

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan masyarakat merupakan salah satu aspek utama dalam pembangunan berkelanjutan. Di era digital saat ini, pendekatan berbasis teknologi informasi semakin penting dalam mendukung pengambilan keputusan di bidang kesehatan. Salah satu pendekatan yang berkembang pesat adalah pemanfaatan data mining dan machine learning, yang mampu mengolah data besar menjadi informasi yang berguna untuk prediksi penyakit, deteksi pola, dan perencanaan strategis. Penyakit menular seperti Demam Berdarah Dengue (DBD), Tuberkulosis (TBC), dan Diare masih menjadi ancaman serius di berbagai daerah di Indonesia, termasuk di Kabupaten Bekasi. Ketiga penyakit ini termasuk dalam kategori penyakit menular yang berkaitan erat dengan faktor lingkungan, kepadatan penduduk, serta tingkat sanitasi dan kualitas layanan kesehatan masyarakat. Ketiganya memiliki dampak besar terhadap produktivitas masyarakat, dan sering kali memuncak pada kondisi tertentu seperti musim hujan atau kemarau. Dalam beberapa tahun terakhir, peningkatan jumlah kasus ketiga penyakit ini mengindikasikan perlunya upaya preventif dan perencanaan kesehatan yang mampu mengidentifikasi pola dan tren penyakit untuk membantu pengambilan keputusan berbasis data.

Kabupaten Bekasi sebagai salah satu wilayah penyangga Ibu Kota mengalami pertumbuhan penduduk yang pesat, tetapi tantangan di sektor kesehatan pun meningkat. Menurut catatan Dinas Kesehatan Kabupaten Bekasi hingga

pertengahan Mei 2024, tercatat 736 kasus DBD melebihi jumlah kasus sepanjang tahun 2023 yang sebanyak 726 kasus. Angka kematian akibat DBD pun tercatat sebanyak empat orang, tersebar di kecamatan Tambun dan Sukatani. Sementara itu, penyakit TBC juga menunjukkan angka yang mengkhawatirkan. Menurut Dinas Kesehatan Kabupaten Bekasi, kasus TBC mencapai 10.000 dalam kurun waktu Januari hingga September 2023, menempatkan daerah ini sebagai peringkat ke 4 tertinggi di Jawa Barat. Tak kalah penting, penyakit diare dan ISPA juga sering mengalami lonjakan pada musim kemarau, yang menunjukkan adanya siklus musiman penyakit yang dapat diprediksi. Peningkatan kasus yang terjadi secara berkala menunjukkan perlunya sistem prediksi berbasis data untuk mengantisipasi lonjakan di masa mendatang. Prediksi jumlah kasus penyakit per kecamatan dengan mempertimbangkan waktu (year-on-year) menjadi langkah penting untuk perencanaan sumber daya, logistik kesehatan, serta intervensi preventif seperti edukasi kebersihan, fogging, penyediaan obat, dan penempatan tenaga medis. Dengan pendekatan prediktif yang akurat, pemerintah daerah dapat menargetkan wilayah rawan secara lebih efisien dan terukur.

Penelitian sebelumnya telah banyak membahas prediksi penyakit berbasis machine learning. Menurut penelitian (Wardhana et al., 2023) menunjukkan bahwa algoritma Decision Tree memiliki akurasi tertinggi (83%) dalam memprediksi kasus DBD menggunakan variabel sosiodemografi. Namun, penelitian tersebut masih terbatas pada klasifikasi keparahan, bukan pada prediksi jumlah kasus aktual. Penelitian oleh (Kusumastuti et al., 2023) menggunakan model deep learning untuk memprediksi penyebaran DBD berdasarkan faktor iklim di DKI Jakarta, dengan

hasil akurasi sangat tinggi. Meski demikian, studi tersebut belum mempertimbangkan demografi lokal seperti usia dan rasio jenis kelamin. Oleh karena itu, penelitian ini hadir dengan pendekatan berbeda, yakni menggunakan Decision Tree Regression untuk memprediksi jumlah kasus aktual per tahun per kecamatan di Kabupaten Bekasi.

Pendekatan ini juga relevan dengan transformasi digital dalam bidang kesehatan yang menekankan pengambilan keputusan berbasis data. Dalam laporan Kementerian Kesehatan, sistem informasi kesehatan nasional mendorong integrasi data spasial dan temporal untuk prediksi penyakit dan perencanaan wilayah rawan wabah. Oleh karena itu, penelitian ini akan memberikan kontribusi nyata dalam menyediakan sistem pendukung keputusan yang dapat digunakan oleh Dinas Kesehatan Kabupaten Bekasi dalam menyusun strategi penanggulangan penyakit.

Dengan memanfaatkan algoritma Decision Tree Regression, penelitian ini bertujuan untuk membangun model prediktif jumlah kasus penyakit DBD, TBC, dan Diare berdasarkan data dari tahun 2020 hingga 2023 per kecamatan di Kabupaten Bekasi. Menurut (I. S. I. Putri et al., 2024) Decision Tree Regression mampu menangani variabel numerik dan bersifat interpretable, memudahkan dalam penyampaian hasil kepada pihak pemerintah daerah. Dibandingkan algoritma lain seperti Random Forest atau ANN yang kompleks, Decision Tree dapat membentuk model dalam bentuk aturan sederhana yang dapat diikuti secara logis. Model ini juga mampu mengidentifikasi variabel paling berpengaruh, seperti kepadatan penduduk, usia pasien, atau rasio jenis kelamin. Dengan kata lain, pemilihan

algoritma ini mempertimbangkan efisiensi, keakuratan, serta kemudahan interpretasi dalam konteks pengambilan keputusan kesehatan masyarakat.

Hasil prediksi dari model ini diharapkan mampu memberikan insight kepada pemerintah daerah dalam hal kebijakan preventif yang lebih tepat, termasuk distribusi fasilitas kesehatan, dan kegiatan promotif seperti edukasi PHBS (Perilaku Hidup Bersih dan Sehat) untuk masyarakat, dengan output berupa estimasi jumlah kasus di tahun mendatang yang dapat dijadikan dasar bagi Dinas Kesehatan dalam menyusun strategi antisipatif dan preventif. Dibandingkan penelitian terdahulu, keunggulan penelitian ini terletak pada cakupan lokal yang lebih spesifik, penggunaan variabel demografis seperti usia dan rasio jenis kelamin, serta output berupa angka prediksi kuantitatif, bukan hanya klasifikasi tingkat risiko. Penelitian ini juga sejalan dengan upaya digitalisasi perencanaan kesehatan yang berbasis data daerah secara berkelanjutan

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana membangun model prediksi jumlah kasus penyakit DBD, TBC, dan Diare per tahun di setiap kecamatan di Kabupaten Bekasi menggunakan algoritma Decision Tree Regression?
2. Seberapa akurat model Decision Tree Regression dalam memprediksi jumlah kasus penyakit menular tersebut berdasarkan variabel demografis dan waktu?

1.3 Batasan Masalah

Agar penelitian lebih terarah, batasan masalah yang ditetapkan dalam penelitian ini meliputi:

1. Penelitian hanya difokuskan pada tiga jenis penyakit menular yaitu Demam Berdarah Dengue (DBD), Tuberkulosis (TBC), dan Diare.
2. Data yang digunakan merupakan data sekunder yang diperoleh dari dinas kesehatan dan sumber resmi lain, dengan cakupan waktu dari tahun 2020 hingga 2023.
3. Variabel input yang digunakan antara lain jumlah penduduk, jumlah fasilitas kesehatan, rasio jenis kelamin pasien, serta rata-rata usia pasien per penyakit.
4. Model prediksi yang dibangun menggunakan algoritma Decision Tree Regression dan tidak dibandingkan secara teknis dengan algoritma lain.

1.4 Tujuan Dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Membangun model prediktif menggunakan algoritma Decision Tree Regression untuk memprediksi jumlah kasus penyakit DBD, TBC, dan Diare per tahun di setiap kecamatan di Kabupaten Bekasi.
2. Mengevaluasi tingkat akurasi model prediktif dalam memproyeksikan jumlah kasus berdasarkan variabel demografis dan waktu.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat diantaranya:

1. Penelitian ini memberikan kontribusi dalam pengembangan penerapan algoritma Decision Tree Regression untuk prediksi penyakit berbasis data wilayah dan waktu, serta memperkaya literatur dalam bidang data mining dan machine learning di sektor kesehatan masyarakat.
2. Model prediktif yang dihasilkan diharapkan dapat digunakan oleh Dinas Kesehatan Kabupaten Bekasi sebagai sistem pendukung keputusan dalam menyusun strategi antisipasi penyakit menular, seperti alokasi sumber daya, distribusi tenaga medis, serta kampanye promotif dan preventif.
3. Hasil dari penelitian ini diharapkan dapat membantu dalam menurunkan tingkat kasus penyakit melalui intervensi dini di wilayah-wilayah yang diprediksi mengalami lonjakan kasus, sehingga berdampak langsung pada peningkatan kualitas hidup masyarakat Kabupaten Bekasi.

1.5 Metode Penelitian

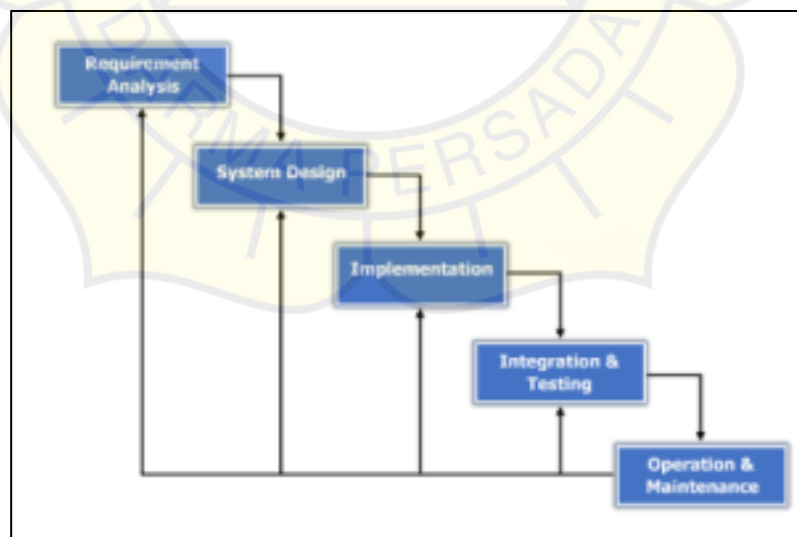
1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode wawancara yaitu dengan mengajukan pertanyaan langsung ke pengurus data sekunder dari Dinas Kesehatan Kabupaten Bekasi. Sumber data utama diperoleh dari Dinas Kesehatan Kabupaten Bekasi berupa data jumlah kasus penyakit DBD, TBC, dan Diare per tahun dan per kecamatan selama periode 2020 hingga 2023. Data tambahan yang bersifat demografis seperti jumlah penduduk per

tahun dan per kecamatan diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Bekasi. Selain itu, data lingkungan per tahun dan per kecamatan seperti curah hujan, suhu rata-rata, dan lainnya diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) sebagai faktor lingkungan yang dapat mempengaruhi penyebaran penyakit. Semua data dikumpulkan dalam format numerik dan akan dijadikan input dalam proses pemodelan prediksi.

1.5.2 Metode Pengembangan Sistem

Metodologi pengembangan sistem yang digunakan adalah metode Waterfall, Menurut (Harjono et al., 2022) Metode Waterfall merupakan metode yang biasanya digunakan dalam pengembangan perangkat lunak software. Perkembangan model ini berkembang secara sistematis dari satu tahap ke tahap lain dalam mode seperti air terjun yang meliputi tahapan berikut:

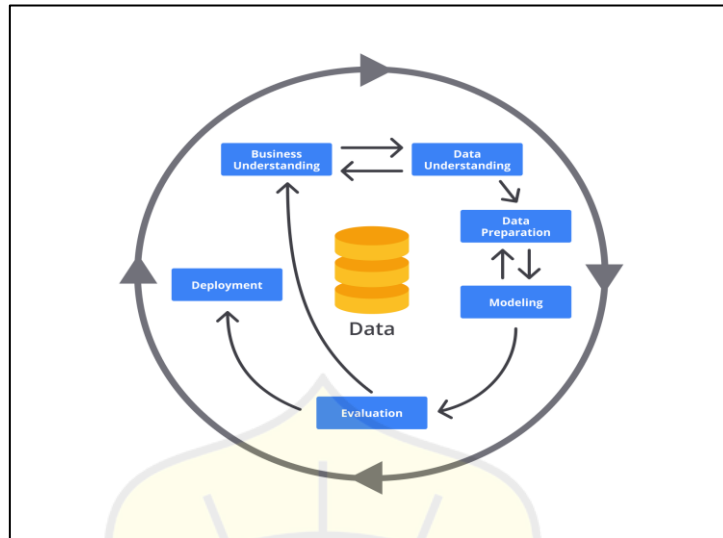


Gambar 1.1 Tahapan Metode Waterfall (Harjono et al., 2022)

1. Requirement Analysis, merupakan tahap-tahap menganalisis hal-hal yang diperlukan dalam pelaksanaan perancangan sistem.
2. System Design/Perancangan, dilakukan perancangan sistem sesuai solusi permasalahan yang terjadi menggunakan bahasan data dan perangkat pemodelan serta struktur.
3. Implementasi dan Kode Program, pada tahap ini akan di implementasikan ke dalam kode program. Kode program disusun berdasarkan prosedur algoritma Decision Tree Regression.
4. Testing/Pengujian, untuk menguji hasil dari program ini dengan melalui data penderita demam berdarah.
5. Deployment and Maintenance, tahap akhir dimana suatu sistem yang sudah selesai dapat mengalami perubahan-perubahan atau permintaan sesuai permintaan pengguna.

1.5.3 Metode Analisis Data

Metode analisis data dalam penelitian ini menggunakan pendekatan CRISP-DM (Cross Industry Standard Process for Data Mining), yang terdiri dari enam tahapan yang terdapat pada gambar 1.1 Metode CRISP-DM dan berikut penjelasannya:



Gambar 1.2 Metode CRISP-DM

1. Business Understanding

Tahap ini bertujuan untuk memahami permasalahan bisnis atau latar belakang penelitian. Permasalahan yang diangkat adalah tingginya jumlah kasus penyakit menular seperti DBD, TBC, dan diare di Kabupaten Bekasi dari tahun ke tahun, yang menimbulkan beban pada fasilitas kesehatan dan mengancam kesehatan masyarakat. Oleh karena itu, diperlukan sistem prediksi yang dapat membantu Dinas Kesehatan dalam mengambil keputusan berbasis data.

2. Data Understanding

Pada tahap ini dilakukan pengumpulan dan eksplorasi awal terhadap data yang digunakan. Data diperoleh dari Dinas Kesehatan Kabupaten Bekasi, mencakup jumlah kasus DBD, TBC, dan diare per kecamatan dari tahun 2020 hingga 2023. Selain itu, data demografis seperti jumlah penduduk, rata-rata usia pasien, dan rasio jenis kelamin pasien juga dikaji. Eksplorasi data dilakukan untuk

memahami struktur data, mendeteksi nilai ekstrem (outlier), dan mengetahui potensi anomali.

3. Data Preparation

Tahap ini mencakup proses pembersihan data dari nilai kosong atau duplikat, normalisasi atau transformasi variabel agar sesuai dengan format input model, serta pembagian data menjadi data latih (training) dan data uji (testing) dengan rasio umum 80:20. Tahap ini sangat penting untuk memastikan kualitas dan kelengkapan data sebelum dilakukan pemodelan.

4. Modeling

Pada tahap ini, algoritma Decision Tree Regression diterapkan menggunakan pustaka Python seperti Scikit-learn. Algoritma ini dipilih karena kemampuannya untuk membangun model prediktif yang interpretable dan dapat divisualisasikan dalam bentuk pohon keputusan. Model dilatih menggunakan data historis untuk mengenali pola dan hubungan antara variabel input dan output (jumlah kasus penyakit).

5. Evaluation

Setelah model terbentuk, dilakukan evaluasi kinerja menggunakan metrik seperti MAE (Mean Absolute Error), RMSE (Root Mean Square Error), dan R² Score. Evaluasi ini bertujuan untuk menilai sejauh mana model mampu memberikan prediksi yang akurat dan layak digunakan dalam konteks pengambilan keputusan nyata.

6. Deployment (Penerapan dan Penyajian)

Tahap terakhir adalah penyajian hasil model dalam bentuk tabel dan grafik visual yang mudah dipahami oleh pemangku kebijakan. Hasil prediksi dapat dimanfaatkan oleh Dinas Kesehatan untuk menentukan alokasi sumber daya kesehatan, distribusi tenaga medis, serta program pencegahan penyakit. Selain itu, sebagai bentuk penerapan sistem, dibangun prototipe aplikasi web menggunakan framework seperti Streamlit atau Flask, yang memungkinkan pengguna menginput data dan memperoleh hasil prediksi secara interaktif.

