

BAB II TINJAUAN PUSTAKA

1.1 Data Mining

Data mining adalah suatu pendekatan yang digunakan untuk menemukan informasi tersembunyi dalam data berukuran besar, yang biasanya dikenal sebagai penemuan pengetahuan (knowledge discovery). Metode ini memanfaatkan teknik analisis data guna mengidentifikasi pola-pola tertentu dalam basis data. Secara garis besar data mining merupakan proses eksplorasi dan ekstraksi informasi bernilai dari volume data yang masif, dengan bantuan metode statistik, pembelajaran mesin dan kecerdasan buatan. Tujuan utamanya adalah untuk mengungkap pengetahuan tersembunyi yang dapat digunakan untuk menunjang proses pengambilan keputusan. Data mining juga merupakan bagian dari proses yang lebih luas, yaitu Knowledge Discover in Database (KDD), yang melibatkan tahap-tahap seperti pengumpulan data, pembersihan atau pra-pemrosesan, analisis data, hingga interpretasi yang diperoleh (Zahra et al., 2024)

Pola dalam data mining dianggap bernilai jika memenuhi sejumlah kriteria penting, seperti tidak bersifat remeh, mengandung makna tersembunyi, belum diketahui sebelumnya, serta memberikan manfaat yang nyata dalam proses pengambilan keputusan. Pola yang ditemukan sebaiknya mudah untuk dipahami, mampu memprediksi data lain dengan tingkat akurasi tertentu, dan memiliki potensi untuk diterapkan di berbagai bidang. Proses pencarian pola ini juga dikenal dengan berbagai istilah lain, seperti Knowledge Discovery in Database (KDD), analisis pola, eksplorasi data, ekstraksi informasi, maupun business intelligence (M. Afriansyah et al., 2024).

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, seperti:

1. Deskripsi

Teknik ini bertujuan untuk menggali dan memaparkan pola-pola yang ada di dalam data, serta memberikan pemahaman tentang tren atau kecenderungan yang muncul. Informasi yang diperoleh dari analisis ini dapat membantu dalam mengidentifikasi hubungan antara variabel-variabel dalam dataset.

2. Estimasi

Estimasi memiliki konsep yang sama dengan klasifikasi, namun variabel target yang digunakan bersifat numerik, bukan kategorikal. Model estimasi dibangun dengan menggunakan data yang memiliki nilai lengkap untuk variabel target, dan model tersebut kemudian digunakan untuk

memperkirakan nilai variabel target, dan model tersebut kemudian digunakan untuk memperkirakan nilai variabel target berdasarkan variabel prediktor lainnya.

3. Prediksi

Prediksi bertujuan untuk meramalkan nilai atau hasil dimasa depan dengan menganalisis pola yang ada dalam data historis. Teknik ini banyak diterapkan dalam berbagai bidang, seperti peramalan penjualan, prediksi stok bahan baku, dan analisis tren jangka panjang.

4. Klasifikasi

Klasifikasi merupakan teknik dalam data mining yang digunakan untuk menglompokkan data ke dalam kategori tertentu berdasarkan sifat atau karakteristiknya. Proses ini sering diterapkan dalam berbagai situasi, seperti deteksi email spam, diagnosis medis dan segmentasi pelanggan.

5. Clustering

Pengelompokan atau clustering merupakan teknik yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik, tanpa memerlukan variabel target. Objek dengan kesamaan tertentu akan dikelompokkan ke dalam klaster yang sama. Teknik ini biasanya sering diterapkan dalam analisis pemasaran, pengelompokan dokumen, dan analisis perilaku pelanggan.

1.2 Prediksi

Prediksi merupakan proses meramalkan atau memperkirakan kondisi di masa depan dengan melakukan analisis berdasarkan data historis yang tersedia (Dewi et al., 2022). Tujuan dari prediksi adalah untuk meminimalkan tingkat ketidakpastian dalam peramalan, sehingga hasil prediksi yang dihasilkan dapat mendekati dengan kondisi yang sebenarnya (Hatta & Fauziah Fitri, 2020).

Pemilihan periode waktu untuk prediksi disesuaikan dengan konteks dan sasaran dari kegiatan prediksi tersebut. Umumnya, periode yang digunakan mencakup harian mingguan, bulanan, per semester, hingga tahunan. Namun, perlu dicatat bahwa semakin panjang jangka waktu prediksi, maka akurasi hasil prediksi biasanya akan semakin berkurang (Akbar & Rakhma Devi, 2022).

Prediksi Juga memiliki pola tersendiri, berikut ini jenis-jenis pola dalam prediksi:

1. Peramalan jangka pendek, sebuah peramalan yang melibatkan periode waktu maksimal 18 bulan, seperti penjadwalan kerja dan perencanaan pembelian material. Pola musiman (Seasonal Pattern), merupakan pola yang berulang secara teratur dalam periode waktu tertentu, seperti bulanan, kuartalan, atau tahunan.
2. Peramalan jangka menengah, peramalan yang melibatkan periode waktu antara 3 hingga 18 bulan, contohnya adalah perencanaan produksi suatu produk.
3. Peramalan jangka panjang, peramalan dengan jangka waktu lebih dari 18 bulan, seperti perencanaan penyediaan fasilitas (Firstiano et al., 2020)

1.3 Time Series

Data deret waktu (time series) merupakan data yang dikumpulkan berdasarkan waktu yang berkelanjutan, dengan nilai tertentu pada setiap titik waktunya. Data ini berkaitan dengan objek tertentu dan dapat dicatat dalam berbagai interval waktu, seperti harian, mingguan, bulanan atau tahunan. Oleh karena itu, data erkala ini berhubungan dengan data statistik yang dicatat dan dianalisis dalam jangka waktu tertentu, seperti data penjualan, harga, persediaan, produksi tenaga kerja, nilai tukar dan harga saham (Fathoni & Wijayanto, 2021).

Metode peramalan deret waktu menggunakan data permintaan historis sebagai dasar dalam menghasilkan prediksi. pendekatan ini didasarkan pada asumsi bahwa data historis merupakan indikator yang baik untuk memperkirakan kondisi dimasa mendatang. Metode ini paling cocok ketika pola dasar permintaan cenderung stabil dan tidak mengalami fluktuasi signifikan dari tahun ke tahun (Asynari et al., 2020).

1.4 Trend Linear

Trend linear merupakan suatu pola perubahan yang menunjukkan kecenderungan meningkat atau menurun dalam jangka waktu tertentu. pola ini diperoleh melalui perhitungan rata-rata perubahan dari waktu ke waktu. Apabila rata-rata perubahan menunjukkan peningkatan, maka disebut sebagai trend positif. Sebaliknya, apabila rata-rata perubahan menunjukkan penurunan, maka disebut sebagai trend negative (Wibowo & Faqih, 2020). DIBagi menjadi empat metode yang biasanya digunakan untuk menyusun atau menentukan trend linear:

1. Metode Semi Rata-rata (Semi Avarage Method)
2. Metode bebas (Freehand Method)
3. Metode Rata-rata Bergerak (Moving Avarage Method)
4. Metode Kuadrat Terkecil (Least Square Method)

1.5 Metode Least Square

Metode Least Square merupakan salah satu metode berupa data deret berkala atau time series, data yang dibutuhkan ialah data historis di masa lampau untuk melakukan peramalan di masa mendatang sehingga dapat dditentukan hasilnya. Least Square biasanya digunakan untuk melihat trend dari data deret waktu (Tulsi et al., 2022).

$$Y' = a + bX \dots\dots\dots (2.1)$$

Keterangan:

- Y' : Hasil Prediksi
- X : Variabel waktu.
- a : Besar nilai trend.
- b : perubahan nilai trend.

Bedasarkan jumlah data, Metode Least Square dibagi menjadi dua kasus, yaitu:

1. Kasus Data Ganjil

Jika data ganjil, maka jarak antara dua waktu diberi nilai 1 satuan. Diatas 0 diberi tanda + dan dibawahnya diberi tanda - (Misal: 0, 1, 2, 3... dan ..., -3, -2, -1, 0).

2. Kasus Data Genap

Jika jumlah data genap, mak jarak antara dua waktu diberi nilai 2 satuan. Diata 0 diberi tanda + dibawahnya tanda - (0, 1, 3, 5, ... dan ... -5, -3, -1, 0) (Adha & Toyib, 2025).

untuk menghitung koefisien a dan b dapat menggunakan persamaan 2.2 dan 2.3.

$$a = \frac{\sum Y}{n} \dots\dots\dots (2.2)$$

$$b = \frac{\sum X Y}{\sum X^2} \dots\dots\dots (2.3)$$

Keterangan:

Y: Data berkala (Time Series).

X: Variabel waktu.

a: Besar nilai trend.

b: perubahan nilai trend.

n: Jumlah total data.

Setelah didapat nilai dari koefisien a dan b maka koefisien tersebut digunakan dalam persamaan 2.1 untuk memprediksi data dimasa yang akan datang.

1.6 Perhitungan Error

Metode perhitungan error dalam analisis data memiliki peranan penting dalam mengevaluasi tingkat dari akurasi suatu model prediksi atau peramalan. Dua metode ini yang sering digunakan untuk menilai besarnya kesalahan dalam hasil peramalan ialah Mean Absolute Deviation (MAD) dan Mean Absolute Percentage Error (MAPE).

1. Mean Absolute Deviation (MAD)

Merupakan metode untuk mengevaluasi metode peramalan menggunakan jumlah dari kesalahan-kesalahan yang absolut. Mean Absolute Deviation (MAD) digunakan untuk mengukur ketepatan ramalan dengan merata-rata kesalahan selisih antara nilai prediksi dan nilai peramalan (nilai absolut masing-masing kesalahan) dalam unit ini yang sama sebagai deret asli. Berikut persamaan yang digunakan untuk MAD:

$$MAD = \frac{\sum_{t=1}^n |A_t - F_t|}{n} \dots\dots\dots (2.4)$$

Keterangan:

A_t : Permintaan Aktual pada periode-t

F_t : Peramalan Permintaan (Forecast) pada

n: Jumlah total data.

2. Mean Absolute Percentage Error (MAPE)

MAPE merupakan rata-rata kesalahan absolut yang dihitung dalam suatu periode tertentu, kemudian dikalikan dengan 100% untuk menghasilkan nilai dalam bentuk persentase. MAPE digunakan terutama untuk ukuran variabel yang diprediksi memiliki peran penting dalam menentukan tingkat akurasi peramalan. Berikut persamaan yang digunakan untuk menghitung MAPE:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \dots\dots\dots (2.5)$$

Keterangan:

A_t : Permintaan Aktual pada periode-t

F_t : Peramalan Permintaan (Forecast) pada

n : Jumlah total data.

1.7 Penelitian Terdahulu

Pecarian artikel merupakan bentuk tulisan yang merangkum penelitian ilmiah terkait dengan suatu topik tertentu. temuan dalam penelitian ini dapat dianggap sebagai ringkasan atau evaluasi terhadap karya ilmiah lainnya. Tujuan utamanya adalah untuk memperdalam pemahaman mengenai suatu topik tertentu.

Tabel 2. 1 Penelitian Terkait

No	Masalah	Metode	Hasil Penelitian	Landasan Literatur
1	Pada Perusahaan Bus Transwisata untuk mengetahui penghasilan yang akan didapatkan pada periode selanjutnya masih menggunakan perkiraan berdasarkan catatan penghasilan pada pendapatan sebelumnya sehingga menyebabkan ketidakakuratan sebuah prediksi tersebut.	<i>Least Square</i>	Bedasarkan hasil penelitian maka didapatkan prediksi penghasilan pada tahun 2020 sebesar Rp. 323.868.188. Memiliki nilai evaluasi akurasi atau nilai Error MAD sebesar 4.316.674,825 dalam rupiah. Sedangkan nilai MSE MSE mendapatkan nilai 2.236 dan nilai MAPE adalah 0.236%.	(Akbar & Rakhma Devi, 2022) “Prediksi Penghasilan Perusahaan Bus Transwisata Menggunakan Metode Leaste Square”
2	Masalah yang terjadi pada pabrik tahu adalah jumlah persediaan bahan baku kedelai yang sangat sulit dikendalikan karena kebutuhan bahan baku kedelai yang	<i>Least Square</i>	Dengan adanya sistem peramalan ini jumlah persediaan bahan baku kedelai dapat kendalikan karena adanya sistem prediksi yang akurat. Hasil perhitungan	(Nuri Fazira et al., 2023). “PREDIKSI KEBUTUHAN BAHAN BAKU KEDELAI DI PABRIK TAHU BUK IYEM MENGGUNAKAN

	naik turun dari waktu ke waktu.		dengan menggunakan metode least square adalah hasil peramalan kebutuhan bahan baku kedelai bulan Januari tahun 2023 sebesar 5.376 kg, Februari 2023 sebesar 5.433 kg dan Maret 2023 sebesar 5.489 kg dengan tingkat kesalahan (error) sebanyak MSE 69196, RMSE 263, MAD 214 dan MAPE 0,04%.	METODE LEAST SQUARE”
3	Permasalahan yang dihadapi pemilik Toko Suryono hanya mempediksi dengan perkiraan tanpa ada perhitungan yang jelas dan akurat, sehingga terkadang Toko Suryono terlalu sedikit mengambil	<i>Least Square</i>	Aplikasi peramalan menggunakan metode Least Square dapat memberikan kemudahan dalam memprediksi penjualan sembako agar dapat meminimalisir penumpukan dan penurunan kualitas	(Maulidya et al., 2021). “METODE LEAST SQUARE SEBAGAI PREDIKSI PENJUALAN SEMBAKO DI TOKO SURYONO”

	<p>persediaan padahal kebutuhan konsumen sedang meningkat dan terkadang Toko Suryono terlalu banyak mengambil persediaan padahal permintaan konsumen sedang menurun.</p>		<p>barang akibat kelebihan persediaan. Dari keseluruhan perhitungan sembako maka didapatkan hasil MAD paling rendah 16,51 dan menghasilkan MAPE sebesar 1,73%.</p>	
4	<p>Diperlukannya prediksi untuk penyediaan stok ayam, serta mengurangi out of stok yang sering terjadi.</p>	<p><i>Single Moving Average (SMA)</i> dan <i>Single Exponential Smoothing (SES)</i>.</p>	<p>MAD yang diperoleh 3,116, MAPE 9% dan MSE 2,762. Sedangkan SES memiliki tingkat kesalahan yaitu MAD 3,134, MAPE 10% dan MSE 2,889.</p>	<p>(Asynari et al., 2020). “ANALISIS PERAMALAN PERMINTAAN PADA GEPREK BENSU MENGGUNAKAN METODE TIME SERIES”</p>
5	<p>Pada coffeshop kopinesia, perekapan data yang dilakukan masih manual yakni dengan menulis sendiri hasil pendapatan yang sudah di dapat</p>	<p><i>Least Square</i></p>	<p>Hasil uji perhitungan metode menggunakan data pendapatan kopinesia bulan januari tahun 2018, untuk memprediksi pendapatan pada</p>	<p>(Firstiano et al., 2020). “FORECASTING OMZET MENGGUNAKAN METODE LEAST SQUARE”</p>

			<p>bulan januari tahun 2019 mendapatkan hasil Rp. 1.649.531, sedangkan hasil data yang sebenarnya sebesar Rp. 1.440.000, menghasilkan presentase hasil 87,29% dan nilai error 12.71%.</p>	
6	<p>Kurangnya persiapan dapat menyebabkan kurangnya kenyamanan bagi para wisatawan yang datang, karena para wisatawan tidak sepenuhnya dapat menikmati sarana dan prasarana yang disediakan.</p>	<p><i>Least Square</i></p>	<p>Bedasarkan hasil penelitian, maka didapatkan jumlah wisatawan mancanegara yang datang ke Ba pada Bulan Januari 2021 – Mei 2022 sebanyak 1.338,62 orang, dengan nilai forecast error MAD sebesar 1.000,8 dan nilai MAPE sebesar 17,4%</p>	<p>(Karisma & Lattu, 2022). “FORECASTING JUMLAH WISATAWAN BUMI MANDIRI DENGAN METODE LEAST SQUARE”</p>
7	<p>Banyaknya keluaran produk model terbaru secara berulang dan</p>	<p><i>Leaste Square</i></p>	<p>Hasil penelitian menunjukkan perhitungan peramalan</p>	<p>(Suhardjono, 2023). “Optimalisasi Pengambilan</p>

	<p>permintaan pelanggan yang tidak tetap, mengakibatkan kebingungan dalam mengisi ulang produk yang tidak pasti jumlahnya tiap bulan.</p>		<p>menghasilkan nilai MAPE sebesar 1% serta terdapat hasil analisis trend penjualan yaitu total kenaikan trend sebesar 29,04% pada jenis barang karcher sehingga analisis trend penjualan dapat dikontribusikan pada tempat penelitian</p>	<p>Keputusan Melalui Analisis Trend Penjualan pada Bisnis Retail Menggunakan Metode Least Square”</p>
8	<p>Masalah yang di hadapi dalam penelitian ini adalah Salah satu daerah penghasil padi di Tuban adalah di kecamatan palang dengan luas sawah mencapai 53,7%. Namun, luasnya lahan ini tidak diimbangi dengan konsistennya hasil panen padi. Hasilnya tidak</p>	<p><i>Least Square</i></p>	<p>Hasil penelitian menunjukkan bahwa hasil perhitungan nilai error peramalan menggunakan Metode MAD (Mean Absolute Deviation) dengan menghitung jumlah kesalahan-kesalahan yang absolut menghasilkan error sebesar 577,058824,</p>	<p>(Sari & Kurniawati, 2020) “PERAMALAN PRODUKSI PADI MENGGUNAKAN METODE LEAST SQUARE DI DESA LERANWETAN KECAMATAN PALANG KABUPATEN TUBAN”</p>

	konsisten di setiap tahunnya.		sedangkan hasil perhitungan MAPE menghasilkan eror yang cukup kecil yaitu 2,71659365%.	
9	Pengrajin ini masih menggunakan sistem manual untuk menentukan jumlah stok yang akan di jual bulan depan, pengrajin sangat kebingungan dalam menentukan stok alat yang sudah jadi yang akan di buat bulan depan	<i>Least Square</i>	Hasil penelitian ini untuk melakukan prediksi penjualan saat pengujian dengan data uji 11 bulan cukup optimal. Akurasi tersebut di ketahui bahwa dengan data uji 11 bulan memiliki hasil 94,59%. Untuk hasil MAD dan MAPE penjualan rebana Al-Banjari hasilnya yaitu 2,04% dan 5,4%.	(Maulana et al., 2024). “Prediksi Penjualan Rebana Al-Banjari Menggunakan Metode Least Square”
10	Toko mengalami kesulitan dalam memperkirakan jumlah barang yang perlu disediakan untuk periode satu bulan ke depan karena adanya	<i>Least Square</i>	Dari hasil penelitian ini, perhitungan penjualan gaun pada bulan September adalah 373 unit, kesalahan dalam perhitungan	(Indah. et al., 2022). “PREDIKSI PEMASARAN PAKAIAN DENGAN MENGGUNAKAN LEAST

ketidakstabilan dalam jumlah penjualan dan barang yang terjual setiap bulannya.		penjualan gaun adalah MAD 136,8 MSE 93571,2, dan MAPE adalah 12%.	SUQUARE METHOD DI TOKO NAURA COLLECTION”
---	--	---	--

