

BAB II

LANDASAN TEORI

Teori penunjang sangat penting sebagai acuan analisis dan perencanaan sistem. Pada bagian ini akan dijelaskan dasar teori yang digunakan untuk pembuatan proyek akhir ini.

2.1 DATA MINING

Data Mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar (Budiarti, 2022) *Data Mining* juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

Secara garis besar proses KDD (*Knowledge Discovery and Data*) terdiri dari 7 tahapan (Harahap, 2019):

1. Pembersihan Data (*Data Cleaning*)

Data cleaning Menghilangkan tidak konsistenan data, noise dan juga data yang kurang relevan. biasanya sebuah data dari database besar memiliki data yang tidak sempurna atau memiliki kekurangan misalnya adanya data yang kurang valid atau salah dalam penginputan data. *Data cleaning* sangat mempengaruhi performasi pada Teknik data mining karena data yang ditangani jumlah dan kompleksitasnya.

2. Integrasi Data (*Data Integration*)

Penggabungan sebuah data dari beragam *database* ke dalam satu *database* lainnya. Integrasi data dilakukan pada atribut –atribut yang mengidentifikasi entitas – entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang

atau bahkan menyesatkan dalam pengambilan aksi nantinya.

3. Seleksi Data (*Data Selection*)

Dalam penyeleksian hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

4. Tranformasi Data (*Data Transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti *analisis asosiasi* dan *clustering* hanya bisa menerima input data kategorikal. Karena data berupa angka numerik yang berlanjut perlu dibagi – bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses Mining

Suatu proses pokok atau utama saat metode diterapkan guna menemukan pengetahuan berharga dan tersembunyi dari data.

6. Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi model-model yang menarik kedalam *knowledge based* yang ditemukan. Dalam tahapan ini hasil akhir dari teknik *data mining* berupa model yang khas maupun pola prediksi dievaluasi untuk menilai hipotesa yang ada.

7. Presentasi Pengetahuan (*Knowledge Presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap akhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.

2.2 FUNGSI DATA MINING

Pada umumnya fungsi data mining terbagi menjadi dua bagian. Yakni deskriptif dan prediktif, tetapi selain itu memiliki fungsi lain seperti *asosiasi*, *klasifikasi*, *clustering*, *forecasting* dan *sequencing* (Mustika & Ardilla, 2021).

1. Deskriptif (*Description*)

Deskriptif lebih kepada merujuk ke fungsi dalam pemahaman data. Tujuan dari proses ini adalah untuk menemukan pola dan karakteristik yang terdapat pada data. Fungsi deskriptif dimanfaatkan sebagai pattern tertentu yang awalnya tidak terlihat di dalam data.

2. Prediktif (*Prediction*)

Fungsi terkait dengan proses yang nantinya digunakan untuk mengetahui pola khusus dari data yang digunakan. Pola ini bisa ditemukan dari beberapa variabel dalam data, ketika sudah menemukan pola, maka pola yang digunakan dipakai untuk memperkirakan variabel lain dan masih belum diketahui nilainya karena itu disebut fungsi prediktif. Contoh prediksi adalah:

- a. Prediksi harga beras dalam tiga bulan yang akan datang.
- b. Prediksi tingkat pengangguran lima tahun akan datang.
- c. Prediksi persentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

3. Asosiasi (*Association*)

Fungsi asosiasi adalah fungsi data mining yang dapat diproses untuk melakukan identifikasi relasi atau hubungan dari setiap data yang ada. Data ini bisa merupakan data dahulu maupun data yang didapat saat ini. Contoh asosiasi adalah :

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons positif terhadap penawaran upgrade layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

4. Klasifikasi (*Classification*)

Klasifikasi dilakukan untuk sebagai cara menyimpulkan beberapa pengertian karakteristik dari suatu grup atau kelompok data. Seperti data pelanggan yang tak lagi menggunakan produk, karena menganggap produk kompetitor memberi manfaat lebih banyak dan customer value untuk para pelanggan. Contoh klasifikasi adalah :

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk penyakit apa.

5. Pengklasteran (*Clustering*)

Clustering merupakan fungsi yang mengarah pada proses identifikasi kelompok, kemudian produk atau barang yang memiliki karakteristik khusus. Biasanya digunakan dalam mengetahui kelompok-kelompok tertentu dalam penyebarannya. Contoh pengklusteran dalam bisnis dan penelitian adalah :

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan
- c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar

6. Peramalan

Disebut juga dengan teknik *forecasting* yang dipakai untuk memperoleh gambaran mengenai nilai dari suatu data di masa mendatang, Forecasting dan peramalan ini bisa dilakukan dengan pengumpulan informasi dalam jumlah yang besar, contoh penerapan forecasting merupakan data terkait peramalan jumlah permintaan terhadap produk tertentu.

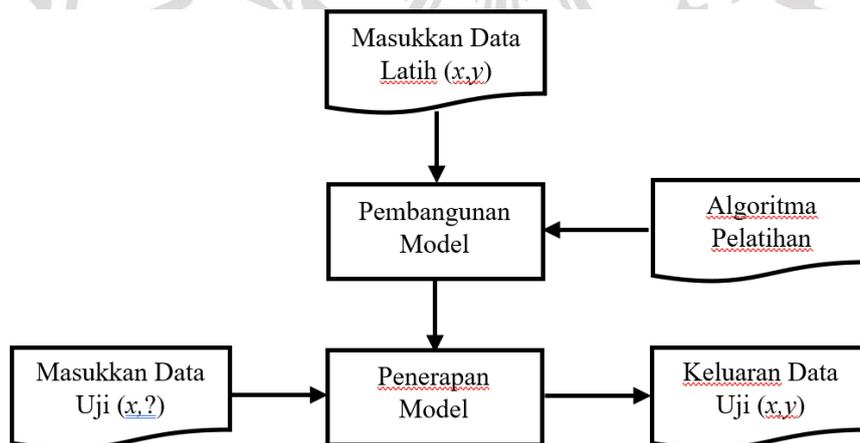
7. *Sequencing*

Sequencing merupakan fungsi terakhir dari data mining, proses identifikasi setiap hubungan yang berbeda dalam periode waktu tertentu. Kemudian sequencing merupakan data pelanggan ketika melakukan repeat purchase suatu produk yang terjadi atau dilakukan secara berulang pada pelanggan.

2.3 KLASIFIKASI

Salah satu tugas yang dapat dilakukan dengan data mining adalah pengklasifikasian. Klasifikasi dapat diartikan secara detail sebagai suatu proses yang melakukan pembelajaran atau pelatihan terhadap fungsi target f yang memetakan setiap vector (set fitur) x ke dalam satu dari sejumlah label kelas y yang tersedia (Amilia, 2021). Proses pelatihan tersebut akan menghasilkan suatu model yang kemudian disimpan sebagai memori.

Dapat juga diartikan sebagai suatu kumpulan data testing digunakan untuk menentukan keakuratan suatu pola. Umumnya, data set yang diberikan dibagi ke dalam sekumpulan data latih dan data uji, dimana data latih digunakan untuk membentuk model dan data uji digunakan untuk menguji.



Proses klasifikasi didasarkan pada empat komponen :

1. Kelas

Variabel terikat yang berupa kategorikal yang mempresentasikan label yang terdapat pada objek. Contohnya : resiko penyakit jantung, resiko kredit.

2. *Predictor*

Variabel bebas yang direpresentasikan oleh karekteristik data. Contohnya: merokok, tabungan.

Training dataset

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.

3. *Testing dataset*

Berisi data yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.4 NORMALISASI DATA

Nilai-nilai atribut data yang variatif perlu dinormalisasikan atau distandarisasikan agar proses *Data Mining* tidak bias. Biasanya normalisasi data dilakukan ke dalam rentang yang kecil, seperti [0,1], sehingga semua atribut akan memiliki bobot yang sama. Teknik normalisasi sangat penting dalam data mining, khususnya klasifikasi dan klasterisasi.

2.3.1 MIN-MAX NORMALIZATION

Metode ini *re-scale* data dari suatu range ke range baru lain. Data di skalakan dalam range 0 dan 1. Metode ini menggunakan nilai minimum dan maksimum untuk melakukan konveksi data secara linier. Misalkan A adalah atribut bertipe numerik, min_A adalah minimum dari atribut A , $maks_A$ adalah maksimum dalam atribut A . Suatu nilai x_i dapat dinormalisasikan menjadi nilai baru x_i^1 yang berada dalam rentang $[new_{min}(A), new_{max}(A)]$ dengan persamaan (1).

$$x_i^1 = \frac{x_i - min_A}{maks_A - min_A} (new_{max}(A) - new_{min}(A)) + new_{min}(A) \quad (2.1)$$

Dimana,

x : atribut data

x^1 : nilai lama dari setiap entri dalam data

$\min(x)$ $\max(x)$: nilai absolut minimal dan maksimal dari x

$new_{\max}(x)$ dan $new_{\min}(x)$: nilai minimal dan maksimal dari range

2.3.2 Z-CORE NORMALIZATION

Teknik ini menggunakan rata-rata dan standar deviasi untuk menormalisasi tiap *input* sesuai dengan persamaan (2).

$$x_i^1 = \frac{x_i - \bar{A}}{\sigma_A} \quad (2.2)$$

Dimana,

x : nilai asli

μ : rata-rata data

σ : standar deviasi data

Dengan memanfaatkan nilai rata-rata dan deviasi standar (σ), metode ini lebih stabil terhadap nilai pencilan maupun adanya nilai-nilai baru yang lebih besar daripada $new_{\max}(A)$ atau lebih kecil daripada $new_{\min}(A)$.

2.5 K-NEAREST NEIGHBOR

K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi yang sangat sederhana dan mudah diimplementasikan hanya dengan mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data/tetangga (*neighbor*) terdekat. Algoritma ini mengklasifikasikan data berdasarkan similarity atau kemiripan atau kedekatannya terhadap data lainnya. Dalam hal ini jumlah data/tetangga terdekat di tentukan oleh user yang dinyatakan

dengan k (Nugraha, 2020). Secara ringkas langkah-langkah *K-Nearest Neighbour* dapat diuraikan sebagai berikut .

1. Masukkan data training, label data training, k , data testing.
2. Hitung jarak pada data testing ke setiap data training.
3. Tentukan nilai k dari data training dengan jarak paling mendekati data testing.
4. Periksa label dari data yang diperoleh dengan nilai jarak terdekat berdasarkan nilai k yang telah ditentukan.
5. Tentukan label dengan frekuensi paling banyak dari data yang diperoleh sejumlah nilai k .
6. Masukkan data testing ke kelas dengan frekuensi paling banyak.
7. Proses klasifikasi berakhir.

Dalam mengukur jarak antara dua objek data terdapat 2 formula yang memiliki atribut numerik, yaitu *Euclidean distance* dan *Manhattan distance*.

Euclidean Distance adalah metode yang dapat digunakan pada *1-dimensional space*, *2-dimensional space*, atau *multi-dimensional space*. *1-dimensional space* berarti perhitungan jarak hanya dengan satu variabel bebas (*independent variable*), *2-dimensional-space* berarti ada dua variabel bebas, dan *multi-dimensional space* berarti ada lebih dari dua variabel.

Secara umum, formula *euclidean distance* pada *1-dimensional space* seperti pada persamaan (3).

$$d(i, j) = \sqrt{\sum_{i=1}^n (x_{i1} - x_{j1})^2} \quad (2.3)$$

Dimana,

x_i : sampel data

y_i : data uji atau data testing

i : variabel data

$d(i, j)$: dissimilarity/jarak

n : dimensi data

Jika ada lebih dari satu, kita dapat menjumlahkannya seperti di pada persamaan (4)

$$d(i, j) = \sqrt{\sum_{i=1}^n (x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (2.4)$$

Dimana,

d : jarak antara x dan y

x : data pusat kluster

y : data pada atribut

i : setiap data

n : jumlah data

x_i : data pada pusat kluster ke- i , ($i = 1, 2, 3, \dots, n$)

y_i : data pada setiap data ke- i , ($i = 1, 2, 3, \dots, n$)

Dimana i dan j adalah dua objek data yang memiliki p atribut bernilai numerik, yang dinyatakan sebagai $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ dan $j = (x_{j1}, x_{j2}, \dots, x_{jp})$

Manhattan Distance atau *city block* adalah ukuran jarak yang mengadaptasi jarak antara dua titik data dalam jalur seperti grid, misalnya 3 blok ke kanan dan 4 blok ke bawah sehingga jarak totalnya adalah 7 blok. Berdasarkan persamaan (5).

$$d(i, j) = \sum_{i=1}^n |x_i - x_j| \quad (2.5)$$

Dimana,

d : jarak antara x dan y

x_i : atribut dari data ke- i , ($i = 1, 2, 3, \dots, n$)

y_i : atribut pusat *Cluster* ke-I ($i = 1, 2, 3, \dots, n$)

Dalam pengaturan parameter k dengan jumlah genap, ada kemungkinan mendapatkan hasil yang sama sehingga KNN kesulitan dalam melakukan klasifikasi. Parameter k berfungsi untuk mengatur tingkat generalisasi terhadap data-data yang akan datang (di luar *data training*). Dengan parameter k yang kecil KNN akan memiliki tingkat generalisasi data yang tinggi. Sebaliknya dengan parameter k yang besar, knn akan memiliki tingkat generalisasi data yang rendah dan mungkin menjadi overfit, hanya mampu mengklasifikasi *data training* saja namun gagal mengklasifikasikan data-data baru yang akan datang.

2.6 MODIFIED K-NEAREST NEIGHBOR

Modified K-NN merupakan pengembangan dari algoritme *K-Nearest Neighbor* dengan menambahkan beberapa alur proses yakni perhitungan nilai validasi data latih dan *Weight Voting* atau pembobotan. Perhitungan nilai validasi bertujuan guna mengatasi permasalahan data yang menyimpang pada algoritme KNN sehingga kurangnya bias dari data latih dan *Weight Voting* untuk menghitung bobot dari data (Hamid, 2019). Algoritme *Modified K-NN* melakukan proses klasifikasi pada data uji dengan menggunakan hasil *Weight Voting* terbesar dari kelas yang ada pada data latih yang datanya sudah dilakukan validasi, berbeda dengan algoritme K-NN yang tidak melakukan proses validasi dari data latih. Dengan ditambahkan proses perhitungan validasi data latih ini dapat menghasilkan akurasi dari sistem yang rata-rata lebih tinggi dari penggunaan algoritme K- NN.

2.7 BLACK BOX TESTING

Black Box Testing adalah salah satu Teknik pengujian perangkat lunak yang dilakukan tanpa mengetahui secara detail bagaimana kode program bekerja di dalamnya. Teknik ini melihat sistem atau program sebagai sebuah kotak hitam

(*Black Box*) dimana input diberikan dan output yang dihasilkan dievaluasi, tanpa memperhatikan bagaimana alur program didalamnya (Rosa dan Salahuddin, 2013).

Tujuan dari *Black Box testing* adalah untuk memastikan bahwa sistem atau program berfungsi dengan benar sesuai dengan persyaratan fungsional dan non fungsional yang telah ditetapkan, serta mengidentifikasi kesalahan atau kegagalan dalam program (Gunawan, 2021)

2.8 Tinjauan Pustaka

Sebagai upaya penguatan topik penelitian yang diambil, penulis melakukan analisis dari hasil riset penelitian sebelumnya yang berkaitan dengan topik penelitian. Berikut ini beberapa hasil dari penelitian sebelumnya dapat dilihat pada tabel 2.1 dibawah ini :

Tabel 2. 1 Penelitian Sebelumnya

No	Peneliti	Judul	Abstrak
1.	Andreas Pratama, Budi Serasi Ginting, Nurhayati.	Penerapan Data Mining Untuk Prediksi Merek Pakain Yang Paling Diminati Dengan Metode K-Nearest Neighbor (Studi Kasus PT.Matahari Departemen Store Binjai) (2021)	Pada penelitian ini Dengan perhitungan data mining menggunakan teknik klasifikasi dengan algoritma K-Nearest Neighbor yang paling mayoritas maka dapat diprediksikan jumlah penjualan pada periode berikutnya mengalami peningkatan dengan rata-rata prediksi setiap bulanya 14,900 dan di dapat merek pakaian yang paling diminati yaitu Cardinal . Menentukan kelompok data hasil uji berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai $k = 3$ maka diambil jarak terkecil yaitu $d_{20} = 1,732$, $d_{25} = 3,873$, $d_{28} = 4,359$. 5. Dengan menggunakan kategori K-Nearest Neighbor yang paling mayoritas maka dapat diprediksikan jumlah penjualan

No	Peneliti	Judul	Abstrak
			pada periode berikutnya mengalami peningkatan dengan rata-rata prediksi setiap bulanya 14,900 dan di dapat merek pakaian yang paling diminati yaitu Cardinal.
2.	Selsa Amelia, Memi Nor Hayati, Surya Prangga.	Penerapan Metode Modified K Nearest Neighbor Pada Pengklasifikasian Status Pembayaran Kredit Barang Elektronik dan Furniture (2022)	Penelitian ini bertujuan untuk mengetahui hasil pengklasifikasian status pembayaran kredit barang elektronik dan furniture serta tingkat akurasi klasifikasi pada metode MKNN. Data yang digunakan adalah data debitur PT.KB Finansia Multi Finance Tahun 2020 dengan status pembayaran kredit lancar dan tidak lancar serta menggunakan 7 variabel bebas yaitu usia, jumlah tanggungan, lama tinggal, pendapatan, masa kerja, besar pembayaran kredit, dan lama peminjaman kredit. Berdasarkan penelitian yang telah dilakukan, diperoleh nilai akurasi sebesar 84,61% dengan K optimal yaitu $K = 5$ pada proporsi 90% : 10%..
3	Amanda Pratiwi1, Ananto Tri Sasongko, Dendy K. Pramudito.	Analisis Prediksi Gilingan Plastik Terlaris Menggunakan Algoritma K-Nearest Neighbor di CV Menembus Batas (2023).	Prediksi Penerapan metode algoritma K-Nearest Neighbor dalam prediksi penjualan membantu perusahaan untuk mengoptimalkan produksi, mengatur stok, dan mengalokasikan sumber daya dengan lebih efisien. Hasil penelitian menunjukkan bahwa Algoritma K-Nearest Neighbor memberikan prediksi yang sangat akurat, dengan nilai akurasi, recall, dan precision mencapai 1.0 dalam klasifikasi produk, sehingga dapat diandalkan dalam mendukung

No	Peneliti	Judul	Abstrak
			<p>keberlanjutan bisnis daur ulang plastik di tengah tantangan global terkait sampah plastik. Berdasarkan hasil analisa dan pengujian terhadap data transaksi penjualan produk yang diambil satu tahun terakhir (Periode bulan April 2022 sampai bulan April 2023), untuk menentukan prediksi gilingan terlaris menggunakan metode K-Nearest Neighbor maka dapat disimpulkan. Metode KNearest Neighbor dapat digunakan untuk prediksi gilingan terlaris dengan menggunakan variabel kode barang 1, kode barang 2, kode barang 3 kode barang 4, qty1, qty2, qty3 dan qty4. Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan K-Nearest Neighbor. Didapatkan hasil prediksi penjualan gilingan plastik terlaris di CV Menembus Batas adalah produk Gilingan ACR. Tools Jupyter notebook dapat digunakan untuk melakukan prediksi gilingan pelastik terlaris dengan penerapan pengujian menggunakan metode Knearest Neighbor. Berdasarkan hasil dari pengolahan dataset, nilai akurasi, recall dan precession terhadap klasifikasi penjualan gilingan plastik terlaris menggunakan K-Nearest Neighbor sebesar akurasi: 1.0, recall: 1.0 dan precision: 1.0</p>
4.	Dina Meilida Meliala, Penda	Perbandingan Algoritma K-Nearest Neighbor Dengan	Memprediksi penjualan sangat penting dalam kemajuan sebuah usaha, terutama dalam penjualan barang yang

No	Peneliti	Judul	Abstrak
	Hasugian.	DecisionTree Dalam Memprediksi Penjualan Makanan Hewan Peliharaan Di Petshop Dore Vet Clinic (2020).	<p>memiliki tanggal kadaluarsa seperti makanan hewan peliharaan. Ada beberapa algoritma yang digunakan untuk menginformasikan prediksi harga penjualan salah satunya algoritma K-Nearest Neighbor dan algoritma Decision Tree. Dengan metode K-nn, dihasilkan kondisi dari 30 data, 6 data diklasifikasikan terlaris sesuai dengan prediksi yang dilakukan dengan metode k-nn, 3 data dari 6 data diprediksi terlaris ternyata tidak terlaris, (data urutan 1, 2, 6). 24 data diprediksi tidak terlaris ternyata 10 data sebelumnya diklasifikasikan terlaris (data urutan 22, 5, 16, 26, 28,19, 17, 20, 23, 24). Dengan metode decision tree algoritma C45, diketahui dari 30 data, merek purina terlaris, ada 4 data daripada royal canin (false negative). Hasil tingkat akurasi decision tree algoritma c45, diketahui true terlaris = 17, false tidak terlaris = 4. False terlaris = 16, true tidak terlaris = 13. Akurasi decision tree algoritma c45 = 83%. Kesimpulan dari hasil penelitian yaitu Penerapan Decision Tree algoritma C45 dapat dipergunakan untuk membantu memprediksi penjualan makanan hewan peliharaan di Petshop Dore Vet Clinic. Hasil Tingkat Akurasi Decision Tree Algoritma C45, Diketahui True terlaris = 17, False Tidak Terlaris = 4. false terlaris = 16, True Tidak Terlaris = 13. Akurasi</p>

No	Peneliti	Judul	Abstrak
			Decision Tree Algoritma C45 = 83%.
5.	Herianto, Nur Syamsiyah , Sapitri Anggraini	Analisa Tingkat Penjualan Produk Menggunakan K-Nearest Neighbor (K-NN) dan K-Means (Studi Kasus Perusahaan Kayu Elang Perkasa) (2021).	Dari penelitian ini akan diuji yang lebih baik apakah menggunakan analisa klasifikasi atau clustering dengan menggunakan data yang sama. Berdasarkan hasil pengujian akurasi kedua metode diperoleh bahwa K-Means memiliki tingkat akurasi yang lebih baik sebesar 78,37% sedangkan K-NN memiliki tingkat akurasi 76,06%. Penelitian ini berhasil membangun dan merancang sebuah analisis penjualan yang dapat digunakan untuk membantu pihak perusahaan dalam menyusun strategi penjualan agar memenuhi kepuasan pelanggan. 2. Berdasarkan hasil pengujian dengan mengukur kinerja kedua metode tersebut menggunakan rumus akurasi, diperoleh bahwa K- Means memiliki tingkat akurasi yang lebih baik yakni sebesar 78,37% sedangkan K-NN memiliki tingkat akurasi yakni 76,06%.
6.	Rismala, Irfan Ali , Ade Rizki Rinaldi .	Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Sepeda Motor Terlaris (2023).	Penggunaan sepeda motordi Indonesia setiap tahunnya selalu meningkat Seiring dengan meningkatnya jumlah kebutuhan akan alat transportasi banyak produsen sepeda motor yang mengeluarkan produk dengan bermacam-macam merek dan desain untuk memenuhi permintaan konsumen.Suatu strategi untuk meningkatkan penjualan produk yang akan dijual sangat dibutuhkan

No	Peneliti	Judul	Abstrak
			<p>pengembang karena tingkat persaingan di dunia bisnis, terutama di insustri penjualan, salah satunya dengan memanfaatkan data penjualan setiap harinya. Guna mengetahui produk mana yang paling banyak di jual (laris) dan produk mana yang kurang dalam penjualannya (tidak laris). Penelitian ini melakukan proses data mining pada data penjualan dari PT. Sumber Rejeki Jabar dari bulanjanuari – desember tahun 2022. Hasil nilai akurasi terhadap data penjualan sepeda motor dari bulan Januari - Desember tahun 2022 di PT. Sumber Rejeki Jabar dengan nilai K sebesar 5 yaitu sebesar 96,15%.Berdasarkan uji coba yang dilakukan menunjukan bahwa algoritma K-NN dapat digunakan dalam klasifikasi data penjualan sepeda motor pada PT. X. Berdasarkan pengujian didapatkan hasil akurasi sebesar 96,15% yang artinya data set dapat digunakan pada tahapan selanjutnya sebagai data yang valid untuk digunakan</p>
7.	Taufik Hidayat, Yuni Handayani, Ahmad Syaifudin.	Implementasi AlgoritmaK-Nearest Neighbor Untuk Meningkatkan Penjualan Produk Meuble Dan Furniture (2023).	Berdasarkan hasil data yang di peroleh dari transaksi penjualan selanjutnya di lakukan proses data mining dan dilanjutkan proses perhitungan dengan algoritma K- Nearest Neighbor hingga akhirnya di temukan hasil dan menjadi bahan pertimbangan dalam strategi penyetokan barang di gudanga, sehingga tidak terjadi penumpukan dalam jumlah

No	Peneliti	Judul	Abstrak
			<p>banyak dan sedikit sehingga dapat mengosongkan tempat dan membuat semuanya lebih efektif dan efisien. Adapun hasil perhitungan yang telah dilakukan dengan menerapkan perhitungan dengan K-Nearest Neighbor memperoleh hasil pengelompokan produk dari 10 jenis barang mebel dengan kategori laris sebanyak 2 jenis barang yaitu Kursi plastik dan Kasur. Sedangkan untuk kategori kurang laris didapatkan hasil sebanyak 8 jenis produk yaitu meja belajar, meja makan, meja tamu, lemari kayu, lemari kaca, lemari plastic, kursi kayu, dan rak piring. Dari hasil nilai akurasi terhadap pengelompokan produk penjualan paling laris dan kurang laris adalah sebesar 80%.</p>
8.	Muhammad Reza Ravi, Indriati, Sigit Adinugroho	Implementasi Algoritme Modified K-Nearest Neighbor(MKNN) Untuk Mengidentifikasi Penyakit Gigi Dan Mulut (2019).	<p>Metode MKNN adalah metode perkembangan dari KNN, terdapat perbedaan dari MKNN dan KNN yaitu MKNN terdapat proses perhitungan validitas dan Weight Voting. Penelitian ini menggunakan 6 kelas yang meliputi Pulpitis, Gingivitis, Karies Gigi, Periodontitis, Deposits, dan Nekrosis Pulpa. Penelitian ini membuktikan bahwa pada data latih sebanyak 70 dan data uji 30 serta nilai $K=60$, metode MKNN dapat melakukan identifikasi jenis penyakit gigi dan mulut dengan mencapai 86,6%. Pada penelitian ini juga membuktikan bahwa metode MKNN cenderung</p>

No	Peneliti	Judul	Abstrak
			<p>lebih tinggi akurasi dibandingkan dengan metode KNN dimana metode MKNN memiliki tingkat akurasi 76,66% sedangkan KNN 43,33%. hal tersebut disebabkan oleh adanya perhitungan nilai validitas yang akan mempengaruhi Weight Voting dan juga akurasi</p>
9.	Rizal setya perdana, Nurul Hidayat	Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam (2020)	<p>Demam merupakan indikator awal untuk beberapa penyakit antara lain demam berdarah, tifoid dan malaria disertai gejala yang mirip, antara lain nyeri otot, gangguan pencernaan, kondisi lidah serta pembesaran pada hati dan limpa. Kemiripan gejala dari masing-masing penyakit sering menimbulkan kesulitan dalam mendapatkan anamnesis (diagnosa sementara) sehingga pasien mendapatkan penanganan awal yang kurang tepat. Untuk menangani masalah tersebut diperlukan teknologi untuk mendapatkan diagnosa sementara dengan menerapkan salah satu metode klasifikasi yaitu Modified KNearest Neighbor (MKNN). Metode tersebut mempelajari pola dari data hasil pemeriksaan sebelumnya berdasarkan 15 gejala penyakit dengan proses perhitungan jarak euclidean, perhitungan nilai validitas dan perhitungan weighted voting yang hasil akhirnya digunakan untuk penetapan kelas klasifikasi berdasarkan nilai K yang telah ditentukan. Berdasarkan hasil pengujian terhadap perubahan</p>

No	Peneliti	Judul	Abstrak
			<p>nilai K, perubahan jumlah data latih dan perubahan komposisi data latih didapatkan rata-rata akurasi untuk pengujian pengaruh nilai K terhadap akurasi sebesar 88.55%. Nilai rata-rata akurasi yang didapatkan dari pengujian pengaruh variasi jumlah data latih adalah 92.42%. Pengujian pengaruh komposisi data latih terhadap akurasi mendapatkan nilai rata-rata akurasi sebesar 87.89%. Pengujian pengaruh komposisi data latih dan data uji terhadap akurasi mendapatkan nilai rata-rata akurasi sebesar 96.35%.</p>
10.	Sri puspita dewi	Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor (2022)	<p>Penerapan Data Mining sangatlah dibutuhkan oleh UD Andar dikarenakan usaha dagang ini menjual berbagai macam jenis produk. Usaha Dagang ini menjual jamu serbuk, kantong plastik, bahan pembuat makanan & minuman, dan makanan frozen yang diminati oleh para konsumen. Dilihat dari banyaknya permintaan konsumen ternyata terdapat beberapa produk terlaris dan tidak terlaris, sehingga berdasarkan data 1 tahun terakhir, maka dibutuhkan sebuah prediksi penjualan produk terlaris, agar mempermudah pihak usaha dagang dalam perencanaan penyedia stok. Dikrenakan sistem yang sedang berjalan saat ini masih manual, untuk itu data yang didapat kurang akurat dan efisien. Maka untuk mengatasi hal ini, diperlukan</p>

No	Peneliti	Judul	Abstrak
			sebuah sistem prediksi penjualan produk terlaris dengan teknik data mining yang menggunakan metode k nearest neighbor. Penelitian ini menghasilkan sebuah sistem algoritma k nearest neighbor pada teknik data mining yang membantu untuk memprediksi penjualan produk terlaris pada UD Andar.

