BAB II

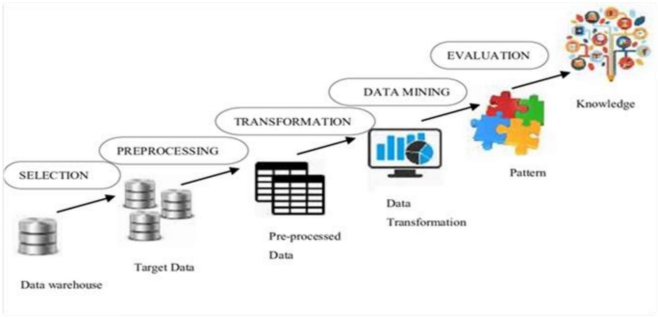
TINJAUAN PUSTAKA

2.1 Teori Dasar

Teori Dasar merupakan sebuah teori umum yang digunakan sebagai landasan penelitian ini, pada tahap teori dasar membahas atau menjelaskan teori–teori yang dipakai berdasarkan pada penelitian**.** Hal–hal yang akan dijelaskan pada bab ini antara lain : *Knowledge Discovery in Database* (KDD), data mining, *K–nearest neighbor*, *software* pendukung, objek penelitian, penelitian terdahulu dan kerangka pemikiran.

2.2 *Knowledge Discovery in Database* (KDD)

Knowledge Discovery in Databases (KDD) sebuah metode yang digunakan untuk mengekstrak informasi berguna dari database seperti pola atau informasi (Dewi & Rahayu, 2022). Prosedur ini memerlukan sejumlah tahapan analisis data yang rumit untuk menemukan wawasan yang dapat diterapkan untuk meramalkan atau membuat pilihan. Langkah–langkah proses KDD dapat di lihat dalam gambar 2.1.



Gambar 2.1 Tahapan Proses *Knowlledge Discovery in Database*

Pada metode KDD terdapat 5 tahapan yang dapat diterapkan untuk prediksi atau peramalan antara lain :

1 Data *Selection*

Pada tahap ini, kumpulan data dipilih sebelum penggalian informasi KDD dimulai. Data yang dipilih tersebut disimpan dalam file yang berbeda dari operasional database untuk digunakan dalam proses data mining.

2 *Pre-processing* atau pembersihan

Data yang menjadi fokus KDD harus dibersihkan sebelum dapat digunakan untuk data mining. Proses pembersihan meliputi, antara lain , pemeriksaan ketidakkonsistenan , perbaikan kesalahan, dan penghapusan data duplikat.

3 *Transformation*

Ini adalah proses yang memungkinkan data yang tidak terstruktur dan tidak teratur diubah menjadi format yang lebih sesuai untuk dijelaskan dan dipahami. Hasil pada tahap ini digunakan pada tahap berikutnya dari proses *Knowledge Discovery in Databases* (KDD).

4 Data Mining

Proses ini adalah pencarian pola atau informasi yang menarik dalam data yang telah dipilih. Ada banyak algoritma, teknik dan metode data mining yang berbeda, dan pemilihan teknik atau algoritma ini sangat bergantung dengan tujuan serta proses *Knowledge Discovery in Databases* (KDD) secara menyeluruh.

5 Interpretation atau Evaluasi

Proses ini menerjemahkan pola dari data mining dan menganalisis apakah informasi atau pola yang ditemukan cocok atau berbeda dengan fakta atau hipotesis sebelumnya.

2.3 Data Mining

Data mining bagian dari teknologi yang tergolong campuran teknik-teknik analisis data dengan algoritma-algoritma untuk memproses data berukuran besar yang digunakan untuk memprediksi nilai yang akan dicapai pada satu periode (Elgohary et al., 2023).

Data mining sebuah proses mengumpulkan dan menemukan informasi penting dan relevan dari sejumlah besar database dengan menggunakan matematika, statistika, AI, dan pembelajaran mesin serta menganalisa data yang berguna untuk pengetahuan dari database yang besar (Elisa, 2022).

Data mining termasuk prosedur yang mengekstrak dan menemukan informasi penting dan pengetahuan terkait dari database besar menggunakan pendekatan statistika, matematika, kecerdasan buatan, dan pembelajaran mesin (Handoko, 2018).

Berdasarkan penjelasan definisi yang telah dijabarkan, maka dapat diambil kesimpulan bahwa data mining merupakan suatu proses pencarian otomatis untuk menemukan model dan pola dalam database yang banyak atau besar.

Adapun kelebihan dan kekurangan yang ada pada data mining antara lain Kelebihan dari data mining (Widaningsih et al., 2022) yaitu :

1) Pencarian Informasi Tersembunyi

   Data mining memungkinkan identifikasi pola dan informasi yang mungkin tidak dapat terdeteksi secara manual, membantu organisasi untuk mendapatkan wawasan yang berharga dari data mereka.

2) Prediksi dan Klasifikasi

   Data mining dapat digunakan untuk memprediksi tren masa depan, mengklasifikasikan data, dan membuat keputusan berdasarkan analisis yang mendalam.

3) Penyesuaian dan Personalisasi

   Dengan data mining, perusahaan dapat memberikan pengalaman yang lebih personal dan relevan kepada pelanggan mereka, misalnya, dalam rekomendasi produk atau penawaran khusus.

Kekurangan dari data mining yaitu :

1) Kualitas Data

   Keberhasilan data mining sangat tergantung pada kualitas data yang digunakan. Data yang buruk atau tidak lengkap dapat menghasilkan hasil yang tidak akurat.

2) *Overfitting*

   Risiko *overfitting*, di mana model data mining "menghafal" data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data baru, perlu diwaspadai.

3) Bias

   Data mining dapat memperkuat bias yang ada dalam data. Jika data yang digunakan sudah bias, hasil dari data mining juga dapat menjadi bias.

2.4 *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* (KNN) termasuk salah satu algoritma dalam machine learning yang digunakan untuk pengklasifikasian dan regresi (Khudhair et al., 2023). Prinsip dasar dari K-NN yaitu bahwa objek-objek yang serupa cenderung berada dekat satu sama lain dalam ruang fitur. Oleh karena itu dengan data tersebut algoritma K-NN memutuskan kategori atau nilai target suatu data berdasarkan data pelatihan yang paling mirip (terdekat), dengan demikian Algoritma K-NN didasarkan pada ruang fitur, jika mayoritas sampel k yang paling dekat dengan sampel tertentu termasuk kategori tertentu, sampel itu sendiri juga termasuk kategori ini (Amalia, 2020).

Pada algoritma *K-Nearest Neighbor* terdapat 5 cara dalam mencari tetangga yang terdekat (Nanglae et al., 2021) yakni:

1. Jarak *Euclidean*
2. Jarak *Manhattan*
3. Jarak *Cosine*
4. Jarak *Correlation*
5. Jarak *Hamming*

Pada penelitian ini penulis hanya menggunakan perhitungan jarak *Euclidean*, sehingga rumus menghitung jarak dengan *Euclidean* adalah sebagai berikut (Bahtiar, 2023) :

$$d\,(x,y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2}$$

Gambar 2.2. Rumus *Euclidean Distance*

Nilai Xi adalah nilai yang ada pada data *training*, nilai Yi adalah nilai yang ada pada data *testing*, sedangkan nilai n merupakan dimensi atribut dan d adalah jarak antara titik pada data *training* x dan titik pada data *testing* y yang akan diklasifikasikan, dimana x=x1,x2,…,xi dan y=y1,y2,…,yi merepresentasikan nilai dari atribut**.**

Selama fase *training,* algoritma ini hanya menyimpan vector vector fitur dan klasifikasi dari contoh data *training.* Pada fase klasifikasi, fitur-fitur yang sama dihitung untuk data *testing* (yang klasifikasinya belum diketahui).

2.5 *Software* Pendukung

Pada penelitian ini penulis menggunakan *software* RapidMiner yang digunakan untuk mengimplementasikan data mining prediksi penjualan produk terlaris. RapidMiner merupakan *software open source* yang digunakan untuk menemukan pengetahuan dan mengolah data. RapidMiner menawarkan lebih dari 400 metode atau prosedur data mining di *software* RapidMiner, termasuk operator untuk visualisasi, prapemrosesan, *input* dan *output* (Yolanda & Fahmi, 2021).

Berikut ini berapa fitur yang ada pada RapidMiner, yaitu :

1. Berlisensi gratis (*open source*).
2. WYSIWYG Design *Environment*

   RapidMiner memiliki antarmuka grafis yang memungkinkan pengguna untuk mendesain alur kerja analisis data menggunakan model *drag-and-drop* (sering disebut sebagai *"What You See Is What You Get"* atau WYSIWYG). Ini memungkinkan pengguna untuk dengan mudah membangun alur kerja analisis tanpa harus menulis kode.

3. Pustaka Algoritma

   RapidMiner menyediakan berbagai algoritma pemodelan prediktif yang siap digunakan. Ini termasuk regresi, klasifikasi, clustering, analisis asosiasi, dan lainnya.

4. Pemrosesan Data

   Platform ini mendukung berbagai tugas pemrosesan data, termasuk transformasi, penggabungan, pemfilteran, dan pengaturan ulang data.

5. Konektivitas Data

   Platform ini dapat menghubungkan dan mengintegrasikan data dari berbagai sumber, termasuk basis data, *spreadsheet*, file teks, dan data *streaming.*

6. Pemrosesan Big Data

   Platform ini mendukung pemrosesan data berukuran besar dan distribusi untuk menangani volume data yang besar.

7. Pengelolaan Proyek

   RapidMiner memungkinkan pengguna untuk mengatur dan mengelola proyek-proyek analisis data mereka dengan baik, termasuk penyimpanan model, data, dan alur kerja.

2.6 Objek Penelitian

Jesindo Mitra Prakarsa adalah Toko mainan anak-anak yang beroperasi di bidang perdagangan dan menawarkan berbagai macam mainan. Pada tanggal 12 Juni 2013, bisnis Jesindo Mitra Prakarsa dibuka yang lokasi di Telaga Indah Blk.N No.8, Sadai, Kec. Bengkong, Kota Batam. Pada awalnya tenaga kerja yang dimiliki Jesindo Mitra Prakarsa hanya terdiri dari lima orang. Dengan segala

kekurangan pada fasilitas serta financial, namun hanya bergantung pada keinginan dan keyakinan dalam kekuasaan Allah SWT, maka berdirilah sebuah Toko yang bernama Jesindo Mitra Prakarsa.

Dan berkat anugerah tuhan dan kompetensi manajemen, keterampilan empiris, dedikasi, ketekunan, dan kerja keras staf Jesindo Mitra Prakarsa, Toko ini dapat bertahan dan terus mempromosikan keberadaan bisnis di Kepulauan Riau Batam. Melihat perkembangan ini, Bapak Yondri Darto, SH menyatakan bahwa selaku pemilik Toko beliau berniat dan berambisi untuk memperluas bisnisnya secara nasional, dan pemilik Toko juga bermaksud mengubah Toko Jesindo Mitra Parakarsa menjadi sebuah perusahaan yang bergerak di bidang sector perdagangan.

Dalam hal ini, Peneliti melakukan analisa dengan tahapan data mining untuk membantu pemilik Toko menghasilkan prediksi penjualan produk terlaris pada Toko Jesindo Mitra Prakarsa, dalam proses pengambilan data peneliti mendapatkan informasi berupa 22 jenis mainan yang di jual pada Toko tersebut.

2.7 Penelitian Terdahulu

Dalam Penelitian ini, penulis mengacu kepada jurnal sebagai referensi yang dapat membantu penulis dalam menyelesaikannya :

Tabel 2.1 Referensi Penelitian Terdahulu

No	Judul Penelitian	Peneliti	Tahun	Hasil Penelitian
1.	Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Produk Terlaris Pada Toko Indah Jaya	Yuni Handayani , Taufik Hidayat , Hanif Arruhama	2023	Dalam penelitian ini melakukan prediksi penjualan makanan kering paling laris di Toko Indah Jaya menggunakan metode K-Nearest Neighbor. Untuk miempriediksi pienjualan yang akan datang digunakan rumus iEuclidiean Distancie, diengan nilai k=3 mienggunakan aplikasi RapidMinier didapatkan hasil priediksi tierlaris untuk 6 bulan kiediepan yaitu rambak kierbau diengan total pienjualan siebanyak 585 pcs. Hasil algoritma yang dipierolieh dari nilai RMSiE untuk produk makanan kiering diengan nilai yang paling miendiekati nol adalah bolu kiering diengan nilai 2,869.
2.	Penerapan Metode *K-Nearest Neighbor* untuk	Mohammad Kafil	2019	Dalam penelitian ini, penjualan diperkirakan berdasarkan hasil pengujian dan implementasi halaman

	Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso			web pada tiga *browser*, yaitu *Mozilla Firefox, Internet Explorer, dan Google Chrome.* Halaman *web* berhasil berjalan di ketiga browser, oleh karena itu dapat dinyatakan bahwa situs *web* dapat beroperasi dengan baik di *browser web* ketiga. Hasil tes akurasi menggunakan 12 data pelatihan dan 12 data pengujian menghasilkan keakuratan 83,3% dan nilai kesalahan 16,7%.
3.	Penerapan Metode *K-Nearest Neighbor* Untuk Sistem Rekomendasi Pemilihan Mobil	Ni Luh Gede Pivin Suwirmayanti	2020	Dalam penelitian ini menerapkan system saran pemilihan kendaraan dengan metode K-NN (*K-Nearest Neighbor*) yang menggabungkan kriteria seperti tujuan pembelian kendaraan, harga kendaraan, tahun pembuatan, kapasitas penumpang, warna, tenaga mesin dan jenis transmisi, serta dapat digunakan sebagai bantuan rekomendasi dalam pemilihan kendaraan pembeli, kemudian perancangan disajikan

				dalam bentuk diagram aliran data untuk perancangan aliran data system dan diagram hubungan entitas untuk perancangan *database*, struktur file, dan perancangan sistem.
4.	Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma *K Nearest Neighbo*r (KNN)	Rofiq Harun	2020	Pada Penelitian ini membahas klasifikasi otomatis dapat dikembangkan dengan menerapkan metode (KNN) berdasarkan hasil analisis data cuaca dalam menentukan apakah cuaca bukan hujan, hujan, atau hujan lebat, dan hasil tes menunjukkan bahwa prediksi cuaca harian dengan algoritma *K-Narest Neighbor* mendapatkan nilai RMSE 9.899 +/- 0.000. Dengan demikian Masyarakat tidak kebingungan dalam mengetahui informasi cuaca harian.
5.	Penerapan Algoritma *K-Nearest Neighbor* Untuk Penentuan Resiko Kredit	Henny Leidiya na	2021	Penelitian ini membahas metode *k-Nearest Neighbor* (kNN) diterapkan pada data pelanggan yang menggunakan layanan keuangan kredit kendaraan bermotor dalam.

	Kepemilikan Kendaraan Bemotor			Hasil testing untuk menguji kinerja algoritma ini termasuk *Curves Cross Validation, Confusion Matrix*, dan ROC, menghasilkan akurasi dan nilai AUC dari 81,46% dan 0,984, masing-masing. Karena nilai AUC berada di kisaran 0,9 hingga 1,0, pendekatan ini dianggap sangat baik. (*excellent*).
6.	Penerapan Algoritma *K-Nearest Neighbor* untuk Memprediksi Penjualan Motor Terlaris Pada PT Daya Anugrah Mandiri	Rozimin, Rahmat Fauzi	2022	Penelitian ini membantu perusahaan dalam membuat keputusan pasokan stok. Pengolahan data penjualan sepeda motor 170 dan tiga kualitas yang hadir dalam pemilihan data menggunakan pendekatan algoritma *K-Nearest Neighbor* menghasilkan proyeksi penjualan motor Honda dengan tipe *metric* yang lebih dicari oleh konsumen daripada jenis olahraga dan CUB. Penelitian ini memakai metode *K-neareast neighbor* yang nilai akurasinya mencapai 97,65%.
7.	*K-Nearest*	Ana María	2023	Dalam penelitian ini membahas

	Neighbor and K-Fold Cross-Validation Used In Wind Turbines For False Alarm Detection	Peco Chac´on, Isaac Segovia Ramírez, Fausto Pedro García M´arquez		mendeteksi alarm kesalahan pada turbin angin yang menggunakan algoritma *K-Nearest Neighbor* yang membandingkan nilai validasi silang *k-fold* yang berbeda untuk deteksi alarm kesalahan. Sebuah skenario kasus nyata yang dibentuk oleh tiga turbin angin nyata disajikan untuk menguji keandalan metodologi. Dataset sinyal dan alarm diperoleh oleh system control dan pengumpulan data pengawas dan *log alarm* sebagai variabel *respons*. Hal ini ditunjukkan bahwa kinerja tiga turbin angin analog dan variasi nilai validasi *k-cross* menunjukkan bahwa akurasi tidak meningkat secara signifikan. Metode yang diusulkan menunjukkan akurasi 98% dan lebih dari 22% peringatan kesalahan terdeteksi dalam studi kasus. Hasil ini menunjukkam ketahanan pendekatan untuk mendeteksi alarm kesalahan.
8.	*Assessing public satisfaction of public service application using supervised machine learning*	Ilham Mustaqim, Hasna Melani Puspasari,Rahmad Syalvei	2024	Pada penelitian ini membahas tentang sentiment publik dan variabel dominan terhadap kepuasan public pada aplikasi layanan public yaitu “Aplikasi Cek Bansos” yang diberikan oleh Kementerian social pada saat pandemic COVID-19 lalu. Penelitian ini menggunakan algoritma , *Naïve*

				Bayes dan K-nearest neighbor yang menghasilkan kinerja yang luar biasa mencapai akurasi 99,21%, dan prediksi sentimen negatif publik mencapai akurasi 83,81%.
9.	*Evaluation of sequential feature selection in improving the K-nearest neighbor classifier for diabetes*	(Rajkumar Govindarajan, Vidhyashree Balaji, Jayanthi Arumugam,radha Mothukuri	2024	Dalam penelitian ini melakukan prediksi terhadap penyakit diabetes dengan menggunakan seleksi fitur sekuensial (SFS) yang beralgoritma *K-Nearest Neighbor.* Hasil yang diperoleh pada penilitian ini menunjukkan efektivitas dengan tingkat akurasinya mencapai 84,41% dan ketika menggunakan algoritma *K-Nearest Neighbor* akurasi meningkat sebesar 2,6% ketika di *training* terhadap fitur optimal yang dipilih dengan SFS.

2.8 Kerangka Pemikiran

Pada tahapan ini menjabarkan bentuk dari kerangka pemikiran yang terdiri dari *input*, proses, dan *output* dalam melakukan implementasi data mining pada prediksi produk terlaris dengan metode.

pada bagian *input* menunjukkan daset dari produk penjualan yang dimana didalamnya terdapat parameter–parameter seperti nama barang, kuantitas, dan bulan penjualan. Data didapatkan dari observasi dan wawancara langsung ke toko tersebut.

Pada bagian proses digunakan untuk membantu dalam mengolah data yang telah diperoleh dengan menggunakan sebuah *software* pendukung dan algoritma *K–Nearest Neighbor*, pada bagian Proses terdapat subjek yang dimana dapat Memprediksikan penjualan produk terlaris untuk membantu pemilik Toko mengelola penyuplaian barang dan mengurangi biaya kerugian dari barang yang terjual. Kemudian pada objek dilakukan di sebuah Toko yang bernama Jesindo Mitra Prakarsa, yang berlokasi di Telaga Indah Blk.N No.8, Sadai, Kec. Bengkong, Kota Batam. Pada bagian Metode yang digunakan menggunakan sebuah algoritma *K–Nearest Neighbor*.

Pada bagian *Output* merupakan hasil dari prediksi yang didapatkan setelah mengolah data-data yang ditemukan dilapangan dengan menggunakan RapidMiner berdasarkan algoritama yang digunakan yaitu *K–Nearest Neighbor*.

Tabel 2.2 Tabel Kerangka Pemikiran