

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digitalisasi saat ini, teknologi informasi telah berkembang pesat dan memberikan dampak signifikan dalam berbagai aspek kehidupan, termasuk dalam bidang kesehatan. Salah satu teknologi yang berperan penting adalah *data mining*, yang mampu mengolah data dalam jumlah besar untuk menghasilkan informasi dan pengetahuan yang bermanfaat. Pemanfaatan *data mining* dalam bidang kesehatan telah terbukti efektif dalam melakukan prediksi dan analisis berbagai masalah kesehatan, termasuk dalam penanganan masalah gizi pada balita (Xsanal Hakim et al., 2024)

Dalam konteks kesehatan balita, *data mining* dapat diimplementasikan untuk menganalisis dan memprediksi berbagai permasalahan gizi, salah satunya adalah *stunting*. *Stunting* merupakan kondisi pertumbuhan anak yang terhambat akibat kekurangan gizi kronis, yang telah menjadi salah satu isu kesehatan masyarakat yang serius di berbagai belahan dunia, termasuk Indonesia (Nurkholik & Rohita, 2024)

Menurut *United Nations International Children's Emergency Fund (UNICEF)*, satu dari tiga anak mengalami *stunting*, dan sekitar 40% anak di daerah pedesaan mengalami pertumbuhan yang terhambat (Ayini Lalu, 2020). Permasalahan gizi yang kerap menjadi tantangan kesehatan pada kelompok balita adalah kondisi gizi buruk atau malnutrisi. Fenomena ini muncul sebagai hasil interaksi antara faktor internal dan eksternal yang tidak mendapat perhatian preventif secara dini. Jika dibiarkan, kondisi gizi buruk dapat memberikan dampak jangka panjang yang serius, tidak hanya pada aspek kesehatan dan pertumbuhan balita saat ini, tetapi juga berpengaruh pada tingkat produktivitas mereka di masa mendatang (Lestari, 2022)

Di Indonesia, angka *stunting* masih cukup tinggi yaitu 21,6% berdasarkan hasil Survei Status Gizi Indonesia (SSGI) tahun 2022, meskipun

mengalami penurunan dari 24,4% pada tahun 2021. Upaya besar masih diperlukan untuk mencapai target penurunan *stunting* sebesar 14% pada tahun 2024 (Kemenkes, 2024). Secara khusus di Kabupaten Kuningan, berdasarkan laporan terbaru pada triwulan II tahun 2024, angka *stunting* tercatat sebesar 8,9 persen atau sekitar 6.115 balita *stunting* dari total 68.408 balita yang diukur tinggi badannya (Almutaqin & Nurhuda, 2024)

Posyandu (Pos Pelayanan Terpadu) adalah wadah komunikasi dalam pelayanan kesehatan masyarakat, yang dikelola oleh masyarakat dan untuk masyarakat, dengan dukungan pelayanan serta pembinaan teknis dari petugas kesehatan. Posyandu merupakan titik temu antara petugas kesehatan dan peran serta masyarakat dalam menanggulangi pelayanan kesehatan dasar (Saepudin et al., 2017)

Posyandu Sakura adalah salah satu pos pelayanan terpadu yang berada di Dusun Pahing, Desa Pancalang, Kecamatan Pancalang, Kabupaten Kuningan. Posyandu ini berperan penting dalam memberikan pelayanan kesehatan bagi ibu dan anak, terutama dalam memantau pertumbuhan dan perkembangan balita. Setiap bulan, Posyandu Sakura melaksanakan kegiatan rutin seperti penimbangan, pengukuran tinggi badan, serta vitamin bagi balita. sebagai unit pelayanan dasar dalam pemantauan kesehatan balita.

Posyandu sakura masih menghadapi berbagai kendala dalam pengelolaan data dan deteksi dini *stunting*. Sistem pencatatan berbasis kertas yang diterapkan menimbulkan beberapa masalah seperti kesulitan pengelolaan data, rentan terhadap kesalahan pencatatan, dan tidak adanya arsip digital. Selain itu, Proses pelaporan data dari Posyandu ke Puskesmas juga mengalami berbagai hambatan. Kader Posyandu harus mengumpulkan dan menyalin ulang data dari buku register ke format laporan bulanan, yang memakan waktu dan berisiko terjadi kesalahan pencatatan. Keterlambatan pengiriman laporan sering terjadi karena proses manual yang panjang dan keterbatasan waktu kader. Petugas kesehatan di Puskesmas juga mengalami kesulitan dalam menganalisis dan memantau tren pertumbuhan balita karena data yang tidak terorganisir

dengan baik dan tidak tersedia dalam format digital yang mudah diolah. sehingga identifikasi dini risiko *stunting* menyebabkan intervensi yang terlambat dan kurang efektif.

Dalam upaya untuk membantu menanggulangi masalah *stunting*, teknologi informasi seperti *data mining* diharapkan dapat dimanfaatkan. *Data mining* merupakan proses eksplorasi informasi bernilai dari kumpulan data berukuran besar. Di bidang komputer, istilah ini mengacu pada metodologi pengolahan data untuk menemukan pola-pola tersembunyi yang dapat menghasilkan wawasan baru dari serangkaian data. *Data mining* memiliki beberapa teknik analisis yang berbeda sesuai dengan tujuan pengolahan datanya, mencakup proses estimasi, prediksi, klasifikasi, pengelompokan (klasterisasi), dan pencarian hubungan (asosiasi) antar data. (Mai et al., 2022)

Salah satu algoritma yang bisa digunakan adalah C4.5, Algoritma C4.5 merupakan sebuah algoritma yang dikembangkan oleh Ross Quinlan untuk menciptakan pohon keputusan. Prinsip fundamental algoritma ini terletak pada pembentukan pohon keputusan dengan memilih atribut yang menjadi prioritas utama, ditentukan dari nilai gain tertinggi berdasarkan perhitungan *entropy* setiap atribut sebagai dasar klasifikasi. Dalam implementasinya, C4.5 beroperasi melalui dua tahap utama: pertama, konstruksi pohon keputusan, dan kedua, pembentukan model aturan. Hasil akhir dari proses ini adalah serangkaian aturan dalam format *if-then* yang diturunkan dari struktur pohon keputusan yang telah terbentuk. yang dapat membantu memprediksi risiko *stunting*. (Ade May Luky Harefa & Mhd. Ramadhan, 2024). Algoritma C4.5 ini akan bekerja dengan membangun pohon keputusan yang membantu mengidentifikasi data - data yang paling mempengaruhi terjadinya *stunting*. Dengan implementasi algoritma C4.5 di posyandu, petugas posyandu dapat memprediksi balita yang berisiko mengalami *stunting* lebih awal dan dapat melakukan intervensi tepat waktu. Studi kasus di Posyandu Sakura akan menjadi landasan penerapan teknologi ini di tingkat dusun, dengan harapan

membantu menurunkan angka *stunting* dan meningkatkan kualitas hidup balita di daerah sekitar posyandu.

Beberapa penelitian terdahulu telah menunjukkan efektivitas penggunaan *data mining*, khususnya algoritma C4.5, dalam prediksi *stunting*. Penelitian di Kota Dumai (Putri et al., 2024). Penelitian ini menerapkan perbandingan algoritma *Decision Tree* dan *Support Vector Machine* (SVM) untuk prediksi *stunting* di sepuluh kelurahan Kota Dumai. Relevansi dengan penelitian saat ini terletak pada penggunaan algoritma *Decision Tree* yang menghasilkan akurasi tinggi (96.15%) dalam memprediksi *stunting*. Namun, penelitian tersebut belum mengimplementasikan sistem dalam bentuk aplikasi yang dapat digunakan langsung oleh petugas posyandu. Dalam penelitian ini, akan mengembangkan sistem yang dapat digunakan oleh petugas posyandu, sehingga memungkinkan deteksi dini dan intervensi yang lebih cepat

Penelitian di Desa Purbayasa Tonjong (Hari Fajar & Rayendra, 2023). Implementasi algoritma C4.5 berbasis website untuk deteksi *stunting* di Pos Kesehatan Desa (PKD) menjadi referensi penting dalam aspek pengembangan sistem. Pendekatan berbasis website yang memudahkan akses sistem. Perbedaannya, penelitian yang akan dilakukan berfokus pada tingkat posyandu dengan karakteristik pengguna dan kebutuhan yang berbeda, serta menambahkan fitur prediksi *stunting*. Penelitian ini akan melengkapi dengan fitur prediksi *stunting* secara terintegrasi dan visualisasi data

Penelitian di Pagar Alam menghasilkan temuan penting dalam penerapan algoritma C4.5 untuk prediksi *stunting* dengan tingkat akurasi mencapai 88.20%. Relevansi dengan penelitian ini dapat dilihat dari keberhasilan penerapan algoritma C4.5 dalam mengategorikan status gizi balita secara otomatis berdasarkan *Z-score* yang telah ditetapkan. Meskipun penelitian tersebut telah berhasil mengembangkan sistem prediksi yang efektif, implementasinya masih terbatas pada penggunaan data yang telah dikumpulkan sebelumnya (data historis). Penelitian yang akan dilakukan akan mengembangkan sistem yang dapat memproses data secara *real-time* seiring

dengan kegiatan pemantauan rutin di posyandu, sehingga memungkinkan deteksi dini dan intervensi yang lebih cepat. Selain itu, penelitian ini akan menambahkan fitur visualisasi data dan *dashboard monitoring* untuk memudahkan kader posyandu dalam memantau perkembangan balita secara berkelanjutan. Penelitian yang akan dilakukan akan mengembangkan sistem yang dapat memproses data secara real-time seiring dengan kegiatan pemantauan rutin di posyandu, sehingga memungkinkan deteksi dini dan intervensi yang lebih cepat, serta menambahkan fitur visualisasi data dan dashboard monitoring. (Xsanal Hakim et al., 2024)

Dengan demikian, penelitian ini tidak hanya melanjutkan keberhasilan penelitian sebelumnya tetapi juga memberikan kontribusi baru dalam pengembangan sistem prediksi *stunting* yang lebih aplikatif di tingkat posyandu dusun. Berdasarkan permasalahan dan kajian penelitian terdahulu tersebut, penulis mengajukan penelitian berjudul **"Implementasi Data Mining Menggunakan Algoritma C4.5 Untuk Sistem Informasi Prediksi Stunting di Posyandu Sakura Desa Pancalang"**. Penelitian ini diharapkan dapat mengembangkan sistem prediksi *stunting* yang efektif dan efisien untuk membantu kader posyandu dalam melakukan deteksi dini dan intervensi tepat waktu, sehingga dapat berkontribusi dalam menurunkan angka *stunting* di tingkat dusun.

Pengembangan sistem ini akan mengadopsi pendekatan Agile untuk memastikan fleksibilitas dan adaptabilitas terhadap perubahan kebutuhan, serta menggunakan kerangka kerja CRISP-DM yang sistematis dalam tahapan data mining, mulai dari pemahaman bisnis hingga deployment

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, beberapa permasalahan yang diidentifikasi dalam penelitian ini adalah sebagai berikut:

1. Proses pencatatan data di Posyandu Sakura masih dilakukan dengan berbasis kertas, yang menyulitkan pengelolaan, pemrosesan data, dan

penentuan intervensi dini. Sehingga menyebabkan keterlambatan dalam mendeteksi balita berisiko, sehingga penanganan dan pencegahan *stunting* sering kali terlambat

2. Masih tingginya angka *stunting* di Kabupaten Kuningan, menunjukkan perlunya penanganan yang lebih terfokus, terutama pada balita di wilayah pedesaan karena belum adanya pemanfaatan teknologi untuk membantu proses deteksi risiko *stunting* pada tingkat posyandu.
3. Pelaporan data dari Posyandu menuju Puskesmas kerap terkendala oleh proses administratif yang rumit. Para kader harus melakukan pengumpulan dan penggandaan data secara manual dari buku register ke dalam format laporan bulanan, yang tidak hanya memerlukan waktu yang panjang, tetapi juga sangat rentan terhadap kesalahan pencatatan. Akibatnya, keterlambatan dalam pengiriman laporan menjadi masalah yang lazim terjadi, disebabkan oleh prosedur kerja yang berbelit-belit dan memakan waktu yang lama.

1.3 Rumusan Masalah

Berdasarkan identifikasi masalah di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan membangun sistem informasi prediksi *stunting* yang dapat membantu pengelolaan data untuk penanganan prediksi risiko *stunting* di Posyandu Sakura ?
2. Bagaimana mengimplementasikan algoritma C4.5 dalam membangun model prediksi *stunting* pada balita di Posyandu Sakura, Desa Pancalang, untuk membantu petugas Posyandu dalam melakukan deteksi dini risiko *stunting*?

1.4 Batasan Masalah

Agar penelitian ini lebih terarah maka diperlukan penetapan batasan masalah. Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya menggunakan data kesehatan balita yang tercatat di Posyandu Sakura, Desa Pancalang dan Kabupaten Kuningan.

2. Data yang digunakan mencakup berat badan, tinggi badan, jenis kelamin dan usia balita
3. Algoritma yang digunakan untuk prediksi adalah algoritma C4.5, tanpa menggabungkan metode atau algoritma lain dan dibuat sistem berbasis web.
4. Penelitian ini hanya mengembangkan model prediksi tanpa melakukan intervensi atau evaluasi dampak terhadap kondisi kesehatan balita
5. Sistem memiliki 3 level hak akses pengguna dengan batasan sebagai berikut:
 - a. Admin memiliki hak akses penuh terhadap sistem meliputi:
 1. Manajemen data pengguna (menambah, mengubah, dan menghapus akun petugas)
 2. Manajemen data master (data balita, data pengukuran, data hasil prediksi)
 3. Melihat dan mengunduh laporan
 - b. Petugas posyandu memiliki hak akses terbatas meliputi:
 1. *Input* data pengukuran balita
 2. Melihat riwayat pengukuran balita
 3. Melakukan prediksi *stunting*
 4. Melihat hasil prediksi
 5. Mencetak laporan sesuai format perbulan dan pertahun
 - c. Petugas Puskesmas memiliki hak akses terbatas meliputi:
 1. Melihat riwayat pengukuran balita
 2. Melihat hasil data prediksi *stunting*
 3. Mencetak laporan sesuai format perbulan dan pertahun

1.5 Tujuan Penelitian

Penelitian ini bertujuan untuk mencapai beberapa hasil berikut:

1. Merancang dan membangun sistem informasi prediksi *stunting* pada balita di Posyandu Sakura, Desa Pancalang, yang dapat membantu pengelolaan dan pemrosesan data untuk mendukung petugas Posyandu dalam melakukan deteksi dini risiko *stunting*.

2. Mengimplementasikan algoritma C4.5 dalam membangun model prediksi *stunting* pada balita di Posyandu Sakura, Desa Pancalang. yang dapat membantu petugas posyandu dalam melakukan deteksi dini risiko *stunting*

1.6 Manfaat Penelitian

1.6.1 Manfaat Teoritis

- a. Penelitian ini diharapkan dapat memperkaya kajian di bidang kesehatan masyarakat, khususnya terkait penerapan *data mining* untuk mendeteksi masalah kesehatan pada balita.
- b. Penelitian ini dapat mendukung pengembangan teori terkait efektivitas algoritma C4.5 dalam prediksi masalah gizi dan kesehatan anak, serta mendorong penelitian lebih lanjut di bidang serupa.

1.6.2 Manfaat Praktis

a. Bagi Posyandu Sakura

1. Membantu petugas posyandu dalam melakukan deteksi dini risiko *stunting* pada balita secara lebih efisien dan akurat
2. Menyediakan sistem pendukung keputusan yang dapat membantu proses pemantauan tumbuh kembang balita
3. Memudahkan pengelolaan dan pengarsipan data pertumbuhan balita

b. Bagi Orang Tua dan Masyarakat

1. Meningkatkan kesadaran tentang pentingnya pemantauan tumbuh kembang balita secara rutin
2. Memberikan informasi dini mengenai risiko *stunting* pada balita sehingga dapat dilakukan pencegahan sejak awal
3. Memudahkan akses informasi terkait status pertumbuhan balita

1.7 Pertanyaan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, berikut adalah pertanyaan penelitian yang harus dijawab dalam penelitian ini:

1. Apakah sistem informasi prediksi *stunting* yang dibangun dapat membantu dalam pengelolaan dan pemrosesan data untuk memprediksi risiko *stunting* pada balita di Posyandu Sakura, Desa Pancalang?

2. Apakah implementasi algoritma C4.5 dapat digunakan untuk memprediksi risiko stunting pada balita sehingga dapat membantu petugas Posyandu Sakura dalam mendeteksi balita yang berisiko stunting?

1.8 Hipotesis Penelitian

Hipotesis yang diajukan dalam penelitian ini adalah:

1. Sistem informasi prediksi stunting yang dibangun dapat membantu petugas Posyandu Sakura dalam mengelola dan memproses data sehingga mampu mendeteksi balita yang berisiko stunting secara lebih dini.
2. Algoritma C4.5 mampu memberikan prediksi risiko dini stunting pada balita, sehingga dapat mendukung petugas Posyandu dalam melakukan deteksi terhadap balita yang berisiko stunting.

1.9 Metodologi Penelitian

1.9.1 Metode Pengumpulan Data

a. Wawancara

Wawancara dengan Ketua Posyandu Sakura, Ibu Casinah, dilakukan untuk memahami lebih dalam proses bisnis yang berjalan saat ini, termasuk alur pelayanan kesehatan balita dan pencatatan data pertumbuhan mereka. Selain itu, wawancara ini juga bertujuan untuk mengidentifikasi kendala yang dihadapi dalam pencatatan dan pemantauan pertumbuhan balita, seperti kesulitan dalam pencatatan manual, ketidakakuratan data, atau keterbatasan sumber daya. Dari hasil wawancara ini, akan digali lebih lanjut mengenai kebutuhan sistem yang dapat dikembangkan guna meningkatkan efisiensi dan akurasi pencatatan, serta mempermudah pemantauan tumbuh kembang balita. Selain itu, data-data yang diperlukan untuk melakukan prediksi stunting juga akan dikumpulkan, seperti berat badan, tinggi badan, usia balita

b. Observasi

Observasi langsung dilakukan untuk memahami secara mendalam bagaimana proses penimbangan dan pengukuran balita berlangsung di Posyandu Sakura. Pengamatan ini mencakup metode yang digunakan, alat-alat yang tersedia, serta potensi kendala yang muncul selama pelaksanaan. Selain itu, sistem pencatatan yang digunakan saat ini juga dianalisis, termasuk bagaimana data balita dicatat, disimpan, dan dikelola oleh petugas

c. Studi Dokumentasi

Pengumpulan dan pengkajian berbagai sumber data dilakukan untuk mendapatkan pemahaman yang komprehensif terkait pertumbuhan balita di Posyandu Sakura. Data historis pertumbuhan balita dianalisis untuk melihat pola perkembangan, mendeteksi tren stunting, serta mengevaluasi efektivitas program pemantauan gizi yang telah berjalan. Buku register posyandu dipelajari guna memahami sistem pencatatan yang telah diterapkan, termasuk kelengkapan dan akurasi data yang tersedia.

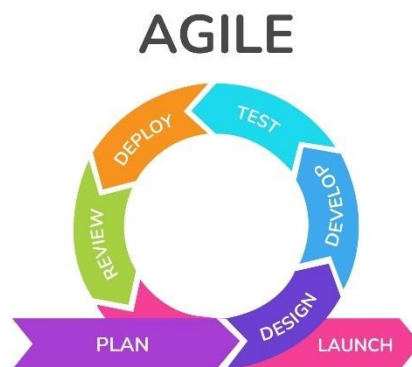
d. Studi Pustaka

Berbagai sumber pustaka digunakan untuk mendukung penelitian ini. Buku memberikan dasar teori, sementara artikel jurnal menyajikan penelitian terkini tentang pemantauan gizi dan stunting. Tesis atau disertasi menawarkan studi kasus serupa, sedangkan laporan penelitian dari lembaga resmi membantu memahami kebijakan dan evaluasi program kesehatan. Selain itu, website atau sumber daring terpercaya seperti *WHO* dan Kementerian Kesehatan menyediakan data dan panduan terbaru yang mendukung validitas penelitian.

1.9.2 Metode Pengembangan Sistem

Agile adalah sebuah metode manajemen proyek yang menggunakan siklus pengembangan yang singkat, atau bisa disebut juga “*sprint*” untuk fokus pada peningkatan berkelanjutan dalam pengembangan suatu produk atau layanan.

Metode *agile Software Development* atau biasa disebut dengan *agile* adalah proses iteratif dalam pembuatan sebuah perangkat lunak. Dalam proses pengembangannya, *agile* dapat dikatakan metode pengembangan yang cepat karena proses utama dari metode pengembangan agile sendiri berfokus pada *design-code-test once day* (Lutfiani et al., 2020)



Gambar 1. 1 Tahapan Proses Agile umum (BINAR, 2023)

Model proses *Agile* menekankan pada adaptabilitas, kolaborasi, dan kecepatan. Model proses Agile adalah siklus yang terdiri dari beberapa tahapan utama:

- a. Perencanaan (*Planning*)
- b. Desain (*Design*)
- c. Pengujian (*Testing*)
- d. *Deploy*
- e. Tinjauan dan Refleksi (*Review & Retrospective*)
- f. Rilis (*Release*)

1.9.3 Metode Penyelesaian Masalah

a. CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi baku yang dirancang untuk mengarahkan proses Data Mining secara sistematis.

Model ini pertama kali digagas oleh tiga perusahaan besar di industri data, yaitu Daimler Chrysler (sebelumnya Daimler-Benz), SPSS (dulu ISL), dan NCR. Seiring waktu, kerangka kerja ini disempurnakan melalui serangkaian lokakarya intensif yang berlangsung dari tahun 1997 hingga 1999.



Gambar 1. 2 Tahapan CRISP-DM (Xsanal Hakim et al., 2024)

Metodologi CRISP-DM terdiri atas enam langkah penting dalam proses penggalian data. Berikut uraian singkat dari setiap tahapannya :

1. *Business Understanding*

Langkah pertama adalah mengenali dan memahami secara mendalam apa yang dibutuhkan oleh pihak bisnis atau pengguna. Pada tahap ini, dilakukan identifikasi tujuan utama, evaluasi kondisi dan sumber daya yang ada, penetapan target pengumpulan data, serta perencanaan strategi proyek secara umum.

2. *Data Understanding*

Tahap kedua berfokus pada pengumpulan serta penelaahan data awal yang relevan. Aktivitas yang dilakukan termasuk mengakses data yang tersedia, melakukan deskripsi umum, mengeksplorasi karakteristik data, dan mengecek kualitasnya untuk memastikan keakuratannya dalam mendukung tujuan proyek.

3. *Data Preparation*

Pada tahap ini, data yang telah dikumpulkan akan diproses agar layak digunakan dalam pembuatan model. Proses ini bisa mencakup pembersihan data, penggabungan beberapa sumber, serta transformasi data ke dalam format yang lebih sesuai dengan kebutuhan analisis lebih lanjut.

4. *Modeling*

Setelah data siap, tahap selanjutnya adalah membangun model menggunakan teknik pemodelan tertentu. Dalam proses ini dilakukan pemilihan metode, perancangan skenario pengujian, pembuatan model itu sendiri, dan kemudian melakukan penilaian awal terhadap performa model tersebut.

5. *Evaluation*

Model yang telah dibuat perlu dievaluasi, tidak hanya dari segi teknis, tetapi juga dari sisi relevansinya terhadap kebutuhan bisnis. Tahapan ini mencakup peninjauan hasil, analisis terhadap keseluruhan proses, serta menentukan apakah model siap diterapkan atau perlu perbaikan.

6. *Deployment*

Langkah terakhir adalah mengintegrasikan hasil analisis ke dalam lingkungan operasional bisnis. Tujuannya adalah agar model yang sudah dikembangkan benar-benar digunakan dan memberikan manfaat nyata, baik melalui sistem otomatis maupun rekomendasi keputusan yang bisa dijalankan.

b. Data mining

Menurut Larose (2005), penambangan data dapat dipahami sebagai suatu proses untuk menemukan hubungan, pola, dan tren baru yang berarti dengan memfilter sejumlah besar data menggunakan berbagai teknik pengenalan pola, termasuk statistik dan matematika. Kegiatan dalam penambangan data umumnya dibagi ke dalam dua kategori utama:

1. Prediktif

Tujuan dari kegiatan prediktif adalah untuk memperkirakan nilai dari suatu atribut berdasarkan nilai atribut lainnya. Atribut yang diperkirakan sering disebut sebagai target atau variabel tergantung, sedangkan atribut yang digunakan untuk memperkirakan disebut sebagai variabel penjelas atau variabel independen.

2. Deskriptif

Tujuan dari kegiatan deskriptif adalah untuk mengidentifikasi pola-pola (korelasi, tren, kluster, jalur, dan anomali) yang merefleksikan hubungan dasar dalam data. Kegiatan penambahan data deskriptif sering kali bersifat eksploratif dan biasanya memerlukan teknik analisis yang lebih mendalam untuk memvalidasi dan menjelaskan temuan yang diperoleh.

c. *Predictive Model*

Dalam dunia sains, perkiraan adalah upaya memprediksi kejadian di masa depan, yang dilakukan secara matematis, bukan lewat cara-cara mistis atau ramalan. Proses ini memanfaatkan data serta pengalaman yang sudah ada. Contohnya, untuk memperkirakan cuaca esok hari di sebuah kota, kita perlu mengumpulkan data iklim dan pola cuaca di wilayah tersebut dari tahun-tahun sebelumnya. Atau, jika kita ingin memperkirakan keuntungan perusahaan bulan depan, kita perlu melihat data penjualan dan pengeluaran perusahaan di masa lalu. Data-data yang jadi dasar perkiraan ini disebut sebagai prediktor.

Hampir sama dengan prediksi, saat ingin memperkirakan sesuatu yang sudah terjadi, kita melakukan estimasi. Misalnya, mengukur tinggi seseorang atau menentukan jarak yang sudah ditempuh sebuah kendaraan.

Model prediktif, sesuai namanya, digunakan untuk memecahkan masalah yang butuh prediksi dan estimasi.

Algoritma *machine learning* akan mencoba menemukan kaitan antara fitur target (yang ingin diprediksi) dan fitur lain dalam data. Penting diingat bahwa secara matematis, tidak ada beda antara prediksi dan estimasi, jadi model yang sama bisa dipakai untuk keduanya.

Selain memprediksi nilai numerik, model prediktif juga bisa dipakai untuk pengelompokan, yaitu menentukan kategori yang tepat untuk suatu contoh (dalam *machine learning*, kategori sering disebut kelas). Prediksi jenis ini disebut klasifikasi, dan modelnya disebut model klasifikasi. Contoh klasifikasi meliputi:

1. Prakiraan cuaca: cerah, berawan, hujan, atau badai.
2. Pengelompokan pelanggan: pelanggan baru, biasa, atau setia
3. Penilaian risiko serangan jantung: rendah atau tinggi.
4. Penyaringan email: spam atau bukan

Agar prediksi bisa dilakukan, model prediktif butuh panduan jelas tentang apa yang harus diperkirakan, serta metode belajar yang diperlukan untuk membangun model tersebut. Seperti siswa yang butuh bimbingan, proses pelatihan ini disebut *supervised learning*. Pengawasan di sini bukan berarti diawasi langsung oleh manusia, tapi lebih ke penyiapan data yang disebut *ground truth*, yaitu data nyata yang dipakai sebagai input dalam proses belajar. Data ini berisi fitur dan atributnya. Tugas mesin adalah menemukan fungsi matematis terbaik untuk menggambarkan hubungan antara fitur dan target.

Hasil yang dikeluarkan model juga harus dinilai oleh manusia agar kita bisa tahu cara meningkatkan kinerja model, dengan mengevaluasi fitur mana yang paling berpengaruh pada hasil model tersebut.

Ada banyak algoritma *machine learning* yang bisa dipakai untuk mengembangkan model prediktif. Jenis target, jumlah fitur, dan tipe data yang bisa dipakai sebagai prediktor akan menentukan algoritma yang paling cocok untuk memecahkan masalah yang ada. (Kurniawan, 2022)

d. Algoritma C4.5

Salah satu cara untuk membuat pohon keputusan adalah dengan memakai algoritma C4. 5. Algoritma C4. 5 ini berupa model yang bentuknya mirip pohon. Di dalam pohon ini, ada yang namanya Node internal (bukan daun), yang mewakili atribut. Setiap cabang menunjukkan hasil dari pengujian atribut tersebut. Sementara itu, setiap daun melambangkan kelas data. Pohon keputusan ini umumnya ditampilkan dalam bentuk tabel yang berisi atribut dan rekaman. Atribut-atribut ini berperan sebagai parameter, yang jadi patokan dalam proses pembuatan pohon keputusan.. (Haqmanullah Pambudi & Darma Setiawan, 2018) Dalam algoritma C4. 5, cara memilih pemecahan kasus yang paling tepat adalah dengan menghitung dan membandingkan gain ratio sampai tercipta Node-Node baru di level berikutnya. Algoritma C4. 5 ini menggunakan data dari sampel pelatihan. Sampel pelatihan ini adalah contoh data yang dipakai untuk membuat pohon yang sudah terbukti benar.

1.10 Jadwal Penelitian

Penulis menyusun jadwal pelaksanaan penelitian yang akan dilaksanakan dapat dilihat pada tabel

Table 1 Jadwal Penelitian

	Bulan 1				Bulan 2				Bulan 3				Bulan 4				Bulan 5			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Proposal																				
Perencanaan Penelitian																				
Design																				
Develop																				
Pengujian																				
Deploy																				
Review																				
Dokumentasi																				

1.11 Sistematika Penelitian

Berikut ini merupakan gambaran inti mengenai sistematika penulisan yang terdiri dari 5 bab, yaitu sebagai berikut :

BAB I : PENDAHULUAN

Bab ini membahas latar belakang masalah, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, pertanyaan penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORITIS

Bab ini membahas tentang teori-teori yang melandasi penulisan skripsi.

BAB III : ANALISA DAN PERANCANGAN

Bab ini membahas perancangan sistem perangkat lunak yang akan dibangun, rancangan *input* dan *output* dari aplikasi.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas tentang spesifikasi dari aplikasi yang akan dibuat, tampilan input dan output serta membahas program yang berhubungan dengan materi skripsi.

BAB V : KESIMPULAN DAN SARAN

Bab ini terdiri dari kesimpulan dan saran yakni kesimpulan terhadap hasil penelitian serta aplikasi yang sudah dibuat secara keseluruhan dan dikemukakan saran-saran (rekomendasi) untuk perbaikan serta pengembangan program.