BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Tinjauan pustaka ini akan membahas teori-teori dan penelitian terkait yang relevan dengan implementasi perbandingan algoritma *Decision Tree* dan *SVM* untuk menentukan barang yang paling diminati. Bab ini akan memberikan dasar teori yang kuat serta konteks penelitian yang mendalam untuk mendukung analisis dan implementasi yang akan dilakukan.

2.1.1 Sistem Informasi

Sistem informasi adalah suatu kombinasi modul yang terorganisir yang terdiri dari komponen-komponen yang terkait dengan *hardware, software,* orang, dan jaringan yang bergantung pada seperangkat komputer yang saling berhubungan atau berinteraksi untuk melakukan pengolahan data menjadi informasi untuk mencapai tujuan tertentu. Sistem yang memiliki kemampuan untuk mengumpulkan, memproses, menyimpan, menganalisis, dan menyebarkan informasi disebut sistem informasi. (Calista et al., 2023)

2.1.2 Decision Support System

Decision Support System (*DSS*) atau sistem pendukung keputusan (SPK) adalah sistem yang memiliki kemampuan untuk memecahkan masalah dan berkomunikasi tentang masalah semi-terstruktur. *DSS* adalah sistem interaktif berbasis komputer yang membantu orang membuat keputusan dengan menggunakan model dan data untuk menyelesaikan masalah yang tidak terstruktur dan semi-terstruktur. (Oktafian, 2020)

2.1.2.1 Sistem Pendukung Pengambilan Keputusan

Istilah sistem keputusan terstruktur (*structured decision system-SDS*) digunakan untuk mendeskripsikan sistem-sistem yang mampu menyelesaikan masalah yang teridentifikasi. Masalah-masalah di bawah garis menyulitkan pemrosesan komputer, dan Gorry dan Scott-Morton menggunakan istilah sistem pendukung pengambilan keputusan (*decision support system-DSS*) untuk menggambarkan sistem yang dapat memberikan dukungan yang dibutuhkan.(Pengampu et al., 2020)

2.1.3 Data Mining

Data mining adalah proses ekstraksi informasi yang bermanfaat dari data besar dengan menggunakan teknik-teknik matematika, statistika, dan kecerdasan buatan. Informasi yang dihasilkan dari proses *data mining* dapat digunakan untuk membantu pengambilan keputusan dan prediksi di berbagai bidang, seperti bisnis, ilmu pengetahuan, teknologi, kesehatan, dan lain-lain. Definisi *data mining* telah berkembang seiring dengan perkembangan teknologi dan penelitian(Kumar, 2024).

Menurut (Hasanah et al., 2021)*Data Mining* suatu proses mencari pola atau informasi dalam kumpulan data yang terpilih dengan menggunakan teknik atau metode tertentu.

Menurut (Firdiansyah & Findawati, 2021)*Data mining* adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstrasi pengetahuan (*knowledge*) secara otomatis.

Menurut (Mamok Andri Senubekti & Lies Anggi Puspita Dewi, 2022)*Data Mining* adalah bidang ilmu multidisiplin, menggambarkan area kerja yang

mencakup teknologi *database*, pembelajaran mesin, statistik, pengenalan pola, pencarian informasi, jaringan saraf tiruan, sistem berbasis pengetahuan, kecerdasan buatan, komputasi kinerja tinggi, dan visualisasi data.

2.1.3.1 Proses Data Mining

Proses *data mining* adalah serangkaian langkah atau tahapan dalam menggali pengetahuan atau informasi dari data yang besar dan kompleks. Proses ini meliputi tahapan mulai dari pemilihan data, *pre-prosesing data*, pemodelan, dan evaluasi hasil. Proses *data mining* yang efektif membutuhkan perencanaan yang matang serta pemilihan teknik dan algoritma yang tepat untuk setiap tahapan.(Muttaqin & Wahyu Wijaya Widiyanto, 2023)

2.1.3.2 Confusion Matrix

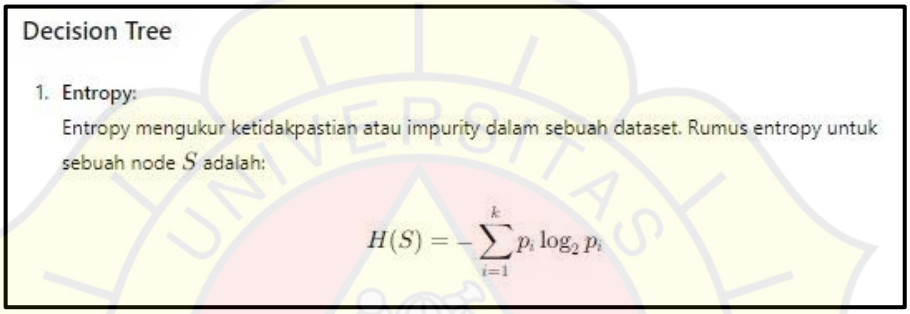
Confusion Matrix merupakan alat visualisasi umum dalam pembelajaran yang diawasi. Setiap kolom dalam matriks mewakili turunan dari kelas predikat, dan setiap baris mewakili turunan aktual dari kelas tersebut. Kinerja sistem biasanya dievaluasi menggunakan *matriks* data yang tersedia. Berikut tabel 2.1 dibawah ini adalah contoh tabel untuk *confusion matrix*(ALVIN JULIANTO, 2024).

2.1.4 Prediksi Barang Yang Paling Diminati

Prediksi barang yang paling diminati adalah proses untuk memprediksi produk atau barang yang akan mendapatkan permintaan tinggi di masa depan berdasarkan data historis dan pola pembelian sebelumnya. (Oktafian, 2020)

2.1.5 Algoritma Decision Tree

Struktur data pohon terdiri dari rusuk (*edge*) dan simpul (*node*). Semua pohon memiliki tiga jenis simpul: simpul akar (akar atau simpul), simpul percabangan atau dalam (cabang atau simpul dalam), dan simpul daun. Dalam pohon keputusan, simpul internal dan simpul akar diberi nama atribut, rusuk-rusuk diberi label nilai atribut yang mungkin, dan simpul daun diberi kelas-kelas yang berbeda. (Nasrullah, 2021)

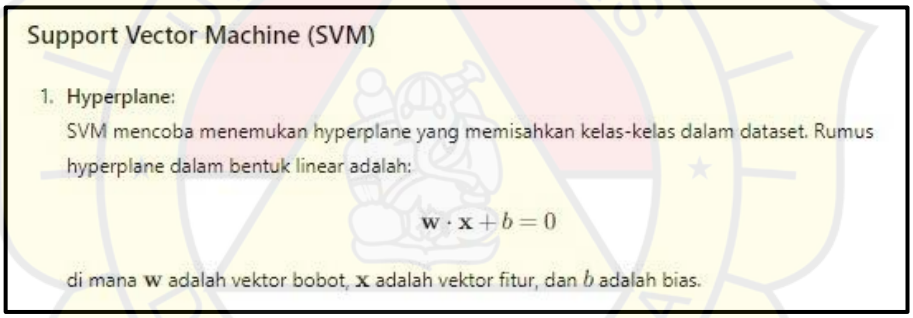


Gambar 2. 1 Rumus Algoritma Decision Tree

Gambar 2.1 adalah Rumus algoritma *decision tree* berfungsi untuk memodelkan dan menghasilkan sebuah pohon keputusan (*decision tree*) berdasarkan data yang diberikan. Tujuan utamanya adalah untuk memecah data menjadi subset yang semakin kecil dan homogen berdasarkan fitur-fitur tertentu, sehingga memungkinkan pengambilan keputusan atau *klasifikasi* data yang lebih baik. Proses ini melibatkan pemilihan fitur terbaik untuk membagi data, serta aturan-aturan keputusan yang dihasilkan dari setiap simpul pohon keputusan untuk memprediksi nilai target atau *klasifikasi* data baru.

2.1.6 Algoritma SVM

Sistem pembelajaran *Support Vector Machine* (*SVM*) menggunakan ruang hipotesis yang terdiri dari fungsi-fungsi linear dalam sifat yang memiliki dimensi yang tinggi. *SVM* dilatih menggunakan algoritma pembelajaran yang berasal dari teori optimasi. *SVM linear* memisahkan data secara linear, yaitu memisahkan dua kelas pada *hyperplane* dengan *margin* yang lembut. Namun, *trik kernel* terhadap ruang yang besar menyebabkan *non-linear*. *Support Vector Machine* menggunakan dua titik, atau *vector*, yang kemudian membentuk garis pembatas, atau sisi pembatas, dari dau buah *vector*. Garis pembatas ini disebut *hyperplane.* (Ariqoh et al., 2023)


Support Vector Machine (SVM)

1. Hyperplane:
   SVM mencoba menemukan hyperplane yang memisahkan kelas-kelas dalam dataset. Rumus hyperplane dalam bentuk linear adalah:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

di mana $\mathbf{w}$ adalah vektor bobot, $\mathbf{x}$ adalah vektor fitur, dan b adalah bias.

Gambar 2. 2 Rumus Algoritma SVM

Gambar 2.2 menunjukkan rumus untuk memisahkan dua kelas data dengan menemukan bidang pemisah, atau *hyperplane,* yang dapat memisahkan kedua kelas tersebut dengan paling efektif. Tujuannya adalah untuk mencari *hyperplane* yang memiliki *margin* (jarak) maksimal dari titik-titik terdekat dari masing-masing kelas, sehingga memaksimalkan ketahanan terhadap data baru yang tidak terlihat (generalisasi) dan meningkatkan akurasi dalam melakukan klasifikasi. *SVM* menggunakan konsep transformasi data ke dimensi yang lebih tinggi (*kernel trick*) untuk menyelesaikan masalah *non-linear* secara efisien.

2.1.6.1 Mengevaluasi Kualitas Model

Metrik evaluasi memberi kita cara untuk mengevaluasi secara *kuantitatif* seberapa baik model kita tampil. Performa model pada data pelatihan dievaluasi untuk mendapatkan akurasi set pelatihan, sedangkan performanya pada data pengujian dievaluasi untuk mendapatkan pengujian akurasi data ketika model memprediksi target dari contoh yang sebelumnya tidak terlihat. Evaluasi pada data pengujian membantu kita mengetahui ukuran kinerja sebenarnya dari model kita.(Sains & Teknologi, 2022)

2.1.7 Bahasa Pemrograman dan Aplikasi yang digunakan

2.1.7.1 Python

Python adalah bahasa pemrograman yang menggunakan *interpreter*, sehingga dapat dijalankan di berbagai platform seperti *Windows* dan *Linux. Python* mendukung berbagai paradigma pemrograman, termasuk prosedural, berorientasi objek, dan fungsional. Hal ini membuatnya fleksibel untuk berbagai jenis proyek. Python sering digunakan untuk pengembangan aplikasi *web*, *desktop*, dan *IoT*, serta mendukung integrasi database dan manipulasi file. Banyak *programmer* dan peneliti memilih *Python* karena kemampuannya dalam menangani data besar dan perhitungan kompleks, serta kemudahan prototyping dan pengembangan cepat.. (Rahman et al., 2023)

2.1.7.2 Jupyter Notebook

Jupyter Notebook adalah *editor Python* berbasis web yang dapat digunakan menggunakan *browser*. *Jupyter Notebook* ini memungkinkan interaksi antara kode dengan *output* dalam satu dokumen. Dokumen yang dihasilkan memiliki *format file* dengan ekstensi "*ipynb*" dan dibuat menggunakan aplikasi *Jupyter Notebook*. Di

dalam dokumen tersebut terdapat kode komputer serta elemen teks yang kaya seperti paragraf, rumus matematika, gambar, dan tautan. (Muzaky et al., 2024)

2.1.7.3 Streamlit

Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa *Python*. Ini cepat dan minimalis, tetapi memiliki tampilan yang cukup baik dan ramah pengguna. *Widget* bawaan yang dapat digunakan oleh pengguna dapat mencakup pengunggahan gambar, penggeser, input teks, dan elemen *hypertext markup language (HTML*) lain yang sudah dikenal, seperti tombol *checkbox* dan tombol *radio*. Setiap kali pengguna berinteraksi dengan aplikasi *Streamlit, script Python* dijalankan dari atas ke bawah. Ini adalah konsep penting yang harus diingat saat mempertimbangkan berbagai status aplikasi yang akan dipilih.(Hastomo et al., 2022)

2.1.7.4 Visual Studio Code

Microsoft mengembangkan editor teks *Visual Studio Code*, yang ringan dan dapat diandalkan untuk berbagai sistem operasi, dan tersedia dalam versi *Linux*, *Mac*, dan *Windows*. Editor teks ini mendukung secara langsung bahasa pemrograman *JavaScript, Typescript,* dan *Node.js*, serta bahasa pemrograman lainnya dengan bantuan plugin yang dapat dipasang melalui *marketplace Visual Studio Code*. Bahasa pemrograman seperti *C++, C#, Python, Go*, dan J*ava* termasuk di antaranya.(Salamah, 2021)

2.1.7.5 Xampp

XAMPP adalah sebuah aplikasi open source terkait pengelolaan server yang dikembangkan oleh *Apache Friends*. Karena bersifat open source, aplikasi ini bisa

Anda digunakan secara gratis. Selain itu, sesuai namanya, *X p*ada *XAMPP* berarti *cross platform*. Artinya, mendukung berbagai platform seperti *Windows, macOS* dan *Linux. XAMPP* sendiri terdiri dari *Apache, MariaDB* (yang dikembangkan dari *MySQL*), *PHP* dan *Perl.*(Kalsum Siregar et al., 2024)

2.2 **Alur Penelitian Terkait**

Tahap penelitian, juga dikenal sebagai jalur penelitian, adalah urutan kronologis tindakan yang dilakukan oleh seorang peneliti dalam pekerjaan penelitian mereka, bukan hanya urutan langkah-langkah yang harus diikuti.

Tabel 2.1 Alur Penelitian

Jurnal	Peneliti	Tahun	Kesimpulan
ANALISIS SENTIMEN PADA PRODUK CUSHION DI WEBSITE FEMALE DAILY MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (*SVM*)	Soultan Ariqoh	2023	analisis sentimen terhadap produk cushion merek Pixy di website Female Daily menggunakan metode Support Vector Machine (*SVM*) menunjukkan sentimen positif terhadap produk

			tersebut. Saran untuk penelitian selanjutnya adalah menambah jumlah dataset dan mencoba algoritma lain seperti K-Nearest Neighbor dan Naïve Bayes.
IMPLEMENTASI ALGORITMA *DECISION TREE* UNTUK KLASIFIKASI	Asmaul Husnah Nasrullah	2021	Dari hasil penelitian yang telah dilakukan untuk data penjualan dan stok barang di UD. Cipta Karya Gorontalo maka diperoleh akurasi sebesar 90% dan nilai AUC 0.709. Dimana nilai ini termasuk dalam

			Fair Classification (Klasifiksi yang cukup). Sehingga dapat disimpulkan bahwa model klasifikasi data mining menggunakan Algoritma Decision Tree C4.5 akurat dalam melakukan klasifikasi untuk produk laris.
ANALISIS DATA MINING ALGORITMA C-4.5 DALAM MENENTUKAN PRODUK YANG	Noni Rahmawati	2023	Kesimpulan dari penelitian ini adalah bahwa penggunaan algoritma data mining C-45 dalam aplikasi

BANYAK DIMINATI PADA TOKO BILLY			Weka dapat membantu pemilik toko Billy untuk mengetahui produk yang diminati oleh konsumen dan meningkatkan pendapatan toko.