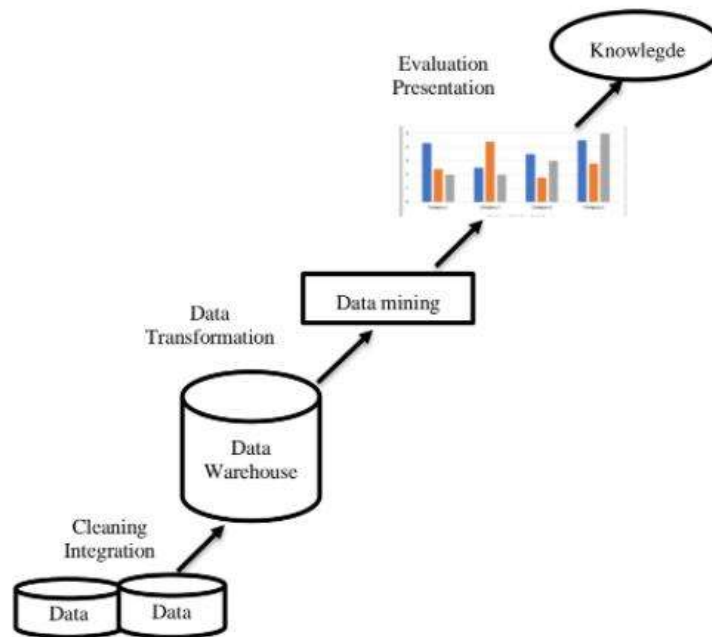


BAB II

TINJAUAN PUSTAKA

2.1 Knowledge Discovery in Database (KDD)

Proses penemuan informasi dalam database dikenal sebagai *Knowledge Discovery in Database* (KDD). Seluruh proses pengorganisasian atau menemukan pola yang mendasari, pengetahuan, dan informasi dari pengumpulan data besar didefinisikan sebagai KDD. Metode KDD memberikan pengetahuan dan informasi praktis, baru, dapat dimengerti, dan penting (Buulolo, 2020: 2). Berikut ini adalah tahapan proses KDD.



Gambar 2.1 Tahapan Proses KDD

Sumber: (Peneliti, 2022)

1. Data

Sebelum memulai proses KDD data harus dipersiapkan terlebih dahulu. Data yang akan dipergunakan adalah data yang telah dipisahkan dengan data operasional.

2. *Selection*

Pemilihan data diperlukan untuk memilah data, karena data yang ada tidak semuanya dapat dipergunakan. Pemrosesan data ini, misalnya, menentukan kumpulan data target, menentukan variabel, memilih sampel data, dan menyimpan data dalam sebuah file.

3. *Pre-processing/cleaning*

Menghapus dan mengoreksi redundansi, inkonsistensi, dan kesalahan dari data yang dipilih dilakukan pada tahap ini. *Enrichment* adalah langkah dalam proses *Pre-processing/cleaning* yang melibatkan penambahan informasi tambahan yang relevan dan dapat diandalkan ke data yang telah dikumpulkan.

4. *Transformation*

Contoh dari tahap ini adalah transformasi dataset sehingga dapat digunakan dalam proses penambangan data. Ini adalah proses kreatif yang sangat bergantung pada pola pencarian database atau jenis informasi yang akan dicari.

5. *Data mining*

Tahapan ini merupakan tahapan utama dari proses KDD, dimana data yang terpilih akan dicari pola atau informasi yang menarik dengan

menggunakan metode dan teknik tertentu. Dalam data *mining*, ada berbagai pendekatan, strategi, dan algoritma untuk dipilih, tergantung pada tujuan atau informasi yang dicari.

6. *Interpretation/Evaluation*

Data *mining* menghasilkan pengetahuan dan informasi yang harus disajikan secara mudah dimengerti oleh mereka yang membutuhkannya. Pengetahuan dan informasi tersebut bisa dipresentasikan kedalam pohon keputusan, grafik atau dalam bentuk rule. Dalam proses KDD, tahap ini termasuk memverifikasi bahwa pola atau informasi yang didapatkan menurut fakta atau asumsi yang telah ditetapkan sebelumnya.

7. *Knowledge*

Tujuan akhir dari rangkaian proses KDD ialah Informasi berguna yang mampu dengan mudah dimengerti oleh beberapa pihak adalah tujuan Pengetahuan atau informasi yang berguna perlu dipraktikkan sesuai dengan nilainya.

2.2 Data Mining

Data *mining* ialah suatu bidang ilmu yang dimanfaatkan dalam menangani permasalahan mendapatkan informasi dari database yang besar dengan menggabungkan teknik dari statistik, mesin pembelajaran, visualisasi data, pengenalan pola, dan juga database. Tujuan dari data *mining* adalah untuk mengekstrak informasi dengan teknik cerdas dari himpunan data yang selanjutnya

mengubah informasi menjadi terstruktur yang dapat dimengerti untuk penggunaan lebih lanjut (Indah Werdiningsih et al., 2020: 17).

Ada banyak cara untuk menambang data, tetapi salah satu yang paling umum adalah melalui penggunaan data *mining*. Data *mining* mencakup berbagai kegiatan, termasuk pengumpulan, ekstraksi, analisis, dan pelaporan data. *Knowledge discovery*, *knowledge extraction*, data / *pattern analysis*, *information harvesting*, dan istilah lainnya digunakan untuk menggambarkan penambangan data (Arhami & Nasir, 2020: 1).

Data *mining* juga menjadi proses logis untuk menemukan informasi yang berguna. Setelah ditemukan informasi dan pola dapat digunakan untuk alat pendukung dalam pengambilan keputusan dalam mengembangkan bisnis. Alat data *mining* dapat memberikan jawaban untuk berbagai pertanyaan yang terkait dengan bisnis dan terlalu sulit untuk diselesaikan. Data *mining* juga dapat digunakan untuk meramalkan tren masa depan, memungkinkan perusahaan untuk membuat pilihan yang lebih proaktif, efisien, dan dinamis.

Data yang telah dianalisis menggunakan metode data *mining* juga dapat mengungkapkan informasi yang sesuai dengan harapan. Misalnya pada bidang kesehatan, cukup banyak data yang dimiliki oleh rumah sakit, seperti data *medical record* dan *radiologi*, tetapi karena belum adanya standar koleksi data maka data-data tersebut sukar untuk diolah sehingga dengan kehadiran data *mining* maka diharapkan data-data yang dimiliki oleh pihak kesehatan dapat diolah sesuai dengan keperluan hingga menghasilkan informasi dan pengetahuan yang dapat dimanfaatkan oleh para pengambil kebijakan terutama pemerintah.

Menurut Siregar (2017: 3) Dengan "menggali" data, data *mining* mampu mengungkap informasi yang sebelumnya tidak diketahui dan memprediksi masa depan. Menggunakan pemodelan sebagai metode, tugas ini selesai. Sebuah "Jawaban" didefinisikan sebagai model yang dibangun dan kemudian digunakan dalam berbagai situasi di mana jawaban yang dicari.

Data *mining* biasanya dilakukan melalui tiga langkah (Arhami & Nasir, 2020: 4), yaitu:

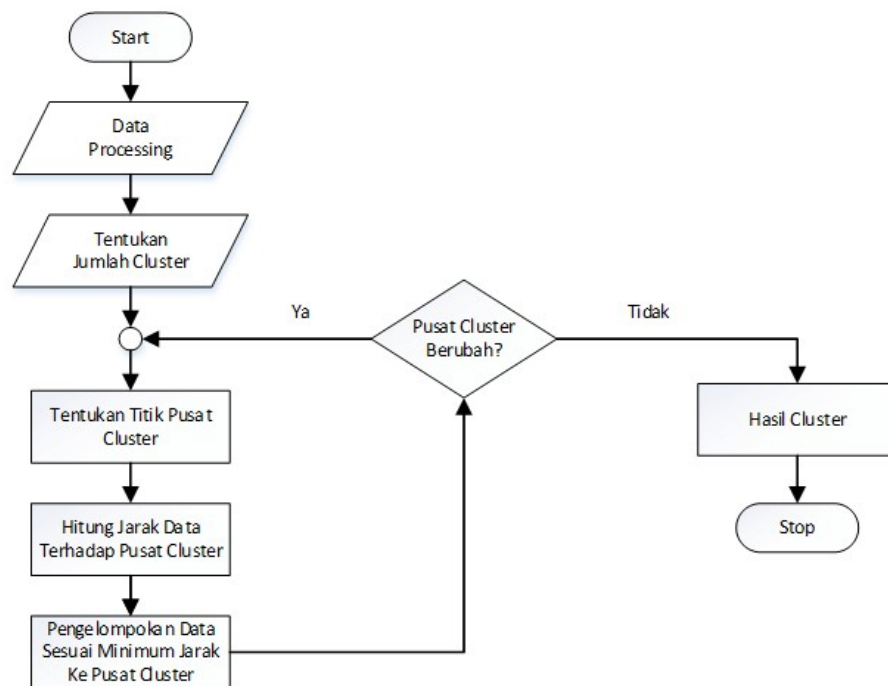
1. Eksplorasi: Hal yang harus dilakukan adalah mempersiapkan data dengan jumlah yang besar, selanjutnya data akan dibersihkan sesuai dengan yang diperlukan, serta menghilangkan data-data duplikat, alhasil data yang tertinggal hanya data yang benar-benar bisa digunakan.
2. Pemodelan atau identifikasi pola: Untuk membuat prediksi terbaik dan paling akurat, buatlah model statistik. Dikarenakan model yang berbeda diterapkan pada dataset yang serupa berulang kali serta hasilnya dibandingkan, prosedur ini bisa memakan waktu lama.
3. Penerapan: Model digunakan pada data pelatihan dan pengujian untuk memberikan perkiraan hasil yang diprediksi pada tahap akhir.

2.3 Metode Data Mining

Data *mining* dipisahkan menjadi beberapa kategori berdasarkan tugas-tugas yang dapat diselesaikan, seperti:

1. *Clustering* (Pengklasteran)

Pada tahun 1976, J.B. Mac Queen adalah orang pertama yang memperkenalkan algoritma *K-Means*. Ketika mengelompokkan data berdasarkan karakteristik yang sama, algoritma ini ialah salah satu algoritma pengelompokan yang paling sering dipakai. Produk akhir yang dikenal sebagai *cluster* terbentuk sebagai hasil dari proses ini (Arhami & Nasir, 2020: 148). *Flowchart* algoritma *K-Means Clustering* dapat dilihat pada Gambar 2.2 berikut:



Gambar 2.2 *Flowchart* algoritma *K-Means Clustering*

Sumber: (Peneliti, 2022)

Clustering adalah pendekatan data *mining* yang berusaha untuk menemukan kelompok item dengan karakteristik serupa yang dapat diurutkan dari set lain dari objek, dengan objek dalam kelompok yang sama yang lebih homogen

daripada objek dalam kelompok yang terpisah (Jollyta et al., 2021: 1). Jumlah dan jenis data objek menentukan jumlah kelompok yang dapat diidentifikasi. Sekelompok data objek dibagi menjadi beberapa kelompok, kelompok-kelompok tersebut mempunyai ciri khas tertentu yang bisa dibedakan antara satu sama lain, tujuannya untuk melakukan interpretasi dan analisis selanjutnya sesuai dengan tujuan penelitian.

Dengan metode *clustering*, dapat kita identifikasikan area yang padat, menemukan hubungan yang menarik antar atribut data dan menemukan pola distribusi secara keseluruhan. Teknik *clustering* digunakan untuk mengelompokkan data dengan kriteria serupa ke wilayah atau grup yang sama, sambil memisahkan data dengan kriteria yang berbeda. Data akan dikelompokkan dalam satu kelompok atau cluster mempunyai tingkat kemiripan yang maksimum dan setiap kelompok mempunyai tingkat kemiripan yang minimum (Ginantra et al., 2021: 76).

Algoritma *Clustering* dibagi atas dua algoritma, yang pertama yaitu Hierarchical *Clustering*, yaitu suatu metode hirarkis untuk membuat kombinasi hirarkis dan diimplementasikan terhadap objek data, seperti *Hierarchical Agglomerative Clustering (HAC)*. *Non-hierarchical Clustering*, juga dikenal dengan algoritma pengelompokan partisi, memberikan jumlah n objek, dan k adalah banyaknya cluster yang dapat dibentuk. Algoritma partisi *clustering* memproses objek menjadi k grup sesuai dengan syarat optimasi tertentu, dimana setiap grup merupakan representasi dari sebuah cluster. Contoh algoritma partisi *clustering* termasuk *K-Means*.

Algoritma *K-means* akan dipakai guna pengelompokan dalam penelitian ini. Pengamatan dibagi menjadi kelompok k menggunakan metode *K-Means clustering* dalam statistik dan machine learning, dan setiap pengamatan ditempatkan dalam kelompok dengan rata-rata terdekat.

Pengelompokan data dengan nonhierarki (sekatan) memiliki beberapa metode salah satu metodenya adalah metode *K-Means* yang berupaya membagi data yang ada menjadi dua atau lebih kelompok. Metode ini memisahkan data ke dalam kategori, dengan setiap kategori termasuk data yang berbeda. Tujuan dari pengelompokan ini adalah untuk mengurangi jumlah spesies di setiap kategori sambil meningkatkan jumlah spesies di antara mereka.

2. *Classification* (Klasifikasi)

Klasifikasi adalah penggalian data yang menetapkan item dalam koleksi untuk menargetkan kelas tertentu. Klasifikasi dimulai dengan pengumpulan data dimana data tersebut terdapat kelas yang telah diketahui. Sebagai contoh, klasifikasi yang untuk mengidentifikasi peringkat kredit, berupa pemohon pinjaman sebagai risiko kredit rendah, sedang, atau tinggi. Data pemohon, seperti riwayat kredit pemohon, riwayat pekerjaan, kepemilikan atau sewa rumah pemohon, tahun tinggal pemohon, jumlah dan jenis investasi yang dilakukan oleh pemohon, dan sebagainya, dapat digunakan sebagai karakteristik. Target/kategori akan menjadi peringkat kredit, dan data untuk setiap aplikasi akan terjadi (Indah Werdiningsih et al., 2020: 19).

3. *Association* (Asosiasi)

Menurut Arhami & Nasir (2020: 160) analisis dengan kaidah asosiasi ini adalah untuk mendapatkan sekumpulan item atau itemset yang muncul secara bersamaan dalam transaksi-transaksi. Suatu itemset terdiri dari i buah item yang disebut dengan i -itemset. Persentase transaksi atau kombinasi yang terdiri dari satu item set disebut dengan *support itemset* dan *confidence* (nilai kepastian) ialah nisbah kekuatan hubungan diantara item dalam peraturan persatuan. Support dan confidence tersebut merupakan dua kriteria dasar dalam kaidah asosiasi.

4. Prediksi

Kegunaan dari metode prediktif adalah membangun model pengetahuan yang dapat digunakan untuk membentuk prediksi. Variabel dalam data dapat digunakan untuk menemukan pola yang berasal dari data. Pola ini dapat digunakan untuk meramalkan variabel yang tidak diketahui (Ginantra et al., 2021:79).

5. Estimasi

Estimasi sangat identik dengan metode klasifikasi; Perbedaannya adalah bahwa variabel target diasumsikan numerik daripada kategori.

6. Deskripsi

Deskripsi ialah teknik guna menggambarkan pola yang dimaksud. Penelitian analitis kadang-kadang bisa menjadi masalah mencoba memahami pola dan tren dalam data. Deskripsi pola dan tren kerap memberikan uraian yang mungkin pada pola atau tren (Tahyudin et al., 2021: 5).

2.4 Algoritma Data Mining

Berikut ini adalah beberapa algoritma yang biasa dikenakan dalam data *mining*, diantaranya:

2.4.1 Algoritma *Apriori*

Algoritma apriori, menurut Buulolo (2020: 11), telah banyak digunakan dalam data *mining* untuk mengidentifikasi pola data atau pola kejadian/frekuensi. Di pasar kecil, pola pembelian pelanggan biasanya ditemukan melalui penggunaan algoritma apriori berdasarkan data transaksi pembelian. Dalam algoritma apriori, istilah "nilai dukungan" mengacu pada persentase dari total data yang muncul di bagian data tertentu.

2.4.2 Algoritma Association Rule

Dalam data *mining*, Association Rule adalah metode untuk membentuk aturan asosiatif antara item dalam item kombinasi. Aturan hubungan ditulis dengan gaya "if then.... Atau jika... kemudian..." Proses Association Rule adalah kelanjutan dari metode sebelumnya, di mana aturan dibentuk dengan menggabungkan 2 item yang memenuhi persyaratan dukungan minimal, artinya yang akan dibuat *rule*-nya hanya kombinasi yang memenuhi *support* minimum.

Tujuan *association rule* adalah untuk mendapatkan hubungan/kekuatan bubungan antar item dalam sebuah kombinasi *itemsets*, contohnya adalah seberapa besar kemungkinan seseorang pembeli teh, juga membeli gula secara bersamaan dan sebaliknya. Kekuatan hubungan/peluang pada *association rule* antara minimal 2 (dua) *itemsets*, nilainya diukur dengan nilai *confidence* (Buulolo, 2020: 19).

2.4.3 Algoritma *Rough Set*

Metode *Rough Set* adalah pendekatan data *mining* yang menggunakan *Decision Systems* (tabel yang dilengkapi dengan *attribute* kriteria dan keputusan) untuk mengekstrak informasi/ pengetahuan dari kelompok data atau tabel. Output dari algoritma *rough set* dalam bentuk *rule* dan dapat digunakan untuk memperkirakan sesuatu yang akan terjadi berdasarkan data-data sebelumnya seperti memprediksi kelulusan mahasiswa berdasarkan nilai yang diperoleh per semester, memprediksi kenaikan harga emas berdasarkan harga emas sebelumnya dan lain sebagainya.

2.4.4 Algoritma C4.5

Algoritma C4.5 adalah alat kategorisasi data yang banyak digunakan dalam data *mining*. Algoritma C4.5 adalah algoritma untuk mengeksplorasi data yaitu untuk mencari hubungan variabel input atau *attribute* kriteria dengan variabel target atau *decision attribute* (atribut keputusan). Hasil dari eksplorasi data diubah dalam bentuk pohon keputusan (*decision tree*) sehingga mudah dipahami dan dimengerti oleh pengguna. Algoritma C4.5 mengubah sejumlah besar data menjadi pohon keputusan, yang kemudian ditafsirkan sebagai rule. Dalam data *mining* banyak algoritma pembentukan pohon keputusan seperti Algoritma *Cart*, Algoritma ID3, Algoritma C4.5 dan Algoritma 05.0.

2.4.5 Algoritma Nearest Neighbor

Sesuai dengan terjemahan ke bahasa Indonesia *nearest* = terdekat, dan *neighbor* = tetangga, berarti tetangga terdekat. Biasanya yang bertetangga selalu ada kemiripan antara rumah tangga yang satu dengan yang lainnya misalnya

model pagar rumah, bentuk rumah, merek TV, bentuk dapur rumah dan lain-lain. Oleh karena itu, algoritma *nearest neighbor* adalah penyelesaian kasus dengan menggunakan pendekatan atau perbandingan kasus lama dengan kasus baru dengan pendekatan atau perbandingan dengan menggunakan bobot nilai setiap kriteria kasus.

Contoh seorang hakim akan menjatuhkan hukuman kepada seorang terdakwa pencuri, selain berdasarkan undang-undang yang berlaku tetapi juga berdasarkan hukuman kepada terdakwa pencuri yang telah dijatuhkan hukuman sebelum-sebelumnya. Bank jika memberikan pinjaman kepada seseorang, akan menggunakan berbagai pertimbangan atau kriteria dan membandingkan dengan data-data nasabah sebelumnya.

2.4.6 Algoritma *Naïve Bayes*

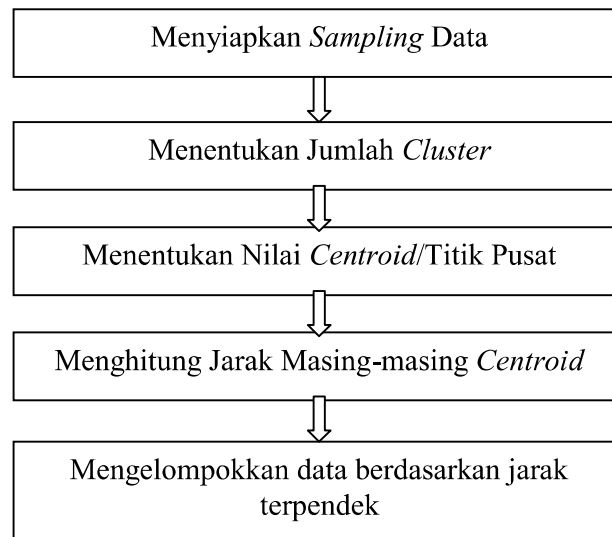
Algoritma Naive Bayes dikembangkan oleh ilmuwan Inggris Thomas Bayes dan memanfaatkan metode probabilitas dan statistik guna mengklasifikasikan data. Nama algoritma *Naive Bayes* mengacu pada *Teorema Bayes*, yang melakukan prediksi peluang masa mendatang bersumber dari pengalaman masa lalu.

Naïve Bayes Classifier ini memiliki asumsi yang sangat kuat bahwasanya setiap kondisi/peristiwa independen. Pengklasifikasi ini berkinerja mengagumkan dibandingkan dengan orang lain dari jenisnya. Untuk menentukan perkiraan parameter yang diperlukan untuk klasifikasi, metode ini menggunakan sejumlah kecil data pelatihan. Varians variabel dapat digunakan untuk menentukan

klasifikasinya, bukan seluruh matriks kovarians, karena diasumsikan bebas (Al-Khowarizmi, 2021: 219).

2.4.7 Algoritma *K-Means*

Algoritma *K-Means* menjadi salah satu teknik dalam data *mining* guna mengklusterkan (*Clustering*) data kedalam beberapa kelompok berdasarkan jarak, kriteria, kondisi atau karakteristik. Data dalam satu kelompok harus memiliki jarak terpendek, kriteria, kondisi atau karakteristik yang serupa atau mendekati sama antara satu dengan yang lain. Algoritma *K-Means* dapat mengelompokkan objek yang memiliki kemiripan. Berikut langkah-langkah dalam pengelompokan data dengan Algoritma *K-Means*:



Gambar 2.3 Cara Kerja Algoritma *K-Means*

Sumber: (Peneliti, 2022)

Untuk menetapkan jarak tiap-tiap *centroid* menurut jarak paling pendek, kriteria, kondisi atau karakteristik yang sama atau hampir sama, menggunakan model *Euclidean* yang dirumuskan berikut:

Rumus 2.1 Rumus *Euclidean*

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Keterangan:

d_{ij} = Jarak dari data ke i ke pusat cluster j

x_{ki} = Data dari ke- i pada *attribute* data ke- k

x_{kj} = Data dari ke- j pada *attribute* data ke- k

2.4.8 Algoritma *K-Nearest Neighbor (K-NN)*

K-NN menjadi salah satu algoritma atau metode untuk klasifikasi yang biasa dikenakan dalam data *mining*. K-NN juga masuk dalam kategori regresi yang juga dapat digunakan untuk memprediksi seperti halnya regresi. K-NN dalam kelompok pembelajaran digolongkan ke dalam *supervised learning* yang berbasis pada jarak. K-NN juga dapat disebut sebagai *Case Based Reasoning* yang merupakan suatu metodologi berbasis penalaran kasus-kasus berbasis pada data latih dengan suatu kasus disimpan, dilatih dan diakses untuk menyelesaikan permasalahan baru. Untuk mendapatkan nilai prediksi dari data baru maka dicari data-data yang sama atau berdekatan (Arhami & Nasir, 2020: 96).

2.4.9 Algoritma *K-Medoids*

Menurut Buulolo (2020: 72) algoritma *K-Medoids*, juga dikenal sebagai algoritma Partitioning Around Medoids (PAM), adalah pendekatan penambangan data untuk mengelompokkan (*clustering*) data ke dalam kelompok berdasarkan jarak, kriteria, kondisi, atau atribut. Algoritma *K-Medoids* dapat mengelompokkan objek yang mempunyai kemiripan. Perbedaannya yaitu algoritma *K-Medoids* ada perwakilan (*medoid*) untuk setiap cluster yang terbentuk, sedangkan algoritma *K-*

Means tidak mempunyai perwakilan karena dalam pembentukan pusat cluster (*centroid*) terbaru menggunakan nilai rata-rata (*mean*) dari cluster data yang terbentuk. Salah satu kelebihan algoritma *K-Medoids* adalah tidak sensitif terhadap *outlier* dan *noise* atau tidak mudah terjadi perubahan *cluster* terhadap perubahan nilai pusat *cluster* (*centroid*), dan sangat berbeda jauh dengan algoritma *K-Means* yang rentang terhadap perubahan *cluster* jika terjadi perubahan nilai pusat *cluster* (*centroid*).

2.5 Davies Bouldin Index

Pada tahun 1979 David L. Davies dan Donald W. Bouldin mengenalkan *Davies Bouldin Index* (DBI), sebuah metode untuk mengevaluasi efisiensi pengelompokan dan mendefinisikan kohesi sebagai ukuran kedekatan data dengan titik pusat cluster. Ada skema evaluasi berdasarkan jumlah dan kedekatan data hasil *clustering* saat menggunakan *Davies Bouldin Index* untuk *clustering* internal (Muningsih et al., 2021: 97)

Davies Bouldin Index menghitung jarak (*dissimilarity*) antar kluster. Algoritma pengelompokan dengan DBI rendah menghasilkan cluster dengan jarak intra-kluster rendah (kesamaan intra-kluster tinggi) dan jarak antar-kluster yang tinggi (kesamaan antar-kluster rendah). Semakin rendah nilai *Davies-Bouldin Index* suatu algoritma *machine learning* yang digunakan untuk klusterisasi, maka semakin tinggi kinerjanya. (Santoso & Azis, 2020: 53). Ini adalah tujuan evaluasi *cluster* untuk menentukan seberapa baik hasil pengelompokan kinerja. DBI adalah

metode yang digunakan untuk mengevaluasi hasil pengelompokan dan menentukan jumlah cluster yang optimal untuk digunakan.

Indeks Davies Bouldin dihitung dengan cara berikut (Manullang, et al. 2020: 1433):

1. *Sum of Square Within-cluster* (SSW)

Sum of Square Within-cluster (SSW) dapat digunakan untuk memperkirakan tingkat kohesi dalam cluster ke-i. Untuk mengukur kohesi, kita melihat berapa banyak titik data yang dekat dengan pusat *cluster* dari *dataset* yang dilacak. Persamaan yang digunakan untuk mendapatkan *Sum of Square Within-cluster* adalah sebagai berikut.

Rumus 2.2 Rumus *Sum of Square Within-cluster*

$$SSW = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i)$$

2. *Sum of Square Between-cluster* (SSB)

Hitung jarak antara cluster menggunakan perhitungan *Sum of Square Between-cluster* (SSB). *Sum of Square Between-cluster* dapat dihitung menggunakan persamaan berikut.

Rumus 2.3 Rumus *Sum of Square Between-cluster*

$$SSB_{i,j} = d(c_i, c_j)$$

3. *Ratio* (Rasio)

Tujuannya adalah untuk menentukan nilai perbandingan antara cluster ke-i dan ke-j. Persamaan berikut digunakan untuk menghitung nilai rasio yang dimiliki oleh setiap *cluster*.

Rumus 2.4 Rumus *Ratio*

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}}$$

4. *Davies Bouldin Index*

Persamaan berikut digunakan untuk menemukan nilai *Davies-Bouldin Index* (DBI) menggunakan nilai yang diperoleh dari perhitungan rasio.

Rumus 2.5 Rumus *Davies Bouldin Index*

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j})$$

Banyaknya cluster yang disimpulkan dari persamaan, adalah k. Semakin kecil nilai DBI (non-negatif ≥ 0), semakin baik cluster yang dibentuk oleh pengelompokan memakai algoritma *clustering* (Butsianto & Saepudin, 2016 : 116).

2.6 Software Pendukung

Penelitian ini dilakukan dengan memanfaatkan software *Microsoft Excel* dan juga *Software RapidMiner* untuk melakukan pengujian data.

2.6.1 Microsoft Excel 2019

Gambar 2.4 Logo *Microsoft Excel* 2019

Sumber: (Peneliti, 2022)

Microsoft Excel adalah program *spreadsheet* (pengolah data) yang digunakan untuk mengotomatisasi manipulasi data menggunakan alat yang ada untuk memudahkan dalam melakukan perhitungan dasar, manipulasi data, pembuatan tabel, manajemen data grafis, dan lainnya. *Microsoft Excel* unggul dalam mengolah berbagai macam data, baik di bidang akuntansi, teknik, statistik, dan bidang lainnya yang membutuhkan perhitungan yang cepat dan akurat.

Pada lokasi penelitian ini pengolahan data menggunakan *Microsoft Excel*, data pengeluaran barang diketik secara manual dan direkap dalam satu *file*. Setelah itu akan disusun kedalam *stock card* berdasarkan item barang dengan menginput tanggal, jumlah item, *project*, dan nomor transaksi.

Software Microsoft Excel dalam penelitian ini akan digunakan sebagai pengolah data awal dengan cara melakukan pembersihan data seperti menghapus data duplikat, mengecek data yang kurang tepat dan lain sebagainya sehingga data siap diolah dengan menggunakan *software RapidMiner*.

2.6.2 *Software RapidMiner*



Gambar 2.5 Logo *RapidMiner*

Sumber: (Peneliti, 2022)

RapidMiner adalah program gratis dan *open-source*. Setiap kali berhubungan dengan data *mining*, *RapidMiner* adalah pilihan yang sangat baik. *RapidMiner* memanfaatkan bermacam-macam teknik deskriptif dan prediksi untuk membantu pengguna membuat keputusan terbaik.

Sejumlah fitur *RapidMiner*, meliputi:

1. Jumlah algoritma data *mining*, misalnya *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, misalnya tumpang tindih diagram histogram, *tree chart* dan *3D Scatter plots*.
3. Jumlah variasi *plugin*, misalnya *text plugin* untuk melakukan analisis teks.
4. Mencadangkan prosedur data *mining* dan machine learning meliputi: ETL (*extraction, transformation, loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi
5. Proses data *mining* terdiri dari operator-operator yang *nestable*, dideskripsikan memakai XML, dan diciptakan memakai GUI
6. Menggabungkan proyek data *mining Weka* dan statistika R.

2.7 Penelitian Terdahulu

Peneliti yang menggunakan data *mining* untuk menentukan tata letak menggunakan penelitian sebelumnya sebagai panduan. Jurnal berikut dikonsultasikan oleh para peneliti:

1. Penelitian oleh (Lestari, 2019) dari STMIK Budi Darma, Jurnal Riset Komputer (JURIKOM), ISSN: 2407-389X, Vol. 6 No. 2, dengan judul

penelitian “**Implementasi Data Mining Untuk Pengaturan Layout Swalayan Delimas Lestari Kencana dengan Menggunakan Algoritma K-Means Clustering**”. Tujuan penelitian ini adalah membuat model tata letak merek produk berdasarkan jenis produk menggunakan data dari tanda terima belanja. Data dibagi menjadi beberapa kelompok menggunakan algoritma pengelompokan *K-means*, yang membagi ke dalam kelompok data dengan ciri khas yang serupa secara bersamaan, dan data dengan karakteristik yang berbeda ke dalam kelompok lain. Sehingga dapat ditentukan berapa penempatan merek produk berdasarkan jenis produk di Supermarket Delima Lestari Kencana, tiga cluster dibuat menggunakan algoritma *K-Means: Cluster C1, Cluster C2, dan Cluster C3*.

2. Penelitian oleh (Syahra et al., 2020) dari STMIK Triguna Dharma, *Journal of Technopreneurship and Information System (JTIS)*, E-ISSN: 2614-3089, Vol. 3 No.1, dengan judul “**Implementasi Data Mining Untuk Penyusunan Tata Letak Data Obat-Obatan Dengan Menggunakan Algoritma K Harmonic Means Pada Apotek Inti Fada Sidamanik**”. Sistem informasi manajemen data obat akan dibangun di Apotek Inti Fada Sidamanik untuk merampingkan serta menyederhanakan sistem informasi yang tidak terstruktur dengan baik. Hasil tes menunjukkan bahwa keanggotaan metode *clustering* metode KHM memiliki tingkat kesepakatan yang sangat bagus, dengan tingkat kesepakatan 92%.
3. Penelitian oleh (Fithri & Wardhana, 2021) dari Universitas Mercu Buana, *Jurnal PILAR Nusa Mandiri*, E-ISSN: 2527-6514, Vol. 17 No. 2, dengan

judul “*Cluster Analysis of Sales Transaction Data using K-Means Clustering at Toko Usaha Mandiri*”. Studi ini menggunakan algoritma pengelompokan *K-Means* untuk menganalisis data penjualan dan membuat kluster penjualan. Berupa kluster1, kluster2 dan kluster3 dengan persentase 62% (11 data), 8% (56 data) dan 30% (11 data), hasil pengelompokan data (25 data). Nilai 0,2 diperoleh dengan memvalidasi algoritma *K-Means Clustering* dengan *Davies Bouldin Index*. Manajemen inventaris dan strategi pemasaran tidak harus dibatasi oleh informasi pengelompokan penjualan.

4. Penelitian oleh (Lusiani & Liperda, 2020) dari Universitas Pertamina, Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi), E-ISSN: 2527-5941, dengan judul “**Perancangan Ulang Tata Letak Produk pada Gudang untuk Mengurangi Biaya Penanganan Material**”. Penelitian ini bertujuan untuk mengurangi biaya penanganan material dengan mendesain ulang tata letak produk gudang. Untuk penelitian ini, empat metode digunakan: *Class Based Storage*, *Dedicated Storage*, *Cube per Order Index*, dan Simulasi. Ada 27 skenario yang mungkin, masing-masing dengan tata letak yang berbeda, arrival, dan jumlah forklift. Tata letak dengan kebijakan Penyimpanan Khusus dan jumlah forklift sama dengan satu *forklift storage* dan satu *forklift delivery* dipilih berdasarkan perhitungan Manajemen Kinerja.
5. Penelitian oleh (Harpendi Bara et al., 2020) dari Institut Teknologi Nasional Malang, JATI (Jurnal Mahasiswa Teknik Informatika), E-ISSN:

2598-828X, Vol. 4 No. 2, dengan judul “**Pengelompokan Data Obat menggunakan Metode *K-Means Clustering* Pada UPT Puskesmas Kondoran Kec. Sangalla**”. Tujuan Penelitian: Mengurangi jumlah waktu yang diperlukan untuk memproses data untuk rencana pembelian obat di masa depan dengan mengembangkan sistem penambangan data yang dapat membantu mengelompokkan data obat ke dalam beberapa kelompok, dan dengan demikian mengurangi jumlah obat yang tidak dibutuhkan yang terakumulasi di gudang obat pusat kesehatan, di mana ia rentan terhadap kedaluwarsa. Sebanyak 183 data masuk ke dalam klaster 1 obat dengan penggunaan lambat dan 21 data dimasukkan ke dalam klaster 2 obat dengan penggunaan cepat, berdasarkan hasil metode pengujian pada 204 data.

6. Penelitian oleh (Hutapea & Fauzi, 2021) dari Universitas Putera Batam, Computer and Science Industrial Engineering (Comasie Journal), E-ISSN: 2715-6265, Vol. 04, No.06, dengan judul “**Data Mining Sistem Tata Letak Material di PT Batam Cyclect**”. Ini adalah tujuan dari penelitian ini untuk memudahkan manajer toko sementara juga mengurangi jumlah waktu yang dibutuhkan untuk menemukan barang. Algoritma apriori memanfaatkan proses ini untuk mengurangi atau mempersempit ruang pencarian untuk frekuensi kandidat item yang ditetapkan. Perangkat lunak RapidMiner digunakan untuk memverifikasi keakuratan algoritma apriori. Setelah penelitian selesai, ditemukan bahwa memesan Marking Tape dan penanda menghasilkan nilai support 5% dan confidence 50%, sementara

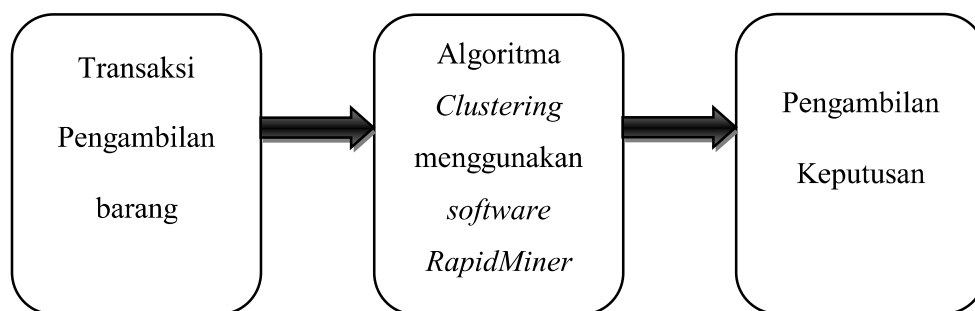
melakukan pemesanan PVC Insulation Tape menghasilkan nilai support 15% dan confidence 89%.

7. Penelitian oleh (Rajagukguk & Sitohang 2021) dari Universitas Putera Batam, Computer and Science Industrial Engineering (Comasie Journal), E-ISSN: 2715-6265, Vol. 5 No. 5, dengan judul “**Penerapan Data Mining dalam Pemilihan Laptop Menggunakan Metode Algoritma C4.5**”. Tujuan dari penelitian ini adalah untuk memanfaatkan perangkat lunak WEKA untuk mengkategorikan laptop berdasarkan harga, kapasitas RAM, kecepatan prosesor, ukuran, dan masa pakai baterai, yang semuanya dianalisis menggunakan algoritma C4.5. Dengan total 80 titik data, ditemukan tiga kriteria, yaitu harga, prosesor, dan daya tahan baterai, memiliki dampak signifikan pada pemilihan laptop. Dari 80 keping data, 53 memilih laptop yang tepat dan 27 memilih laptop yang salah.
8. Penelitian oleh (Fajrin & Handoko, 2018) dari Universitas Putera Batam, Jurnal Ilmiah Informatika (JIF), E-ISSN: 2615-1049, Vol. 06, No. 02, dengan judul “**Penerapan Data Mining Untuk Mengolah Tata Letak Buku Dengan Metode Association Rule**”. Tujuan dari penelitian ini adalah untuk menggunakan data transaksi pinjaman buku perpustakaan ABC untuk memecahkan tantangan pemrosesan data saat ini, mencari tahu buku mana yang sering dipinjam, dan membantu penempatan buku tergantung pada jenisnya. Pakai. Untuk membantu perpustakaan dalam meningkatkan kualitas layanan, tingkat pinjaman telah meningkat. Support minimum = 0,032, confidence minimum = 0,50, berdasarkan tes yang

dilakukan pada data dari 659 transaksi pinjaman buku dengan total 182 atribut. RapidMiner digunakan dalam uji ini.

2.8 Kerangka Pemikiran

Kerangka pemikiran ialah garis pemikiran yang masuk akal apabila membuat kesimpulan. Melalui pemanfaatan data *mining* algoritma *clustering* susunan letak barang di PT Lautan Lestari Shipyard dapat ditentukan susunannya, tahapan yang dibutuhkan untuk mendapatkan hasil rancangan dengan fungsi yang baik sebagai berikut:



Gambar 2.6 Tahapan Kerangka Pemikiran

Sumber: (Peneliti, 2022)

Seperti yang terlihat dari *flowchart*, langkah awal dari penelitian ini adalah:

1. Dihimpun lalu dilaksanakan *cleaning data*, *transformation* guna mengetahui pola pengambilan barang. Tahapan tersebut menghasilkan variabel masukan berwujud sejumlah pengambilan barang dan tipe item.

2. Tahap ini ialah proses pengkategorisasi data dengan memanfaatkan metode aturan *clustering*. Selanjutnya dilakukan pengujian dengan *software RapidMiner* berdasarkan data-data yang diperoleh.
3. Temuan penelitian ini adalah aturan untuk membuat keputusan dimana aturan yang digunakan sebagai panduan ketika memutuskan tata letak barang.