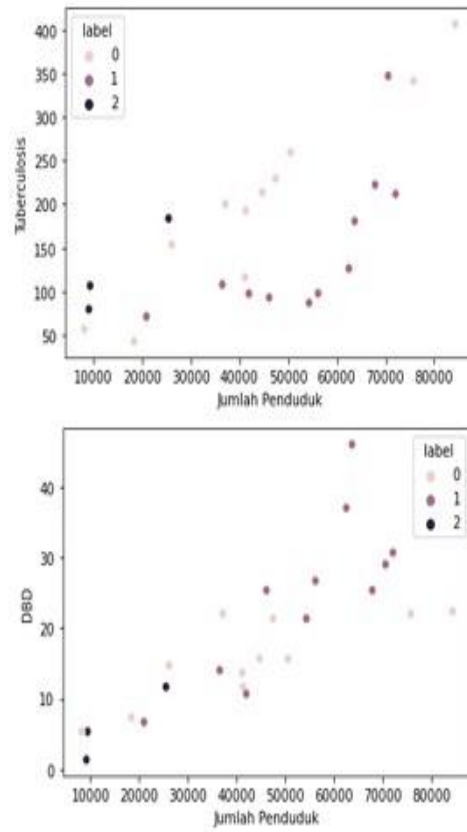


Gambar 34. Hubungan Clustering Jumlah Penduduk dan Diare

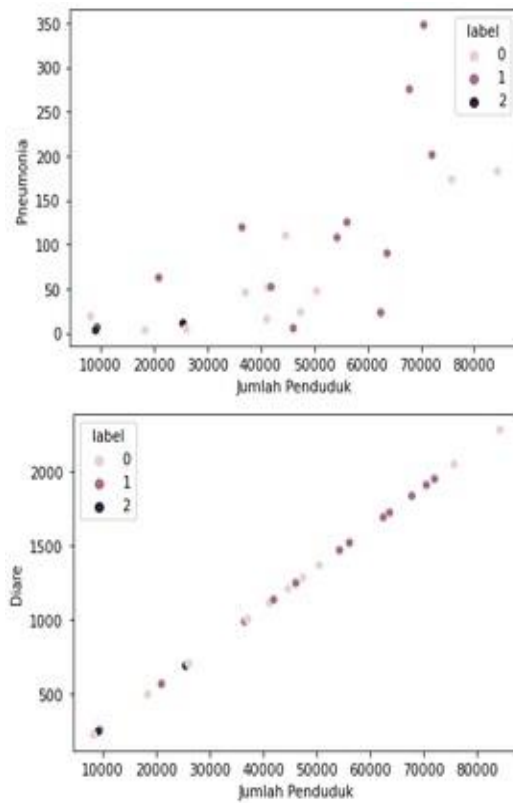
Berdasarkan Gambar 31 dalam bentuk visualisasi data menggunakan rapidminer menunjukkan bahwa sebagian besar bahwa penderita penyakit Pneumonia yang tinggi dan jumlah penduduk yang tinggi berada dalam cluster 2, sedangkan untuk cluster 1 di dominasi oleh penyakit Pneumonia sedang dan jumlah penduduk sedang, untuk cluster 3 didominasi oleh penyakit Pneumonia rendah dan jumlah penduduk rendah. Dalam gambar 32 cluster 2 di dominasi oleh penderita penyakit DBD tinggi dan jumlah penduduk tinggi, sedangkan pada cluster 1 di dominasi oleh penderita penyakit DBD sedang dan jumlah penduduk sedang, untuk cluster 3 di dominasi oleh penderita penyakit DBD rendah dan jumlah penduduk rendah. Dalam gambar 33 menunjukkan bahwa cluster 2 di dominasi oleh penderita penyakit Tuberculosis tinggi dan jumlah penduduk yang tinggi, sedangkan pada cluster 1 di dominasi oleh penderita penyakit Tuberculosis sedang dan jumlah penduduk sedang, untuk cluster 3 di dominasi oleh penderita penyakit tuberculosis rendah dan jumlah penduduk rendah. Dalam gambar 34 cluster 2 di dominasi oleh penderita penyakit Diare yang tinggi dan jumlah penduduk yang tinggi, sedangkan untuk cluster 1 di dominasi oleh penderita penyakit Diare yang sedang dan jumlah penduduk yang sedang, untuk cluster 3 di dominasi oleh penderita penyakit Diare yang rendah dan jumlah penduduk yang rendah. Untuk hubungan variabel kategori jumlah penduduk dan penyakit Pneumonia, DBD, Tuberculosis, dan Diare pada cluster 1 di dominasi oleh tingkat jumlah penduduk yang sedang dan jumlah tingkat penderita penyakit yang sedang, sedangkan pada cluster 2 di dominasi oleh tingkat jumlah penduduk yang tinggi dan jumlah tingkat penderita penyakit yang tinggi, dan untuk cluster 3 di dominasi oleh tingkat jumlah penduduk yang rendah dan jumlah tingkat penderita penyakit yang rendah.

5.3.2 Proses Keseluruhan Klasterisasi *K-Medoids*

Sebaran data keseluruhan hasil *cluster* menggunakan metode *K-Medoids* dapat dilihat pada Gambar 35 dan 37.



Gambar 35. Scatter Plot Hasil *Clustering* TB dan DBD



Gambar 36. Scatter Plot Hasil *Clustering* Pneumonia dan Diare

Berdasarkan Gambar 35 dan 36, Hasil menunjukkan bahwa hubungan variabel x (jumlah penduduk) dan y (tuberkulosis), semakin banyak jumlah penduduk dan tuberkulosis maka hasil clustering terdapat pada cluster 1, sedangkan semakin sedikit jumlah penduduk dan kasus tuberkulosis maka hasil clustering terdapat pada cluster 3, sedangkan untuk cluster 2 di dominasi pada pertengahan jumlah penduduk dan tuberkulosis. Pada variabel x (jumlah penduduk) dan y (dbd) pada cluster 1 di dominasi oleh jumlah penduduk dan pneumonia yang memiliki data di tengah – tengah nilai diare dan jumlah penduduk, cluster 2 di dominasi oleh jumlah penduduk dan pneumonia yang tinggi, sedangkan cluster 3 di dominasi oleh jumlah penduduk dan pneumonia yang rendah. Untuk gambar 33 dapat dijelaskan bahwa variabel x (jumlah penduduk) dan y (pneumonia) pada cluster 1 di dominasi oleh jumlah penduduk dan penderita dbd yang memiliki nilai sedang, untuk cluster 2 di dominasi oleh jumlah penduduk dan dbd yang tinggi, sedangkan untuk cluster 3 cenderung rendah. Pada variabel x (jumlah penduduk) dan y (diare) cluster 1 di dominasi oleh tingkat jumlah penduduk dan diare yang tinggi, cluster 2 di dominasi pada jumlah penduduk dan diare yang sedangkan, sedangkan cluster 3 cenderung rendah.

5.3.3 Proses Pembagian Data Latih dan Data Uji

Data latih digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*, model ini merupakan representasi dari pengetahuan yang akan digunakan untuk memprediksi pada kelas baru yang belum pernah ada, semakin besar data latih yang digunakan semakin baik juga sistem dalam memahami pola data. Sedangkan, data uji digunakan untuk mengukur sejauh mana klasifikasi berhasil dilakukan dengan benar. Data yang digunakan untuk data latih dan data uji adalah data yang sudah memiliki kelas (Ines, 2019). Pembagian data latih dan data uji dilakukan dengan mengimplementasikan 10% data uji dari 25 data yang dilatih yaitu dengan data latih sebanyak 22 sedangkan data ujinya sebanyak 3.

```
In [6]: #melatih data model dan tes menggunakan 10% data dari 25 data yang diat
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1
x_train.shape, x_test.shape, y_train.shape, y_test.shape
```

```
Out[6]: ((22, 6), (3, 6), (22,), (3,))
```

```
In [6]: #Data Uji
x_test
```

```
Out[6]:
```

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
5	6	180.67	46.00	90.00	1719.67	63695.67
2	3	259.33	15.67	47.33	1364.67	50533.33
19	20	98.00	26.67	125.00	1517.00	56183.67

Gambar 37. Pembagian Data Latih dan Data Uji

5.3.4 Pembuatan Model *Naïve Bayes*

Proses ini dilakukan pembuatan model *Naïve Bayes* dengan *ipeline*, proses pembuatan model pada algoritma *Naïve Bayes* dapat dilihat pada gambar 39.

```
In [8]: #Pergabungan data preprocessor dengan algoritma Gaussian NB
from sklearn.naive_bayes import GaussianNB
pipeline = Pipeline([
    ('prep', preprocessor),
    ('algo', GaussianNB())
])

In [9]: #Pembentukan Model Naïve Bayes
pipeline.fit(x_train, y_train)

Out[9]: Pipeline(memory=None,
                 steps=[('prep',
                        ColumnTransformer(n_jobs=None, remainder='drop',
                                           sparse_threshold=0.3,
                                           transformer_weights=None,
                                           transformers=[('numeric',
                                                         Pipeline(memory=None,
                                                                    steps=[('imputer',
                                                                              simpleImputer

(add_indicator=False,
copy=True,
fill_value=None,
missing_values=nan,
strategy='median',
verbose=0)]),

(verbose=False),
['Tuberculosis', 'DBD',
 'Pneumonia', 'Diare',
 'Jumlah Penduduk'])],
(verbose=False)),
('algo', GaussianNB(priors=None, var_smoothing=1e-09))],
verbose=False)

In [10]: pipeline.score(x_train, y_train)
Out[10]: 0.6818181818181818
```

Gambar 38. Pembuatan Model *Naïve Bayes*

5.3.5 Hasil Presentase Infeksi per tahun

Proses dilakukan dengan menggunakan metode *Naïve Bayes* untuk mendapatkan hasil presentase infeksi pertahun dalam rentang waktu dari tahun 2019 sampai 2021. Dapat dilihat pada Gambar 40.

```
In [57]: #Hasil Presentase Infeksi Per Tahun
PrediksiPuskesmas = {'Puskesmas' : ["Bogor Selatan", "Mulyaharja", "Cipaku", "Bondongan", "Lawang Gintung", "Bogor Timur", "Pulo Army", "E
    'Persentase Prediksi 2019' : [PersentaseP12019, PersentaseP22019, PersentaseP32019, PersentaseP42019, PersentaseP52019, Pe
    'Persentase Prediksi 2020' : [PersentaseP12020, PersentaseP22020, PersentaseP32020, PersentaseP42020, PersentaseP52020, Pe
    'Persentase Prediksi 2021' : [PersentaseP12021, PersentaseP22021, PersentaseP32021, PersentaseP42021, PersentaseP52021, Pe
    }
Prediksi = pd.DataFrame(PrediksiPuskesmas)
Prediksi
```

```
Out[57]:
```

	Puskesmas	Persentase Prediksi 2019	Persentase Prediksi 2020	Persentase Prediksi 2021
0	Bogor Selatan	4.140584	3.282257	4.214038
1	Mulyaharja	3.754521	3.546098	3.121254
2	Cipaku	3.385595	3.255626	3.425314
3	Bondongan	3.377011	3.748005	3.750954
4	Lawang Gintung	2.812698	3.177376	3.076844
5	Bogor Timur	4.591130	4.182837	3.803625
6	Pulo Army	4.810786	4.802685	2.786393
7	Bogor Utara	4.578084	4.057896	4.936369
8	Warung Jambu	4.269749	3.718217	4.037374
9	Tegal Gundil	4.594338	3.002092	2.810090
10	Bogor Tengah	4.296032	5.134962	4.999088
11	Merdeka	3.646663	4.487903	3.132393
12	Gang Aut	3.559262	2.415681	3.145571
13	Belong	3.388552	2.957458	4.401739
14	Sempur	4.144158	2.637220	2.411120

Gambar 39. Hasil Presentase Infeksi pertahun

Pada tahap prediksi dengan menggunakan metode *Naïve Bayes* proses yang telah dibuat sebelumnya berdasarkan hasil presentase infeksi pertahun maka mendapatkan hasil presentase prediksi infeksi tiap puskesmas pada tahun 2022. Dapat dilihat pada Gambar 41.

```
In [59]: #Hasil Presentase Prediksi Infeksi Tiap Puskesmas Tahun 2022
PrediksiP = {'Puskesmas' : ["Bogor Selatan", "Mulyaharja", "Cipaku", "Bondongan", "Lawang Gintung", "Bogor Timur", "Pulo Armin", "Bogor Utara", "Bogor Tengah", "Merdeka", "Gang Aut", "Belong", "Sempur", "Pasir Mulya", "Semplak", "Pancasan", "Sindang Barang", "Gang Kelor"],
'Presentase Prediksi 2022' : [PrediksiP12022, PrediksiP22022, PrediksiP32022, PrediksiP42022, PrediksiP52022, PrediksiP62022, PrediksiP72022, PrediksiP82022, PrediksiP92022, PrediksiP102022, PrediksiP112022, PrediksiP122022, PrediksiP132022, PrediksiP142022, PrediksiP152022, PrediksiP162022, PrediksiP172022, PrediksiP182022, PrediksiP192022]}
Prediksi2021 = pd.DataFrame (PrediksiP)
Prediksi2021
```

```
Out[59]:
```

	Puskesmas	Persentase Prediksi 2022
0	Bogor Selatan	5.165816
1	Mulyaharja	2.696409
2	Cipaku	3.595002
3	Bondongan	3.753902
4	Lawang Gintung	2.976311
5	Bogor Timur	3.424413
6	Pulo Armin	0.770100
7	Bogor Utara	5.814842
8	Warung Jambu	4.366531
9	Tegal Gundil	2.218088
10	Bogor Tengah	4.863234
11	Merdeka	1.776883
12	Gang Aut	3.875461
13	Belong	5.846020
14	Sempur	2.185020
15	Pasir Mulya	3.890822
16	Semplak	3.811462
17	Pancasan	2.635409
18	Sindang Barang	13.514759
19	Gang Kelor	5.659492

Gambar 40. Hasil presentasi prediksi infeksi tiap puskesmas tahun 2022

5.4 Pembahasan

Pada tahap pengolahan data algoritma *Naïve Bayes* analisis *big data* untuk *clustering*, klasifikasi dan prediksi pada dinas kesehatan kota Bogor. Proses pada data training dan data uji merupakan inti dari pembahasan pengolahan data yang dibuat. Data training digunakan untuk menyimpan kasus lama yang sudah diketahui hasil akhirnya, sehingga ketika ada data uji yang dimasukkan dapat diprediksi hasil akhirnya berdasarkan ukuran kedekatan. Dalam penelitian kali ini menggunakan 22 data training untuk membuat suatu model lalu di uji coba ke dalam data training sebanyak 22 data selanjutnya untuk data uji menggunakan sebanyak 3 data. Hasil dari 22 data training tersebut dilakukan klasifikasi dan di bandingkan dengan data asli yaitu sebanyak 22 data training. Sama halnya dengan 3 data uji yang baru dibandingkan dengan data uji sebelumnya. Untuk lebih jelasnya dapat di lihat pada gambar 42.

```
In [47]: #Membuat klasifikasi baru menggunakan naive bayes dengan data dengan data training sebanyak 22
data["Infeksi"] = pipeline.predict(data)
data
```

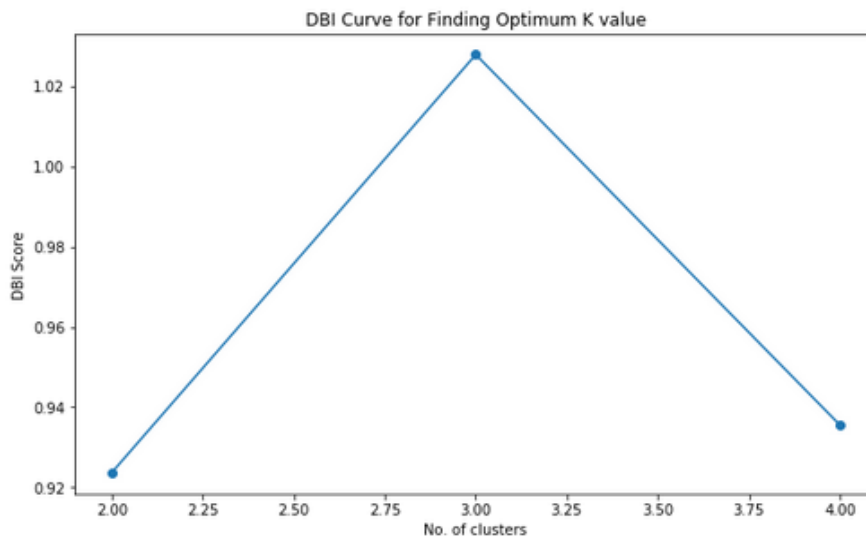
Out[47]:

	Puskesmas	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Infeksi
0	1	183.67	11.67	11.00	688.00	25485.33	2
1	2	192.67	11.67	51.33	1115.67	41313.33	0
2	3	269.33	15.67	47.33	1364.67	50533.33	0
3	4	229.67	21.33	23.67	1281.67	47455.67	0
4	5	118.33	13.67	15.67	1111.67	41166.67	0
5	6	180.67	46.00	90.00	1719.67	63695.67	1
6	7	213.67	15.67	110.00	1207.00	44708.00	0
7	8	211.67	30.67	201.00	1947.00	72104.67	1
8	9	341.00	22.00	173.33	2046.00	75781.33	1
9	10	87.00	21.33	107.67	1467.00	54335.33	1
10	11	106.67	5.33	6.00	253.33	9372.67	2
11	12	153.67	14.67	4.33	705.00	26110.67	2
12	13	43.00	7.33	3.00	496.00	18369.33	2
13	14	79.67	1.33	3.00	246.00	9118.00	2
14	15	97.67	10.67	52.00	1133.33	41970.67	0
15	16	108.00	14.00	119.33	986.00	36526.67	0
16	17	93.00	25.33	5.33	1245.33	46128.00	1
17	18	200.33	22.00	45.67	1002.67	37137.00	0
18	19	347.00	29.00	347.67	1905.67	70580.00	1
19	20	98.00	26.67	125.00	1517.00	56183.67	1
20	21	56.67	5.33	19.00	220.67	8170.00	2
21	22	71.33	6.67	62.33	565.67	20951.33	0
22	23	222.33	25.33	275.00	1832.33	67873.00	1
23	24	406.33	22.33	182.67	2277.67	84362.00	0
24	25	126.33	37.00	23.00	1688.00	62513.00	1

Gambar 41. Hasil Keseluruhan Klasifikasi Prediksi

5.4.1 Uji Akurasi K-Medoids dengan menggunakan Davies Bouldin Index (DBI)

Berdasarkan hasil implementasi, diperoleh hasil nilai DBI berdasarkan masing-masing jumlah cluster yang direpresentasikan dalam grafik, dapat dilihat pada Gambar 43.



Gambar 42. Grafik Nilai DBI Menggunakan Algoritma K-Medoids

Proses penentuan jumlah *cluster* yang optimal dapat dilakukan dengan cara pengujian validasi menggunakan metode *Davies Bouldin Index* (DBI). Jumlah *cluster* dengan nilai DBI paling kecil dapat disimpulkan sebagai jumlah *cluster* optimal. Untuk melihat hasil nilai DBI secara detail dapat dilihat pada Tabel 30 untuk metode *K-Medoids*.

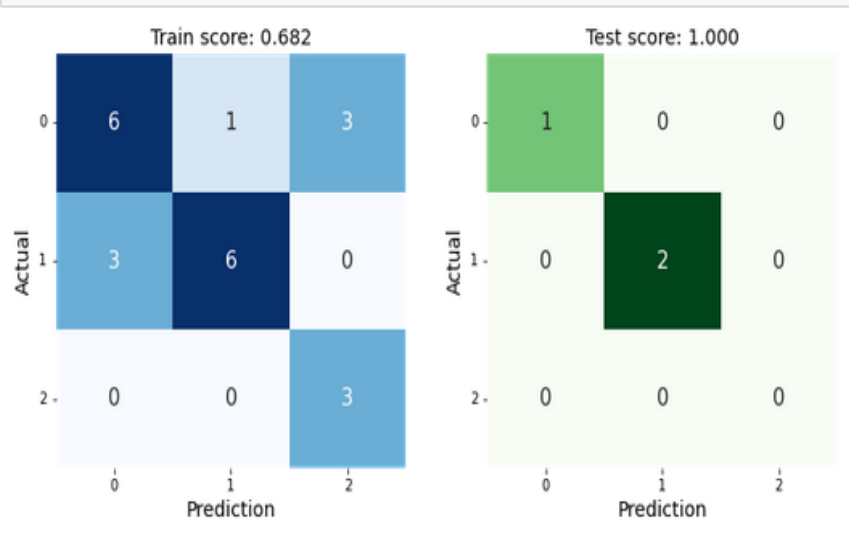
Tabel 30. Nilai DBI *K-Medoids*

Jumlah Cluster	Nilai DBI
2	0,923
3	1,027
4	0,935

Berdasarkan tabel 18, menunjukan hasil nilai *cluster* yang telah dilakukan pengujian validasi menggunakan DBI bahwa dari metode *K-Medoids* berdasarkan hasil perhitungan, dapat dilihat bahwa nilai DBI terendah dari metode *K-Medoids* terdapat pada *cluster*. Nilai DBI didapat dari label yang sudah di tentukan oleh *K-Medoids*. Setelah dilakukan percobaan dengan pemodelan K=2, K=3 dan K=4, didapatkan nilai *cluster* optimal terdapat pada K=2 untuk metode *K-Medoids* dengan nilai DBI 0,923. Berdasarkan nilai DBI menunjukkan hasil *cluster* yang paling optimal yaitu k= 2 sebab semakin kecil nilai DBI maka semakin baik pula hasil *cluster* tersebut.

5.4.2 Uji Akurasi *Naïve Bayes* dengan menggunakan *Confusion Matrix*

Uji akurasi hasil klasifikasi dilakukan untuk menguji tingkat akurasi klasifikasi yang telah dihitung berdasarkan model *Naïve Bayes* yang telah dibuat. Berdasarkan perhitungan confusion matrix tersebut maka klasifikasi data uji mendapatkan 100% dengan melakukan pengujian data latih dengan data uji sebesar 90/10. Nilai akurasi dengan hasil data prediksi klasifikasi sebanyak 3 dan data aktual sebanyak 3. Hasil akurasi ini didapat dengan cara menghitung menggunakan rumus akurasi yaitu $= (TP+TN)/(TP+FP+FN+TN)$. Akurasi yang didapatkan data uji sangat baik karena perbandingan data aktual dan prediksi cukup sedikit. sedangkan untuk data training mendapatkan 68.2% dengan data prediksi klasifikasi sebanyak 22 dan data aktual sebanyak 22. Untuk melihat data aktual dan data prediksi pada confusion matrix dapat dilihat pada Gambar 44.



Gambar 43. Hasil *Confusion Matrix*

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan model analisis klasifikasi, prediksi dan *clustering K-Medoids* dan *Naïve Bayes* untuk analisis *big data* pada penyakit dengan pengambilan data berasal dari dinas kesehatan kota Bogor dari rentang waktu 2019-2021 yang dibagi menjadi 6 kategori yaitu puskesmas, jumlah kasus, tahun, dan jumlah penduduk dengan jenis penyakit yang akan digunakan yaitu tuberkulosis (TB), diare, demam berdarah dengue (DBD), dan pneumonia. Hasil menunjukkan bahwa dominan yang bergerombol terdapat *cluster* 2 didominasi oleh tingkat infeksi penyakit sedang sebesar 50% dengan terdiri dari 12 puskesmas. *Cluster* 1 didominasi oleh tingkat infeksi penyakit tinggi sebesar 35% dengan terdiri dari 9 puskesmas *Cluster* 3 didominasi oleh tingkat infeksi penyakit tinggi sebesar 15% dengan terdiri dari 5 puskesmas. Pengujian validasi menggunakan DBI dimana setelah dilakukan percobaan dengan pemodelan $K=2$, $K=3$ dan $K=4$, didapatkan nilai *cluster* optimal terdapat pada $K=2$ untuk metode *K-Medoid* dengan nilai DBI 0,923. Berdasarkan nilai DBI menunjukkan hasil *cluster* yang paling optimal yaitu $k=2$ sebab semakin kecil nilai DBI maka semakin baik pula hasil *cluster* tersebut.

Penelitian ini menggunakan 22 data training untuk membuat suatu model lalu di uji coba ke dalam data training sebanyak 22 data selanjutnya untuk data uji menggunakan sebanyak 3 data. Hasil dari 22 data training tersebut dilakukan klasifikasi dan di bandingkan dengan data asli yaitu sebanyak 22 data training. Sama halnya dengan 3 data uji yang baru dibandingkan dengan data uji sebelumnya. Maka hasil prediksi dapat dihasilkan bahwa puskesmas dengan tingkat infeksi yang tinggi di tahun 2019 berdasarkan kriteria penyakit dan jumlah penduduk adalah puskesmas kayu manis dengan persentase tingkat infeksi sebesar 30,078%, di tahun 2020 puskesmas kayu manis dengan persentase tingkat infeksi sebesar 28,254%, di tahun 2021 puskesmas kayu manis dengan persentase tingkat infeksi sebesar 30,65%, dan pada 2022 prediksi tingkat infeksi tertinggi adalah puskesmas kayu manis dengan persentase tingkat infeksi sebesar 33,046%.

6.2 Saran

Penelitian ini dapat dikembangkan kembali dengan pembandingan atau penambahan data penyakit yang lebih kompleks dan penggabungan metode data mining lainnya seperti *SVM*, *K-Prototype*, ataupun yang lainnya sehingga menghasilkan pengelompokan data yang lebih optimal dan klasifikasi yg lebih akurat. Untuk Sistem ini bisa dikembangkan dari segi aplikasi yang lebih lengkap, pemetaan penyakit apa saja yang rawan terkena pada skala besar dengan GIS dan juga dapat memonitoring tingkatan penyakit yang sedang naik setiap bulannya dengan aplikasi mobile sehingga pembaca dapat mencegah ataupun merawat diri dari penyakit tersebut.

DAFTAR PUSTAKA

- Armansyah. 2021. *Model Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes*. Vol. 7.
- Dinkes,A. (2016). Profil Kesehatan Kabupaten Bogor 2016. Dinas kesehatan Kab.Bogor
- Dinkes, A. (2017). Profil Kesehatan Kabupaten Kabupaten Bogor tahun 2017. Dinas Kesehatan Kab.Bogor
- Emir Ramon. 2022. “Klasifikasi Gtatus Gazi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Verctor Machine (SVM).” *Jurnal Sistem Informasi dan Informatika (Simika)*
- Fajrl Muhammad Bhakti. 2022. “Klasterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Menggunakan K-Means Clustering.” *Jurnal of Information Technology Ampere* 3(3): 317-334
- Gullo, Francesco. 2016. “From Patterns in Data to Knowledge Discovery: What Data Mining Can Do.” *Physics Procedia* 62: 18–22.
- Gustientiedinaa, M.Hasmil Adiyaa, and Yenny Desnelitab. 2019. “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru.” *JURNAL NASIONAL TEKNOLOGI DAN SISTEM* 5 (1): 17–34
- Juninda, Tri, Mustasim, and Elvia Andri. 2019. “Penerapan Algoritma K-Medoids Untuk Pengelompokan Penyakit Di Pekanbaru Riau.” *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri* 11 (1): 42–49.
- Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques* 4th Edition. Elsevier. 2022
- Putri Dyan Dyanmita, Furqon M. Tanzil. Perdana Rizal Setya, 2018. “Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDTSVM) (Studi Kasus: Puskesmas Dinoyo Kota Malang).” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 2 (5): 1912-1920
- Rohmawati, Fitri, M Ghofar Rohman, and Siti Mujilawati. 2017. “Sistem Prediksi Jumlah Pengunjung Wisata Wego Kec.Sugio Kab.Lamongan Menggunakan Metode Fuzzy Time Series.” *Jouticla* 2 (2).
- Septiani, Ike Wahyu. 2022. “Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru.” *Jurnal Sistem Komputer Informatika* 3 (4): 556-566
- Syamfithriani Tri Septiar, Mirantika Nita, Trisudarmo Ragel. 2022. “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pemetaan Daerah Penanganan Diare Pada Balita di Kabupaten Kuningan.” *Jurnal Sistem Informasi Bisnis* : 132 – 139.
- Wicaksono, Anggoro Eko. 2016. “Implementasi Data Mining Dalam Pengelompokan Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means.” *Jurnal Teknologi Rekayasa* 21 (3):206–16.

LAMPIRAN

Lampiran 1. Data Penelitian

Puskesmas	Tahun 2019				
	Jenis Penyakit				
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Bogor Selatan	88	19	19	684	25,333
Mulyaharja	70	22	106	1087	40,265
Cipaku	57	22	129	1340	49,619
Bondongan	85	33	29	1291	47,797
Lawang Gintung	33	25	12	1103	40,855
Bogor Timur	76	94	118	1695	62,772
Pulo Armin	99	19	277	1201	44,487
Bogor Utara	187	41	295	1914	70,877
Warung Jambu	157	41	288	1999	74,039
Tegal Gundil	100	35	270	1466	54,284
Bogor Tengah	28	10	6	258	9,541
Merdeka	56	22	10	707	26,185
Gang Aut	55	11	2	500	18,518
Belong	36	1	5	250	9,271
Sempur	105	18	143	1119	41,432
Pasir Mulya	86	18	293	992	36,756
Semplak	65	35	8	1210	44,811
Pancasan	83	33	116	1000	37,034
Sindang Barang	124	24	752	1869	69,211
Gang Kelor	76	20	42	1498	55,481
Tanah Sareal	20	11	8	224	8,297
Pondok Rumput	56	6	108	573	21,240
Kedung Badak	136	20	398	1825	67,589
Kayu Manis	95	18	22	2184	80,892
Mekar Wangi	82	23	43	1627	60,242

Puskesmas	Tahun 2020				
	Jenis Penyakit				
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Bogor Selatan	198	2	9	688	25490
Mulyaharja	205	7	37	1116	41316
Cipaku	354	8	1	1365	50541
Bondongan	217	14	25	1282	47467
Lawang Gintung	190	9	3	1112	41174
Bogor Timur	203	21	85	1720	63704
Pulo Armin	406	15	42	1207	44717
Bogor Utara	113	14	170	1947	72111
Warung Jambu	409	3	143	2046	75786
Tegal Gundil	95	9	43	1467	54349
Bogor Tengah	133	2	10	253	9373

Merdeka	333	7	0	705	26118
Gang Aut	17	3	1	496	18374
Belong	76	0	2	246	9120
Sempur	83	7	6	1133	41977
Pasir Mulya	175	1	47	986	36536
Semplak	118	8	2	1245	46126
Pancasan	264	14	19	1003	37146
Sindang Barang	351	23	207	1906	70589
Gang Kelor	170	2	233	1517	56193
Tanah Sareal	66	3	1	221	8173
Pondok Rumput	80	1	76	566	20959
Kedung Badak	370	12	268	1833	67902
Kayu Manis	480	14	392	2277	84341
Mekar Wangi	160	42	19	1687	62499

Puskesmas	Tahun 2021				
	Jenis Penyakit				
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Bogor Selatan	265	14	5	692	25633
Mulyaharja	303	6	11	1144	42359
Cipaku	367	17	12	1389	51440
Bondongan	387	17	17	1272	47103
Lawang Gintung	126	7	32	1120	41471
Bogor Timur	263	23	67	1744	64611
Pulo Armyn	136	13	11	1213	44920
Bogor Utara	335	37	138	1980	73326
Warung Jambu	457	22	89	2093	77519
Tegal Gundil	66	20	10	1468	54373
Bogor Tengah	159	4	2	249	9204
Merdeka	72	15	3	703	26029
Gang Aut	57	8	6	492	18216
Belong	127	3	2	242	8963
Sempur	105	7	7	1148	42503
Pasir Mulya	63	23	18	980	36288
Semplak	96	33	6	1281	47447
Pancasan	254	19	2	1005	37231
Sindang Barang	566	40	84	1942	71940
Gang Kelor	48	58	100	1536	56877
Tanah Sareal	84	2	48	217	8040
Pondok Rumput	78	13	3	558	20655
Kedung Badak	161	44	159	1839	68128
Kayu Manis	644	35	134	2372	87853
Mekar Wangi	137	46	7	1750	64798

Lampiran 2. Perhitungan Manual

K-Medoid

- **Langkah pertama**, kita tentukan jumlah k sebanyak 3
- **Langkah kedua**, kita pilih data 1 sebagai medoid 1, data 2 sebagai medoid 2 dan data 3 sebagai medoid ke 3.
- **Langkah ketiga** adalah menghitung jarak objek ke masing masing medoid yang telah di Pilih. Kita akan gunakan rumus jarak paling sering dipakai yaitu Euclidean.

Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 1

Data	Jarak objek ke-1 medoid	Jarak objek ke-2 medoid	Jarak objek ke-3 medoid	Terdekat	Klaster yang diikuti
1	0	$D_{1m2} = \sqrt{(30-42)^2 + (29-17)^2 + (20-15)^2 + (1330-1709)^2 + (17921-36069)^2} = \sqrt{329493858} = 18.151,966$	$D_{1m3} = \sqrt{(30-74)^2 + (29-6)^2 + (20-25)^2 + (1330-1747)^2 + (17921-168894)^2} = \sqrt{22793023108} = 150.973,584$	0	1
2	$D_{2m1} = \sqrt{(42-30)^2 + (17-29)^2 + (15-20)^2 + (1709-1330)^2 + (36069-17921)^2} = \sqrt{329493858} = 18.151,966$	0	$D_{2m3} = \sqrt{(42-74)^2 + (17-6)^2 + (15-25)^2 + (1709-1747)^2 + (36069-168894)^2} = \sqrt{17642483314} = 132.825,010$	0	2
3	$D_{3m1} = \sqrt{(74-30)^2 + (6-29)^2 + (25-20)^2 + (1747-1330)^2 + (168894-17921)^2} = \sqrt{22793023108} = 150.973,584$	$D_{3m2} = \sqrt{(74-42)^2 + (6-17)^2 + (25-15)^2 + (1747-1709)^2 + (168894-36069)^2} = \sqrt{17642483314} = 132.825,010$	0	0	3
4	$D_{4m1} = \sqrt{(14-30)^2 + (14-29)^2 + (13-20)^2 + (1036-1330)^2 + (25949-17921)^2} = \sqrt{64535750} = 8.033,415$	$D_{4m2} = \sqrt{(14-42)^2 + (14-17)^2 + (13-15)^2 + (1036-1709)^2 + (25949-36069)^2} = \sqrt{102868126} = 10.142,392$	$D_{4m3} = \sqrt{(14-74)^2 + (14-6)^2 + (13-25)^2 + (1036-1747)^2 + (25949-168894)^2} = \sqrt{20433782354} = 142.946,781$	8.033,415	1
5	$D_{5m1} = \sqrt{(17-30)^2 + (23-29)^2 + (18-20)^2 + (1085-1330)^2 + (12477-17921)^2} = \sqrt{29697370} = 5.449,530$	$D_{5m2} = \sqrt{(17-42)^2 + (23-17)^2 + (18-15)^2 + (1085-1709)^2 + (12477-36069)^2} = \sqrt{556972510} = 23.600,265$	$D_{5m3} = \sqrt{(17-74)^2 + (23-6)^2 + (18-25)^2 + (1085-1747)^2 + (12477-168894)^2} = \sqrt{24466719720} = 156.418,412$	0	1
6	$D_{6m1} = \sqrt{(29-30)^2 + (31-29)^2 + (19-20)^2 + (1423-1330)^2 + (18355-17921)^2} = \sqrt{197011} = 443,860$	$D_{6m2} = \sqrt{(29-42)^2 + (31-17)^2 + (19-15)^2 + (1423-1709)^2 + (18355-36069)^2} = \sqrt{313867973} = 17.716,319$	$D_{6m3} = \sqrt{(29-74)^2 + (31-6)^2 + (19-25)^2 + (1423-1747)^2 + (18355-168894)^2} = \sqrt{22662098183} = 150.539,357$	443,860	1
7	$D_{7m1} = \sqrt{(46-30)^2 + (19-29)^2 + (15-20)^2 + (1793-1330)^2 + (37127-17921)^2} =$	$D_{7m2} = \sqrt{(46-42)^2 + (19-17)^2 + (15-15)^2 + (1793-1709)^2 + (37127-36069)^2} =$	$D_{7m3} = \sqrt{(46-74)^2 + (19-6)^2 + (15-25)^2 + (1793-1747)^2 + (37127-168894)^2} =$	1.061,338	2

	$\sqrt{369085186} = 19.211,590$	$\sqrt{1126440} = 1.061,338$	$\sqrt{17362545458} = 131.767,012$		
8	$D8m1 = \sqrt{(87-30)^2 + (8-29)^2 + (26-20)^2 + (1826-1330)^2 + (170920-17921)^2} = \sqrt{23408943743} = 152.999,816$	$D8m2 = \sqrt{(87-42)^2 + (8-17)^2 + (26-15)^2 + (1826-1709)^2 + (170920-36069)^2} = \sqrt{18184808117} = 134.851,059$	$D8m3 = \sqrt{(87-74)^2 + (8-6)^2 + (26-25)^2 + (1826-1747)^2 + (170920-168894)^2} = \sqrt{4111091} = 2.027,582$	2.027,582	3
9	$D9m1 = \sqrt{(30-30)^2 + (16-29)^2 + (15-20)^2 + (1102-1330)^2 + (26809-17921)^2} = \sqrt{79048722} = 8.890,935$	$D9m2 = \sqrt{(30-42)^2 + (16-17)^2 + (15-15)^2 + (1102-1709)^2 + (26809-36069)^2} = \sqrt{86116194} = 9.279,881$	$D9m3 = \sqrt{(30-74)^2 + (16-6)^2 + (15-25)^2 + (1102-1747)^2 + (26809-168894)^2} = \sqrt{20188565386} = 142.086,471$	8.890,935	1
10	$D10m1 = \sqrt{(18-30)^2 + (25-29)^2 + (20-20)^2 + (1126-1330)^2 + (12894-17921)^2} = \sqrt{25312505} = 5.031,153$	$D10m2 = \sqrt{(18-42)^2 + (25-17)^2 + (20-15)^2 + (1126-1709)^2 + (12894-36069)^2} = \sqrt{537421179} = 23.182,346$	$D10m3 = \sqrt{(18-74)^2 + (25-6)^2 + (20-25)^2 + (1126-1747)^2 + (12894-168894)^2} = \sqrt{24336389163} = 156.001,247$	5.031,153	1
11	$D11m1 = \sqrt{(32-30)^2 + (34-29)^2 + (23-20)^2 + (1487-1330)^2 + (18863-17921)^2} = \sqrt{912051} = 955,014$	$D11m2 = \sqrt{(32-42)^2 + (34-17)^2 + (23-15)^2 + (1487-1709)^2 + (18863-36069)^2} = \sqrt{296096173} = 17.207,445$	$D11m3 = \sqrt{(32-74)^2 + (34-6)^2 + (23-25)^2 + (1487-1747)^2 + (18863-168894)^2} = \sqrt{22509371113} = 150.031,234$	955,014	1
12	$D12m1 = \sqrt{(50-30)^2 + (22-29)^2 + (17-20)^2 + (1884-1330)^2 + (39210-17921)^2} = \sqrt{453528895} = 21.296,218$	$D12m2 = \sqrt{(50-42)^2 + (22-17)^2 + (17-15)^2 + (1884-1709)^2 + (39210-36069)^2} = \sqrt{9896599} = 3.145,886$	$D12m3 = \sqrt{(50-74)^2 + (22-6)^2 + (17-25)^2 + (1884-1747)^2 + (39210-168894)^2} = \sqrt{16817959521} = 129.684,076$	3.145,886	2
13	$D13m1 = \sqrt{(93-30)^2 + (7-29)^2 + (28-20)^2 + (1860-1330)^2 + (179679-17921)^2} = \sqrt{26165935981} = 161.758,882$	$D13m2 = \sqrt{(93-42)^2 + (7-17)^2 + (28-15)^2 + (1860-1709)^2 + (179679-36069)^2} = \sqrt{20623857771} = 143.610,089$	$D13m3 = \sqrt{(93-74)^2 + (7-6)^2 + (28-25)^2 + (1860-1747)^2 + (179679-168894)^2} = \sqrt{116329365} = 10.785,609$	10.785,609	3
14	$D14m1 = \sqrt{(33-30)^2 + (18-29)^2 + (17-20)^2 + (1166-1330)^2 + (27438-17921)^2} = \sqrt{90600324} = 9.518,420$	$D14m2 = \sqrt{(33-42)^2 + (18-17)^2 + (17-15)^2 + (1166-1709)^2 + (27438-36069)^2} = \sqrt{74789096} = 8.648,069$	$D14m3 = \sqrt{(33-74)^2 + (18-6)^2 + (17-25)^2 + (1166-1747)^2 + (27438-168894)^2} = \sqrt{20010139386} = 141.457,200$	8.648,069	2
15	$D15m1 = \sqrt{(20-30)^2 + (26-29)^2 + (21-20)^2 + (1189-1330)^2 + (13132-17921)^2} = \sqrt{22954512} = 4.791,087$	$D15m2 = \sqrt{(20-42)^2 + (26-17)^2 + (21-15)^2 + (1189-1709)^2 + (13132-36069)^2} = \sqrt{526376970} = 22.942,907$	$D15m3 = \sqrt{(20-74)^2 + (26-6)^2 + (21-25)^2 + (1189-1747)^2 + (13132-168894)^2} = \sqrt{24262115340} = 155.763,010$	4.791,087	1
Jumlah	567.505,47	566.364,972	1.753.306,585		
Total	2.887.177,027				

- **Langkah keempat** adalah melakukan iterasi pada medoid yang baru dan akan berhenti jika total simpangannya lebih besar dari 0. Kita pilih responden 13 sebagai medoid 1, responden 14 sebagai medoid 2 dan responden 15 sebagai medoid 3.

Hasil Perhitungan Jarak Ke Setiap Centroid Baru Hasil Iterasi 2

Data	Jarak objek ke-1 medoid	Jarak objek ke-2 medoid	Jarak objek ke-3 medoid	Terdekat	Kluster yang diikuti
1	$D_{1m1} = \sqrt{(30-93)^2 + (29-7)^2 + (20-28)^2 + (1330-1860)^2 + (17921-179679)^2} = \sqrt{26165935981} = 161.758,882$	$D_{1m2} = \sqrt{(30-33)^2 + (29-18)^2 + (20-17)^2 + (1330-1166)^2 + (17921-27438)^2} = \sqrt{90600324} = 9.518,420$	$D_{1m3} = \sqrt{(30-20)^2 + (29-26)^2 + (20-21)^2 + (1330-1189)^2 + (17921-13132)^2} = \sqrt{22954512} = 4.791,087$	4.791,087	3
2	$D_{2m1} = \sqrt{(42-93)^2 + (17-7)^2 + (15-28)^2 + (1709-1860)^2 + (36069-179679)^2} = \sqrt{20623857771} = 143.610,089$	$D_{2m2} = \sqrt{(42-33)^2 + (17-18)^2 + (15-17)^2 + (1709-1166)^2 + (36069-27438)^2} = \sqrt{74789096} = 8.648,069$	$D_{2m3} = \sqrt{(42-20)^2 + (17-26)^2 + (15-21)^2 + (1709-1189)^2 + (36069-13132)^2} = \sqrt{526376970} = 22.942,907$	8.648,069	2
3	$D_{3m1} = \sqrt{(74-93)^2 + (6-7)^2 + (25-28)^2 + (1747-1860)^2 + (168894-179679)^2} = \sqrt{116329365} = 10.785,609$	$D_{3m2} = \sqrt{(74-33)^2 + (6-18)^2 + (25-17)^2 + (1747-1166)^2 + (168894-27438)^2} = \sqrt{20010139386} = 141.457,200$	$D_{3m3} = \sqrt{(74-20)^2 + (6-26)^2 + (25-21)^2 + (1747-1189)^2 + (168894-13132)^2} = \sqrt{24262115340} = 155.763,010$	10.785,609	1
4	$D_{4m1} = \sqrt{(14-93)^2 + (14-7)^2 + (13-28)^2 + (1036-1860)^2 + (25949-179679)^2} = \sqrt{23633598391} = 153.732,230$	$D_{4m2} = \sqrt{(14-33)^2 + (14-18)^2 + (13-17)^2 + (1036-1166)^2 + (25949-27438)^2} = \sqrt{2234414} = 1.494,796$	$D_{4m3} = \sqrt{(14-20)^2 + (14-26)^2 + (13-21)^2 + (1036-1189)^2 + (25949-13132)^2} = \sqrt{164299142} = 12.817,923$	1.494,796	2
5	$D_{5m1} = \sqrt{(17-93)^2 + (23-7)^2 + (18-28)^2 + (1085-1860)^2 + (12477-179679)^2} = \sqrt{27957115561} = 167.203,814$	$D_{5m2} = \sqrt{(17-33)^2 + (23-18)^2 + (18-17)^2 + (1085-1166)^2 + (12477-27438)^2} = \sqrt{223838364} = 14.961,229$	$D_{5m3} = \sqrt{(17-20)^2 + (23-26)^2 + (18-21)^2 + (1085-1189)^2 + (12477-13132)^2} = \sqrt{439868} = 663,225$	663,225	3
6	$D_{6m1} = \sqrt{(29-93)^2 + (31-7)^2 + (19-28)^2 + (1423-1860)^2 + (18355-179679)^2} = \sqrt{26025628698} = 161.324,607$	$D_{6m2} = \sqrt{(29-33)^2 + (31-18)^2 + (19-17)^2 + (1423-1166)^2 + (18355-27438)^2} = \sqrt{82567127} = 9.086,645$	$D_{6m3} = \sqrt{(29-20)^2 + (31-26)^2 + (19-21)^2 + (1423-1189)^2 + (18355-13132)^2} = \sqrt{27334595} = 5.228,250$	5.228,250	3
7	$D_{7m1} = \sqrt{(46-93)^2 + (19-7)^2 + (15-28)^2 + (1793-1860)^2 + (37127-179679)^2} = \sqrt{20321079715} = 142.552,024$	$D_{7m2} = \sqrt{(46-33)^2 + (19-18)^2 + (15-17)^2 + (1793-1166)^2 + (37127-27438)^2} = \sqrt{94270024} = 9.709,275$	$D_{7m3} = \sqrt{(46-20)^2 + (19-26)^2 + (15-21)^2 + (1793-1189)^2 + (37127-13132)^2} = \sqrt{576125602} = 24.002,617$	9.709,275	2

8	$D8m1 = \sqrt{(87-93)^2 + (8-7)^2 + (26-28)^2 + (1826-1860)^2 + (170920-179679)^2} = \sqrt{76721278} = 8.759,068$	$D8m2 = \sqrt{(87-33)^2 + (8-18)^2 + (26-17)^2 + (1826-1166)^2 + (170920-27438)^2} = \sqrt{20587523021} = 143.483,529$	$D8m3 = \sqrt{(87-20)^2 + (8-26)^2 + (26-21)^2 + (1826-1189)^2 + (170920-13132)^2} = \sqrt{24897463551} = 157.789,301$	8.759,068	1
9	$D9m1 = \sqrt{(30-93)^2 + (16-7)^2 + (15-28)^2 + (1102-1860)^2 + (26809-179679)^2} = \sqrt{23369815683} = 152.871,893$	$D9m2 = \sqrt{(30-33)^2 + (16-18)^2 + (15-17)^2 + (1102-1166)^2 + (26809-27438)^2} = \sqrt{399754} = 632,261$	$D9m3 = \sqrt{(30-20)^2 + (16-26)^2 + (15-21)^2 + (1102-1189)^2 + (26809-13132)^2} = \sqrt{187068134} = 13.677,285$	632,261	2
10	$D10m1 = \sqrt{(18-93)^2 + (25-7)^2 + (20-28)^2 + (1126-1860)^2 + (12894-179679)^2} = \sqrt{27817780994} = 166.786,633$	$D10m2 = \sqrt{(18-33)^2 + (25-18)^2 + (20-17)^2 + (1126-1166)^2 + (12894-27438)^2} = \sqrt{211529819} = 14.544,064$	$D10m3 = \sqrt{(18-20)^2 + (25-26)^2 + (20-21)^2 + (1126-1189)^2 + (12894-13132)^2} = \sqrt{60619} = 246,209$	246,209	3
11	$D11m1 = \sqrt{(32-93)^2 + (34-7)^2 + (23-28)^2 + (1487-1860)^2 + (18863-179679)^2} = \sqrt{25861929460} = 160.816,446$	$D11m2 = \sqrt{(32-33)^2 + (34-18)^2 + (23-17)^2 + (1487-1166)^2 + (18863-27438)^2} = \sqrt{73633959} = 8.581,023$	$D11m3 = \sqrt{(32-20)^2 + (34-26)^2 + (23-21)^2 + (1487-1189)^2 + (18863-13132)^2} = \sqrt{32933377} = 5.738,761$	5.738,761	3
12	$D12m1 = \sqrt{(50-93)^2 + (22-7)^2 + (17-28)^2 + (1884-1860)^2 + (39210-179679)^2} = \sqrt{19731542732} = 140.469,010$	$D12m2 = \sqrt{(50-33)^2 + (22-18)^2 + (17-17)^2 + (1884-1166)^2 + (39210-27438)^2} = \sqrt{139095813} = 11.793,889$	$D12m3 = \sqrt{(50-20)^2 + (22-26)^2 + (17-21)^2 + (1884-1189)^2 + (39210-13132)^2} = \sqrt{680546041} = 26.087,277$	11.793,889	2
13	0	$D13m2 = \sqrt{(93-33)^2 + (7-18)^2 + (28-17)^2 + (1860-1166)^2 + (179679-27438)^2} = \sqrt{23177807559} = 152.242,594$	$D13m3 = \sqrt{(93-20)^2 + (7-26)^2 + (28-21)^2 + (1860-1189)^2 + (179679-13132)^2} = \sqrt{27738359189} = 166.548,369$	0	1
14	$D14m1 = \sqrt{(33-93)^2 + (18-7)^2 + (17-28)^2 + (1166-1860)^2 + (27438-179679)^2} = \sqrt{23177807559} = 152.242,594$	0	$D14m3 = \sqrt{(33-20)^2 + (18-26)^2 + (17-21)^2 + (1166-1189)^2 + (27438-13132)^2} = \sqrt{204662414} = 14.306,027$	0	2
15	$D15m1 = \sqrt{(20-93)^2 + (26-7)^2 + (21-28)^2 + (1189-1860)^2 + (13132-179679)^2} = \sqrt{27738359189} = 166.548,369$	$D15m2 = \sqrt{(20-33)^2 + (26-18)^2 + (21-17)^2 + (1189-1166)^2 + (13132-27438)^2} = \sqrt{204662414} = 14.306,027$	0	0	3
Jumlah	1.889.461,223	540.459,021	610.602,248		
Total	3.040.522,492				

- Langkah kelima adalah menghitung total simpangan

$$S = B - A$$

$$S = 3.040.522,492 - 2.887.177,027$$

$$S = 153.345,465$$

Karena $B > A$ maka iterasi di hentikan sehingga, anggota cluster yang terbentuk pada masing masing medoid itu adalah :

Cluster 1 = Puskesmas Bogor Utara

Cluster 2 = Puskesmas Bogor Timur (Baranang Siang & Katulampa) dan Kedung Badak

Cluster 3 = Puskesmas Mulyaharja dan Kayu Manis

2. Klasifikasi

- **Langkah 1** : Baca Data Training

Data set diambil dari data Dinas Kesehatan pada tahun 2019 – 2021 pada tabel

Data Training

Puskesmas	Kriteria Tahun 2019					Kriteria Tahun 2020					Kriteria Tahun 2021				
	Jenis Penyakit					Jenis Penyakit					Jenis Penyakit				
	Tubercu losis	D B D	Pne umonia	Diare	Ju ml ah Pe nd ud uk	Tubercu losis	D B D	Pne umonia	Diare	Ju ml ah Pe nd ud uk	Tubercu losis	D B D	Pne umonia	Diare	Ju ml ah Pe nd ud uk
Mulyaharja	30	29	20	1330	17921	29	31	19	1423	18355	32	34	23	1487	18863
Bogor Timur (Baranang Siang & Katulampa)	42	17	15	1709	36069	46	19	15	1793	37127	50	22	17	1884	39210
Bogor Utara	74	6	25	1747	168894	87	8	26	1826	170920	93	7	28	1860	179679
Kedung Badak	14	14	13	1036	25949	30	16	15	1102	26809	33	18	17	1253	27438
Kayu Manis	17	23	18	1085	12477	18	25	20	1126	12894	20	26	21	1189	13132

- **Langkah 2** mencari nilai *Mean* dan Standar Deviasi dari data training

a. Nilai *Mean*

$$\mu(\text{Tinggi|Tuberculosis}) = (30+17+29+18+32+20)/(3)/(2) = 24,333$$

$$\mu(\text{Rendah|Tuberculosis}) = (74+87+94)/(3) = 85$$

$$\mu(\text{Sedang|Tuberculosis}) = (42+14+46+30+50+33)/(3)/(2) = 35,833$$

$$\mu(\text{Tinggi|DBD}) = (29+23+31+25+34+26)/(3)/(2) = 28$$

$$\mu(\text{Rendah|DBD}) = (6+8+7)/(3) = 7$$

$$\mu(\text{Sedang|DBD}) = (17+14+19+16+22+18)/(3)/(2) = 17,667$$

$$\mu(\text{Tinggi|Pneumonia}) = (20+18+19+20+23+21)/(3)/(2) = 20,167$$

$$\mu(\text{Rendah|Pneumonia}) = (25+26+28)/(3) = 26,333$$

$$\mu(\text{Sedang|Pneumonia}) = (15+13+15+15+22+18)/(3)/(2) = 16,333$$

$$\mu(\text{Tinggi|Diare}) = (1330+1085+1423+1126+1487+1189)/(3)/(2) = 1273,333$$

$$\mu(\text{Rendah|Diare}) = (1747+1826+1860)/(3) = 1811$$

$$\mu(\text{Sedang|Diare}) = (1709+1036+1793+1102+1884+1253)/(3)/(2) = 1462,833$$

$\mu(\text{Tinggi}|\text{Jumlah Penduduk})$

$$= (17921+12477+18355+12894+18863+13132)/(3)/(2) = 15607$$

$$\mu(\text{Rendah}|\text{Jumlah Penduduk}) = (168894+170920+179679)/(3) = 173164,333$$

$\mu(\text{Sedang}|\text{Jumlah Penduduk})$

$$= (36069+25949+37127+26809+39210+27438)/(3)/(2) = 32100,333$$

Berikut adalah nilai mean untuk seluruh kriteria, dapat dilihat pada tabel

Mean Seluruh Kriteria

Kelas	Jumlah	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Tinggi	2	24,333	28	20,167	1273,333	15607
Rendah	1	85	7	26,333	1811	173164,333
Sedang	2	35,833	17,667	16,333	1462,833	32100,333

b. Nilai Standar Deviasi

$$\begin{aligned} \sigma(\text{Tinggi} | \text{Tuberculosis}) &= \sqrt{((30,333 - 24,333)^2 + (18,333 - 24,333)^2) / 2} \\ &= \sqrt{72/2} = \sqrt{36} = 6 \end{aligned}$$

$$\sigma(\text{Rendah} | \text{Tuberculosis}) = \sqrt{(85 - 85)^2 / 1} = 0$$

$$\begin{aligned} \sigma(\text{Sedang} | \text{Tuberculosis}) &= \sqrt{((46 - 35,833)^2 + (25,667 - 35,833)^2) / 2} \\ &= \sqrt{206,715/2} = \sqrt{103,357} = 10,166 \end{aligned}$$

$$\begin{aligned} \sigma(\text{Tinggi} | \text{DBD}) &= \sqrt{((31,333 - 28)^2 + (24,667 - 28)^2) / 2} \\ &= \sqrt{22,218/2} = \sqrt{11,109} = 3,333 \end{aligned}$$

$$\sigma(\text{Rendah} | \text{DBD}) = \sqrt{(7 - 7)^2 / 1} = 0$$

$$\begin{aligned} \sigma(\text{Sedang} | \text{DBD}) &= \sqrt{((19,333 - 17,667)^2 + (16 - 17,667)^2) / 2} \\ &= \sqrt{5,554/2} = \sqrt{2,777} = 1,666 \end{aligned}$$

$$\begin{aligned} \sigma(\text{Tinggi} | \text{Pneumonia}) &= \sqrt{((20,667 - 20,167)^2 + (19,667 - 20,167)^2) / 2} \\ &= \sqrt{0,5/2} = \sqrt{0,25} = 0,5 \end{aligned}$$

$$\sigma(\text{Rendah} | \text{Pneumonia}) = \sqrt{(26,333 - 26,333)^2 / 1} = 0$$

$$\begin{aligned} \sigma(\text{Sedang} | \text{Pneumonia}) &= \sqrt{((17,333 - 16,333)^2 + (15,333 - 16,333)^2) / 2} \\ &= \sqrt{2/2} = \sqrt{1} = 1 \end{aligned}$$

$\sigma(\text{Tinggi} | \text{Diare})$

$$\begin{aligned} &= \sqrt{((1413,333 - 1273,333)^2 + (1133,333 - 1273,333)^2) / 2} \\ &= \sqrt{39200/2} = \sqrt{19600} = 140 \end{aligned}$$

$$\sigma(\text{Rendah} | \text{Diare}) = \sqrt{((1811 - 1811)^2) / 1} = 0$$

$\sigma(\text{Sedang} | \text{Diare})$

$$\begin{aligned} &= \sqrt{((1795,333 - 1462,833)^2 + (1130,333 - 1462,833)^2) / 2} \\ &= \sqrt{221112,5/2} = \sqrt{110556,25} = 332,5 \end{aligned}$$

$\sigma(\text{Tinggi} | \text{Jumlah Penduduk})$

$$= \sqrt{((18392 - 15607)^2 + (12804,5 - 15607)^2) / 2}$$

$$= \sqrt{15610231,25/2} = \sqrt{7854008,25} = 2802,5$$

$$\sigma(\text{Rendah} \mid \text{Jumlah Penduduk}) = \sqrt{((173164,333 - 173164,333)^2 / 1) = 0}$$

$$\sigma(\text{Sedang} \mid \text{Jumlah Penduduk})$$

$$= \sqrt{((37639,5 - 32100,333)^2 + (26693,5 - 32100,333)^2) / 2}$$

$$= \sqrt{59916214,144/2} = \sqrt{29958107,072} = 5473,4$$

Berikut adalah nilai standar deviasi dari seluruh kriteria, dapat dilihat pada tabel

Standar Deviasi Seluruh Kriteria

Prediksi	Jumlah	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Tinggi	2	6	3,333	0,5	140	2802,5
Rendah	1	0	0	0	0	0
Sedang	2	10,166	1,666	1	332,5	5473,4

- Langkah 3 : Mencari nilai *probabilitas*

Setelah menghitung nilai mean dan standar deviasi untuk data numerik, maka selanjutnya menentukan nilai

probabilitas setiap fitur pada setiap kelasnya.

$$P(\text{Tinggi}) = 2/5 = 0,4$$

$$P(\text{Rendah}) = 1/5 = 0,2$$

$$P(\text{Sedang}) = 2/5 = 0,4$$

- Langkah 4 : Nilai Distribusi Gaussian

Langkah selanjutnya adalah menghitung nilai probabilitas untuk Data Testing puskesmas lainnya yang diambil dari tahun 2019-2021. Dikarenakan semua data yang digunakan adalah data numerik, maka nilai probabilitas harus dihitung semua sesuai dengan Data Testing yang digunakan. Berikut adalah data Testing yang digunakan, dapat dilihat pada tabel

Data Testing

Puskesmas	Kriteria Tahun 2019					Kriteria Tahun 2020					Kriteria Tahun 2021					Kelas Infeksi
	Jenis Penyakit					Jenis Penyakit					Jenis Penyakit					
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	
Lawang Gintung	68	10	12	12	7029	72	11	14	13	7378	74	13	16	14	7728	?
Pulo Arman	28	22	26	44	40182	30	23	27	44	4079	32	25	29	50	39987	?
Tegal Gundil	36	16	21	13	40198	45	26	16	62	41029	48	27	22	17	41875	?

Untuk mengetahui nilai probabilitas data testing, maka harus mencari nilai distribusi gaussian terlebih dahulu. Berikut adalah proses menghitung nilai distribusi gaussian :

$$P(\text{TC Lawang} = 71,333 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi * 6}} * e^{-\frac{(71,333-24,333)^2}{(2*36)}}$$

$$= \frac{1}{15,036} * 2,718281^{-30,6806} = \frac{1}{15,036} * 0,00000000000000474 = 0,00000000000000315$$

$$P(\text{TC Lawang} = 71,333 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi * 0}} * e^{-\frac{(71,333-85)^2}{(2*0)}} = 0$$

$$P(\text{TC Lawang} = 71,333 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi * 10,166}} * e^{-\frac{(71,333-35,833)^2}{(2*103,357)}}$$

$$= \frac{1}{25,476} * 2,718281^{-6,097} = \frac{1}{25,476} * 0,002249611 = 0,0000883034$$

$$P(\text{DBD Lawang} = 11,333 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi * 3,333}} * e^{-\frac{(11,333- 28)^2}{(2*11,109)}}$$

$$= \frac{1}{8,3525} * 2,718281^{-12,503} = \frac{1}{8,3525} * 0,0000037155 = 0,0000004448$$

$$P(\text{DBD Lawang} = 11,333 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi * 0}} * e^{-\frac{(11,333- 7)^2}{(2*0)}} = 0$$

$$P(\text{DBD Lawang} = 11,333 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi * 1,666}} * e^{-\frac{(11,333- 17,667)^2}{(2*2,777)}}$$

$$= \frac{1}{4,175} * 2,718281^{-7,22354} = \frac{1}{4,175} * 0,000729218 = 0,000174663$$

$$P(\text{PN Lawang} = 14 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi * 0,5}} * e^{-\frac{(14 - 20,167)^2}{(2*0,25)}}$$

$$= \frac{1}{1,253} * 2,718281^{-76,064} = \frac{1}{1,253} * 0,000924$$

$$= 0,000737$$

$$P(\text{PN Lawang} = 14 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi * 0}} * e^{-\frac{(14- 26,333)^2}{(2*0)}} = 0$$

$$P(\text{PN Lawang} = 14 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi * 1}} * e^{-\frac{(14- 16,333)^2}{(2*1)}}$$

$$= \frac{1}{2,506} * 2,718281^{-2,7214445} = \frac{1}{2,506} * 0,065779722 = 0,026248891$$

$$P(\text{Diare Lawang} = 1375 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi * 140}} * e^{-\frac{(1375- 1273,333)^2}{(2*19600)}}$$

$$= \frac{1}{350,839} * 2,718281^{-0,263678033} = \frac{1}{350,839} * 0,768220903 = 0,002189668$$

$$P(\text{Diare Lawang} = 1375 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi * 0}} * e^{-\frac{(1375- 1811)^2}{(2*0)}} = 0$$

$$P(\text{Diare Lawang} = 1375 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi * 332,5}} * e^{-\frac{(1375- 1462,833)^2}{(2*110556,25)}}$$

$$= \frac{1}{833,243} * 2,718281^{-0,034890094} = \frac{1}{833,243} * 0,965711558 = 0,00115898$$

$$P(\text{JP Lawang} = 7377,667 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi * 2802,5}} * e^{-\frac{(7377,667-15607)^2}{(2*7854008,25)}}$$

$$= \frac{1}{7023,045} * 2,718281^{-4,311296823} = \frac{1}{7023,045} * 0,013416158 = 0,0000019103$$

$$\begin{aligned}
P(\text{JP Lawang} = 7377,667 \mid \text{Rendah}) &= \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(7377,667 - 173164,333)^2}{(2*0)}} = 0 \\
P(\text{JP Lawang} = 7377,667 \mid \text{Sedang}) &= \frac{1}{\sqrt{2\pi} * 5473,4} * e^{-\frac{(7377,667 - 32100,333)^2}{(2*29958107,072)}} \\
&= \frac{1}{13716,301} * 2,718281^{-10,20108201} = \frac{1}{13716,301} * 0,0000371302 = 0.000000002707 \\
P(\text{TC Pulo Armyn} = 30 \mid \text{Tinggi}) &= \frac{1}{\sqrt{2\pi} * 6} * e^{-\frac{(30-24,333)^2}{(2*36)}} \\
&= \frac{1}{15,036} * 2,718281^{-0,446040125} = \frac{1}{15,036} * 0,6401581722 = 0,042575031 \\
P(\text{TC Pulo Armyn} = 30 \mid \text{Rendah}) &= \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(30-85)^2}{(2*0)}} = 0 \\
P(\text{TC Pulo Armyn} = 30 \mid \text{Sedang}) &= \frac{1}{\sqrt{2\pi} * 10,166} * e^{-\frac{(30-35,833)^2}{(2*103,357)}} \\
&= \frac{1}{25,476} * 2,718281^{-0,1645940236} = \frac{1}{25,476} * 0,8482380413 = 0,033295574 \\
P(\text{DBD Armyn} = 23,333 \mid \text{Tinggi}) &= \frac{1}{\sqrt{2\pi} * 3,333} * e^{-\frac{(23,333 - 28)^2}{(2*11,109)}} \\
&= \frac{1}{8,3525} * 2,718281^{-0,9803262669} = \frac{1}{8,3525} * 0,3751887793 = 0,044919339 \\
P(\text{DBD Armyn} = 23,333 \mid \text{Rendah}) &= \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(23,333 - 7)^2}{(2*0)}} = 0 \\
P(\text{DBD Armyn} = 23,333 \mid \text{Sedang}) &= \frac{1}{\sqrt{2\pi} * 1,666} * e^{-\frac{(23,333 - 17,667)^2}{(2*2,777)}} \\
&= \frac{1}{4,175} * 2,718281^{-5,78025855239} = \frac{1}{4,175} * 0,0030879224 = 0,000739622 \\
P(\text{PN Armyn} = 27,333 \mid \text{Tinggi}) &= \frac{1}{\sqrt{2\pi} * 0,5} * e^{-\frac{(27,333 - 20,167)^2}{(2*0,25)}} \\
&= \frac{1}{1,253} * 2,718281^{-102,70311200} = \frac{1}{1,253} * 2,49241E-45 = 1,98915E-45 \\
P(\text{PN Armyn} = 27,333 \mid \text{Rendah}) &= \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(27,333 - 26,333)^2}{(2*0)}} = 0 \\
P(\text{PN Armyn} = 27,333 \mid \text{Sedang}) &= \frac{1}{\sqrt{2\pi} * 1} * e^{-\frac{(27,333 - 16,333)^2}{(2*1)}} \\
&= \frac{1}{350,839} * 2,718281^{-60,5} = \frac{1}{350,839} * 5,31119E-27 = 1,51385E-29 \\
P(\text{Diare Armyn} = 458 \mid \text{Tinggi}) &= \frac{1}{\sqrt{2\pi} * 140} * e^{-\frac{(458 - 1273,333)^2}{(2*19600)}} \\
&= \frac{1}{350,839} * 2,718281^{-16,9583648185} = \frac{1}{350,839} * 0,0000000432 = 0,00000000123018 \\
P(\text{Diare Armyn} = 458 \mid \text{Rendah}) &= \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(458 - 1811)^2}{(2*0)}} = 0 \\
P(\text{Diare Armyn} = 458 \mid \text{Sedang}) &= \frac{1}{\sqrt{2\pi} * 332,5} * e^{-\frac{(458 - 1462,833)^2}{(2*110556,25)}} \\
&= \frac{1}{833,243} * 2,718281^{-4,56640559845} = \frac{1}{833,243} * 0,0103952718 = 0,0000124757 \\
P(\text{JP Armyn} = 4026 \mid \text{Tinggi}) &= \frac{1}{\sqrt{2\pi} * 2802,5} * e^{-\frac{(4026 - 15607)^2}{(2*7854008,25)}}
\end{aligned}$$

$$= \frac{1}{7023,045} * 2,718281^{-8.5382875043} = \frac{1}{7023,045} * 0.0001958258 = 0,0000000278833$$

$$P(\text{JP Armyn} = 4026 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(4026 - 173164,333)^2}{(2*0)}} = 0$$

$$P(\text{JP Armyn} = 4026 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi} * 5473,4} * e^{-\frac{(4026 - 32100,333)^2}{(2*29958107,072)}}$$

$$= \frac{1}{13716,301} * 2,718281^{-13,1545055817} = \frac{1}{13716,301} * 0.0000019367 =$$

$$P(\text{TC Tegal Gundil} = 43 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi} * 6} * e^{-\frac{(43 - 24,333)^2}{(2*36)}}$$

$$= \frac{1}{15,036} * 2,718281^{-4.839679} = \frac{1}{15,036} * 0.0079096042 = 0,000526044$$

$$P(\text{TC Tegal Gundil} = 43 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(43 - 85)^2}{(2*0)}} = 0$$

$$P(\text{TC Tegal Gundil} = 43 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi} * 10,166} * e^{-\frac{(43 - 35,833)^2}{(2*103,357)}}$$

$$= \frac{1}{25,476} * 2,718281^{-0.248487712} = \frac{1}{25,476} * 0,7799795038 = 0,030616247$$

$$P(\text{DBD Tegal Gundil} = 23 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi} * 3,333} * e^{-\frac{(23 - 28)^2}{(2*11,109)}}$$

$$= \frac{1}{8,3525} * 2,718281^{-1.125213791} = \frac{1}{8,3525} * 0,3245831784 = 0,038860602$$

$$P(\text{DBD Tegal Gundil} = 23 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(23 - 7)^2}{(2*0)}} = 0$$

$$P(\text{DBD Tegal Gundil} = 23 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi} * 1,666} * e^{-\frac{(23 - 17,667)^2}{(2*2,777)}}$$

$$= \frac{1}{4,175} * 2,718281^{-5.120793842} = \frac{1}{4,175} * 0,0059712901 = 0,001430249$$

$$P(\text{PN Tegal Gundil} = 19,667 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi} * 0,5} * e^{-\frac{(19,667 - 20,167)^2}{(2*0,25)}}$$

$$= \frac{1}{1,253} * 2,718281^{-0.5} = \frac{1}{1,253} * 0,6065307521 = 0,484062851$$

$$P(\text{PN Tegal Gundil} = 19,667 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(19,667 - 26,333)^2}{(2*0)}} = 0$$

$$P(\text{PN Tegal Gundil} = 19,667 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi} * 1} * e^{-\frac{(19,667 - 16,333)^2}{(2*1)}}$$

$$= \frac{1}{2,506} * 2,718281^{-5.557778} = \frac{1}{2,506} * 0,0038573444 = 0,001539244$$

$$P(\text{Diare Tegal Gundil} = 1549,667 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi} * 140} * e^{-\frac{(1549,667 - 1273,333)^2}{(2*19600)}}$$

$$= \frac{1}{350,839} * 2,718281^{-1,947971417} = \frac{1}{350,839} * 0,1425630639 = 0,000406349$$

$$P(\text{Diare Tegal Gundil} = 1549,667 \mid \text{Rendah}) = \frac{1}{\sqrt{2\pi} * 0} * e^{-\frac{(1549,667 - 1811)^2}{(2*0)}} = 0$$

$$P(\text{Diare Tegal Gundil} = 1549,667 \mid \text{Sedang}) = \frac{1}{\sqrt{2\pi} * 332,5} * e^{-\frac{(1549,667 - 1462,833)^2}{(2*110556,25)}}$$

$$= \frac{1}{833,243} * 2,718281^{-0.034100938} = \frac{1}{833,243} * 0,9664739562 = 0,001159894$$

$$P(\text{JP Tegal Gundil} = 41033 \mid \text{Tinggi}) = \frac{1}{\sqrt{2\pi} * 2802,5} * e^{-\frac{(41033 - 15607)^2}{(2*7854008,25)}}$$

$$P(\text{Pulo Armyn}|\text{Tinggi}) = 0,042575031 * 0,044919339 * 1,98915E-45 * 0,000000000123018 * 0,0000000278833 = 1,30487E-65$$

$$P(\text{Pulo Armyn}|\text{Rendah}) = 0 * 0 * 0 * 0 = 0$$

$$P(\text{Pulo Armyn}|\text{Sedang}) = 0,033295574 * 0,000739622 * 1,51385E-29 * 0,0000124757 * 0,0000000001412 = 6.56718E-49$$

$$P(\text{Tegal Gundil}|\text{Tinggi}) = 0,000526044 * 0,038860602 * 0,484062851 * 0,000406349 * 0,00000000000000000000019 = 7.63987E-31$$

$$P(\text{Tegal Gundil}|\text{Rendah}) = 0 * 0 * 0 * 0 = 0$$

$$P(\text{Tegal Gundil}|\text{Sedang}) = 0.030616247 * 0.001430249 * 0.001539244 * 0.001159894 * 0.0000192 = 1.50103E-15$$

Hasil Probabilitas akhir setiap kelas untuk seluruh puskesmas dapat dilihat pada tabel

Hasil Probabilitas Akhir

No	Puskesmas	Tinggi	Rendah	Sedang	Kelas Infeksi
1	Lawang	4,31939E-63	0	1.27015E-21	Sedang
2	Pulo Armyn	1,30487E-65	0	6.56718E-49	Sedang
3	Tegal Gundil	7.63987E-31	0	1.50103E-15	Sedang

1. Prediksi

- **Langkah 1** : Menghitung $P(C_i)$ untuk setiap puskesmas

Menghitung $P(C_i)$ tiap penyakit dan jumlah penduduk dari tiap puskesmas pada tahun 2019.

$$P(C_i) = \text{Total Tuberculosis/Kriteria}$$

$$P(C_i) = 177/5 = 35,4$$

$$P(\text{Puskesmas Mulyaharja}) = 30/19330 = 0,001552$$

$$P(\text{Puskesmas Bogor Timur}) = 42/37825 = 0,001110$$

$$P(\text{Puskesmas Bogor Utara}) = 74/170746 = 0,000433$$

$$P(\text{Puskesmas Kedung Badak}) = 14/27026 = 0,000518$$

$$P(\text{Puskesmas Kayu Manis}) = 17/13620 = 0,001248$$

$$P(C_i) = \text{Total DBD/Kriteria}$$

$$P(C_i) = 89/5 = 17,8$$

$$P(\text{Puskesmas Mulyaharja}) = 29/19330 = 0,001500$$

$$P(\text{Puskesmas Bogor Timur}) = 17/37852 = 0,000449$$

$$P(\text{Puskesmas Bogor Utara}) = 6/170746 = 0,000035$$

$$P(\text{Puskesmas Kedung Badak}) = 14/27026 = 0,000518$$

$$P(\text{Puskesmas Kayu Manis}) = 23/13620 = 0,001688$$

$$P(C_i) = \text{Total Pneumonia/Kriteria}$$

$$P(C_i) = 91/5 = 18,2$$

$$P(\text{Puskesmas Mulyaharja}) = 20/19330 = 0,001034$$

$$P(\text{Puskesmas Bogor Timur}) = 15/37852 = 0,000396$$

$$P(\text{Puskesmas Bogor Utara}) = 25/170746 = 0,000146$$

$$P(\text{Puskesmas Kedung Badak}) = 13/27026 = 0,000481$$

$$P(\text{Puskesmas Kayu Manis}) = 18/13620 = 0,001321$$

$$P(C_i) = \text{Total Diare/Kriteria}$$

$$P(C_i) = 6907/5 = 1381,4$$

$$P(\text{Puskesmas Mulyaharja}) = 1330/19330 = 0,068805$$

$$P(\text{Puskesmas Bogor Timur}) = 1709/37852 = 0,045149$$

$$P(\text{Puskesmas Bogor Utara}) = 1747/170746 = 0,010231$$

$$P(\text{Puskesmas Kedung Badak}) = 1036/27026 = 0,038333$$

$$P(\text{Puskesmas Kayu Manis}) = 1085/13620 = 0,079662$$

$$P(C_i) = \text{Total Jumlah Penduduk/Kriteria}$$

$$P(C_i) = 261310/5 = 52262$$

$$P(\text{Puskesmas Mulyaharja}) = 17921/19330 = 0,927108$$

$$P(\text{Puskesmas Bogor Timur}) = 36069/37852 = 0,952895$$

$$P(\text{Puskesmas Bogor Utara}) = 168894/170746 = 0,989153$$

$$P(\text{Puskesmas Kedung Badak}) = 25949/27026 = 0,960149$$

$$P(\text{Puskesmas Kayu Manis}) = 12477/13620 = 0,916079$$

Menghitung $P(C_i)$ tiap penyakit dari tiap puskesmas pada tahun 2020.

$$P(C_i) = \text{Total Tuberculosis/Kriteria}$$

$$P(C_i) = 210/5 = 42$$

$$P(\text{Puskesmas Mulyaharja}) = 29/19857 = 0,001460$$

$$P(\text{Puskesmas Bogor Timur}) = 46/39000 = 0,001179$$

$$P(\text{Puskesmas Bogor Utara}) = 87/172867 = 0,000503$$

$$P(\text{Puskesmas Kedung Badak}) = 30/27972 = 0,001072$$

$$P(\text{Puskesmas Kayu Manis}) = 18/14083 = 0,001278$$

$$P(C_i) = \text{Total DBD/Kriteria}$$

$$P(C_i) = 99/5 = 19,8$$

$$P(\text{Puskesmas Mulyaharja}) = 31/19857 = 0,001561$$

$$P(\text{Puskesmas Bogor Timur}) = 19/39000 = 0,000487$$

$$P(\text{Puskesmas Bogor Utara}) = 8/172867 = 0,000046$$

$$P(\text{Puskesmas Kedung Badak}) = 16/27972 = 0,000572$$

$$P(\text{Puskesmas Kayu Manis}) = 25/14083 = 0,001775$$

$P(C_i) = \text{Total Pneumonia/Kriteria}$

$$P(C_i) = 95/5 = 19$$

$$P(\text{Puskesmas Mulyaharja}) = 19/19857 = 0,000957$$

$$P(\text{Puskesmas Bogor Timur}) = 15/39000 = 0,000385$$

$$P(\text{Puskesmas Bogor Utara}) = 26/172867 = 0,000150$$

$$P(\text{Puskesmas Kedung Badak}) = 15/27972 = 0,000536$$

$$P(\text{Puskesmas Kayu Manis}) = 20/14083 = 0,001420$$

$P(C_i) = \text{Total Diare/Kriteria}$

$$P(C_i) = 7270/5 = 1454$$

$$P(\text{Puskesmas Mulyaharja}) = 1423/19857 = 0,071662$$

$$P(\text{Puskesmas Bogor Timur}) = 1793/39000 = 0,045974$$

$$P(\text{Puskesmas Bogor Utara}) = 1826/172867 = 0,010563$$

$$P(\text{Puskesmas Kedung Badak}) = 1102/27972 = 0,039396$$

$$P(\text{Puskesmas Kayu Manis}) = 1126/14083 = 0,079954$$

$P(C_i) = \text{Total Jumlah Penduduk/Kriteria}$

$$P(C_i) = 266105/5 = 53221$$

$$P(\text{Puskesmas Mulyaharja}) = 18355/19857 = 0,924359$$

$$P(\text{Puskesmas Bogor Timur}) = 37127/39000 = 0,951974$$

$$P(\text{Puskesmas Bogor Utara}) = 170920/172867 = 0,988737$$

$$P(\text{Puskesmas Kedung Badak}) = 26809/27972 = 0,958422$$

$$P(\text{Puskesmas Kayu Manis}) = 12894/14083 = 0,915572$$

Menghitung $P(C_i)$ tiap penyakit dari tiap puskesmas pada tahun 2021.

$P(C_i) = \text{Total Tuberculosis/Kriteria}$

$$P(C_i) = 228/5 = 45,6$$

$$P(\text{Puskesmas Mulyaharja}) = 32/20439 = 0,001566$$

$$P(\text{Puskesmas Bogor Timur}) = 50/41183 = 0,001214$$

$$P(\text{Puskesmas Bogor Utara}) = 93/181667 = 0,000512$$

$$P(\text{Puskesmas Kedung Badak}) = 33/28759 = 0,001147$$

$$P(\text{Puskesmas Kayu Manis}) = 20/14388 = 0,001390$$

$P(C_i) = \text{Total DBD/Kriteria}$

$$P(C_i) = 107/5 = 21,4$$

$$P(\text{Puskesmas Mulyaharja}) = 34/20439 = 0,001663$$

$$P(\text{Puskesmas Bogor Timur}) = 22/41183 = 0,000534$$

$$P(\text{Puskesmas Bogor Utara}) = 7/181667 = 0,000038$$

$$P(\text{Puskesmas Kedung Badak}) = 18/28759 = 0,000626$$

$$P(\text{Puskesmas Kayu Manis}) = 26/14388 = 0,001807$$

$$P(C_i) = \text{Total Pneumonia/Kriteria}$$

$$P(C_i) = 106/5 = 21,2$$

$$P(\text{Puskesmas Mulyaharja}) = 23/20439 = 0,001125$$

$$P(\text{Puskesmas Bogor Timur}) = 17/41183 = 0,000413$$

$$P(\text{Puskesmas Bogor Utara}) = 28/181667 = 0,000154$$

$$P(\text{Puskesmas Kedung Badak}) = 17/28759 = 0,000591$$

$$P(\text{Puskesmas Kayu Manis}) = 21/14388 = 0,001459$$

$$P(C_i) = \text{Total Diare/Kriteria}$$

$$P(C_i) = 7673/5 = 1534,6$$

$$P(\text{Puskesmas Mulyaharja}) = 1487/20439 = 0,072753$$

$$P(\text{Puskesmas Bogor Timur}) = 1884/41183 = 0,045747$$

$$P(\text{Puskesmas Bogor Utara}) = 1860/181667 = 0,010238$$

$$P(\text{Puskesmas Kedung Badak}) = 1253/28759 = 0,043569$$

$$P(\text{Puskesmas Kayu Manis}) = 1189/14388 = 0,082638$$

$$P(C_i) = \text{Total Jumlah Penduduk/Kriteria}$$

$$P(C_i) = 278322/5 = 55664,4$$

$$P(\text{Puskesmas Mulyaharja}) = 18863/20439 = 0,922892$$

$$P(\text{Puskesmas Bogor Timur}) = 39210/41183 = 0,952092$$

$$P(\text{Puskesmas Bogor Utara}) = 179679/181667 = 0,989057$$

$$P(\text{Puskesmas Kedung Badak}) = 27438/28759 = 0,954066$$

$$P(\text{Puskesmas Kayu Manis}) = 13132/14388 = 0,912705$$

Untuk memudahkan dalam pembacaan data, nilai distribusi probabilitas yang telah didapatkan dari pencarian di atas selanjutnya akan ditampilkan ke dalam bentuk tabel seperti yang terlihat pada tabel

Nilai Distribusi Probabilitas

Puskesmas	Kriteria Tahun 2019					Kriteria Tahun 2020					Kriteria Tahun 2021				
	Jenis Penyakit					Jenis Penyakit					Jenis Penyakit				
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk
Mulyaharja	0,01552	0,0000	0,0014	0,00605	0,9278	0,001460	0,00061	0,00097	0,00162	0,9359	0,001566	0,000635	0,001253	0,00753	0,922892

Bogor Timur (Baranang Siang & Katulampa)	0,0 011 10	0, 00 04	0,0 00 39	0, 04 51	0,9 52 89	0,0 011 79	0, 00 04	0,0 00 38	0, 04 59	0,9 51 97	0,0 012 14	0, 00 05	0,0 00 41	0, 04 57	0,9 52 09
Bogor Utara	0,0 004 33	0, 00 00	0,0 00 14	0, 01 02	0,9 89 15	0,0 005 03	0, 00 00	0,0 00 15	0, 01 05	0,9 88 73	0,0 005 12	0, 00 00	0,0 00 15	0, 01 02	0,9 89 05
Kedung Badak	0,0 005 18	0, 00 05	0,0 00 48	0, 03 83	0,9 60 14	0,0 010 72	0, 00 05	0,0 00 53	0, 03 93	0,9 58 42	0,0 011 47	0, 00 06	0,0 00 59	0, 04 35	0,9 54 06
Kayu Manis	0,0 012 48	0, 00 16	0,0 01 32	0, 07 96	0,9 16 07	0,0 012 78	0, 00 17	0,0 01 42	0, 07 99	0,9 15 57	0,0 013 90	0, 00 18	0,0 01 45	0, 08 26	0,9 12 70

- Langkah 3 : Menghitung $P(X|Ci)$ untuk setiap kriteria dan kelas

Untuk mempermudah proses perhitungan maka data penyakit diberi singkatan TB, DBD, PN, DI dan untuk jumlah penduduk yaitu JP. Untuk puskesmas diberi singkatan MU (Mulyaharja), BT (Bogor Timur), BU (Bogor Utara), KB (Kedung Badak), KM (Kayu Manis).

Menghitung nilai $P(X|Ci)$ penyakit pada tahun 2019 :

1. Penyakit Tuberculosis :

$$\begin{aligned}
 (TB|MU) &= \frac{TB \times MU}{(TB \times MU) + (TB \times BT) + (TB \times BU) + (TB \times KB) + (TB \times KM)} \\
 &= \frac{35,4 \times 0,001552}{(35,4 \times 0,001552) + (35,4 \times 0,001110) + (35,4 \times 0,000433) + (35,4 \times 0,000518) + (35,4 \times 0,001248)} \\
 &= \frac{0,054941}{0,172079} = 0,319278
 \end{aligned}$$

$$\begin{aligned}
 (TB|BT) &= \frac{TB \times BT}{(TB \times MU) + (TB \times BT) + (TB \times BU) + (TB \times KB) + (TB \times KM)} \\
 &= \frac{35,4 \times 0,001110}{(35,4 \times 0,001552) + (35,4 \times 0,001110) + (35,4 \times 0,000433) + (35,4 \times 0,000518) + (35,4 \times 0,001248)} \\
 &= \frac{0,039294}{0,172079} = 0,228348
 \end{aligned}$$

$$\begin{aligned}
 (TB|BU) &= \frac{TB \times BU}{(TB \times MU) + (TB \times BT) + (TB \times BU) + (TB \times KB) + (TB \times KM)} \\
 &= \frac{35,4 \times 0,000433}{(35,4 \times 0,001552) + (35,4 \times 0,001110) + (35,4 \times 0,000433) + (35,4 \times 0,000518) + (35,4 \times 0,001248)} \\
 &= \frac{0,015328}{0,172079} = 0,089075
 \end{aligned}$$

$$\begin{aligned}
 (TB|KB) &= \frac{TB \times KB}{(TB \times MU) + (TB \times BT) + (TB \times BU) + (TB \times KB) + (TB \times KM)} \\
 &= \frac{35,4 \times 0,000518}{(35,4 \times 0,001552) + (35,4 \times 0,001110) + (35,4 \times 0,000433) + (35,4 \times 0,000518) + (35,4 \times 0,001248)} \\
 &= \frac{0,018337}{0,172079} = 0,106562
 \end{aligned}$$

$$(TB|KM) = \frac{TB \times KM}{(TB \times MU) + (TB \times BT) + (TB \times BU) + (TB \times KB) + (TB \times KM)}$$

$$= \frac{35,4 \times 0,001248}{(35,4 \times 0,001552) + (35,4 \times 0,001110) + (35,4 \times 0,000433) + (35,4 \times 0,000518) + (35,4 \times 0,001248)}$$

$$= \frac{0,044179}{0,172079} = 0,256738$$

9. Penyakit DBD :

(DBD|MU)

$$= \frac{17,8 \times 0,001500}{(17,8 \times 0,001500) + (17,8 \times 0,000449) + (17,8 \times 0,000035) + (17,8 \times 0,000518) + (17,8 \times 0,001688)}$$

$$= \frac{0,0267}{0,074582} = 0,357995$$

(DBD|BT)

$$= \frac{17,8 \times 0,000449}{(17,8 \times 0,001500) + (17,8 \times 0,000449) + (17,8 \times 0,000035) + (17,8 \times 0,000518) + (17,8 \times 0,001688)}$$

$$= \frac{0,007992}{0,074582} = 0,107159$$

(DBD|BU)

$$= \frac{17,8 \times 0,000035}{(17,8 \times 0,001500) + (17,8 \times 0,000449) + (17,8 \times 0,000035) + (17,8 \times 0,000518) + (17,8 \times 0,001688)}$$

$$= \frac{0,000623}{0,074582} = 0,008353$$

(DBD|KB)

$$= \frac{17,8 \times 0,000518}{(17,8 \times 0,001500) + (17,8 \times 0,000449) + (17,8 \times 0,000035) + (17,8 \times 0,000518) + (17,8 \times 0,001688)}$$

$$= \frac{0,009220}{0,074582} = 0,123628$$

(DBD|KM)

$$= \frac{17,8 \times 0,001688}{(17,8 \times 0,001500) + (17,8 \times 0,000449) + (17,8 \times 0,000035) + (17,8 \times 0,000518) + (17,8 \times 0,001688)}$$

$$= \frac{0,030046}{0,074582} = 0,402858$$

10. Penyakit Pneumonia :

$$(PN|MU) = \frac{18,2 \times 0,001034}{(18,2 \times 0,001034) + (18,2 \times 0,000396) + (18,2 \times 0,000146) + (18,2 \times 0,000481) + (18,2 \times 0,001321)}$$

$$= \frac{0,018819}{0,074582} = 0,252326$$

$$(PN|BT) = \frac{18,2 \times 0,000396}{(18,2 \times 0,001034) + (18,2 \times 0,000396) + (18,2 \times 0,000146) + (18,2 \times 0,000481) + (18,2 \times 0,001321)}$$

$$= \frac{0,007207}{0,074582} = 0,096632$$

$$(PN|BU) = \frac{18,2 \times 0,000146}{(18,2 \times 0,001034) + (18,2 \times 0,000396) + (18,2 \times 0,000146) + (18,2 \times 0,000481) + (18,2 \times 0,001321)}$$

$$= \frac{0,002657}{0,074582} = 0,035625$$

$$(PN|KB) = \frac{18,2 \times 0,000481}{(18,2 \times 0,001034) + (18,2 \times 0,000396) + (18,2 \times 0,000146) + (18,2 \times 0,000481) + (18,2 \times 0,001321)}$$

$$= \frac{0,008754}{0,074582} = 0,117374$$

$$(PN|KM) = \frac{18,2 \times 0,001321}{(18,2 \times 0,001034) + (18,2 \times 0,000396) + (18,2 \times 0,000146) + (18,2 \times 0,000481) + (18,2 \times 0,001321)}$$

$$= \frac{0,024042}{0,074582} = 0,322357$$

11. Penyakit Diare :

(DI|MU)

$$= \frac{1381,4 \times 0,068805}{(1381,4 \times 0,068805) + (1381,4 \times 0,045149) + (1381,4 \times 0,010231) + (1381,4 \times 0,038333) + (1381,4 \times 0,079662)}$$

$$= \frac{95,047227}{334,547452} = 0,284107$$

(DI|BT)

$$= \frac{1381,4 \times 0,045149}{(1381,4 \times 0,068805) + (1381,4 \times 0,045149) + (1381,4 \times 0,010231) + (1381,4 \times 0,038333) + (1381,4 \times 0,079662)}$$

$$= \frac{62,368829}{334,547452} = 0,186427$$

(DI|BU)

$$= \frac{1381,4 \times 0,010231}{(1381,4 \times 0,068805) + (1381,4 \times 0,045149) + (1381,4 \times 0,010231) + (1381,4 \times 0,038333) + (1381,4 \times 0,079662)}$$

$$= \frac{14,133103}{334,547452} = 0,042245$$

(DI|KB)

$$= \frac{1381,4 \times 0,038333}{(1381,4 \times 0,068805) + (1381,4 \times 0,045149) + (1381,4 \times 0,010231) + (1381,4 \times 0,038333) + (1381,4 \times 0,079662)}$$

$$= \frac{52,953206}{334,547452} = 0,158283$$

(DI|KM)

$$= \frac{1381,4 \times 0,079662}{(1381,4 \times 0,068805) + (1381,4 \times 0,045149) + (1381,4 \times 0,010231) + (1381,4 \times 0,038333) + (1381,4 \times 0,079662)}$$

$$= \frac{110,045087}{334,547452} = 0,328937$$

12. Jumlah Penduduk :

(JP|MU)

$$= \frac{52262 \times 0,927108}{(52262 \times 0,927108) + (52262 \times 0,952895) + (52262 \times 0,989153) + (52262 \times 0,960149) + (52262 \times 0,916079)}$$

$$= \frac{48452,518296}{248003,258608} = 0,195370$$

(JP|BT)

$$= \frac{52262 \times 0,952895}{(52262 \times 0,927108) + (52262 \times 0,952895) + (52262 \times 0,989153) + (52262 \times 0,960149) + (52262 \times 0,916079)}$$

$$= \frac{49800,19849}{248003,258608} = 0,200805$$

(JP|BU)

$$= \frac{52262 \times 0,989153}{(52262 \times 0,927108) + (52262 \times 0,952895) + (52262 \times 0,989153) + (52262 \times 0,960149) + (52262 \times 0,916079)}$$

$$= \frac{51695,114086}{248003,258608} = 0,208445$$

(JP|KB)

$$= \frac{52262 \times 0,960149}{(52262 \times 0,927108) + (52262 \times 0,952895) + (52262 \times 0,989153) + (52262 \times 0,960149) + (52262 \times 0,916079)}$$

$$= \frac{50179,307038}{248003,258608} = 0,202333$$

(JP|KM)

$$= \frac{52262 \times 0,916079}{(52262 \times 0,927108) + (52262 \times 0,952895) + (52262 \times 0,989153) + (52262 \times 0,960149) + (52262 \times 0,916079)}$$

$$= \frac{47876,120698}{248003,258608} = 0,193046$$

Menghitung nilai P(X|Ci) penyakit pada tahun 2020 :

1. Penyakit Tuberculosis :

$$(TB|MU) = \frac{42 \times 0,001460}{(42 \times 0,001460) + (42 \times 0,001179) + (42 \times 0,000503) + (42 \times 0,001072) + (42 \times 0,001278)}$$

$$= \frac{0,06132}{0,230664} = 0,265841$$

$$(TB|BT) = \frac{42 \times 0,001179}{(42 \times 0,001460) + (42 \times 0,001179) + (42 \times 0,000503) + (42 \times 0,001072) + (42 \times 0,001278)}$$

$$= \frac{0,049518}{0,230664} = 0,214676$$

$$(TB|BU) = \frac{42 \times 0,000503}{(42 \times 0,001460) + (42 \times 0,001179) + (42 \times 0,000503) + (42 \times 0,001072) + (42 \times 0,001278)}$$

$$= \frac{0,021126}{0,230664} = 0,091588$$

$$(TB|KB) = \frac{42 \times 0,001072}{(42 \times 0,001460) + (42 \times 0,001179) + (42 \times 0,000503) + (42 \times 0,001072) + (42 \times 0,001278)}$$

$$= \frac{0,045024}{0,230664} = 0,195193$$

$$(TB|KM) = \frac{42 \times 0,001278}{(42 \times 0,001460) + (42 \times 0,001179) + (42 \times 0,000503) + (42 \times 0,001072) + (42 \times 0,001278)}$$

$$= \frac{0,053676}{0,230664} = 0,232702$$

2. Penyakit DBD :

$$(DBD|MU) = \frac{19,8 \times 0,001561}{(19,8 \times 0,001561) + (19,8 \times 0,000487) + (19,8 \times 0,000046) + (19,8 \times 0,000572) + (19,8 \times 0,001775)}$$

$$= \frac{0,030908}{0,087932} = 0,351499$$

$$(DBD|BT) = \frac{19,8 \times 0,000487}{(19,8 \times 0,001561) + (19,8 \times 0,000487) + (19,8 \times 0,000046) + (19,8 \times 0,000572) + (19,8 \times 0,001775)}$$

$$= \frac{0,0096426}{0,087932} = 0,109659$$

$$(DBD|BU) = \frac{19,8 \times 0,000046}{(19,8 \times 0,001561) + (19,8 \times 0,000487) + (19,8 \times 0,000046) + (19,8 \times 0,000572) + (19,8 \times 0,001775)}$$

$$= \frac{0,0009108}{0,087932} = 0,010358$$

$$(DBD|KB) = \frac{19,8 \times 0,000572}{(19,8 \times 0,001561) + (19,8 \times 0,000487) + (19,8 \times 0,000046) + (19,8 \times 0,000572) + (19,8 \times 0,001775)}$$

$$= \frac{0,0113256}{0,087932} = 0,128799$$

(DBD|KM)

$$= \frac{19,8 \times 0,001775}{(19,8 \times 0,001561) + (19,8 \times 0,000487) + (19,8 \times 0,000046) + (19,8 \times 0,000572) + (19,8 \times 0,001775)}$$

$$= \frac{0,035145}{0,087932} = 0,399684$$

3. Penyakit Pneumonia :

$$\begin{aligned} (PN|MU) &= \frac{19x0,000957}{(19x0,000957)+(19x0,000385)+(19x0,000150)+(19x0,000536)+(19x0,001420)} \\ &= \frac{0,018183}{0,065512} = 0,277552 \end{aligned}$$

$$\begin{aligned} (PN|BT) &= \frac{19x0,000385}{(19x0,000957)+(19x0,000385)+(19x0,000150)+(19x0,000536)+(19x0,001420)} \\ &= \frac{0,007315}{0,065512} = 0,111659 \end{aligned}$$

$$\begin{aligned} (PN|BU) &= \frac{19x0,000150}{(19x0,000957)+(19x0,000385)+(19x0,000150)+(19x0,000536)+(19x0,001420)} \\ &= \frac{0,00285}{0,065512} = 0,043503 \end{aligned}$$

$$\begin{aligned} (PN|KB) &= \frac{19x0,000536}{(19x0,000957)+(19x0,000385)+(19x0,000150)+(19x0,000536)+(19x0,001420)} \\ &= \frac{0,010184}{0,065512} = 0,155452 \end{aligned}$$

$$\begin{aligned} (PN|KM) &= \frac{19x0,001420}{(19x0,000957)+(19x0,000385)+(19x0,000150)+(19x0,000536)+(19x0,001420)} \\ &= \frac{0,02698}{0,065512} = 0,264151 \end{aligned}$$

4. Penyakit Diare :

$$\begin{aligned} (DI|MU) &= \frac{1454x0,071662}{(1454x0,071662)+(1454x0,045974)+(1454x0,010563)+(1454x0,039396)+(1454x0,079954)} \\ &= \frac{104,196548}{359,936246} = 0,289486 \end{aligned}$$

$$\begin{aligned} (DI|BT) &= \frac{1454x0,045974}{(1454x0,071662)+(1454x0,045974)+(1454x0,010563)+(1454x0,039396)+(1454x0,079954)} \\ &= \frac{66,846196}{359,936246} = 0,185717 \end{aligned}$$

$$\begin{aligned} (DI|BU) &= \frac{1454x0,010563}{(1454x0,071662)+(1454x0,045974)+(1454x0,010563)+(1454x0,039396)+(1454x0,079954)} \\ &= \frac{15,358602}{359,936246} = 0,042670 \end{aligned}$$

$$\begin{aligned} (DI|KB) &= \frac{1454x0,039396}{(1454x0,071662)+(1454x0,045974)+(1454x0,010563)+(1454x0,039396)+(1454x0,079954)} \\ &= \frac{57,281784}{359,936246} = 0,159144 \end{aligned}$$

$$\begin{aligned} (DI|KM) &= \frac{1454x0,079954}{(1454x0,071662)+(1454x0,045974)+(1454x0,010563)+(1454x0,039396)+(1454x0,079954)} \\ &= \frac{116,253116}{359,936246} = 0,322982 \end{aligned}$$

5, Jumlah Penduduk :

(JP|MU)

$$= \frac{53221x0,924359}{(53221x0,924359)+(53221x0,951974)+(53221x0,988737)+(53221x0,958422)+(53221x0,915572)}$$

$$= \frac{49195,31034}{252217,7251} = 0,195051$$

(JP|BT)

$$= \frac{53221x0,951974}{(53221x0,924359)+(53221x0,951974)+(53221x0,988737)+(53221x0,958422)+(53221x0,915572)}$$

$$= \frac{50665,00825}{252217,7251} = 0,200878$$

(JP|BU)

$$= \frac{53221x0,988737}{(53221x0,924359)+(53221x0,951974)+(53221x0,988737)+(53221x0,958422)+(53221x0,915572)}$$

$$= \frac{52621,57188}{252217,7251} = 0,208635$$

(JP|KB)

$$= \frac{53221x0,958422}{(53221x0,924359)+(53221x0,951974)+(53221x0,988737)+(53221x0,958422)+(53221x0,915572)}$$

$$= \frac{51008,17726}{252217,7251} = 0,202239$$

(JP|KM)

$$= \frac{53221x0,915572}{(53221x0,924359)+(53221x0,951974)+(53221x0,988737)+(53221x0,958422)+(53221x0,915572)}$$

$$= \frac{48727,65741}{252217,7251} = 0,193197$$

Menghitung nilai P(X|Ci) penyakit pada tahun 2021 :

1. Penyakit Tuberculosis :

$$(TB|MU) = \frac{45,6x0,001566}{(45,6x0,001566)+(45,6x0,001214)+(45,6x0,000512)+(45,6x0,001147)+(45,6x0,001390)}$$

$$= \frac{0,0714096}{0,2658024} = 0,268657$$

$$(TB|BT) = \frac{45,6x0,001214}{(45,6x0,001566)+(45,6x0,001214)+(45,6x0,000512)+(45,6x0,001147)+(45,6x0,001390)}$$

$$= \frac{0,055358}{0,2658024} = 0,208267$$

$$(TB|BU) = \frac{45,6x0,000512}{(45,6x0,001566)+(45,6x0,001214)+(45,6x0,000512)+(45,6x0,001147)+(45,6x0,001390)}$$

$$= \frac{0,0233472}{0,2658024} = 0,087837$$

$$(TB|KB) = \frac{45,6x0,001147}{(45,6x0,001566)+(45,6x0,001214)+(45,6x0,000512)+(45,6x0,001147)+(45,6x0,001390)}$$

$$= \frac{0,0523032}{0,2658024} = 0,196775$$

$$(TB|KM) = \frac{45,6x0,001390}{(45,6x0,001566)+(45,6x0,001214)+(45,6x0,000512)+(45,6x0,001147)+(45,6x0,001390)}$$

$$= \frac{0,063384}{0,2658024} = 0,238463$$

2. Penyakit DBD :

(DBD|MU)

$$\begin{aligned} &= \frac{21,4x0,001663}{(21,4x0,001663)+(21,4x0,000534)+(21,4x0,000038)+(21,4x0,000626)+(21,4x0,001807)} \\ &= \frac{0,0355882}{0,0998952} = 0,356255 \end{aligned}$$

$$\begin{aligned} \text{(DBD|BT)} &= \frac{21,4x0,000534}{(21,4x0,001663)+(21,4x0,000534)+(21,4x0,000038)+(21,4x0,000626)+(21,4x0,001807)} \\ &= \frac{0,0114276}{0,0998952} = 0,114396 \end{aligned}$$

(DBD|BU)

$$\begin{aligned} &= \frac{21,4x0,000038}{(21,4x0,001663)+(21,4x0,000534)+(21,4x0,000038)+(21,4x0,000626)+(21,4x0,001807)} \\ &= \frac{0,0008132}{0,0998952} = 0,008140 \end{aligned}$$

(DBD|KB)

$$\begin{aligned} &= \frac{21,4x0,000626}{(21,4x0,001663)+(21,4x0,000534)+(21,4x0,000038)+(21,4x0,000626)+(21,4x0,001807)} \\ &= \frac{0,0133964}{0,0998952} = 0,134104 \end{aligned}$$

(DBD|KM)

$$\begin{aligned} &= \frac{21,4x0,001807}{(21,4x0,001663)+(21,4x0,000534)+(21,4x0,000038)+(21,4x0,000626)+(21,4x0,001807)} \\ &= \frac{0,0386698}{0,0998952} = 0,387104 \end{aligned}$$

3. Penyakit Pneumonia :

$$\begin{aligned} \text{(PN|MU)} &= \frac{21,2x0,001125}{(21,2x0,001125)+(21,2x0,000413)+(21,2x0,000154)+(21,2x0,000591)+(21,2x0,001459)} \\ &= \frac{0,02385}{0,0793304} = 0,300641 \end{aligned}$$

$$\begin{aligned} \text{(PN|BT)} &= \frac{21,2x0,000413}{(21,2x0,001125)+(21,2x0,000413)+(21,2x0,000154)+(21,2x0,000591)+(21,2x0,001459)} \\ &= \frac{0,0087556}{0,0793304} = 0,110369 \end{aligned}$$

$$\begin{aligned} \text{(PN|BU)} &= \frac{21,2x0,000154}{(21,2x0,001125)+(21,2x0,000413)+(21,2x0,000154)+(21,2x0,000591)+(21,2x0,001459)} \\ &= \frac{0,0032648}{0,0793304} = 0,041154 \end{aligned}$$

$$\begin{aligned} \text{(PN|KB)} &= \frac{21,2x0,000591}{(21,2x0,001125)+(21,2x0,000413)+(21,2x0,000154)+(21,2x0,000591)+(21,2x0,001459)} \\ &= \frac{0,0125292}{0,0793304} = 0,157937 \end{aligned}$$

$$\begin{aligned} \text{(PN|KM)} &= \frac{21,2x0,001459}{(21,2x0,001125)+(21,2x0,000413)+(21,2x0,000154)+(21,2x0,000591)+(21,2x0,001459)} \\ &= \frac{0,0309308}{0,0793304} = 0,389898 \end{aligned}$$

4. Penyakit Diare :

(DI|MU)

$$\begin{aligned} &= \frac{1534,6x0,072753}{(1534,6x0,072753)+(1534,6x0,045747)+(1534,6x0,010238)+(1534,6x0,043569)+(1534,6x0,082638)} \\ &= \frac{111,6467538}{391,238597} = 0,285367 \end{aligned}$$

(DI|BT)

$$\begin{aligned} &= \frac{1534,6x0,045747}{(1534,6x0,072753)+(1534,6x0,045747)+(1534,6x0,010238)+(1534,6x0,043569)+(1534,6x0,082638)} \\ &= \frac{70,2033462}{391,238597} = 0,179439 \end{aligned}$$

(DI|BU)

$$\begin{aligned} &= \frac{1534,6x0,010238}{(1534,6x0,072753)+(1534,6x0,045747)+(1534,6x0,010238)+(1534,6x0,043569)+(1534,6x0,082638)} \\ &= \frac{15,7112348}{391,238597} = 0,040158 \end{aligned}$$

(DI|KB)

$$\begin{aligned} &= \frac{1534,6x0,043569}{(1534,6x0,072753)+(1534,6x0,045747)+(1534,6x0,010238)+(1534,6x0,043569)+(1534,6x0,082638)} \\ &= \frac{66,8609874}{391,238597} = 0,170896 \end{aligned}$$

(DI|KM)

$$\begin{aligned} &= \frac{1534,6x0,082638}{(1534,6x0,072753)+(1534,6x0,045747)+(1534,6x0,010238)+(1534,6x0,043569)+(1534,6x0,082638)} \\ &= \frac{126,8162748}{391,238597} = 0,324140 \end{aligned}$$

5. Jumlah Penduduk :

(JP|MU)

$$\begin{aligned} &= \frac{55664,4x0,922892}{(55664,4x0,922892)+(55664,4x0,952092)+(55664,4x0,989057)+(55664,4x0,954066)+(55664,4x0,912705)} \\ &= \frac{51372,2294448}{263337,8115} = 0,195081 \end{aligned}$$

(JP|BT)

$$\begin{aligned} &= \frac{55664,4x0,952092}{(55664,4x0,922892)+(55664,4x0,952092)+(55664,4x0,989057)+(55664,4x0,954066)+(55664,4x0,912705)} \\ &= \frac{52997,6299248}{263337,8115} = 0,201253 \end{aligned}$$

(JP|BU)

$$\begin{aligned} &= \frac{55664,4x0,989057}{(55664,4x0,922892)+(55664,4x0,952092)+(55664,4x0,989057)+(55664,4x0,954066)+(55664,4x0,912705)} \\ &= \frac{55055,2644708}{263337,8115} = 0,209067 \end{aligned}$$

(JP|KB)

=

$$\frac{55664,4 \times 0,954066}{(55664,4 \times 0,922892) + (55664,4 \times 0,952092) + (55664,4 \times 0,989057) + (55664,4 \times 0,954066) + (55664,4 \times 0,912705)}$$

$$= \frac{53107,5114504}{263337,8115} = 0,201671$$

(JP|KM)

$$\frac{55664,4 \times 0,912705}{(55664,4 \times 0,922892) + (55664,4 \times 0,952092) + (55664,4 \times 0,989057) + (55664,4 \times 0,954066) + (55664,4 \times 0,912705)}$$

$$= \frac{50805,176202}{263337,8115} = 0,192928$$

Untuk memudahkan dalam pembacaan data, nilai P(X|Ci) Probabilitas yang telah didapatkan dari pencarian di atas selanjutnya akan ditampilkan ke dalam bentuk tabel seperti yang terlihat pada tabel

Nilai P(X|Ci)

Puskesmas	Kriteria Tahun 2019					Tuberculosis	1
	Jenis Penyakit						
	Tuberculosis	DBD	Pneumonia	Diare	Jumlah Penduduk		
Mulyaharja	0,319278	0,357995	0,252326	0,284107	0,195370	0,265841	0,3
Bogor Timur (Baranang Siang & Katulampa)	0,228348	0,107159	0,096632	0,186427	0,200805	0,214676	0,3
Bogor Utara	0,089075	0,008353	0,035625	0,042245	0,208445	0,091588	0,0
Kedung Badak	0,106562	0,123628	0,117374	0,158283	0,202233	0,195193	0,3
Kayu Manis	0,256738	0,402858	0,322357	0,328937	0,193046	0,232702	0,3

- **Langkah 4** : Menghitung Persentase P(X|Ci) setiap kriteria.

Mencari persentasi dari penyakit Tuberculosis pada tahun 2019 :

$$\text{Tuberculosis Mulyaharja} = \frac{0,319278}{1} \times 100\% = 31,93$$

$$\text{Tuberculosis Bogor Timur} = \frac{0,228348}{1} \times 100\% = 22,83$$

$$\text{Tuberculosis Bogor Utara} = \frac{0,089075}{1} \times 100\% = 8,91$$

$$\text{Tuberculosis Kedung Badak} = \frac{0,106562}{1} \times 100\% = 10,66$$

$$\text{Tuberculosis Kayu Manis} = \frac{0,256738}{1} \times 100\% = 25,67$$

Mencari persentasi dari penyakit DBD pada tahun 2019 :

$$\text{DBD Mulyaharja} = \frac{0,357995}{1} \times 100\% = 35,80$$

$$\text{DBD Bogor Timur} = \frac{0,107159}{1} \times 100\% = 10,72$$

$$\text{DBD Bogor Utara} = \frac{0,008353}{1} \times 100\% = 0,83$$

$$\text{DBD Kedung Badak} = \frac{0,123628}{1} \times 100\% = 12,36$$

$$\text{DBD Kayu Manis} = \frac{0,402858}{1} \times 100\% = 40,29$$

Mencari persentasi dari penyakit Pneumonia pada tahun 2019 :

$$\text{Pneumonia Mulyaharja} = \frac{0,252326}{1} \times 100\% = 25,23$$

$$\text{Pneumonia Bogor Timur} = \frac{0,096632}{1} \times 100\% = 9,66$$

$$\text{Pneumonia Bogor Utara} = \frac{0,035625}{1} \times 100\% = 3,56$$

$$\text{Pneumonia Kedung Badak} = \frac{0,117374}{1} \times 100\% = 11,74$$

$$\text{Pneumonia Kayu Manis} = \frac{0,322357}{1} \times 100\% = 32,24$$

Mencari persentasi dari penyakit Diare pada tahun 2019 :

$$\text{Diare Mulyaharja} = \frac{0,284107}{1} \times 100\% = 28,41$$

$$\text{Diare Bogor Timur} = \frac{0,186427}{1} \times 100\% = 18,64$$

$$\text{Diare Bogor Utara} = \frac{0,042245}{1} \times 100\% = 4,22$$

$$\text{Diare Kedung Badak} = \frac{0,158283}{1} \times 100\% = 15,82$$

$$\text{Diare Kayu Manis} = \frac{0,328937}{1} \times 100\% = 32,89$$

Mencari persentasi dari Jumlah Penduduk pada tahun 2019 :

$$\text{Jumlah Penduduk Mulyaharja} = \frac{0,195370}{1} \times 100\% = 19,54$$

$$\text{Jumlah Penduduk Bogor Timur} = \frac{0,200805}{1} \times 100\% = 20,08$$

$$\text{Jumlah Penduduk Bogor Utara} = \frac{0,208445}{1} \times 100\% = 20,84$$

$$\text{Jumlah Penduduk Kedung Badak} = \frac{0,202233}{1} \times 100\% = 20,22$$

$$\text{Jumlah Penduduk Kayu Manis} = \frac{0,193046}{1} \times 100\% = 19,30$$

Mencari persentasi dari penyakit Tuberculosis pada tahun 2020 :

$$\text{Tuberculosis Mulyaharja} = \frac{0,265841}{1} \times 100\% = 26,58$$

$$\text{Tuberculosis Bogor Timur} = \frac{0,214676}{1} \times 100\% = 21,47$$

$$\text{Tuberculosis Bogor Utara} = \frac{0,091588}{1} \times 100\% = 9,15$$

$$\text{Tuberculosis Kedung Badak} = \frac{0,195193}{1} \times 100\% = 19,52$$

$$\text{Tuberculosis Kayu Manis} = \frac{0,232702}{1} \times 100\% = 23,27$$

Mencari persentasi dari penyakit DBD pada tahun 2020 :

$$\text{DBD Mulyaharja} = \frac{0,351499}{1} \times 100\% = 35,15$$

$$\text{DBD Bogor Timur} = \frac{0,109659}{1} \times 100\% = 10,97$$

$$\text{DBD Bogor Utara} = \frac{0,010358}{1} \times 100\% = 1,04$$

$$\text{DBD Kedung Badak} = \frac{0,128799}{1} \times 100\% = 12,88$$

$$\text{DBD Kayu Manis} = \frac{0,399684}{1} \times 100\% = 39,97$$

Mencari persentasi dari penyakit Pneumonia pada tahun 2020 :

$$\text{Pneumonia Mulyaharja} = \frac{0,277552}{1} \times 100\% = 27,75$$

$$\text{Pneumonia Bogor Timur} = \frac{0,111659}{1} \times 100\% = 11,17$$

$$\text{Pneumonia Bogor Utara} = \frac{0,043503}{1} \times 100\% = 4,35$$

$$\text{Pneumonia Kedung Badak} = \frac{0,155452}{1} \times 100\% = 15,54$$

$$\text{Pneumonia Kayu Manis} = \frac{0,264151}{1} \times 100\% = 26,41$$

Mencari persentasi dari penyakit Diare pada tahun 2020 :

$$\text{Diare Mulyaharja} = \frac{0,289486}{1} \times 100\% = 28,95$$

$$\text{Diare Bogor Timur} = \frac{0,185717}{1} \times 100\% = 18,57$$

$$\text{Diare Bogor Utara} = \frac{0,042670}{1} \times 100\% = 4,27$$

$$\text{Diare Kedung Badak} = \frac{0,159144}{1} \times 100\% = 15,91$$

$$\text{Diare Kayu Manis} = \frac{0,322982}{1} \times 100\% = 32,30$$

Mencari persentasi dari Jumlah Penduduk pada tahun 2020 :

$$\text{Jumlah Penduduk Mulyaharja} = \frac{0,195051}{1} \times 100\% = 19,50$$

$$\text{Jumlah Penduduk Bogor Timur} = \frac{0,200878}{1} \times 100\% = 20,09$$

$$\text{Jumlah Penduduk Bogor Utara} = \frac{0,208635}{1} \times 100\% = 20,86$$

$$\text{Jumlah Penduduk Kedung Badak} = \frac{0,202239}{1} \times 100\% = 20,22$$

$$\text{Jumlah Penduduk Kayu Manis} = \frac{0,193197}{1} \times 100\% = 19,32$$

Mencari persentasi dari penyakit Tuberculosis pada tahun 2021 :

$$\text{Tuberculosis Mulyaharja} = \frac{0,268657}{1} \times 100\% = 26,87$$

$$\text{Tuberculosis Bogor Timur} = \frac{0,208267}{1} \times 100\% = 20,83$$

$$\text{Tuberculosis Bogor Utara} = \frac{0,087837}{1} \times 100\% = 8,78$$

$$\text{Tuberculosis Kedung Badak} = \frac{0,196775}{1} \times 100\% = 19,68$$

$$\text{Tuberculosis Kayu Manis} = \frac{0,238463}{1} \times 100\% = 23,85$$

Mencari persentasi dari penyakit DBD pada tahun 2021 :

$$\text{DBD Mulyaharja} = \frac{0,356255}{1} \times 100\% = 35,62$$

$$\text{DBD Bogor Timur} = \frac{0,114396}{1} \times 100\% = 11,44$$

$$\text{DBD Bogor Utara} = \frac{0,008140}{1} \times 100\% = 0,81$$

$$\text{DBD Kedung Badak} = \frac{0,134014}{1} \times 100\% = 13,40$$

$$\text{DBD Kayu Manis} = \frac{0,387104}{1} \times 100\% = 38,71$$

Mencari persentasi dari penyakit Pneumonia pada tahun 2021 :

$$\text{Pneumonia Mulyaharja} = \frac{0,300641}{1} \times 100\% = 30,06$$

$$\text{Pneumonia Bogor Timur} = \frac{0,110369}{1} \times 100\% = 11,04$$

$$\text{Pneumonia Bogor Utara} = \frac{0,041154}{1} \times 100\% = 4,11$$

$$\text{Pneumonia Kedung Badak} = \frac{0,157937}{1} \times 100\% = 15,79$$

$$\text{Pneumonia Kayu Manis} = \frac{0,389898}{1} \times 100\% = 38,99$$

Mencari persentasi dari penyakit Diare pada tahun 2021 :

$$\text{Diare Mulyaharja} = \frac{0,285367}{1} \times 100\% = 28,54$$

$$\text{Diare Bogor Timur} = \frac{0,179439}{1} \times 100\% = 17,94$$

$$\text{Diare Bogor Utara} = \frac{0,040158}{1} \times 100\% = 4,02$$

$$\text{Diare Kedung Badak} = \frac{0,170896}{1} \times 100\% = 17,09$$

$$\text{Diare Kayu Manis} = \frac{0,324140}{1} \times 100\% = 32,41$$

Mencari persentasi dari Jumlah Penduduk pada tahun 2021 :

$$\text{Jumlah Penduduk Mulyaharja} = \frac{0,195081}{1} \times 100\% = 19,51$$

$$\text{Jumlah Penduduk Bogor Timur} = \frac{0,201253}{1} \times 100\% = 20,12$$

$$\text{Jumlah Penduduk Bogor Utara} = \frac{0,209067}{1} \times 100\% = 20,91$$

$$\text{Jumlah Penduduk Kedung Badak} = \frac{0,201671}{1} \times 100\% = 20,17$$

$$\text{Jumlah Penduduk Kayu Manis} = \frac{0,192928}{1} \times 100\% = 19,29$$

Untuk mempermudah dalam membaca hasil persentase, maka dibuat didalam bentuk tabel 9.

Hasil Persentase P(X|Ci)

Puskesmas	Kriteria Tahun 2019					Kriteria Tahun 2020					Kriteria Tahun 2021				
	Jenis Penyakit					Jenis Penyakit					Jenis Penyakit				
	Tubercu losis	D B D	Pne um onia	D ia re	Ju ml ah Pe nd ud uk	Tubercu losis	D B D	Pne um onia	D ia re	Ju ml ah Pe nd ud uk	Tubercu losis	D B D	Pne um onia	D ia re	Ju ml ah Pe nd ud uk
Mulyaharja	31,93	5,80	25,23	8,41	19,54	26,58	5,15	27,75	8,95	19,50	26,87	5,62	30,06	8,54	19,51
Bogor Timur (Baranang Siang & Katulampa)	22,83	1,72	9,66	8,64	20,08	21,47	1,77	11,17	8,57	20,09	20,83	1,44	11,04	7,94	20,12
Bogor Utara	8,91	0,83	3,56	4,22	20,84	9,15	1,04	4,35	4,27	20,86	8,78	0,81	4,11	4,02	20,91
Kedung Badak	10,66	1,23	11,74	5,82	20,22	19,52	2,88	15,54	5,91	20,22	19,68	3,40	15,79	7,09	20,17
Kayu Manis	25,67	4,02	32,24	2,89	19,30	23,27	3,97	26,41	2,30	19,32	23,85	3,71	38,99	2,41	19,29

Untuk menentukan hasil akhir dari persentase maka dilakukan penambahan masing-masing dari hasil persentase setiap kriteria.

Penambahan hasil persentase data tahun 2019 :

$$\text{Puskesmas Mulyaharja} = \frac{140,91}{5} = 28,182 \%$$

$$\text{Puskesmas Bogor Timur} = \frac{81,93}{5} = 16,386 \%$$

$$\text{Puskesmas Bogor Utara} = \frac{38,36}{5} = 7,672 \%$$

$$\text{Puskesmas Kedung Badak} = \frac{70,8}{5} = 14,16 \%$$

$$\text{Puskesmas Kayu Manis} = \frac{150,39}{5} = 30,078 \%$$

Penambahan hasil persentase data tahun 2020 :

$$\text{Puskesmas Mulyaharja} = \frac{137,93}{5} = 27,586$$

$$\text{Puskesmas Bogor Timur} = \frac{82,27}{5} = 16,454 \%$$

$$\text{Puskesmas Bogor Utara} = \frac{39,67}{5} = 7,934 \%$$

$$\text{Puskesmas Kedung Badak} = \frac{84,07}{5} = 16,814 \%$$

$$\text{Puskesmas Kayu Manis} = \frac{141,27}{5} = 28,254 \%$$

Penambahan hasil persentase data tahun 2021 :

$$\text{Puskesmas Mulyaharja} = \frac{140,6}{5} = 28,12 \%$$

$$\text{Puskesmas Bogor Timur} = \frac{81,37}{5} = 16,274 \%$$

$$\text{Puskesmas Bogor Utara} = \frac{38,63}{5} = 7,726 \%$$

$$\text{Puskesmas Kedung Badak} = \frac{86,13}{5} = 17,226 \%$$

$$\text{Puskesmas Kayu Manis} = \frac{153,25}{5} = 30,65 \%$$

Untuk mempermudah membaca hasil persentase pertahun nya maka dibuat dalam bentuk tabel

. Hasil Persentase Setiap Kriteria

Puskesmas	Persentase Akhir Tahun 2019	Persentase Akhir Tahun 2020	Persentase Akhir Tahun 2021	Tingkat Infeksi
Mulyaharja	28,182 %	27,586	28,12 %	Tinggi
Bogor Timur	16,386 %	16,454 %	16,274 %	Sedang
Bogor Utara	7,672 %	7,934 %	7,726 %	Rendah
Kedung Badak	14,16 %	16,814 %	17,226 %	Sedang
Kayu Manis	30,078 %	28,254 %	30,65 %	Tinggi

Dari hasil proses perhitungan persentase akhir yang dapat dilihat pada tabel 10. Kemudian untuk memprediksi data di tahun 2022 dilakukan dengan ketentuan, jika nilai pengurangan data di tahun 2021 nilainya lebih tinggi dari nilai tahun 2020 maka persentasenya ditambah dengan hasil pengurangan, dan jika nilai persentase di tahun 2021 lebih rendah dari nilai di tahun 2020 maka persentasenya dikurangi dengan hasil pengurangan.

$$\begin{aligned} \text{Puskesmas Mulyaharja} &= 27,586 - 28,12 = -0,534 \\ &= 28,12 - (-0,534) = 28,654\% \end{aligned}$$

$$\begin{aligned} \text{Puskesmas Bogor Timur} &= 16,454 - 16,274 = 0,18 \\ &= 16,274 - 0,18 = 16,094\% \end{aligned}$$

$$\begin{aligned} \text{Puskesmas Bogor Utara} &= 7,934 - 7,726 = 0,208 \\ &= 7,726 - 0,208 = 7,518\% \end{aligned}$$

$$\begin{aligned} \text{Puskesmas Kedung Badak} &= 16,814 - 17,226 = -0,412 \\ &= 17,226 - (-0,412) = 17,638\% \end{aligned}$$

$$\begin{aligned} \text{Puskesmas Kayu Manis} &= 28,254 - 30,65 = -2,396 \\ &= 30,65 - (-2,396) = 33,046\% \end{aligned}$$

Dari hasil perhitungan di atas agar mudah dipahami hasil prediksi perhitungan datanya maka dibuat dalam bentuk Tabel 11.

Hasil Prediksi

Puskesmas	Persentase Akhir Tahun 2022
Mulyaharja	28,654%
Bogor Timur	16,094%
Bogor Utara	7,518%
Kedung Badak	17,638%
Kayu Manis	33,046%