

BAB II

LANDASAN TEORI

2.1 Perkiraan

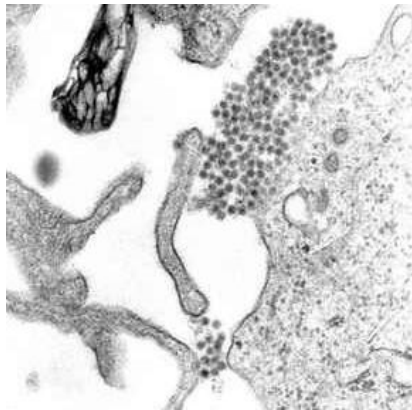
Perkiraan atau peramalan merupakan ilmu yang digunakan untuk memprediksi suatu peristiwa peristiwa yang akan terjadi dimasa yang akan datang.[4] Prediksi dapat dilakukan dilakukan dengan melalui berbagai macam metode yang terkait, hal ini juga dapat dilakukan dengan menggabungkan dua atau lebih metode untuk memperoleh hasil prediksi yang lebih akurat.[5] Prediksi merupakan sebuah proses sistematis yang digunakan untuk memperkiraan hasil yang paling mendekati kemungkinan berdasarkan informasi saat ini dan sebelumnya dengan tujuan mengurangi kesalahan dalam perkiraan. Jika hasil peramalan terlalu tinggi atau rendah dibandingkan dengan kenyataan yang sebenarnya, maka hasil tersebut dikatakan biasa.[6]

Prediksi bisa ilmiah atau subjektif. Sebagai contoh, prediksi cuaca selalu didasarkan pada data dan informasi yang terbaru, termasuk pengamatan satelit. Begitupun dengan prediksi gempa, gunung meletus ataupun bencana lainnya. Namun ketika melakukan prediksi seperti perbandingan sepak bola, olahraga lainnya pada umumnya didasarkan pada pandangan subjektif dengan sudut pandang sendiri yang memprediksinya.[7]

2.2 Demam Berdarah

Penyakit DBD merupakan penyakit yang disebabkan oleh virus dengue yang dibawa oleh nyamuk *Aedes Aegypti* betina lewat air liur gigitan saat

menghisap darah manusia yang menyebabkan gangguan pada system pembekuan darah, sehingga mengakibatkan pendarahan. Penyakit DBD ini ditandai dengan ciri ciri demam tinggi mendadak selama 2-7 hari tanpa sebab yang jelas disertai manifestasi seperti petekie, epistaxis kadang disertai muntah darah, kesadaran menurun, bertandensi menimbulkan senjata (shock) dan juga dapat menyebabkan kematian.[3] Penyakit demam berdarah merupakan wabah yang disebabkan oleh virus *dengue* yang termasuk ke dalam kelompok *B Arthropod Borne Virus (Arboviroses)* yang pada saat ini dikenal sebagai *genus Flavivirus, family Falaviviridea* seperti yang ditunjukkan pada Gambar 1. Dan mempunya 4 jenis tipe serotype diantaranya adalah DEN-1, DEN-2, DEN-3, DEN-4.



Gambar 2.2 Perkembangbiakan jentik nyamuk

Perkembangbiakan jentik nyamuk *Aedes Aegypti* merupakan jenis nyamuk yang menyebabkan timbulnya penyakit Demam Berdarah Dengue (DBD) dipengaruhi oleh faktor iklim seperti curah hujan, suhu, dan kelembapan. Selain itu factor geogrfis seperti lingkungan rumah juga menjadi salah satu penyebab penyakit demam berdarah dengue (DBD). Kepadatan tempat tinggal dan lingkungan yang tidak bersih serta terdapatnya genangan air yang menyebabkan

berkembangbiaknya jenis nyamuk tersebut. Kepadatan penduduk yang tinggi meningkatkan risiko penularan DBD karena jarak antara individu yang terjangkit dengan yang sehat lebih dekat. Area dengan kepadatan penduduk di atas 1000/km² cenderung memiliki risiko lebih tinggi untuk penyebaran DBD. Sedangkan curah hujan lebih dari 100 mm/ bulan memiliki risiko yang lebih besar terutama pada musim hujan dan tingkat kelembaban udara di atas 70% mendukung aktivitas nyamuk bertahan hidup lebih lama dan mempercepat siklus reproduksi[8]

Menurut WHO, daerah dengan populasi padat lebih rentan terhadap penularan virus dengue karena tingginya tingkat interaksi antara manusia dengan nyamuk *Aedes aegypti*. Penelitian oleh Gubler et al.(1998) juga menunjukkan korelasi antara kepadatan penduduk dengan lonjakan kasus demam berdarah di daerah tropis dan subtropis. Oleh karena itu, kebijakan kesehatan masyarakat yang efektif harus mempertimbangkan aspek demografi dalam pencegahan dan pengendalian dengue. Curah hujan merupakan faktor lingkungan utama yang mempengaruhi siklus hidup nyamuk *Aedes Aegypti*. Hujan yang tinggi dapat meningkatkan populasi nyamuk dalam waktu yang cukup singkat akan tetapi, dalam beberapa kasus ekstrem hujan yang sangat deras bisa mengurangi tempat berkembang biak nyamuk dengan membersihkan wadah penampungan air. Oleh karena itu, hubungan antara curah hujan dan kasus dengue bervariasi tergantung pada pola curah hujan di suatu wilayah. Sedangkan menurut Morin et al.(2022) menemukan bahwa kelembaban udara di atas 70% dapat meningkatkan frekuensi

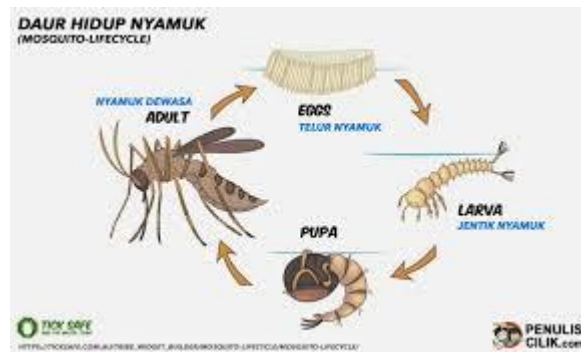
gigitan nyamuk. Kelembaan udara mempengaruhi aktivitas nyamuk dan daya tahan nyamuk.

Penyakit Demam Berdarah Dengue (DBD) juga disebabkan oleh kondisi kesehatan lingkungan yang tidak baik atau buruk. Seperti perilaku masyarakat yang sering membuang sampah di tempat yang tidak boleh atau tidak disarankan, dimana sampah menjadi tempat nyamuk berkembang biak dan membawa virus Dengue fever ke tempat yang lebih tinggi lagi. Tingkat kelembapan udara juga mempengaruhi kejadian demam berdarah hal ini yang menyebabkan populasi nyamuk *Aedes Aegypti* semakin meningkat.[9]

Demam Berdarah *Dengue* (DBD) merupakan penyakit yang terdapat pada anak-anak dan juga orang dewasa dengan gejala utama yang dirasakan yakni demam, nyeri otot dan sendi yang biasanya memburuk setelah dua hari pertama. Penyakit *DBD* ditandai dengan ciri demam yang mendadak tinggi selama 2-7 hari tanpa sebab yang jelas dan disertai manifestasi seperti *petekie*, *epistaxis* kadang juga disertai muntah darah, kedarasan menurun, bertandensi menimbulkan senjata (shock) dan kematian.[10]

2.2.1 Siklus Hidup Nyamuk *Aedes Egypti*

Nyamuk *Aedes Aegypti* mengalami metamorfosis yang cukup sempurna, dimana siklus hidup nyamuk terdiri dari empat fase diantaranya meliputi: telur, larva, pupa, dan nyamuk dewasa. Stadium telur, larva dan pupa hidup didalam air. Sedangkan nyamuk dewasa hidup di darat.[11]



Gambar 2.2.1 Siklus Hidup Nyamuk

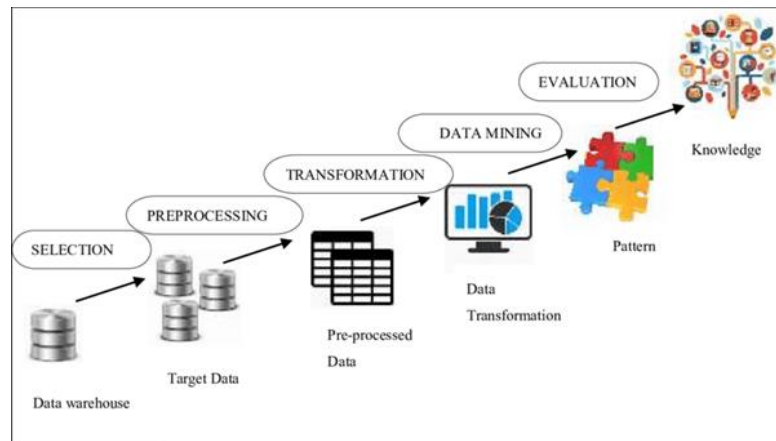
Nyamuk jenis ini mempunyai siklus hidup yang cukup sempurna. Nyamuk meletakkan telurnya pada kondisi permukaan air yang bersih secara individual. Telur nyamuk berbentuk elips, berwarna hitam dan juga terpisah satu dengan yang lain. Telurnya dapat menetas dalam waktu 1-2 hari kemudian akan berubah menjadi jentik. Nyamuk terdiri dari 4 tahap dalam perkembangan jentik yang dikenal sebagai instar. Perkembangan instar 1 hingga ke 4 membutuhkan waktu sekitar 5 hari. Selanjutnya untuk sampai pada instar ke 4, larva ini berubah menjadi pupa dimana jentik tersebut telah memasuki masa dorman.[12]

2.3 Knowledge Discovery in Database (KDD)

Data mining merupakan salah satu dari tahapan yang terdapat dalam proses *knowledge discovery in database* (KDD) yaitu pencarian informasi-informasi baru dan bernilai yang terdapat dalam suatu kumpulan data atau sebuah database.[13].Serangkaian dari proses data tersebut mempunyai beberapa tahapan, adapun tahapan tersebut diantaranya adalah sebagai berikut:

1. Pembersihan data (untuk membersihkan data yang tidak sesuai)
2. Integritas data (beberapa sumber data yang digabungkan)
3. Transformasi data (data dapat dirubah bentuk yang sesuai untuk dimining)

4. Aplikasi teknik data mining, pola data yang ada di proses ekstraksi
5. Evaluasi pola yang ditemukan (pola proses interpretasi menjadi pengetahuan dapat digunakan untuk mengambil sebuah keputusan)
6. Presentasi pengetahuan (dengan teknik visualisasi)[14]



Gambar 2.3 Tahapan *Knowledge Discovery in Database*

Sumber: (<https://sis.binus.ac.id/>)

Berikut ini adalah penjelasan dari beberapa tahapan *knowledge discovery in database* (KDD)

1. *Selection and Addition*

Pada tahap ini merupakan tahap pengumpulan data. Data yang telah dikumpulkan kemudian dilakukan seleksi atribut dan hasil dari seleksi tersebut diintegrasikan menjadi sebuah dataset. Proses pembangunan dataset merupakan suatu proses yang penting karena proses pembelajaran data mining dan penemuan pola baru didasarkan paa dataset yang telah dibentuk.

2. *Preprocessing and Data Cleaning*

Pada tahap ini merupakan tahap melakukan pembersihan data untuk meningkatkan keandalan data. Pembersihan data dilakukan dengan menangani

nilai kosong, menangani data yang tidak relevan dan menghilangkan noise atau outlier. Proses persiapan awal ini dapat melibatkan metode statistik yang kompleks atau menggunakan algoritma data mining yang spesifik.

3. *Transformation*

Pada tahap ini merupakan tahap pengembangan data sehingga data dipersiapkan dengan lebih baik dan siap untuk dilakukan pemodelan data mining. Hal yang dapat dilakukan untuk mempersiapkan data menjadi lebih baik adalah melakukan reduksi dimensi seperti pemilihan fitur dan ekstraksi sampel data. Selain itu, dapat juga dilakukan transformasi atribut seperti mengubah atribut numeric menjadi atribut diskrit dan transformasi fungsional.

4. Data Mining

Tahap data mining terdiri dari tiga tahap diantaranya adalah pemilihan model data mining, pemilihan algoritma data mining dan penggunaan data mining. Model data mining yang dipilih merupakan model yang sesuai dengan kebutuhan diantaranya adalah klasifikasi, regresi atau pengelompokan. Algoritma yang dipilih disesuaikan dengan model data yang dipilih dengan mempertimbangkan kelebihan dan kekurangan dari algoritma tersebut. Setelah dilakukan pemilihan model dan algoritma data mining, kemudian data mining digunakan untuk menemukan pola atau aturan yang baru. Algoritma data mining dimungkinkan untuk digunakan berulang kali hingga memperoleh hasil yang sesuai. Pada penggunaan data mining juga dilakukan pengaturan parameter control algoritma.

5. *Evaluation and Interpretation*

Pada tahap ini dilakukan evaluasi dan penafsiran pengetahuan yang berupa pola hasil penggunaan data mining. Selain itu juga dilakukan pendokumentasian yang ditemukan untuk penggunaan yang lebih lanjut lagi.

6. *Discovered Knowledge*

Pada tahap ini, pengetahuan yang ditemukan telah siap untuk diimplementasikan kedalam sebuah system. Tahap ini menentukan keaktifan dari keseluruhan proses yang telah dilakukan.[15]

2.4 Data Mining

Data mining merupakan salah satu disiplin ilmu yang digunakan untuk menentukan suatu informasi tertentu dalam sekumpulan data sebagai pendukung pengambilan keputusan. Data mining juga sebagai salah satu proses untuk memperoleh informasi yang memiliki nilai guna dari sekumpulan sebuah data.[3] Adapun pengertian lain mengenai data mining, dimana data mining merupakan proses dalam mencari suatu pola yang dapat menggunakan teknik statistic, matematika, dan kecerdasan buatan untuk mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait *database* yang begitu besar. Namun secara umum data mining juga dapat disebut sebagai serangkaian proses yang digunakan untuk menemukan nilai yang tidak dapat diketahui secara langsung dari adanya sebuah data.[17] Proses data mining merupakan proses dimana dengan melakukan pengelolaan data dengan menerapkan algoritma dan tujuan kita.[18] Data mining juga meliputi langkah-langkah dalam menentukan variabel atau fitur yang penting yang digunakan dalam klasifikasi dan regresi. *Data mining*

memegang peran yang cukup penting dalam bidang industry, keuangan, cuaca, ilmu pengetahuan dan teknologi. Selain itu *data mining* juga berkenaan dengan pengelolaan data dalam skala yang cukup besar.[19]

Dalam konteks *data mining*, pengelompokan data juga dapat dilakukan. Teknik ini memungkinkan pengelompokan data yang berdasarkan pola atau karakteristik tertentu, memberikan pemahaman lebih lanjut mengenai struktur dan hubungan data. Sehingga *data mining* menjadi alat yang sangat berharga dalam menggali wawasan dari data besar untuk mendukung proses pengambilan sebuah keputusan.[20]

2.5 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan prinsip Teorema Bayes untuk memprediksi kelas suatu data. Algoritma ini disebut “Bayes” hal ini dikarenakan diasumsikan bahwa setiap fitur dalam data set bersifat independen satu sama lain, meskipun asumsi independen sering kali tidak realistis, algoritma Naïve Bayes tetap memberikan kinerja yang baik dalam banyak kasus klasifikasi, terutama pada dataset yang memiliki fitur besar atau ketika kecepatan komputasi menjadi factor yang sangat penting. Asumsi ini dikenal dengan istilah *class conditional independence*. Asumsi ini dibuat untuk menyederhanakan perhitungan yang terkait, sehingga disebut sebagai “Naïve”. [21] Naïve Bayes Classifier merupakan metode klasifikasi probabilitas tinggi antar atribut. Klasifikasi merupakan proses menemukan model yang dapat membagi data berdasarkan kelas, dan terbagi menjadi dua fase diantaranya adalah : fase pelatihan (pembelajaran) dan pengujian

untuk memahami apakah kategori data yang diketahui dapat ditingkatkan.[22]
Salah satu keunggulan menggunakan metode Naïve Bayes Classifier bahwa metode ini hanya memerlukan sedikit data pelatihan untuk mengestimasi parameter yang dibutuhkan dalam proses pengklasifikasiannya.[23]

Defenisi lain mengatakan jika *naïve bayes* merupakan proses pengklasifikasian dengan menggunakan metode probabilitas dan statistic yang dikemukakan oleh ilmuan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa lalu atau dimasa sebelumnya.[19]

Persamaan dari *Teorema Bayes* diantaranya adalah:

$$P(C / F) = \frac{P(C) \times P(F | C)}{P(C)} \quad (2.5)$$

Keterangan:

F : Data dengan kelas yang belum diketahui

C : Hipotesis data merupakan suatu kelas spesifik

$P(C / F)$: Probabilitas hipotesis dengan syarat F (probabilitas posterior)

$P(C)$: Prtobabilitas hipotesis C (probabilitas prior)

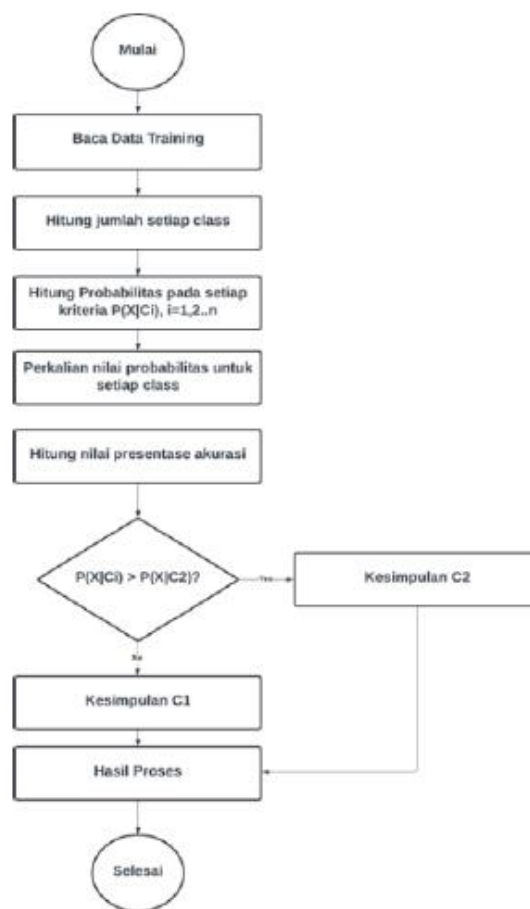
$P(F / C)$: Probabilitas hipotesis F dengan syarat C

$P(F)$: Probabilitas hipotesis F

Untuk menjelaskan metode *naïve bayes*, perlu diketahui jika proses klasifikasi memerlukan beberapa petunjuk yang digunakan untuk menentukan kelas apa yang cocok bagi sampel yang akan di analisis tersebut. Maka dari itu, *naïve bayes* diatas disesuaikan sebagai berikut:

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

Dimana variable C mempresentasikan kelas, sementara variabel $F_1 \dots F_n$ mempresentasikan krakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka dari itu rumus tersebut menjelaskan bahwa peluang masuknya sebuah sampel dengan karkteristik tertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *pior*), dikalikan dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C atau biasa disebut juga *likelihood*, dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara umum atau yang disebut juga sebagai *evidence*.



Gambar 2.5 Flowchart Naïve Bayes

2.6 K- Nearest Neighbor

Algoritma K-NN merupakan suatu metode klasifikasi data berdasarkan pembelajaran data yang sudah terklasifikasi sebelumnya. KNN termasuk kedalam golongan *supervised learning*, dimana hasil *query instance* yang baru diklasifikasi berdasarkan mayoritas kedekatan jarak dari kategori yang ada dalam metode KNN. Dimana nantinya, kelas yang baru dari suatu data akan dipilih berdasarkan grup kelas yang paling dekat jarak vektornya. Klasifikasi menggunakan *voting algoritma K-Nearest Neighbors* menggunakan klasifikasi ketanggaaan sebagai nilai prediksi dari *query instance* yang baru.[24] Tujuan algoritma K-NN ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan training samples.[23] Prinsip kerja dari metode K-NN ini adalah dengan cara menentukan K objek yang berbeda pada data pelatihan dengan jarak yang mendekati atau paling dekat dengan objek atau data test. Kelebihan dari metode K-NN ini adalah memiliki tingkat akurasi yang tinggi, tidak ada asumsi tentang data serta intensif pada outlier. Langkah pertama yang dapat dilakukan pada metode ini ialah dengan menentukan nilai K terlebih dahulu.[25]

Metode *K-Nearest Neighbors* sangatlah sederhana, bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana, masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai jika kelas c merupakan klasifikasi yang banyak ditemui pada k buah tetangga terdekat dari titik tertentu.[24]

Implementasi algoritma K-NN dalam penelitian ini dilakukan melalui pengambilan *dataset* sebagai data *training*, penentuan nilai k , memasukkan nilai *Euclidean distance*, pengurutan (*sorting*) data dari yang paling kecil ke besar, data diambil sesuai jumlah k , dan nilai yang mayoritas keluar diambil sebagai hasil prediksi. Nilai k merupakan jumlah tetangga terdekat yang digunakan untuk menentukan hasil prediksi. Nilai jarak dari ruang kelas dihitung dari *dataset* yang ditandai dengan $x^i_{training}$ dan $x^y_{testing}$ dari nilai data baru yang dimasukkan.[25]

Proses perhitungan didalam algoritma K-NN ini merupakan sebuah perhitungan *Euclidean Distance* yaitu metode pencarian antara dua titik variabel, semakin dekat dan mirip maka semakin kecil jarak antara dua titik tersebut.[26] Perhitungan jarak dilakukan dengan rumus *Euclidean Distance* yang ditunjukkan pada persamaan berikut:

$$d(x,y) = \sqrt{\sum_i^n (x^i_{training} - y^i_{testing})^2} \quad (2.6)$$

Keterangan:

$d(x,y)$ = jarak

$x^i_{training}$ = Data testing

$y^i_{testing}$ = Data training

i = Variabel data

n = Dimensi data

Hasil perhitungan *Euclidean Distance* dapat digunakan untuk mengetahui nilai jarak antara data *training* dan data masukkan baru (*testing*). Setelah itu

dilakukan pengurutan data dari data yang paling kecil hingga yang terbesar, kemudian menggunakan nilai k untuk batas data yang diambil berdasarkan nilai k , nilai yang mayoritas diambil sebagai hasil prediksi.[25]

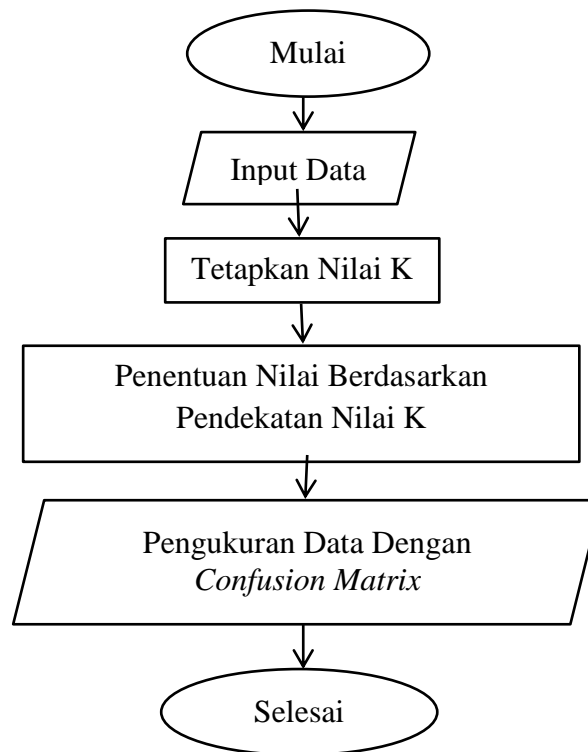
Berikut ini adalah kelebihan dari algoritma *K-Nearest Neighbor*:

1. Algoritma *K-Nearest Neighbor* sederhana dan mudah dipahami dan diimplementasikan.
2. *K-Nearest Neighbor* adalah algoritma non parametric, yang berarti tidak memerlukan pelatihan yang rumit. Hal ini dikarenakan model *K-NN* tersebut dapat “belajar” langsung dari data pelatihan.
3. Dapat menangani hubungan non linear diantara fitur karena tidak membuat asumsi mengenai bentuk fungsi keputusan.

Adapun beberapa kekurangan dari algoritma *K-Nearest Neighbor* diantaranya adalah sebagai berikut:

1. *K-Nearest Neighbor* memerlukan data pelatihan yang berlabel untuk mengklasifikasikan data baru. Tanpa label yang sesuai, algoritma ini tidak dapat memberikan prediksi yang baik.
2. Performa *K-Nearest Neighbor* sangat bergantung pada pemilihan parameter K
3. Penentuan jarak tidak jelas dan atribut apa yang harus digunakan untuk memperoleh hasil terbaik.
4. Biaya komputasi yang cukup tinggi karena diperlukan perhitungan jarak dari setiap data latih dengan data ujinya.

Berikut merupakan gambar dari *Flowchart* algoritma *K-Nearest Neighbor*:



Gambar 2.6 Flowchart Algoritma K-NN

2.7 Tools Pendukung Dalam Penelitian

2.7.1 Rapidminer



Gambar 2.7 Logo *Rapidminer*

Sumber: RapidMiner.com

Rapid Miner merupakan perangkat lunak yang digunakan untuk pengolahan data serta merupakan sebuah solusi untuk melakukan analisis terhadap

data mining, text mining dan analisis prediksi. Aplikasi ini memudahkan *user* melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator yang berfungsi untuk melakukan modifikasi data. Data dihubungkan melalui node-node pada operator kemudian dihubungkan ke node hasil untuk menampilkan hasil akhir sehingga hasil dapat ditampilkan secara visual dengan grafik.[27]

Rapid Miner menyediakan metode mulai dari *association, classification, clustering* dan lain sebagainya. *Rapid Miner* merupakan software atau data dalam ilmu pengetahuan dan dikembangkan oleh sebuah perusahaan yang sama dengan nama aplikasi tersebut dan menyediakan lingkungan yang terpadu untuk mesin pembelajaran (*meachine learning*), pembelajaran yang mendalam (*deep learning*), dan analisis prediktif (*predictive analystics*), serta penambahan text (*text mining*). Aplikasi ini digunakan untuk penelitian, pelatihan, pendidikan, koersial, bisnis dan juga untuk membuat prototype serta pengembangan aplikasi dan mendukung semua langkah proses pembelajaran mesin termasuk untuk persiapan validasi data, pengoptimalan dan visualisasi hasil.[14]

2.7.2 Microsoft Excel



Gambar 2.7.2 Ikon Microsoft Excel
Sumber : Tekno Kompas.com

Microsoft excel dapat disebut juga dengan excel merupakan lembar kerja yang berfungsi mengolah data atau angka dengan otomatis. *Microsoft Excel* di distribusiikan dan dibuat oleh *Microsoft corporation* dan dapat dijalankan melalui *Mac OS* dan *Mirosoft Windows*. Aplikasi ini merupakan bbgan dari *microsoft office system*. Selain digunakan untuk mengolah data, icrosoft excel juga digunakan untuk memanajemen data, penggunaan rumus tertentu, perhitungan dasar serta dapat digunakan untuk membuat sebuah grafik.[14]

2.8 Metodologi Penelitian

2.8.1 Penelitian Terdahulu

Untuk melakukan penelitian dalam memprediksi penyebaran kasus *Demam Berdarah* dengan menggunakan metode *naive bayes* dan *K-nearest neighbor* maka penulis menjadikan beberapa jurnal sebagai bahan referensi. Adapun jurnal referensi tersebut seperti tabel yang ada dibawah ini:

Referensi	1
Judul	Prediksi Penyebaran Kasus Demam Berdarah <i>Dengeu</i> (DBD) Dengan menggunakan Metode K-Nearest Neighbor Di Kabupaten Maluku Tengah
Penulis	S. Latuconsina ¹ dan M. Y. Matdoan ²
Tahun	2022
Hasil	Tujuan dari penelitian ini yaitu agar dapat segera untuk mengetahui wilayah

	<p>yang tingkat penyebarannya tinggi hingga rendah untuk melakukan tindakan dan pencegahan yang sesuai, sehingga mengurangi dampak dari penyakit epidemic DBD tersebut. Penelitian ini diperoleh hasil bahwa penyebaran kasus DBD di Kabupaten Maluku Tengah dengan rincian 4 kecamatan yang berpotensi penyebaran DBD tinggi yaitu Kecamatan Banda, Telutih, Salahutu dan Kota Masohi. Kecamatan yang berpotensi penyebaran DBD sedang yaitu Kecamatan Tehoru, Amahai, Saparua, Nusalaut, Pulau Haruku, Leihitu, Leihitu Barat dan Teon Nila Serua. Selanjutnya kecamatan yang berpotensi penyebaran DBDB rendah yaitu Kecamatan Teluk Elpaputih, Seram Utara, Seram Utara Barat, Seram Utara Timur Kobi dan Seram Utara Timur Seti.</p>
Referensi	2
Penulis	Bimo Aditya Wahyudi ¹ , Hannie ² , Aries

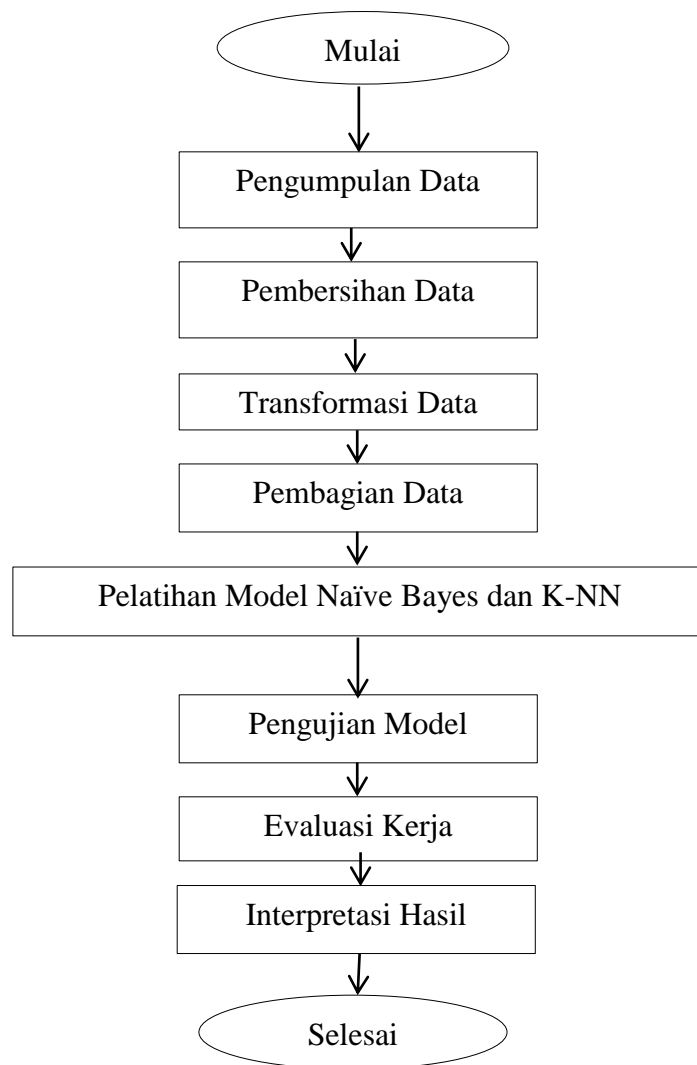
	Soeharso
Judul	Prediksi Penyebaran Kausus Demam Berdarah Dengue (DBD) Dengan Metode K-Nearest Neighbor di Kabupaten Karawang
Tahun	2021
Hasil	Tujuan dari penelitian ini yakni agar dapat dengan segera mengetahui wilayah mana yang tingkat persebaran paling tinggi hingga paling rendah untuk melakukan tindakan dan pencegahan yang sesuai. Pada referensi penelitian ini menggunakan data pada tahun 2016-2020 menggunakan 5 data parameter sebagai input dengan total 150 records. Hasil percobaan menghasilkan <i>performance</i> dengan nilai <i>accuracy</i> terbaik sebesar 81.27% menggunakan nilai K=5 dan 10-Fold Cross Validation. Hasil Prediksi ditunjukkan dengan petaan wilayah Kabupaten Karawang berdasarkan kecamatan sebanyak 18 kecamatan

	dengan tingkat persebaran tinggi, 7 kecamatan dengan tingkat persebaran sedang dan 5 kecamatan dengan tingkat persebaran rendah.
Referensi	3
Penulis	Kusumastuti, Evelyn Anggita, Mauridhi Hery Purnomo, dan Eko Mulyanto Yuniarno
Judul	Prediksi Penyebaran Kasus Demam Berdarah Berdasarkan Tingkat Keparahan di Provinsi DKI Jakarta Menggunakan Metode <i>Deep Learning</i>
Hasil	Tujuan Dari penggunaan metode ini adalah agar dapat menghasilkan prediksi yang lebih akurat terhadap daerah rawan peningkatan kasus DBD guna mendapatkan tindakan preventif yang lebih. Hal ini dapat dilihat dari akurasi hasil prediksi probabilitas stasiun Meteorologi Kemayoran sebesar 95,24; Stasiun Meteorologi Halim Perdana Kusuma sebesar 84,32; Stasiun Maritim Meteorologi Tanjung Priok

	sebesar 82,61.
--	----------------

2.9 Kerangka Kerja Penelitian

Untuk mempermudah penelitian ini maka peneliti membuat flowchart sebagai berikut:



Gambar 2.9 Kerangka Kerja Penelitian

Pada Gambar 2.9 diatas terlihat jelas bahwa penelitian ini dilakukan secara systematik dan bertahap, adapun penjelasan mengenai tahapan tersebut diantaranya adalah sebagai berikut:

1. Pengumpulan Data

Data dikumpulkan dari laporan history kasus DBD 1-2 tahun, data ini dapat mencakup jumlah pasien yang terkena DBD, wilayah tempat tinggal, umur, dan lain sebagainya.

2. Pembersihan Data

Pada tahap ini, data diperiksa untuk memastikan kualitasnya, dengan menghapus data yang tidak lengkap, duplikat, atau tidak relevan. Pembersihan ini dilakukan untuk meminimalisasi kesalahan analisis data yang tidak konsisten.

3. Transformasi Data

Data dikonversi ke dalam bentuk numerik atau skala yang seragam, memastikan bahwa variabel yang berbeda dapat dibandingkan secara langsung dalam model prediksi.

4. Pembagian Data:

Dataset dibagi menjadi set pelatihan (training set) dan set pengujian (testing set), dengan proporsi seperti 80:20 untuk memastikan bahwa model dapat diuji pada data baru.

5. Pelatihan Model:

Algoritma Naïve Bayes dilatih dengan menggunakan data pelatihan untuk menghitung probabilitas prior dan likelihood dari setiap variabel.

6. Pengujian Model:

Model diuji menggunakan set pengujian untuk menghasilkan prediksi pada data yang belum pernah digunakan dalam pelatihan.

7. Evaluasi Model:

Model dievaluasi dengan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur performa klasifikasi.