

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka**

##### **2.1.1 Kesehatan Masyarakat dan Penyakit Menular**

Kesehatan masyarakat merupakan cabang ilmu kesehatan yang fokus pada upaya kolektif untuk meningkatkan kualitas hidup dan kesejahteraan masyarakat secara menyeluruh. Menurut Turnock (2020), kesehatan masyarakat adalah seni dan ilmu dalam mencegah penyakit, memperpanjang harapan hidup, dan meningkatkan kesehatan melalui usaha terorganisir dari masyarakat, baik melalui edukasi, kebijakan, maupun layanan kesehatan. Kesehatan masyarakat tidak hanya berfokus pada individu, tetapi mencakup populasi secara keseluruhan, dengan tujuan menciptakan lingkungan yang sehat dan sistem layanan yang merata.

Penyakit menular adalah jenis penyakit yang disebabkan oleh agen infeksius seperti virus, bakteri, atau parasit, dan dapat menyebar dari satu individu ke individu lain, baik secara langsung maupun tidak langsung. Penyakit menular seperti Demam Berdarah Dengue (DBD), Tuberkulosis (TBC), dan Diare ketiganya memiliki karakteristik epidemiologis yang berbeda, tetapi sama-sama memiliki hubungan erat dengan faktor lingkungan, perilaku hidup bersih, sanitasi, dan kepadatan penduduk. Menurut Notoatmodjo (2021), “penyakit menular merupakan penyakit yang disebabkan oleh agen biologis, yang dapat menyebar secara langsung atau tidak langsung dari satu individu ke individu lainnya melalui media

lingkungan”. Penyakit ini sering kali berkaitan erat dengan sanitasi lingkungan, kualitas air, dan akses terhadap fasilitas kesehatan.

#### **2.1.1.1 Definisi dan Karakteristik Demam Berdarah Dengue (DBD)**

Demam Berdarah Dengue (DBD) merupakan penyakit tular vektor yang disebabkan oleh virus dengue (DENV). Virus dengue termasuk ke dalam famili Flaviviridae dengan genus Flavivirus yang terbagi menjadi serotipe DENV-1, DENV-2, DEN-3, DENV-4. Penyakit ini ditularkan melalui perantara vektor nyamuk *Aedes aegypti*. 1–4 DBD menjadi salah satu masalah kesehatan masyarakat secara global dengan tingginya insiden, tingkat keparahan penyakit, dan banyaknya daerah endemik. (Salamah et al., 2023)

Menurut (Nugraheni et al., 2023) Gejala DBD meliputi demam tinggi mendadak, nyeri otot dan sendi, sakit kepala, mual, muntah, ruam kulit, serta pendarahan ringan seperti mimisan atau gusi berdarah. Faktor-faktor yang menyebabkan tingginya kasus DBD meliputi lingkungan tempat tinggal yang lembap dan tidak bersih, kondisi cuaca seperti musim hujan yang memperbanyak genangan air (tempat berkembang biaknya nyamuk), kepadatan penduduk, serta kurangnya kesadaran masyarakat dalam menerapkan PHBS (Perilaku Hidup Bersih dan Sehat). Selain itu, minimnya program fogging dan kurangnya sistem drainase yang baik juga menjadi pemicu peningkatan penyebaran virus dengue. (Santoso et al., 2025)

### **2.1.1.2 Definisi dan Karakteristik TBC**

Tuberkulosis (TBC) adalah penyakit menular yang disebabkan oleh infeksi bakteri *Mycobacterium tuberculosis*. Bakteri ini terutama menyerang paru-paru, namun dapat juga menyebar ke organ lain melalui aliran darah atau sistem limfatik. Penularan terjadi melalui droplet udara saat penderita batuk atau bersin (Kaka et al., 2021)

Gejala umum TBC meliputi batuk berdahak yang berlangsung lebih dari dua minggu, demam, keringat malam, dan penurunan berat badan. Diagnosis TBC biasanya dilakukan melalui pemeriksaan dahak, rontgen dada, dan uji tuberkulin. Pengobatan melibatkan regimen antibiotik jangka panjang untuk mencegah resistensi obat,

Menurut Direktorat P2P Kemenkes RI (2022) Faktor utama penyebab penularan TBC antara lain adalah ventilasi udara yang buruk, tingginya kepadatan tempat tinggal, serta kekebalan tubuh yang rendah, Kondisi sosial ekonomi seperti kemiskinan dan sanitasi lingkungan yang rendah juga turut memperburuk penyebaran TBC. Selain itu, akses terbatas ke fasilitas kesehatan menyebabkan keterlambatan dalam diagnosis dan pengobatan, yang mempermudah penularan.

### **2.1.1.3 Definisi dan Karakteristik Diare**

Diare adalah kondisi yang ditandai dengan peningkatan frekuensi buang air besar (tiga kali atau lebih dalam sehari) dengan konsistensi tinja yang lebih cair dari biasanya. Penyebab diare dapat berupa infeksi oleh bakteri, virus, atau parasit, serta faktor non-infeksi seperti intoleransi makanan atau efek samping obat (Mustain et al., 2024)

Menurut World Health Organization (WHO), Salah satu penyebab morbiditas dan mortalitas tertinggi di dunia, terutama di negara berkembang, adalah diare. Sebagian besar diare terjadi karena makanan dan air tercemar. WHO melaporkan bahwa 2,5 miliar orang tidak memiliki sanitasi yang baik dan 780 juta orang memiliki akses yang buruk untuk sumber air minum. Selama tahun 2018, ada 3 Riset Kesehatan Dasar yang mencatat kasus diare. 1.017.290 kasus diare dilaporkan

### **2.1.2 Machine Learning**

Machine Learning (ML) merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem untuk belajar dari data tanpa diprogram secara eksplisit. Menurut (Géron, 2019) machine learning adalah sistem komputer dilatih untuk mengenali pola, membuat prediksi, dan mengambil keputusan berdasarkan pengalaman (data historis).

Dalam konteks kesehatan, machine learning digunakan secara luas untuk analisis prediktif, klasifikasi penyakit, serta pengenalan pola tren epidemiologi. algoritma machine learning juga dapat meningkatkan efisiensi prediksi penyebaran penyakit DBD dengan akurasi yang lebih tinggi dibandingkan metode statistik konvensional. Hal ini dikarenakan machine learning mampu menangani data dengan banyak variabel (high dimensional data) dan mampu menemukan pola yang kompleks antar variabel.(Tsan et al., 2022)

Machine learning terbagi menjadi beberapa pendekatan utama, salah satunya adalah supervised learning, di mana model dilatih menggunakan data yang sudah diberi label (misalnya, data jumlah kasus penyakit di masa lalu sebagai label).

Dalam supervised learning, terdapat dua jenis tugas utama yaitu klasifikasi (classification) dan regresi (regression). Menurut (Zhang et al., 2024), regresi lebih cocok untuk prediksi nilai numerik, seperti jumlah kasus penyakit, sedangkan klasifikasi digunakan untuk memetakan tingkat risiko (rendah, sedang, tinggi).

Seiring dengan perkembangan teknologi, implementasi machine learning dalam bidang kesehatan telah mencakup berbagai jenis algoritma seperti Decision Tree, Random Forest, Support Vector Machine, hingga Deep Learning. Dalam studi oleh (Nasir et al., 2024), penggunaan machine learning tidak hanya meningkatkan ketepatan prediksi, tetapi juga mempercepat waktu respon dalam perencanaan intervensi kesehatan masyarakat.

### **2.1.3 Data Mining**

Menurut Yahya (2022), *data mining* merupakan proses untuk menggali informasi penting dan pola tersembunyi dari kumpulan data yang besar, dengan menggunakan teknik statistik, kecerdasan buatan, dan pembelajaran mesin. Proses ini tidak hanya mencakup pencarian pola, tetapi juga analisis prediksi, klasifikasi, pengelompokan (clustering), dan asosiasi. Data mining menjadi salah satu komponen utama dalam sistem pengambilan keputusan berbasis data karena mampu menghasilkan pengetahuan yang bermanfaat dari data yang kompleks. Proses ini terbagi menjadi dua pendekatan utama, yaitu *supervised learning*, yang memerlukan pembuatan model dan variabel target seperti dalam klasifikasi dan regresi, serta *unsupervised learning*, yang tidak memerlukan variabel target dan bekerja berdasarkan struktur alami data.

#### **2.1.4 Prediksi**

Prediksi merupakan bagian penting dalam data mining dan pembelajaran mesin (machine learning) yang digunakan untuk memperkirakan nilai atau kejadian masa depan berdasarkan pola yang ditemukan dalam data historis. Dalam konteks ilmu komputer, prediksi sering diterapkan dalam berbagai bidang seperti keuangan, kesehatan, cuaca, dan lainnya, termasuk untuk memprediksi jumlah kasus penyakit di wilayah tertentu.

Menurut Han, Pei, dan Kamber (2022), prediksi adalah proses menggunakan model yang dibangun dari data historis untuk memperkirakan nilai masa depan dari atribut tertentu. Prediksi tidak selalu membutuhkan label kelas seperti dalam klasifikasi, melainkan bisa juga berfokus pada nilai numerik berkelanjutan, yang umumnya dilakukan dengan teknik regresi. Ini menjadikan prediksi sebagai bagian dari pendekatan supervised learning.

Proses prediksi dilakukan setelah model pohon keputusan selesai dibangun, dengan menggabungkan hasil dari setiap pohon. Untuk masalah klasifikasi, prediksi akhir ditentukan berdasarkan suara terbanyak (voting) dari seluruh pohon. Sementara itu, dalam kasus regresi, hasil prediksi diperoleh dengan menghitung rata-rata (mean) atau nilai tengah (median) dari semua prediksi yang dihasilkan oleh pohon-pohon tersebut (Rofiani et al., 2024)

#### **2.1.5 Algoritma Decision Tree Regression**

Decision Tree Regression adalah metode supervised learning non-parametrik yang digunakan untuk tugas prediksi nilai kontinu. Berbeda dengan classification

tree yang memprediksi label kelas, regression tree memprediksi nilai numerik riil. Model ini bekerja dengan cara mempartisi ruang fitur secara rekursif menjadi sejumlah persegi panjang (atau hyper-rectangle) yang lebih kecil dan sederhana.

Tujuan utama dari proses partisi ini adalah untuk meminimalkan variasi (variance) dari variabel target di dalam setiap subset (region) yang terbentuk. Untuk setiap region, nilai prediksi yang diberikan adalah nilai rata-rata dari semua sampel data training yang jatuh ke dalam region tersebut.

Proses pemilihan fitur dan titik pemisahan (split point) terbaik didasarkan pada kriteria yang dapat meminimalkan error prediksi. Metrik yang paling umum digunakan untuk mengukur error atau impuritas pada sebuah node dalam regression tree adalah Mean Squared Error (MSE).

$$\text{MSE}(D) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

**Keterangan:**

- $D$  adalah himpunan data pada node yang dianalisis.
- $n$  adalah jumlah total sampel data pada node  $D$ .
- $y_i$  adalah nilai target dari sampel data ke- $i$ .
- $\bar{y}$  adalah nilai rata-rata dari semua nilai target ( $y$ ) pada node  $D$ . Nilai  $\bar{y}$  ini juga merupakan nilai prediksi untuk node tersebut.

Setelah menghitung Mean Squared Error (MSE), langkah berikutnya adalah Algoritma akan memilih fitur dan titik pemisahan yang memberikan penurunan MSE paling besar. Penurunan ini dikenal sebagai Variance Reduction (VR). Split terbaik adalah yang mampu menghasilkan node-node anak (child nodes) dengan

total MSE tertimbang yang paling kecil dibandingkan MSE node induknya (parent node). Rumus perhitungannya adalah:

$$VR(D, \text{split}) = \text{MSE}(D_{\text{parent}}) - \left( \frac{n_{\text{left}}}{n_{\text{parent}}} \text{MSE}(D_{\text{left}}) + \frac{n_{\text{right}}}{n_{\text{parent}}} \text{MSE}(D_{\text{right}}) \right)$$

**Keterangan:**

- $\text{MSE}(D_{\text{parent}})$  adalah MSE dari node sebelum dipisah.
- $\text{MSE}(D_{\text{left}})$  dan  $\text{MSE}(D_{\text{right}})$  adalah MSE dari node anak kiri dan kanan setelah dipisah.
- $n_{\text{parent}}$ ,  $n_{\text{left}}$ , dan  $n_{\text{right}}$  adalah jumlah sampel di masing-masing node.

Proses ini diulang secara rekursif untuk setiap node anak hingga mencapai kondisi berhenti, seperti kedalaman maksimum pohon (`max_depth`) atau jumlah minimum sampel per daun (`min_samples_leaf`). Pendekatan ini memungkinkan model untuk mempelajari pola data secara hierarkis dan menghasilkan prediksi yang akurat.

Menurut (I. S. I. Putri et al., 2024) algoritma Decision Tree Regression mampu menangani data numerik dengan sangat baik dan menghasilkan model yang dapat dijelaskan secara visual dan logis, sehingga sangat cocok digunakan dalam konteks kebijakan publik.

### 2.1.6 CRISP-DM (Cross Industry Standard Process for Data Mining)

CRISP-DM merupakan metodologi yang paling banyak digunakan dalam proyek data mining. Proses ini terdiri dari enam tahap: business understanding, data understanding, data preparation, modeling, evaluation, dan deployment.

Menurut Chapman et al. (2020), “CRISP-DM menyediakan pendekatan sistematis dan fleksibel untuk membangun solusi data mining yang efektif dan efisien di berbagai sektor, termasuk kesehatan”.

### **2.1.7 Perbedaan Konseptual Regresi dan Decision Tree Regression**

Dalam bidang *machine learning*, penting untuk membedakan antara jenis masalah yang dihadapi dengan algoritma yang digunakan untuk menyelesaikannya. Dalam konteks penelitian ini, perlu dipahami perbedaan antara Regresi dan Decision Tree Regression.

Regresi bukanlah sebuah algoritma, melainkan salah satu jenis tugas pemodelan prediktif. Tujuan dari tugas regresi adalah untuk memprediksi atau mengestimasi sebuah nilai keluaran (*output*) yang bersifat numerik dan kontinu. Contoh dari nilai kontinu adalah jumlah kasus penyakit, harga saham, suhu udara, atau data numerik lain yang nilainya tidak terbatas pada kategori tertentu. Tugas ini berbeda dengan klasifikasi, yang bertujuan untuk memprediksi label kategori diskrit (misalnya, 'Tinggi', 'Sedang', 'Rendah').

Sementara itu, Decision Tree Regression adalah nama dari salah satu algoritma atau metode spesifik yang digunakan untuk menyelesaikan tugas regresi. Algoritma ini bekerja dengan membangun model prediktif dalam bentuk struktur pohon untuk memetakan hubungan antara fitur input dengan hasil numerik kontinu.

Secara analogi, jika "Regresi" adalah masalahnya (misalnya, "memprediksi jumlah kasus penyakit"), maka "Decision Tree Regression" adalah salah satu solusinya. Terdapat berbagai algoritma lain yang juga dapat digunakan untuk

menyelesaikan tugas regresi, seperti Regresi Linear (*Linear Regression*), *Support Vector Regression* (SVR), dan *Random Forest Regression*.





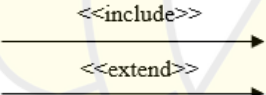
### **2.1.8 Pemodelan UML**

Unified Modeling Language (UML) adalah sebuah bahasa standar yang digunakan untuk memodelkan perangkat lunak (Khairunnisa et al., 2024), UML merupakan hasil integrasi dari beberapa bahasa pemodelan sebelumnya, seperti Booch, Object Modeling Technique (OMT), dan Object-Oriented Software Engineering (OOSE). Metode ini mendukung proses analisis dan desain yang dilakukan secara iteratif, termasuk langkah-langkah seperti identifikasi kelas dan objek, penentuan semantik hubungan antarobjek dan kelas, serta pengembangan antarmuka dan implementasi (Voutama, 2022).

#### **2.1.8.1 Use Case Diagram**

*Use case diagram* merupakan pemodelan untuk kelakuan (behavior) sistem informasi yang akan dibuat (Hendini, 2020), *Use Case Diagram* digunakan untuk menggambarkan perilaku sistem informasi yang akan dikembangkan. Diagram ini bekerja dengan menjelaskan pola interaksi khas antara pengguna (*user*) dan sistem tersebut melalui narasi atau skenario yang menggambarkan cara sistem digunakan. Diagram ini membantu memvisualisasikan hubungan antara aktor (pengguna atau sistem eksternal) dan fungsionalitas utama sistem, sehingga memudahkan pemahaman kebutuhan sistem (Wira et al., 2019). Berikut adalah komponen yang terdapat pada Use Case Diagram:








Table 2.1 Komponen komponen Use Case (Fitria et al., 2020).

Simbol	Nama Komponen	Fungsi
	Aktor ( <i>Actor</i> )	untuk memberikan input ke sistem, menerima informasi dari sistem, atau melakukan keduanya—mengirim dan menerima informasi.
	Dependency	Untuk menghubungkan antara dua elemen, di mana perubahan pada satu elemen memengaruhi elemen lainnya
	Association	hubungan antara elemen-elemen struktural yang saling terhubung dalam sebuah sistem.
	Generalization	Untuk menggambarkan relasi khusus antara objek anak ( <i>child</i> ) dan induk ( <i>parent</i> ).
	Include dan extends	include menggambarkan perilaku yang wajib dipenuhi agar suatu <i>event</i> dapat terjadi extends menunjukkan perilaku tambahan yang hanya terjadi di bawah kondisi tertentu

### 2.1.8.2 Diagram Activity

Menurut (Wira et al., 2019), Activity Diagram adalah diagram yang digunakan untuk menggambarkan alur kerja (workflow) atau aktivitas dalam suatu sistem perangkat lunak. Diagram ini memvisualisasikan urutan aktivitas, keputusan, dan aksi, memberikan pemahaman yang jelas tentang proses sistem yang sedang berjalan Berikut komponen yang terdapat dalam Activity Diagram:

Table 2.2 Komponen dalam activity diagram (Fitria et al., 2020).

Simbol	Nama Komponen	Fungsi
 	<ul style="list-style-type: none"> <li>• <i>Start State</i></li> <li>• <i>End State</i></li> </ul>	<ul style="list-style-type: none"> <li>• Sebagai tanda awal proses dari <i>activity diagram</i>.</li> <li>• menggambarkan akhir atau terminal dari pada sebuah <i>activity diagram</i></li> </ul>
	<i>Activity</i>	mengambarkan sebuah pekerjaan atau tugas dalam workflow
	<i>Decision</i>	mengindikasikan suatu kondisi kemungkinan perbedaan transisi
	<i>State Transition</i>	menunjukkan kegiatan apa berikutnya setelah suatu kegiatan sebelumnya
	<i>Fork/percabangan</i>	menunjukkan kegiatan yang dilakukan secara paralel atau untuk menggabungkan dua kegiatan paralel menjadi satu
	<i>Join (penggabungan)</i>	menunjukkan adanya dekomposisi

### 2.1.8.3 Sequence Diagram

*Sequence Diagram* digunakan untuk menggambarkan perilaku objek dalam suatu *use case* dengan memvisualisasikan waktu hidup objek serta pesan-pesan yang dikirimkan dan diterima antar objek. Diagram ini dibuat berdasarkan jumlah *use case* yang telah didefinisikan, terutama yang memiliki proses spesifik atau

signifikan. Setiap interaksi pesan dalam *use case* harus tercakup dalam *sequence diagram*. Oleh karena itu, semakin banyak *use case* yang didefinisikan, semakin banyak pula *sequence diagram* yang harus dibuat untuk mencakup semua interaksi sistem secara menyeluruh.(Wira et al., 2019).

### **2.1.9 Database dan DBMS**

Database adalah struktur sistematis yang menggabungkan "basis" sebagai pusat operasi untuk mengelola informasi, dan "data" sebagai fakta yang digunakan dalam pengambilan keputusan. Kombinasi ini memungkinkan penyimpanan, pengelolaan, dan akses informasi secara efisien (Widiarti et al., 2024).

Menurut (Taipalus, 2024), DBMS (Database Management System) adalah perangkat lunak yang digunakan untuk mengelola dan mengatur database, yang merupakan kumpulan data yang saling terkait. DBMS memungkinkan aplikasi perangkat lunak untuk berinteraksi dengan database, dan secara kolektif, DBMS menyediakan antarmuka dan alat untuk mengatur struktur basis data, memanipulasi data, dan menjaga keamanan dan integritas data (Siregar et al., 2024).

#### **2.1.9.1 SQL (Structured Query Language)**

Language Structured Query Language atau yang disingkat SQL Menurut (Siregar et al., 2024), SQL merupakan bahasa pemrograman yang digunakan untuk mengakses basis data relasional dan memudahkan praktisi data dalam mengolah data. Saat ini hampir semua server basis data yang ada mendukung bahasa ini untuk melakukan manajemen data dalam RDBMS

### 2.1.9.2 MySQL

MySQL adalah salah satu jenis database server yang sangat populer dan banyak digunakan untuk membangun aplikasi web yang memanfaatkan database sebagai sumber dan pengolah data. Sebagai Database Management System (DBMS) yang bersifat *open source*, MySQL mendukung fitur multiuser, *multithreaded*, dan tersedia secara gratis. Awalnya, MySQL didukung oleh bahasa pemrograman berbasis *script* untuk internet, seperti PHP dan perl. dengan fitur-fitur ini, MySQL menjadi alat yang andal dan fleksibel untuk pengembangan aplikasi berbasis data (Noviana, 2022; Siregar et al., 2024).

### 2.1.9.3 XAMPP

XAMPP adalah perangkat lunak berbasis *web server* yang bersifat *open-source* dan dapat digunakan pada berbagai platform, termasuk Windows, Linux, dan macOS. Nama XAMPP merupakan akronim dari (X-platform, Apache, MySQL, PHP, Perl). Alat ini sering digunakan sebagai *standalone server* atau yang dikenal dengan istilah *localhost*. Fungsinya mempermudah proses pengeditan, desain, serta pengembangan aplikasi.

Penggunaan XAMPP sangat penting dalam mengembangkan perangkat lunak atau antarmuka website secara lebih terstruktur, cepat, dan efisien. Alat ini terdiri dari tiga komponen utama: *htdocs*, *Control Panel*, dan *PhpMyAdmin*. XAMPP dapat dimanfaatkan sebagai alat bantu untuk memulai belajar proses pengembangan perangkat lunak sesuai kebutuhan proyek atau bisnis.(Noviana, 2022)

### **2.1.10 Software Dan Pemrograman Terkait**

Perancangan sistem prediksi berbasis machine learning membutuhkan dukungan perangkat lunak dan bahasa pemrograman tertentu. Dalam penelitian ini, digunakan Python sebagai bahasa pemrograman utama dengan bantuan beberapa pustaka (library) populer seperti Scikit-learn, Pandas, dan Matplotlib untuk membangun dan mengevaluasi model prediksi. Selain itu, editor seperti Jupyter Notebook digunakan untuk mengembangkan dan mendokumentasikan kode secara interaktif. Sedangkan penyajian hasil dilakukan melalui aplikasi web berbasis PHP.

#### **2.1.10.1 Editor Jupyter**

Jupyter Notebook adalah lingkungan komputasi interaktif berbasis web yang mendukung live code, visualisasi data, dan dokumentasi teks dalam satu dokumen (notebook). Jupyter Notebook merupakan *open-source interactive environment* yang mendukung bahasa pemrograman Python dan banyak digunakan dalam proyek *data science*.

Menurut (Rule et al., 2019), Jupyter telah menjadi standar de facto dalam penelitian data science karena kemampuannya mengintegrasikan kode, output, dan narasi secara dinamis. Penelitian ini menggunakan Jupyter untuk pengembangan model Decision Tree Regression dan eksplorasi data. Menurut (Kluyver et al., 2016) Jupyter memungkinkan pengguna untuk menulis kode, menjalankan analisis, dan menampilkan visualisasi data dalam satu dokumen interaktif. Editor ini digunakan dalam penelitian ini untuk proses pelatihan model Decision Tree Regression.

### 2.1.10.2 Bahasa Pemrograman Python dan Library Terkait

Python telah menjadi bahasa pemrograman dominan dalam bidang machine learning dan analisis data karena beberapa keunggulan mendasar. Sebagai bahasa tingkat tinggi dengan sintaks yang intuitif dan mudah dipelajari, Python memungkinkan peneliti untuk fokus pada penyelesaian masalah daripada menghabiskan waktu untuk masalah teknis pemrograman. (Virtanen et al., 2020) menjelaskan bahwa Python, didukung oleh ekosistem perpustakaan yang kaya, telah merevolusi komputasi ilmiah dengan membuat analisis data kompleks menjadi lebih mudah diakses dan efisien. Fleksibilitas Python dalam menangani berbagai tugas, mulai dari pemrosesan data hingga pembangunan model prediktif, menjadikannya pilihan utama dalam penelitian ini.

Untuk kebutuhan manipulasi data tabular, penelitian ini memanfaatkan library Pandas. (McKinney, 2022) dalam bukunya menjelaskan bahwa Pandas menyediakan struktur data DataFrame yang memungkinkan operasi seperti pembersihan data, transformasi, dan agregasi dilakukan dengan sintaks yang ringkas namun ekspresif. Kemampuan Pandas dalam menangani data missing value, melakukan merge dan join dataset, serta operasi time series sangat mendukung tahap pra-pemrosesan data dalam penelitian ini.

Sementara untuk komputasi numerik, NumPy menjadi pilihan utama. (Harris et al., 2020) menyoroiti bahwa NumPy tidak hanya menyediakan array multidimensi yang efisien, tetapi juga menjadi fondasi bagi banyak library ilmiah Python lainnya. Optimasi yang dilakukan NumPy pada operasi vektor dan matriks memungkinkan

komputasi numerik berjalan dengan performa tinggi, yang sangat krusial dalam proses pelatihan model machine learning.

Visualisasi data sebagai bagian penting dalam analisis eksploratori dan presentasi hasil dilakukan menggunakan Matplotlib dan Seaborn. (Waskom, 2021) mengembangkan Seaborn di atas Matplotlib untuk menyediakan antarmuka tingkat tinggi yang lebih intuitif dalam membuat visualisasi statistik yang menarik secara visual. Kombinasi kedua library ini digunakan dalam penelitian untuk mengeksplorasi distribusi data, hubungan antar variabel, serta mempresentasikan hasil prediksi model.

### **2.1.10.3 Library Scikit-learn**

Scikit-learn adalah salah satu library machine learning paling populer di Python. Library ini menyediakan berbagai algoritma klasifikasi, regresi, klustering, serta alat evaluasi model. library *machine learning* Python yang menyediakan alat untuk pra-pemrosesan data, pelatihan model, dan evaluasi.

(Buitinck et al., 2023) menekankan bahwa scikit-learn tetap menjadi pilihan utama untuk implementasi model *supervised learning* seperti *Decision Tree Regression* karena stabilitas dan optimasinya. Dalam penelitian ini, modul `sklearn.tree.DecisionTreeRegressor` digunakan untuk membangun model prediktif.

#### **2.1.10.4 Bahasa Pemrograman PHP**

PHP adalah (PHP Hypertext Preprocessor) adalah bahasa pemrograman web berbasis *server-side scripting* yang dirancang untuk pengembangan aplikasi web, namun juga dapat digunakan sebagai bahasa pemrograman umum (Noviana, 2022),

Penggunaan bahasa pemrograman PHP memungkinkan pengembang membuat aplikasi berbasis web secara lebih dinamis sehingga dalam mengelola dan perbaikan bias menjadi lebih mudah, efektif dan efisien (Ferizal et al., 2021).

### **2.2 Kajian Penelitian Terdahulu**

Penelitian terkait dan yang relevan dengan topik "Prediksi Jumlah Kasus Penyakit DBD, TBC, dan ISPA Berdasarkan Data Per Kecamatan dan Waktu di Kabupaten Bekasi Menggunakan Algoritma Decision Tree Regression" Berikut beberapa paper penelitiannya:

#### **2.2.1 Paper 1**

Penelitian yang dilakukan oleh (I. S. I. Putri et al., 2024) berjudul "*Decision Tree Regression untuk Prediksi Prevalensi Stunting di Provinsi Nusa Tenggara Timur*" yang di publikasikan pada Jurnal Teknologi Informatika dan Komputer MH. Thamrin juga telah terakreditasi Sinta 4 Penelitian ini membahas penggunaan algoritma Decision Tree Regression untuk memprediksi prevalensi stunting pada balita di Provinsi Nusa Tenggara Timur (NTT), daerah dengan tingkat stunting tertinggi di Indonesia menurut Riskesdas dan data nasional lainnya. Latar belakang penelitian ini menunjukkan bahwa stunting merupakan permasalahan serius di Indonesia yang disebabkan oleh berbagai faktor seperti kekurangan gizi, infeksi

kronis, sanitasi buruk, dan akses layanan kesehatan yang terbatas. Model prediksi ini dibuat sebagai alat bantu untuk memahami dan menanggulangi stunting secara lebih dini, berbasis data aktual.

Penelitian ini sangat relevan dengan skripsi ini karena memiliki kesamaan dalam hal penggunaan algoritma Decision Tree Regression sebagai metode yang diterapkan juga tujuan utamanya adalah prediksi nilai numerik kesehatan masyarakat (angka prevalensi), dan Fokus pada analisis data spasial dan temporal berdasarkan wilayah administratif (dalam hal ini: Provinsi NTT).

Penelitian ini menunjukkan bagaimana Decision Tree dapat digunakan untuk memahami faktor-faktor utama yang memengaruhi kondisi kesehatan, mirip seperti yang ingin dicapai dalam prediksi jumlah kasus DBD, TBC, dan Diare per kecamatan di Kabupaten Bekasi.

#### **2.2.1.1 Tujuan Penelitian**

Tujuan penelitian yang dilakukan oleh (I. S. I. Putri et al., 2024) untuk membuat model prediksi menggunakan Decision Tree Regression untuk memprediksi prevalensi stunting di NTT. Dengan demikian, penelitian ini selain menghasilkan model prediksi juga dapat memberikan pemahaman yang lebih komperhensif mengenai faktor-faktor yang mempengaruhi tingkat stunting di NTT dan mendukung upaya untuk menurunkan angka prevalensinya di provinsi tersebut. Untuk menguji model prediksi yang dihasilkan, penelitian ini menggunakan metrik RMSE.

### **2.2.1.2 Metodologi Yang Digunakan**

Penelitian ini menggunakan data prevalensi stunting yang diperoleh dari Badan Pusat Statistik (BPS) dan Kelompok Kerja (Pokja) Penanganan Stunting NTT tahun 2022. Metode yang diterapkan adalah Decision Tree Regression, dengan tahap preprocessing data meliputi normalisasi menggunakan Orange Data Mining untuk menstandarisasi rentang nilai fitur. Evaluasi kinerja model dilakukan dengan metrik RMSE (Root Mean Square Error), yang mengukur tingkat kesalahan prediksi antara nilai aktual dan nilai hasil prediksi. Hasil RMSE yang rendah menunjukkan bahwa model yang dibangun memiliki akurasi yang baik dalam memprediksi prevalensi stunting di Provinsi Nusa Tenggara Timur.

### **2.2.1.3 Temuan Utama**

Hasil dari penelitian ini menunjukkan bahwa model Decision Tree Regression memiliki nilai RMSE sebesar 0,093, yang berarti tingkat kesalahan prediksi relatif kecil. Hal ini menandakan bahwa model mampu memberikan prediksi yang cukup akurat. Selain itu, penelitian juga berhasil mengungkapkan bahwa faktor-faktor seperti akses sanitasi yang buruk, kurangnya pendidikan gizi, dan rendahnya kualitas pelayanan kesehatan merupakan determinan penting dari tingginya angka stunting di wilayah tersebut.

Model juga memberikan keunggulan dalam interpretabilitas karena Decision Tree Regression dapat menunjukkan jalur pengambilan keputusan yang jelas, misalnya: daerah dengan akses sanitasi rendah dan gizi buruk cenderung memiliki prevalensi stunting tinggi.

#### **2.2.1.4 Kesimpulan penelitian**

Penelitian ini membuktikan bahwa algoritma Decision Tree Regression sangat efektif dalam memprediksi nilai numerik berbasis data kesehatan masyarakat. Studi ini memanfaatkan data sekunder dari BPS dan Pokja Penanganan Stunting, serta menggunakan berbagai indikator seperti kemiskinan, tingkat imunisasi, dan bantuan persalinan. Dengan membangun model prediksi berbasis Python menggunakan library scikit-learn, mereka berhasil memperoleh nilai *Root Mean Square Error (RMSE)* yang rendah, yaitu 0,093, menandakan prediksi yang sangat akurat. Yang membuat penelitian ini sangat relevan dengan skripsi adalah bentuk output-nya yang berupa angka prediksi kasus serta pemanfaatan visualisasi pohon keputusan untuk mengidentifikasi faktor paling dominan dalam memengaruhi hasil. Dalam konteks skripsi ini, pendekatan tersebut dapat diaplikasikan secara serupa untuk memprediksi jumlah kasus penyakit DBD, TBC, dan Diare berdasarkan karakteristik tiap kecamatan di Kabupaten Bekasi.

#### **2.2.2 Paper 2**

Pada penelitian (Wardhana et al., 2023) yang berjudul "*Penerapan Machine Learning dalam Prediksi Tingkat Kasus Penyakit di Indonesia*", di Publikasikan pada Journal of Information System Management (JOISM), Bina Nusantara University Tahun 2023 yang terakreditasi Sinta 4 ini dilakukan perbandingan performa berbagai algoritma Machine Learning dalam memprediksi tingkat keparahan penyakit menular seperti DBD. Penelitian ini menunjukkan bahwa dari beberapa algoritma yang diuji, termasuk Random Forest dan XGBoost, algoritma

Decision Tree memberikan hasil akurasi tertinggi yaitu sebesar 83%. Meskipun model ini digunakan untuk klasifikasi tingkat keparahan, bukan prediksi jumlah kasus aktual, temuan ini memberikan landasan kuat bahwa Decision Tree sangat cocok digunakan untuk konteks data kesehatan. Relevansi terhadap skripsi ini terletak pada validasi bahwa Decision Tree memiliki kinerja unggul dan bersifat interpretable, sehingga sangat tepat diterapkan dalam sistem pendukung keputusan berbasis data untuk Dinas Kesehatan Kabupaten Bekasi.

#### **2.2.2.1 Tujuan Penelitian**

Penelitian ini bertujuan untuk membangun model klasifikasi tingkat kasus penyakit menular di Indonesia dengan menggunakan algoritma machine learning. Data yang digunakan berasal dari Profil Kesehatan Indonesia, yang mencakup variabel seperti tingkat pelayanan kesehatan, sosio-demografi, vaksinasi, dan prevalensi penyakit. Hasil model diharapkan bisa mendukung pengambilan keputusan cepat oleh pembuat kebijakan kesehatan.

#### **2.2.2.2 Metodologi Yang Digunakan**

Penelitian menggunakan pendekatan supervised learning, khususnya dengan algoritma Decision Tree Classifier. Data diambil dari Kementerian Kesehatan dan dinormalisasi ke dalam 5 kelas (skor 1–5) yang merepresentasikan tingkat penyebaran penyakit. Selain Decision Tree, model lain juga dibandingkan (Naive Bayes, ANN), tetapi Decision Tree menunjukkan hasil paling interpretatif dan akurat. Framework pengolahan data dilakukan dengan Python, melalui tahapan: data preparation, data modeling, training dan evaluasi akurasi model menggunakan data historis dari seluruh provinsi di Indonesia.

### **2.2.2.3 Temuan Utama**

Model Decision Tree yang dibangun dalam penelitian ini memberikan akurasi prediksi yang tinggi, dan dianggap efektif dalam mengelompokkan wilayah berdasarkan tingkat kasus penyakit. Hasil model mampu memberikan insight terhadap wilayah dengan tingkat kasus tinggi, dan variabel-variabel penting yang memengaruhinya. Model ini juga memberikan kelebihan dalam hal interpretabilitas karena pengguna dapat melihat struktur pohon keputusan yang menunjukkan jalur dan aturan klasifikasi.

### **2.2.2.4 Kesimpulan Penelitian**

Kesimpulan dari penelitian ini menyatakan bahwa machine learning, khususnya algoritma Decision Tree, memiliki potensi besar untuk digunakan dalam prediksi kasus penyakit menular sebagai bagian dari strategi transformasi digital kesehatan.

Relevansi terhadap penelitian skripsi ini sangat kuat karena sama-sama menggunakan pendekatan Decision Tree, meskipun dalam bentuk regresi, sama-sama mengandalkan data numerik dan spasial-temporal dari sektor kesehatan masyarakat, dan memiliki tujuan akhir yang sama: mendukung pengambilan keputusan kebijakan kesehatan daerah berdasarkan prediksi jumlah kasus penyakit.

Dengan menggunakan bentuk regresi (bukan klasifikasi seperti pada paper ini), penelitian skripsi ini justru menyempurnakan pendekatan dengan memperkirakan angka absolut jumlah kasus, bukan hanya kelompok risiko,

sehingga dapat digunakan langsung untuk perencanaan sumber daya kesehatan di Kabupaten Bekasi.

### **2.2.3 Paper 3**

Dalam kajian lain yang dilakukan oleh (N. G. Putri et al., 2019) penelitiannya yang berjudul "*Peramalan Jumlah Kasus Penyakit Hipertensi di Kabupaten Jember dengan Metode Time Series*", di Publikasi pada Journal of Health Science and Prevention Tahun 2019 dengan Klasifikasi jurnal Sinta 3 berfokus pada prediksi kasus hipertensi menggunakan pendekatan deret waktu (*time series*). Meski tidak menggunakan algoritma Machine Learning, penelitian ini tetap relevan karena menunjukkan pentingnya pemodelan prediksi berbasis waktu untuk perencanaan intervensi kesehatan jangka panjang. Namun, pendekatan time series cenderung hanya memperhatikan variabel waktu dan kurang mempertimbangkan faktor-faktor demografis lainnya seperti usia atau jenis kelamin. Dalam skripsi ini, pendekatan regresi menggunakan Decision Tree lebih unggul karena mampu memasukkan banyak variabel independen dalam satu model yang lebih kompleks namun tetap mudah diinterpretasikan.

#### **2.2.3.1 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk melakukan peramalan jumlah kasus hipertensi dalam kurun waktu 36 bulan mendatang di Kabupaten Jember. Prediksi ini ditujukan sebagai dasar untuk pengambilan keputusan dalam perencanaan pengendalian dan pengelolaan strategi pelayanan kesehatan terhadap penyakit hipertensi.

### **2.2.3.2 Metodologi Yang Digunakan**

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode peramalan deret waktu (time series). Teknik pemodelan dilakukan dengan bantuan aplikasi POM-QM, menggunakan model exponential smoothing. Data dikumpulkan dari catatan kasus hipertensi Dinas Kesehatan Jember, dengan pengelompokan berdasarkan jenis kelamin dan kelompok usia. Evaluasi hasil peramalan dilakukan dengan menggunakan MAPE (Mean Absolute Percentage Error) sebagai indikator ketepatan prediksi.

### **2.2.3.3 Temuan Utama**

Hasil peramalan dalam penelitian ini menunjukkan bahwa pola jumlah kasus hipertensi di Kabupaten Jember cenderung stabil, meskipun terdapat sedikit variasi berdasarkan waktu dan kelompok usia. Ketepatan model diukur menggunakan nilai MAPE (Mean Absolute Percentage Error), yang bervariasi antar kelompok. Untuk kelompok laki-laki, tingkat kesalahan prediksi berada pada angka 25,71%, sedangkan pada perempuan lebih rendah, yaitu 19,63%. Hasil yang jauh lebih akurat ditemukan pada kelompok usia 45 tahun ke atas, dengan nilai MAPE sebesar 13,98%. Sebaliknya, prediksi untuk kelompok usia 44 tahun ke bawah menunjukkan hasil yang kurang akurat, dengan nilai MAPE yang sangat tinggi mencapai 98,42%. Temuan ini mengindikasikan bahwa metode time series exponential smoothing cukup efektif digunakan untuk memprediksi jumlah kasus pada kelompok usia yang lebih tua, di mana pola data cenderung lebih stabil. Namun, metode ini kurang optimal untuk kelompok data yang bersifat fluktuatif atau tidak konsisten, seperti pada kelompok usia yang lebih muda.

#### **2.2.3.4 Kesimpulan Penelitian**

Penelitian ini menyimpulkan bahwa metode time series exponential smoothing dapat digunakan sebagai alat bantu dalam meramalkan jumlah kasus penyakit berdasarkan data historis. Namun, efektivitas model bergantung pada kestabilan pola data.

Relevansi terhadap penelitian skripsi ini terletak pada pendekatan prediktifnya yang sama-sama bertujuan menghasilkan informasi kuantitatif jumlah kasus penyakit dalam ruang dan waktu tertentu. Namun, penelitian skripsi ini menawarkan pendekatan yang lebih kuat dan fleksibel melalui Decision Tree Regression, yang dapat menangani banyak variabel input, non-linearitas, dan variabel iklim maupun non-klimatis, serta memberikan insight lebih detail dibanding metode time series konvensional.

#### **2.2.4 Paper 4**

Selanjutnya, penelitian (Kusumastuti et al., 2023) yang berjudul "*Prediksi Penyebaran Kasus Demam Berdarah Berdasarkan Tingkat Keparahan di Provinsi DKI Jakarta Menggunakan Metode Deep Learning*", Publikasi di Jurnal Teknik ITS Tahun 2023 Klasifikasi jurnal Sinta 4 penelitian ini menggunakan pendekatan Artificial Neural Network (ANN) untuk memprediksi tingkat keparahan kasus DBD dengan input faktor iklim seperti curah hujan, suhu, dan kelembapan. Hasilnya sangat akurat, dengan tingkat akurasi hingga 95% untuk beberapa wilayah di Jakarta. Namun, kelemahan dari pendekatan ini adalah model yang kompleks dan sulit diinterpretasikan secara langsung oleh pengambil kebijakan non-teknis.

Dalam konteks skripsi, penggunaan Decision Tree Regression justru lebih disarankan karena mampu menghasilkan model yang lebih transparan, berbentuk pohon keputusan yang mudah dipahami, serta cocok untuk diterapkan dalam pengambilan keputusan di instansi seperti Dinas Kesehatan.

#### **2.2.4.1 Tujuan Penelitian**

Penelitian ini bertujuan untuk memprediksi wilayah rawan lonjakan kasus DBD di DKI Jakarta berdasarkan tingkat keparahan kasus dan parameter iklim (curah hujan, suhu udara, kelembapan), menggunakan model ANN. Tujuannya adalah untuk menghasilkan sistem pendukung pengambilan keputusan dalam pencegahan kasus luar biasa (KLB) DBD.

#### **2.2.4.2 Metodologi Yang Digunakan**

Metodologi penelitian melibatkan pemodelan menggunakan Deep Learning dengan arsitektur Artificial Neural Network (ANN) yang terdiri dari input layer, beberapa hidden layer, dan output layer. Dataset yang digunakan berasal dari data iklim dan jumlah kasus DBD tahun 2016–2020 di lima wilayah Jakarta dan Kepulauan Seribu. Model dievaluasi menggunakan akurasi prediksi probabilitas di beberapa stasiun meteorologi (Kemayoran, Halim, Tanjung Priok), dan hasil menunjukkan tingkat akurasi tinggi: 95,24% di Kemayoran, 84,32% di Halim, dan 82,61% di Tanjung Priok.

#### **2.2.4.3 Temuan Utama**

Model ANN terbukti sangat efektif dalam memetakan potensi peningkatan kasus DBD berdasarkan parameter iklim dan spasial. Penelitian ini juga

mengonfirmasi bahwa faktor cuaca sangat berpengaruh signifikan terhadap dinamika penyebaran penyakit, khususnya pada daerah padat penduduk seperti Jakarta. Selain itu, model ini menunjukkan kekuatan ANN dalam mengenali pola non-linier dan kompleks.

#### **2.2.4.4 Kesimpulan**

Penelitian ini menyimpulkan bahwa pemodelan prediksi kasus DBD dengan Deep Learning dapat menghasilkan akurasi tinggi dalam mengidentifikasi wilayah rawan lonjakan kasus, namun membutuhkan volume data yang besar dan pelatihan model yang intensif. Relevansinya terhadap penelitian skripsi ini terletak pada Fokus penyakit yang sama (DBD), tujuan prediktif spasial-temporal yang sejalan, dan penggunaan data iklim sebagai fitur model.

Namun, pendekatan skripsi ini menggunakan Decision Tree Regression, yang lebih sederhana dan interpretatif dibanding ANN. Hal ini menjadi keunggulan tersendiri karena hasil model dapat lebih mudah dipahami dan digunakan oleh pembuat kebijakan, tanpa mengorbankan presisi prediktif. Selain itu, skripsi ini mencakup lebih banyak variabel non-iklim, seperti sanitasi, populasi, dan akses kesehatan, sehingga cakupannya lebih komprehensif dalam konteks perencanaan kesehatan daerah.

#### **2.2.5 Paper 5**

Terakhir, paper yang ditulis oleh (Mamenun et al., 2021) berjudul "*Kemajuan Penelitian Pemodelan Prediksi Demam Berdarah Dengue Menggunakan Faktor*

*Iklm di Indonesia: A Systematic Literature Review*", yang dipublikasikan pada Buletin Penelitian Kesehatan tahun 2021 dan juga telah terakreditasi Sinta 2 penelitian ini merupakan studi tinjauan pustaka sistematis (Systematic Literature Review/SLR) yang membahas berbagai pendekatan prediktif terhadap penyebaran kasus Demam Berdarah Dengue (DBD) di Indonesia, khususnya yang menggunakan variabel iklim seperti suhu, kelembapan, dan curah hujan. Penelitian ini sangat relevan sebagai referensi karena menyoroti kekuatan dan kelemahan berbagai algoritma prediksi (baik statistik maupun machine learning), termasuk Decision Tree, SVM, Random Forest, dan model statistik.

Fokusnya terhadap analisis spasial dan temporal, serta pengaruh iklim terhadap tren penyakit, selaras dengan skripsi ini yang memprediksi kasus penyakit DBD, TBC, dan Diare berdasarkan data kecamatan dan tahun, termasuk faktor iklim yang turut memengaruhi

#### **2.2.5.1 Tujuan Penelitian**

Tujuan utama dari kajian ini adalah untuk mengidentifikasi tren perkembangan penelitian prediksi kasus DBD di Indonesia, termasuk metode, model, serta faktor-faktor apa saja yang umum digunakan dalam pemodelan prediksi. Selain itu, kajian ini juga ingin menyoroti kesenjangan penelitian di Indonesia dibanding global, serta memberikan rekomendasi pengembangan riset ke depan, seperti integrasi faktor mobilitas, topografi, dan proyeksi jangka panjang.

#### **2.2.5.2 Metodologi Yang Digunakan**

Penelitian ini menggunakan metode Systematic Literature Review (SLR) yang dilakukan dengan menelusuri literatur melalui PubMed dan Google Scholar.

Kriteria seleksi meliputi publikasi ilmiah yang membahas prediksi kasus DBD di Indonesia dengan menggunakan variabel iklim. Sebanyak 32 artikel ilmiah yang diterbitkan dari berbagai daerah di Indonesia, seperti Jawa, Sumatera, Kalimantan, dan Nusa Tenggara Barat, diulas.

Pendekatan yang digunakan dalam kajian meliputi:

- Korelasi antara variabel iklim dan jumlah kasus DBD (dengan Spearman dan Pearson),
- Model prediksi berbasis regresi linier dan non-linier,
- Algoritma machine learning seperti SVM, Random Forest, dan Decision Tree.

### **2.2.5.3 Temuan Utama**

Penelitian menunjukkan bahwa variabel iklim seperti suhu, kelembapan, dan curah hujan secara konsisten berpengaruh terhadap naik-turunnya kasus DBD di berbagai wilayah. Mayoritas penelitian menggunakan data bulanan dan skala kota/kabupaten. beberapa algoritma machine learning seperti Decision Tree, Random Forest, dan SVM telah diterapkan dan terbukti memberikan performa prediksi yang baik, terutama jika data cukup bersih dan lengkap.

Namun, kajian ini juga mencatat bahwa penelitian di wilayah Indonesia timur masih sangat terbatas, dan faktor penting lain seperti mobilitas manusia, urbanisasi, serta proyeksi perubahan iklim ke depan belum banyak digunakan dalam model prediksi.

#### **2.2.5.4 Kesimpulan Penelitian**

Kesimpulan dari kajian ini menyatakan bahwa penelitian prediksi DBD di Indonesia telah berkembang dengan memanfaatkan berbagai metode statistik dan machine learning, khususnya dengan menggunakan variabel iklim seperti suhu, curah hujan, dan kelembapan. Namun, penelitian-penelitian tersebut masih memiliki keterbatasan karena minimnya integrasi faktor non-iklimatis, seperti kondisi sanitasi, kepadatan penduduk, dan akses terhadap layanan kesehatan.

Penelitian skripsi ini memiliki relevansi kuat dengan kajian tersebut, karena berupaya untuk melengkapi dan menyempurnakan pendekatan prediktif yang ada dengan cara menggabungkan variabel iklim dan non-iklim ke dalam satu model prediksi berbasis Decision Tree Regression. Dengan begitu, model yang dibangun tidak hanya mempertimbangkan pengaruh cuaca, tetapi juga faktor struktural dan demografis yang secara nyata memengaruhi penyebaran penyakit DBD, TBC, dan Diare di Kabupaten Bekasi. Pendekatan ini diharapkan menghasilkan prediksi yang lebih akurat dan kontekstual sesuai kondisi lokal.

Kelima penelitian ini secara keseluruhan memberikan dasar teoritis dan metodologis yang kuat untuk mendukung penelitian skripsi kamu. Pemilihan algoritma Decision Tree Regression terbukti tepat karena mampu menghasilkan model yang akurat, mudah dijelaskan, serta dapat digunakan sebagai dasar kebijakan yang konkret bagi pemerintah daerah dalam upaya preventif dan pengendalian penyakit menular di Kabupaten Bekasi.