

BAB II

LANDASAN TEORI

2.1. Data Science

Data Science adalah disiplin yang mencakup sejumlah metodologi, teknik, dan alat untuk mengekstrak pengetahuan dan informasi yang berharga dari data [1]. Dalam konteks ini, Data Mining muncul sebagai salah satu cabang penting dari Data Science. Data Mining adalah proses penemuan pola yang bermanfaat, informasi tersembunyi, dan pengetahuan yang dapat diekstraksi dari dataset besar. Tujuan utamanya adalah untuk mengungkap wawasan yang dapat digunakan untuk membuat keputusan bisnis yang lebih baik, mendeteksi tren, dan memahami perilaku konsumen.

Data Mining melibatkan penerapan berbagai metode statistik, matematika, dan kecerdasan buatan untuk mengeksplorasi dan menganalisis data secara mendalam [2]. Teknik-teknik ini mencakup klasifikasi, klustering, regresi, asosiasi, dan lainnya. Dengan menggunakan algoritma-algoritma ini, para profesional Data Science dapat mengidentifikasi pola yang mungkin tidak terlihat secara langsung, memahami hubungan antara variabel-variabel, dan membuat prediksi berdasarkan data historis.

Pentingnya Data Mining dalam Data Science terletak pada kemampuannya untuk mengubah data mentah menjadi pengetahuan yang dapat diaplikasikan. Hal ini memungkinkan perusahaan dan organisasi untuk membuat keputusan yang lebih informasional dan cerdas. Selain itu, Data Mining juga dapat membantu mengoptimalkan proses bisnis, meningkatkan efisiensi operasional, dan

mengidentifikasi peluang baru. Oleh karena itu, integrasi Data Mining dalam praktik Data Science menjadi kunci dalam meraih nilai strategis dari kumpulan data yang semakin berkembang pesat.

2.1.1. Data Mining

Data mining adalah suatu proses ekstraksi informasi yang relevan, pola, dan pengetahuan yang terpendam dalam suatu set data besar [3]. Tujuan utama dari data mining adalah mengidentifikasi hubungan yang tidak terlihat dan pola yang signifikan, sehingga dapat memberikan wawasan yang berharga dan mendukung pengambilan keputusan [4]. Proses ini melibatkan penerapan berbagai teknik matematika, statistika, dan machine learning untuk menganalisis data secara mendalam.

Salah satu aspek utama dari data mining adalah identifikasi pola. Ini dapat mencakup pola asosiasi, di mana hubungan antara berbagai variabel diidentifikasi, atau pola sekuen, yang melibatkan urutan kejadian atau tindakan. Data mining juga mencakup teknik klasifikasi, di mana model prediktif dikembangkan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut tertentu. Selain itu, clustering atau pengelompokan data adalah bagian penting dari data mining. Ini melibatkan pembagian data menjadi kelompok-kelompok homogen berdasarkan kesamaan karakteristik tertentu. Teknik ini dapat membantu dalam mengidentifikasi struktur internal data dan mengelompokkan entitas yang serupa.

Data mining banyak digunakan dalam berbagai industri, termasuk perdagangan, keuangan, kesehatan, dan penelitian ilmiah [5]. Contohnya termasuk

analisis pelanggan untuk meramalkan perilaku pembelian, deteksi penipuan dalam transaksi keuangan, serta penemuan obat baru dalam penelitian medis. Sementara data mining memberikan manfaat besar dalam mengeksplorasi dan menggali potensi informasi yang tersembunyi dalam data, penting untuk mempertimbangkan aspek etika dan privasi. Penggunaan data mining harus mematuhi norma-norma dan regulasi yang berlaku untuk melindungi hak privasi individu dan memastikan keamanan data. Secara keseluruhan, data mining merupakan alat yang kuat untuk mendapatkan nilai tambah dari data besar dan kompleks. Dengan kemampuannya untuk mengidentifikasi pola yang signifikan, data mining menjadi elemen kunci dalam mendukung pengambilan keputusan yang cerdas dan strategi bisnis yang efektif [6].

2.1.2. Database dan Data Processing

Database dan pengolahan data merupakan dua konsep yang sangat terkait dan memainkan peran kunci dalam berbagai aspek teknologi informasi dan bisnis modern. Sebuah database adalah koleksi terstruktur dari data yang dapat diakses, dimanipulasi, dan dikelola secara efisien. Sementara itu, pengolahan data mencakup serangkaian aktivitas yang melibatkan penyimpanan, pengambilan, transformasi, dan analisis data untuk mendukung pengambilan keputusan dan pemahaman lebih lanjut. Sebuah database adalah wadah terorganisir yang menyimpan data dalam struktur yang terstruktur dan dapat diakses. Database memungkinkan pengguna untuk menyimpan, mengambil, dan mengelola data dengan efisien. Mereka terdiri dari tabel-tabel yang berisi baris dan kolom, dengan setiap baris mewakili satu entitas dan setiap kolom mewakili atribut atau

informasi tertentu. Database memungkinkan data disimpan dalam hubungan yang jelas antar entitas, memudahkan integritas data dan konsistensi.

Pengelolaan database melibatkan perangkat lunak sistem manajemen database (DBMS) yang memungkinkan pengguna membuat, mengakses, dan memperbarui data dengan cara yang terstruktur dan aman. Beberapa contoh DBMS termasuk MySQL, PostgreSQL, Microsoft SQL Server, dan Oracle Database. Keamanan data, integritas referensial, dan kinerja query adalah aspek kritis dalam pengelolaan database yang efisien.

2.1.3. Visualisation

Visualisasi data adalah metode penting dalam menyajikan informasi secara grafis untuk membantu pemahaman dan interpretasi data. Dalam konteks ini, visualisasi dapat mencakup berbagai teknik dan alat untuk menggambarkan pola, hubungan, dan tren dalam dataset dengan cara yang mudah dipahami. Visualisasi bukan hanya alat untuk membuat data lebih menarik, tetapi juga merupakan instrumen efektif untuk mengungkapkan wawasan yang mungkin sulit dipahami melalui presentasi data dalam bentuk tabel atau angka.

1. Visualisasi 3D dan Interaktif

Visualisasi 3D adalah bentuk visualisasi data yang mengintegrasikan dimensi ketiga, memberikan dimensi ekstra untuk pemahaman dan interpretasi data. Dengan menambahkan dimensi ini, visualisasi 3D memungkinkan pengguna untuk melihat data dari berbagai sudut pandang, memungkinkan representasi yang lebih dinamis dan mendalam. Ini dapat berguna terutama dalam konteks data spasial, simulasi, atau representasi objek kompleks. Salah satu keuntungan

visualisasi 3D adalah kemampuannya untuk menyajikan struktur dan relasi yang kompleks dengan cara yang lebih mudah dimengerti. Misalnya, dalam ilmu kimia, model molekul 3D dapat memberikan gambaran yang lebih jelas tentang struktur molekuler daripada representasi 2D. Dalam rekayasa, model 3D dapat digunakan untuk memvisualisasikan prototipe atau desain produk dengan lebih akurat.

2. Tujuan Visualisasi

Tujuan utama dari visualisasi adalah menyampaikan informasi secara grafis untuk mempermudah pemahaman dan interpretasi data. Dalam berbagai konteks, visualisasi memiliki beberapa tujuan utama yang bertujuan untuk membantu pemirsa memahami pola, tren, dan hubungan dalam data dengan cara yang lebih jelas dan intuitif.

3. Peran Visualisasi dalam Data Mining

Peran visualisasi dalam data mining sangat penting karena visualisasi memfasilitasi pemahaman dan interpretasi yang lebih baik terhadap pola dan informasi yang tersembunyi dalam dataset yang besar dan kompleks. Visualisasi membantu mengidentifikasi pola dan hubungan dalam data dengan cara yang intuitif. Dengan menggunakan grafik, diagram, atau plot, analisis data dapat dengan cepat melihat struktur data dan mencari pola yang mungkin tidak terdeteksi melalui analisis teks atau angka saja. Visualisasi memungkinkan analisis untuk menggambarkan dan menjelaskan pola sepanjang waktu, perbandingan antar variabel, atau distribusi data.

6. Pentingnya Kesesuaian dan Kreativitas

Kesesuaian dan kreativitas dalam visualisasi data adalah dua aspek penting yang saling melengkapi untuk mencapai efektivitas maksimal dalam

menyampaikan informasi. Kesesuaian mengacu pada pemilihan jenis visualisasi yang sesuai dengan jenis data dan pesan yang ingin disampaikan, sementara kreativitas menyoroti cara yang unik atau inovatif untuk menyajikan data tersebut. Kombinasi kedua elemen ini menciptakan visualisasi yang dapat memikat, membuka wawasan, dan meningkatkan pemahaman pemirsa terhadap informasi yang disajikan.

2.1.4. Statistik

Statistik adalah cabang ilmu matematika yang mencakup pengumpulan, analisis, interpretasi, presentasi, dan pengorganisasian data. Tujuan utama statistik adalah menyajikan informasi dalam bentuk yang dapat dimengerti, memungkinkan pengambilan keputusan yang lebih baik dan pengungkapan pola atau tren yang terdapat dalam data. Statistik digunakan di berbagai bidang, mulai dari ilmu pengetahuan hingga bisnis, dan merupakan alat penting dalam mendukung penelitian dan pengembangan pengetahuan.

2.1.7. Pattern Recognition

Pengenalan pola, atau pattern recognition, adalah disiplin ilmu yang berkaitan dengan identifikasi pola atau struktur dalam data. Tujuan utama dari pengenalan pola adalah mengembangkan model atau algoritma yang dapat mengklasifikasikan atau mengidentifikasi objek atau data berdasarkan karakteristik tertentu yang teridentifikasi dalam pola tersebut. Bidang ini memiliki berbagai aplikasi luas, dari pengenalan wajah dan suara hingga pengenalan tulisan tangan dan pengolahan citra medis.

2.2. Model Klasifikasi

Model klasifikasi adalah suatu pendekatan yang penting dalam pembelajaran mesin yang memiliki tujuan untuk mengelompokkan atau mengklasifikasikan data ke dalam kategori atau kelas yang sesuai. Proses pembentukan model dimulai dengan menggunakan dataset pelatihan yang telah diberi label, di mana model mempelajari pola dan karakteristik yang membedakan setiap kelas. Algoritma klasifikasi, seperti Decision Trees, Support Vector Machines (SVM), atau Neural Networks, digunakan untuk melatih model ini dan menghasilkan suatu aturan yang dapat digunakan untuk mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Salah satu aspek kritis dalam pengembangan model klasifikasi adalah pemilihan algoritma yang paling sesuai dengan karakteristik data dan tujuan spesifik pengklasifikasian. Setiap algoritma memiliki kelebihan dan kekurangan masing-masing, dan keberhasilan model sangat tergantung pada kecocokan antara algoritma dan sifat data yang dihadapi. Misalnya, SVM efektif dalam menangani data dengan batas keputusan yang kompleks, sementara Naive Bayes dapat bekerja dengan baik pada data teks atau kategori.

2.3. Algoritma Naïve Bayes

Algoritma Naive Bayes adalah metode klasifikasi probabilistik yang sangat populer dan sering digunakan dalam berbagai aplikasi seperti analisis sentimen, klasifikasi teks, dan sistem rekomendasi [7]. Algoritma ini didasarkan pada teorema Bayes dan mengasumsikan independensi antar-fitur, yang menjadikannya "naive" atau sederhana. Meskipun asumsi ini terkadang tidak sepenuhnya sesuai dengan dunia nyata [8], Naive Bayes tetap efektif dan efisien dalam banyak kasus.

Model Naive Bayes sering digunakan untuk mengklasifikasikan teks, seperti email spam atau analisis sentimen pada ulasan produk. Penggunaan yang luas dan kecepatan komputasional yang tinggi membuat Naive Bayes menjadi pilihan yang menarik untuk tugas klasifikasi dalam domain data mining dan pembelajaran mesin. Meskipun sederhana [9], algoritma Naive Bayes terbukti cukup andal dan mampu memberikan hasil yang baik, terutama ketika data terdiri dari banyak fitur atau ketika jumlah data terbatas.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Keterangan:

X : Data dengan jelas yang belum diketahui

C : Hipotesis dan X merupakan suatu kelas spesifik

P(C|X) : Probabilitas hipotesis C berdasarkan kondisi X (probabilitas posterior)

P(C) : Probabilitas hipotesis C (probabilitas prior)

P(X|C) : Probabilitas X berdasarkan kondisi pada hipotesis C

P(X) : Probabilitas X

2.3.1. Uji Performa

Confusion matrix dalam konteks data mining, terutama ketika menggunakan metode Naive Bayes, memberikan gambaran yang berharga tentang keefektifan model klasifikasi. Metode Naive Bayes adalah pendekatan probabilitas yang sering digunakan untuk klasifikasi dalam data mining. Confusion matrix pada Naive Bayes mencakup True Positive (TP), True Negative (TN), False Positive

(FP), dan False Negative (FN). Dengan menganalisis matriks kebingungan ini, para praktisi data mining dapat mengevaluasi sejauh mana model Naive Bayes dapat mengklasifikasikan instance-data dengan akurat berdasarkan probabilitas dan asumsi independensi yang diterapkan oleh metode ini. Dengan informasi yang dihasilkan oleh confusion matrix, mereka dapat menilai kehandalan model dalam memprediksi kelas yang benar, mengidentifikasi potensi kesalahan, dan memperbaiki parameter model untuk meningkatkan performa klasifikasi. Confusion matrix menjadi instrumen kunci dalam mengevaluasi kesesuaian model Naive Bayes dalam konteks data mining, memungkinkan para analis untuk membuat keputusan yang lebih cerdas dan mendalam terkait dengan pola dan pengetahuan yang ada dalam dataset.

Tabel 2.3. 1. Confusion Matrix Metode Naïve Bayes

		Kelas Prediksi		
		Kelas	Benar	Salah
Kelas Atribut	Benar	Benar	True Positive (TP)	False Positive (FP)
	Salah	Salah	False Negative (FN)	True Negative (TN)

Dimana tabel ini berisi:

- TP (True Positive), yaitu jumlah data positif yang memiliki nilai benar.
- TN (True Negative), yaitu jumlah data negatif yang memiliki nilai benar.
- FN (False Negative), yaitu jumlah data negatif tetapi yang memiliki nilai salah.

- FP (False Positive), yaitu jumlah data yang positif tetapi yang memiliki nilai salah.

$$Acuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Presisi = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

2.4. Alat Bantu Program/Tools Pendukung

2.4.1. Orange

Orange adalah platform analisis data visual yang kuat dan bersifat open-source, dirancang untuk membantu pengguna dari berbagai latar belakang mengatasi tugas-tugas analisis data. Aplikasi ini menawarkan antarmuka pengguna grafis yang intuitif yang memungkinkan pengguna untuk membangun dan mengevaluasi model pembelajaran mesin tanpa perlu pengetahuan pemrograman mendalam. Orange menyediakan beragam alat dan komponen visual yang dapat digunakan untuk eksplorasi data, preprocessing, dan pemodelan, serta memberikan dukungan untuk visualisasi hasil analisis. Salah satu keunggulan utama Orange adalah fleksibilitasnya dalam menangani berbagai jenis tugas analisis data. Aplikasi ini dapat digunakan untuk eksplorasi data sederhana, analisis statistik, dan pembuatan model prediktif yang kompleks. Orange juga menyediakan berbagai algoritma pembelajaran mesin yang dapat diintegrasikan

dalam alur kerja analisis, termasuk pohon keputusan, regresi, pengelompokan, dan banyak lagi.

Platform ini sering digunakan dalam konteks riset akademis, pembelajaran mesin, dan analisis data bisnis. Bagi para peneliti, Orange menyediakan alat untuk menguji hipotesis, membuat visualisasi interaktif, dan menghasilkan model prediktif untuk eksperimen tertentu. Di bidang bisnis, aplikasi ini dapat digunakan untuk analisis pelanggan, prediksi penjualan, dan pengambilan keputusan berbasis data. Selain itu, Orange juga mendukung ekosistem tambahan melalui add-on dan plugin, memperluas fungsionalitasnya. Pengguna dapat mengakses berbagai dataset contoh dan mengimpor data dari berbagai sumber, termasuk spreadsheet, database, dan file teks. Orange juga mendukung ekspor model yang dihasilkan ke format umum seperti PMML (Predictive Model Markup Language) untuk penggunaan di platform atau aplikasi lain. Dengan kombinasi antarmuka yang ramah pengguna, kekuatan analisis data, dan fleksibilitasnya yang tinggi, Orange memainkan peran penting dalam mempercepat proses analisis data untuk pemula dan ahli data. Aplikasi ini menyediakan solusi yang efektif dan efisien untuk tugas analisis data yang kompleks, memungkinkan pengguna untuk menjelajahi, memahami, dan mengoptimalkan informasi dari dataset mereka dengan lebih mudah dan cepat.

2.5. Metodologi Penelitian

2.5.1. Penelitian Terdahulu

Tabel 2.6. 1. Penelitian Terdahulu

Referensi Penelitian	1
Judul	Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics
Nama Penulis	Dr. V. Suma,
Tahun	2020
Hasil	<p>Penerapan data mining dalam konteks prediksi permintaan barang elektronik rekondisi telah memberikan hasil prediksi yang akurat. Dengan memanfaatkan teknik analisis data yang canggih, data mining memungkinkan organisasi untuk mengeksplorasi pola-pola kompleks dan hubungan tersembunyi dalam dataset permintaan barang elektronik rekondisi. Melalui proses ini, model dapat memperoleh pemahaman mendalam tentang faktor-faktor yang mempengaruhi permintaan, seperti tren pasar, musim, dan preferensi pelanggan. Hasil prediksi yang akurat yang dihasilkan oleh data mining dapat membantu perusahaan untuk mengoptimalkan persediaan, mengelola rantai pasok, dan meningkatkan kepuasan pelanggan dengan menyediakan produk yang sesuai dengan tingkat permintaan aktual. Dengan demikian, penerapan data</p>

	<p>mining tidak hanya memungkinkan analisis yang lebih mendalam tetapi juga memberikan landasan yang kokoh untuk pengambilan keputusan strategis dalam mengelola pasokan dan permintaan barang elektronik rekondisi [10].</p>
Referensi Penelitian	2
Judul	Intelligent Recommendation System Based on Mathematical Modeling in Personalized Data Mining
Nama Penulis	Yimin Cui
Tahun	2021
Hasil	<p>Dalam era data besar, penambangan data menjadi teknologi kunci dalam penelitian dan bisnis. Makalah ini fokus pada penambangan data dengan pendekatan sistem rekomendasi cerdas, menggunakan pemodelan matematis berdasarkan aturan asosiasi. Dengan memanfaatkan teknologi Java 2, Edisi Perusahaan, arsitektur sistem dibagi menjadi lapisan presentasi, logika bisnis, dan data. Modul rekomendasi, terbagi menjadi representasi data, pembelajaran model, dan mesin rekomendasi, dioptimalkan menggunakan algoritma fuzzy clustering. Evaluasi kinerja sistem</p>

	<p>mencakup akurasi, cakupan, dan waktu respons. Pengujian operasi di platform e-commerce menunjukkan peningkatan signifikan dalam rasio klik-tayang dan konversi pembelian produk yang direkomendasikan. Data eksperimen memperlihatkan akurasi tertinggi sekitar 0,73, dengan tingkat cakupan sistem mencapai 0,75. Rata-rata rasio klik-tayang meningkat 11,04%, sementara tingkat pembelian meningkat 9,35%, menunjukkan bahwa sistem ini mampu merekomendasikan produk yang sesuai dengan kebutuhan pengguna dan meningkatkan interaksi, meskipun aspek kepuasan pengguna masih dapat ditingkatkan [11].</p>
Referensi Penelitian	3
Judul	Implementasi Metode Naive Bayes Pada Perancangan Aplikasi Sistem Pakar Diagnosa Bronkiektasis
Nama Penulis	Hamzah Muhar Siregar
Tahun	2020
Hasil	Penerapan metode Naive Bayes dalam sistem pakar untuk diagnosa penyakit Bronkiektasis membuktikan efektivitasnya sebagai alat konsultasi yang dapat

	<p>memberikan informasi kesehatan secara cepat dan efisien. Dalam konteks penyakit ini, di mana diagnosa membutuhkan sumber pengetahuan dari seorang pakar medis, Naive Bayes membantu mengatasi keterbatasan waktu untuk berkonsultasi langsung dengan dokter pakar Bronkiektasis. Metode ini memanfaatkan fitur independen, yang berarti bahwa setiap fitur dalam data dianggap tidak berkaitan dengan keberadaan atau tidaknya fitur lain dalam data yang sama. Hasil klasifikasi didasarkan pada perbandingan nilai posterior dari kelas-kelas yang ada, dengan nilai posterior tertinggi menjadi hasil klasifikasi. Dengan demikian, pengguna dapat dengan cepat mendapatkan informasi tentang kemungkinan diagnosis penyakit Bronkiektasis melalui sistem pakar Naive Bayes tanpa perlu bertemu langsung dengan seorang pakar medis, memberikan solusi praktis untuk meningkatkan aksesibilitas konsultasi kesehatan [12].</p>
Referensi Penelitian	4
Judul	Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode Naive Bayes
Nama Penulis	Rahmania Aulia Ikimah Putri ¹ , Tacbir Hendro

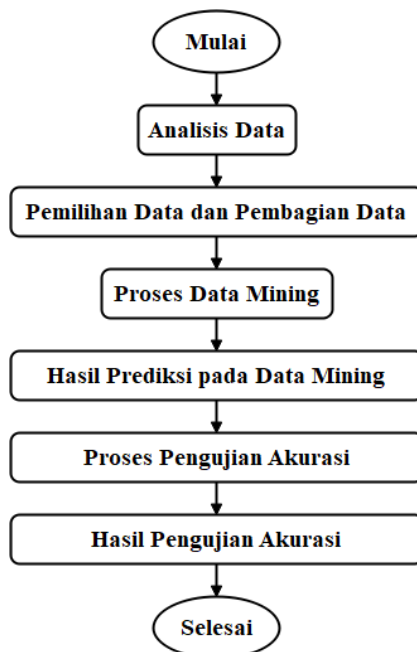
	Pudjiantoro ²
Tahun	2020
Hasil	<p>Penggunaan metode Naive Bayes dalam mengukur prestasi siswa di Sekolah XYZ telah terbukti sebagai alat prediksi yang akurat, memfasilitasi Bagian Bimbingan Konseling (BK) dalam memberikan arahan kepada siswa terkait pilihan masuk Perguruan Tinggi Negeri (PTN) melalui Seleksi Nasional Masuk Perguruan Tinggi (SNMPTN). Prestasi siswa, diukur dalam bentuk hasil belajar yang termanifestasi dalam prestasi akademik, menjadi indikator penting. Naive Bayes mampu memprediksi kemungkinan siswa diterima di PTN dengan cukup akurat, mengandalkan informasi dan data sebelumnya yang diterima. Hal ini memungkinkan BK untuk memberikan panduan yang lebih terarah kepada siswa, membantu mereka memilih strategi yang optimal untuk sukses dalam proses seleksi masuk PTN, dan pada akhirnya menciptakan jalan menuju prestasi akademik yang lebih tinggi [13].</p>
Referensi Penelitian	5
Judul	Analisis dan Implementasi Diagnosis Penyakit Sawit

	dengan Metode Convolutional Neural Network (CNN)
Nama Penulis	Errissya Rasywir ¹ , Rudolf Sinaga ² , Yovi Pratama ³
Tahun	2020
Hasil	<p>Studi ini menerapkan metode neural network, khususnya Convolutional Neural Network (CNN), untuk melakukan diagnosis penyakit pada tanaman kelapa sawit di Provinsi Jambi. Meskipun kelapa sawit menjadi komoditas andalan di wilayah tersebut, keterbatasan pemahaman petani terhadap hama dan penyakit dapat mempengaruhi produktivitas. Dalam upaya meningkatkan diagnosis tanaman secara efisien, pembelajaran mendalam digunakan dengan memanfaatkan data citra kelapa sawit dari Dinas Perkebunan Provinsi Jambi. Hasil penelitian menunjukkan bahwa model CNN mencapai akurasi tinggi dalam mengenali 11 kategori penyakit pada kelapa sawit, dengan akurasi tertinggi mencapai 0,89 dan rata-rata akurasi sebesar 0,87. Keberhasilan ini memberikan indikasi positif terhadap pengembangan sistem otomatis dan mobile untuk klasifikasi penyakit kelapa sawit, yang dapat menjadi alat yang sangat bermanfaat bagi para petani dalam meningkatkan</p>

	manajemen penyakit dan produktivitas tanaman kelapa sawit [14].
Referensi Penelitian	6
Judul	Deteksi Penyakit pada Daun Kentang Menggunakan Pengolahan Citra dengan Metode Convolutional Neural Network
Nama Penulis	Abdul Jalil Rozaqi ^{*1} , Andi Sunyoto ² , Rudyanto Arief ³
Tahun	2021
Hasil	<p>Penelitian ini mengusulkan penggunaan metode neural network, khususnya Convolutional Neural Network (CNN), untuk mendeteksi penyakit pada daun kentang. Kentang, sebagai makanan utama, memiliki peran krusial dalam pertanian. Namun, kendala utama yang dihadapi adalah serangan penyakit seperti late blight dan early blight yang dapat merugikan produksi dan menyebabkan kegagalan panen jika tidak ditangani dengan tepat. Proses identifikasi manual gejala pada daun kentang seringkali memakan waktu, mengakibatkan penanganan yang lambat dan peningkatan biaya perawatan. Melalui teknologi pengolahan citra digital dan metode Deep Learning</p>

	<p>menggunakan CNN, penelitian ini berhasil mengklasifikasikan tiga kelas, yaitu daun sehat, early blight, dan late blight. Hasilnya menunjukkan tingkat akurasi yang baik, dengan training akurasi mencapai 95% dan validation accuracy 94% pada epoch ke-10 dengan batch size 20. Pendekatan ini memberikan solusi efisien dalam mendeteksi dini penyakit pada daun kentang, memungkinkan petani untuk mengambil langkah-langkah pencegahan yang cepat dan tepat [15].</p>
--	--

2.5.2. Kerangka Kerja Penelitian



1. Analisis Data

Analisis data dalam konteks data mining merupakan proses penyelidikan dan penafsiran pola, hubungan, dan informasi yang terkandung dalam dataset

yang besar dan kompleks. Tujuan utamanya adalah mengungkapkan pengetahuan yang berguna dan pemahaman mendalam tentang pola tersembunyi atau tren yang dapat digunakan untuk pengambilan keputusan. Proses analisis data melibatkan langkah-langkah seperti pemrosesan data, pemilihan atribut yang relevan, identifikasi pola, dan pembuatan model prediktif. Teknik-teknik seperti clustering, klasifikasi, regresi, dan asosiasi digunakan untuk menggali informasi yang berharga dari data yang ada. Analisis data dalam data mining membantu mengidentifikasi insight yang tidak terlihat secara langsung, memungkinkan organisasi untuk membuat keputusan yang lebih informasional dan mendukung strategi bisnis yang lebih cerdas.

2. Pemilihan Data dan Pembagian Data

Pemilihan data dan pembagian data adalah dua tahap kritis dalam proses data mining yang menentukan keberhasilan analisis. Pemilihan data melibatkan identifikasi dan ekstraksi subset data yang relevan untuk tujuan analisis tertentu. Langkah ini mencakup penentuan atribut yang paling berpengaruh, eliminasi data yang tidak relevan, dan penanganan nilai yang hilang. Pembagian data, di sisi lain, merupakan langkah esensial untuk menguji dan mengukur kinerja model yang dikembangkan. Proses ini melibatkan pemisahan dataset menjadi set pelatihan dan set pengujian, di mana set pelatihan digunakan untuk melatih model, sementara set pengujian digunakan untuk mengevaluasi sejauh mana model tersebut dapat menggeneralisasi pada data yang belum pernah dilihat sebelumnya. Dengan pemilihan data yang bijaksana dan pembagian data yang tepat, proses data mining

dapat memberikan hasil yang lebih akurat dan dapat diandalkan, serta mendukung pengambilan keputusan yang lebih baik dalam berbagai bidang, mulai dari bisnis hingga ilmu pengetahuan.

3. Proses Data Mining

Proses data mining adalah serangkaian langkah sistematis yang dilakukan untuk menggali pengetahuan yang berharga dari dataset yang kompleks. Tahap pertama dalam proses ini adalah pemahaman bisnis dan pemahaman data, di mana tujuan dan karakteristik data yang relevan diidentifikasi. Langkah selanjutnya melibatkan pemilihan dan pemrosesan data, termasuk pembersihan data, penanganan nilai yang hilang, dan transformasi atribut. Setelah itu, tahap pemodelan melibatkan pemilihan metode analisis yang sesuai, seperti klasifikasi, regresi, clustering, atau asosiasi, dan pengembangan model berdasarkan dataset yang telah diproses sebelumnya. Tahap evaluasi dan validasi dilakukan untuk menilai kinerja model terhadap data yang tidak terlihat sebelumnya, dan jika hasilnya memuaskan, model tersebut dapat diimplementasikan untuk pengambilan keputusan atau prediksi. Terakhir, tahap implementasi dan pemeliharaan melibatkan integrasi model ke dalam sistem yang ada dan pemantauan kinerjanya seiring waktu. Proses ini memberikan kerangka kerja yang sistematis dan efektif untuk mendapatkan wawasan berharga dari data dan mendukung pengambilan keputusan yang informasional.

4. Hasil Prediksi Pada Data Mining

Hasil prediksi pada data mining merupakan output dari model analisis yang telah dikembangkan untuk memahami pola dan tren dalam dataset. Hasil ini memberikan gambaran tentang kemungkinan kejadian atau nilai di masa depan

berdasarkan pembelajaran dari data historis. Dalam konteks klasifikasi, hasil prediksi dapat berupa penentuan kategori atau label tertentu untuk suatu observasi. Di regresi, hasilnya dapat berupa nilai prediksi berkelanjutan. Interpretasi hasil prediksi memegang peranan kunci dalam pengambilan keputusan, membantu organisasi untuk merencanakan strategi berbasis data, mengidentifikasi peluang atau risiko, dan meningkatkan efisiensi operasional. Evaluasi yang cermat terhadap akurasi dan kinerja model penting untuk memastikan keandalan hasil prediksi, yang pada gilirannya dapat memberikan manfaat signifikan dalam berbagai konteks, seperti bisnis, keuangan, kesehatan, dan penelitian ilmiah.

5. Proses Pengujian Akurasi

Proses pengujian akurasi pada data mining merupakan tahap penting dalam mengevaluasi sejauh mana model yang dikembangkan mampu menggeneralisasi dan memberikan prediksi yang tepat pada data yang belum pernah dilihat sebelumnya. Langkah awal melibatkan pembagian dataset menjadi dua bagian: set pelatihan untuk melatih model dan set pengujian untuk menguji performa model. Selama pengujian, model diuji menggunakan data yang tidak terlibat dalam proses pelatihan untuk menghindari bias. Metrik-metrik evaluasi seperti akurasi, presisi, recall, dan F1-score digunakan untuk mengukur seberapa baik model dapat mengklasifikasikan atau memprediksi data. Proses ini membantu menentukan sejauh mana model dapat diandalkan dan relevan dalam mengatasi tantangan data baru. Melalui pengujian akurasi yang cermat, pengembang dan pengguna model

dapat memahami kekuatan dan kelemahan model, serta membuat penyesuaian yang diperlukan untuk meningkatkan kinerja dan ketepatan prediksi.

6. Hasil Pengujian Akurasi

Hasil pengujian akurasi pada data mining menyediakan gambaran kritis tentang kinerja model analisis terhadap data yang tidak terlibat dalam proses pelatihan. Metode ini memungkinkan evaluasi objektif terhadap sejauh mana model dapat menggeneralisasi dan memberikan prediksi yang akurat pada situasi dunia nyata. Hasil pengujian akurasi diukur melalui berbagai metrik evaluasi, seperti akurasi keseluruhan, presisi, recall, F1-score, dan area di bawah kurva ROC, yang memberikan wawasan mendalam tentang kemampuan model dalam mengklasifikasikan atau memprediksi data. Interpretasi hasil pengujian membantu pengambil keputusan dan analis data untuk menilai kehandalan model, mengidentifikasi potensi perbaikan, dan membuat keputusan strategis. Dengan memahami dan mengoptimalkan hasil pengujian akurasi, organisasi dapat memastikan bahwa model data mining yang dikembangkan dapat memberikan informasi yang dapat diandalkan untuk mendukung proses pengambilan keputusan yang lebih baik.