BAB II

TINJAUAN PUSTAKA

2.1 Pengertian Data

Menurut Sutabri dalam jurnal Situmorang (2019:36) berpendapat bahwa data adalah kenyataan yang menggambarkan suatu kejadian-kejadian dan kesatuan nyata. Kejadian adalah sesuatu yang terjadi saat tertentu di dalam dunia bisnis.

Menurut Yakun dalam jurnal Situmorang (2019:37) berpendapat bahwa data adalah deskripsi kenyataan yang menggambarkan adanya suatu kejadian *(event),* data terdiri dari fakta *(fact)* dan angka yang secara relatif tidak berarti bagi pemakai.

2.1.1 Jenis-jenis Data

Menurut Sugiyono dalam jurnal penelitian Pratiwi (2019:211), ada dua jenis data pada penelitian, diantaranya adalah:

1. Data kualitatif adalah data yang dinyatakan dalam bentuk kata, kalimat, dan gambar.
2. Data kuantitatif adalah data sebagai metode penelitian yang berlandaskan pada filsafat *positivism,* digunakan untuk meneliti pada populasi atau sampel tertentu, teknik pengambil sampel pada umumnya dilakukan secara random, pengumpulan data menggunakan instrument penelitian, analisis data bersifat kuatitatif/statistik dengan tujuan untuk menguji hipotesis yang telah ditetapkan.

2.1.2 Teknik Pengumpulan Data

Menurut jurnal penelitian Pratiwi (2019:212-213), ada tiga jenis teknik pengumpulan data penelitian, diantaranya adalah:

1. Wawancara adalah pertemuan dua orang untuk bertukar informasi dan ide melalui tanya jawab, sehingga dapat dikonstruksikan makna dalam suatu topik tertentu.
2. Observasi adalah proses yang kompleks, suatu proses yang tersusun dari berbagai proses biologis dan psikologis, dua diantara yang terpenting adalah prosesproses pengamatan dan ingatan.
3. Dokumentasi merupakan catatan peristiwa yang sudah berlalu. Dokumen bisa berbentuk tulisan, gambar, atau karya-karya monumental dari seseorang. Dokumen yang digunakan merupakan data pendukung terhadap hasil pengamatan dan wawancara berkaitan dengan bentuk pesan verbal dan non verbal dan juga hambatanhambatan yang ditemui oleh peneliti.

2.2 Pengertian *Mining*

Menurut Kamus Bahasa Inggris yang diterjemahkan ke Bahasa Indonesia, arti kata mining adalah peranjauan. Arti lainnya dari mining adalah pertambangan. Kata *Mining* masuk ke dalam bahasa inggris yaitu bahasa Jermanik yang pertama kali dituturkan di Inggris pada Abad Pertengahan Awal dan saat ini merupakan bahasa yang paling umum digunakan di seluruh dunia.

Menurut jurnal penelitian Marisa hal 92, Mining merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

2.3 Pengertian Data Mining

Data mining merupakan proses analisa data dari sudut yang berbeda dan mengolahnya menjadi informasi - informasi penting yang bisa digunakan untuk meningkatkan keuntungan. Secara teknis, data mining dapat disebut juga sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field. (Asmira, 2019:22).

Data mining merupakan pemisahan dari model informasi yang bermanfaat dalam penyimpanan database. Data mining saja merupakan urutan proses yang berfungsi untuk mencari informasi tambahan yang belum ada didalam database serta model-model yang bermaksud mencarian model data yang dapat berubah menjadi informasi yang penting dari hasil pemisahan serta pengenalan model yang bermanfaat atau menarik dari data yang terdapat dalam database.(Zulfa,Ira, Rayuwati Rayuwati dan Khaidir Koko,2020:71).

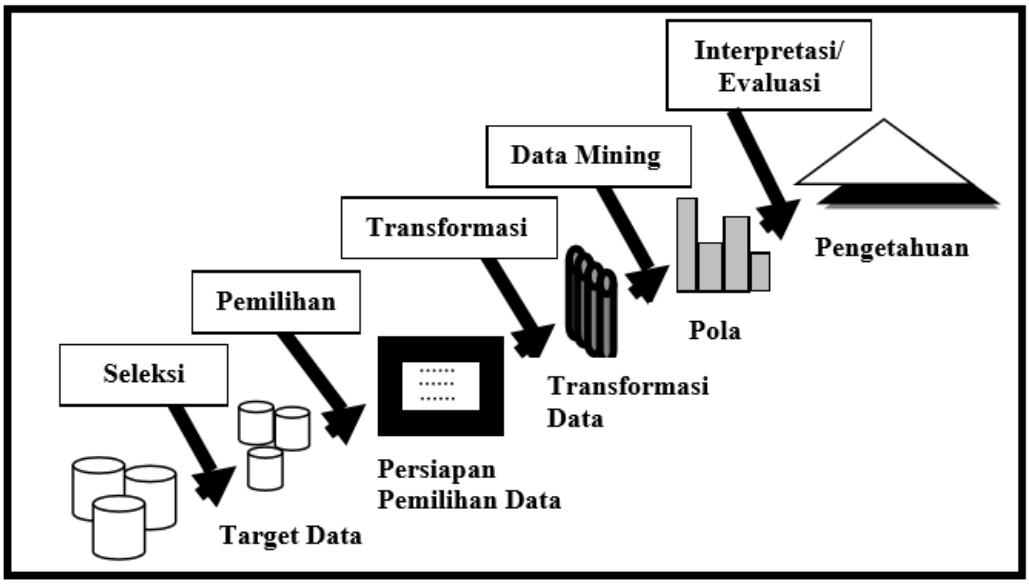
Ada beberapa definisi data mining menurut jurnal Asmira (2019:23), diantaranya:

1. Mengekstrak atau “mining” pengetahuan dari kumpulan data yg sangat besar Ekstraksi informasi yg berguna dari data, dimana sebelumnya tidak diharapkan, tidak dikenal & implisit.
2. Eksplorasi & analisis, secara otomatis atau semiotomatis dari sekumpulan data yg sangat besar untuk memperoleh pola 2 data yg berarti.
3. Proses analisis database yg besar secara semiotomatis untk menemukan pola yang valid, baru, berguna dan dapat dipahami manusia.

Data Mining merupakan inti dari *Knowledge Discovery in Database*

(KDD), KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan mudah dimengerti dari kumpulan data yang besar dan kompleks. Data Mining adalah proses ektraksi kumpulan data-data untuk mendapatkan informasi penting yang sebelumnya tidak diketahui. Data Mining dilakukan untuk mencari pola di dalam data yang nantinya dapat menghasilkan prediksi yang akurat pada data di masa yang akan datang. (Chailes, Anwan, Aditiya Hermawan, dan Didi Kurnaedi, 2020:4).

Pada proses Data Mining yang biasa disebut *Knowledge Discovery Database* (KDD). *Knowledge Discovery Database* (KDD) adalah penerapan metode saintifik pada data mining. Pada konteks ini data mining merupakan satu langkah dari proses KDD, terdapat beberapa proses seperti terlihat pada gambar di bawah ini:



Gambar 2.1 Tahapan Proses Data Mining

sumber:Syahril, Muhammad ,Kamil Erwansyah dan Milfa Yetri,2020.

2.3.1 Tahapan Data Mining

Menurut jurnal penelitian dari Syahril, Muhammad, Kamil Erwansyah dan Milfa Yetri (2020:120), ada beberapa tahapan dalam data mining diantaranya adalah sebagai berikut:

1. Seleksi Data *(Selection)*

   Selection (seleksi/pemilihan) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *Knowledge Discovery Database* (KDD) dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pemilihan Data *(Preprocessing/Cleaning)*

   Proses *Preprocessing* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses "memperkaya" data yang sudah ada dengan data atauinformasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. Transformasi (Transformation)

   Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas ke dalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining.

4. Data Mining

   Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.

5. Interpretasi/Evaluasi *(Interpratation/Evaluation)*

Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses Data Mining pola informasi.

2.3.2 Karakteristik Data Mining

Menurut jurnal penelitian Saputra, Ramadani, Alexander J.P. Sibarani (2020:264), ada 3 karakteristik dari data mining, diantaranya adalah:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

2.3.3 Langkah-langkah Proses Data Mining

Menurut jurnal penelitian Amrin (2017) ada 3 langkah proses yang ada pada data mining, diantaranya adalah:

1. *Data Preparation*

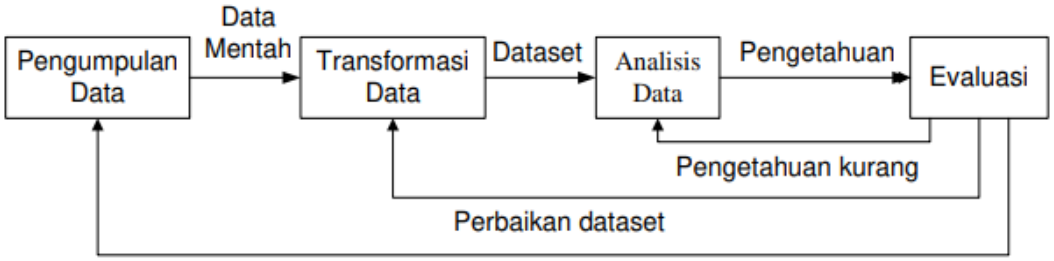
   Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

2. Algoritma Data Mining

   Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

3. Fase Analisa

   data Keluaran dari data mining dievaluasi untuk melihat apakah *knowledge domain* ditemukan dalam bentuk rule yang telah diekstrak dari jaringan.



Gambar 2.2 Langkah-langkah Proses Data Mining

sumber: Firdaus,2017.

2.3.4 Kelompok Kerja Data Mining

Menurut jurnal penelitian Gerung (2018), Data mining dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, yaitu:

1. Deskripsi

   Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

   Hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan baris data *(record)* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel

target dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

   Hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

   Terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Pengklasteran

   Merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas obyek-obyek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan record dalam klaster yang lain. Berbeda dengan klasifikasi, pada pengklasteran tidak ada variabel target. Pengklasteran tidak melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target, akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

6. Asosiasi

Tugas asosiasi dalam data mining adalah untuk menemukan atribut yang muncul dalam satu waktu. Salah satu implementasi dari asosiasi adalah market basket analysis sebagaimana yang akan dibahas dalam penelitian ini. Dalam bidang keilmuan data mining, terdapat suatu metode yang dinamakan *association rule*. Metode ini sering juga dinamakan dengan *market basket analysis*. *Association rule mining* adalah suatu prosedur untuk mencari hubungan antar item dalam suatu data set yang ditentukan.

2.3.5 Kegunaan Data Mining

Menurut jurnal penelitian Zulfaa,Ira, Rayuwati Rayuwati, dan Khaidir Koko (2020:71), Penggalian data *(data mining)* pilihan serta dirubah dalam database menggunakan macam-macam teknik dalam pemrosesan data mining. Algoritma dalam data mining sangat banyak macamnya, maka dari itu harus dipilih yang sesuai dengan permasalahan yang ingin diselesaikan Pencarian model serta tampilan pengetahuan bagian ini memproses tampilan dari pengetahuan seperti pengecekkan model informasi yang didapatkan sudah sesuai dengan kenyataan atau aturan yang sudah dibuat dahulu. Tahapan akhir adalah KDD yang berfungsi mengimplementasikan pengetahuan dengan model yang sangat mudah dimengerti oleh pejadi. Data mining memiliki banyak kegunaan, diantaranya sebagai berikut:

1. Gambaran dalam sebuah model serta memungkinkan memilih item set dengan definisi yang sesuai.
2. Perkiraan dan pengelompokan yang sama, kecuali variabel perkiraan banyak kepada perhitungan disbanding pegelompokan.

3. Pemisahan merupakan alur untuk mendapatkan pola atau kegunaan untuk mendeskripsikan dan membuat perbedaan kelompok item set atau prinsip dari arah untuk menebak kelompok itemset yang tidak memilki kelompoknya.
4. Pemisahan untuk mengumpulkan record, pemantauan, atau melihat dari kelompok subjek-subjek yang mempunyai kesamaan. Kluster merupakan gabungan record yang mempunyai kesamaan serta mempunyai ketidaksamaan dengan beberapa record didalam kluster lain.
5. Asosiasi pada data mining merupakan symbol yang ada didalam suatu waktu. Pada dunia bisnis sering disebut juga sebagai wadah analisis.

2.3.6 Tujuan Data Mining

Menurut jurnal penelitian Napitupul, dkk (2019:168), tujuan dari data mining, diantaranya adalah:

1. Data mining merupakan suatu prosedur otomatis yang menghasilkan prediksi berdasarkan data yang sudah ada.
2. Data yang akan dianalisis yaitu berupa kumpulan data yang kompleks.
3. Data mining bertujuan untuk menemukan relasi yang memungkinkan dapat menghasilkan manifestasi yang bermanfaat.

2.3.7 Kelebihan Data Mining

Kelebihan data mining menurut Turban (2007) sebagai berikut:

1. Data mining mampu menangani data dalam jumlah besar dan kompleks.
2. Data mining dapat menangani data dengan berbagai macam tipe atribut.
3. Data mining mampu mencari dan mengolah data secara semi otomatis, diperlukan parameter yang harus di input oleh user secara manual.

4. Data mining dapat menggunakan pengalaman ataupun kesalahan terlebih dahulu untuk meningkatkan kualitas dan hasil analisa sehingga mendapatkan hasil yang terbaik.

2.3.8 Manfaat Data Mining

Menurut jurnal penelitian Kamil (2019), Pemanfaatan Data Mining dilihat dari dua sudut pandang, yaitu sudut pandang komersial dan sudut pandang keilmuan.

1. Sudut Pandang Komersial

   Sudut pandang komersial pemanfaatan data mining dapat digunakan untuk menangani meledaknya volume data, dengan menggunkan teknik komputasi dapat digunakan untuk menghasilkan informasi-informasi yang dibutuhkan merupakan asset yang dapat meningkatkan daya saing suatu institusi.

2. Sudut Pandang Keilmuan

   Sudut pandang keilmuan data mining dapat digunakan untuk mengcapture, menganalisis serta menyimpan data yang bersifat *real time* dan sangat besar.

2.3.9 Pengertian Algoritma

Menurut Maulana dalam jurnal penelitian Retta, dkk (2016), algoritma merupakan kumpulan perintah untuk menyelesaikan suatu masalah dimana masalah tersebut diselesaikan dituntut secara sistematis, terstruktur dan logis.

Menurut Barakbah dalam jurnal penelitian Retta,dkk (2016), pengertian algoritma sangat lekat dengan kata logika, yaitu kemampuan seorang manusia untuk berpikir dengan akal tentang suatu permasalahan menghasilkan sebuah

kebenaran, dibuktikan dan dapat diterima akal, logika seringkali dihubungkan dengan kecerdasan, seseorang yang mampu berlogika dengan baik sering orang menyebutnya sebagai pribadi yang cerdas. Dalam menyelesaikan suatu masalahpun logika mutlak diperlukan. Logika identik dengan masuk akal dan penalaran. Pertimbangan dalam penerapan algoritma adalah:

1. Algoritma haruslah benar, artinya algoritma akan memberikan keluaran yang dikehendaki dari sejumlah masukan yang diberikan. Tidak peduli sebagus apapun algoritma, kalau memberikan keluaran yang salah, pastilah algoritma tersebut bukanlah algoritma yang baik.
2. Algoritma yang baik harus mampu memberikan hasil yang sedekat mungkin dengan nilai yang sebenarnya. Kita harus mengetahui seberapa baik hasil yang dicapai oleh algoritma tersebut. Hal ini penting terutama pada algoritma untuk menyelesaikan masalah yang memerlukan aproksimasi hasil (hasil yang hanya berupa pendekatan).
3. Efisiensi algoritma, semisal algoritma itu benar (mendekati kebenaran), tetapi memakan waktu yang lama dalam mendapatkan kebenaran algoritma, untuk apa algoritma tersebut dipakai, Karena inti dari algoritma yang baik adalah mendapatkan jawaban kebenaran (mendekati kebenaran) dengan cepat.

2.4 Algoritma *K-Means*

Algoritma *K-Means* merupakan metode *non-heirarchial* yang pada awalnya mengambil sebagian dari banyak nya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada tahap ini pusat *cluster* di pilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* dapat menguji masing-masing

komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah di definisikan tergantung dari jaarak minimum antar komponen dengan tiap-tiap pusat *cluster.* Posisi pusat *cluster* akan di hitung kembali sampai semua komponen data dikatakan ke dalam tiap-tiap pusat cluster dan terakhir terbentuk posisi pusat cluster baru. Beberapa alternatif penerapan K-Means dengan beberapa pengembangan teori-teori penghitungan terkait telah di usulkan. Hal ini termasuk pemilihan Selain itu adapun penelitian menyatakan bahwa kelebihan dengan menggunakan metode *K-Means* ini adalah kecepatan yang lebih tinggi dari dari pengelompokan objek. Kekurangan nya adalah menentukan jumlah *cluster* sebelum percobaan. Dengan berlandaskan teori pada penelitian sebelumnya atau bahkan sumber media lainnya. Peneliti berharap dapat memberikan hasil yang maksimal. Serta memberikan data yang validitasnya dapat di pertanggung jawabkan. (Novianto, 2019).

Algoritma *K-Means* pada dasarnya melakukan dua proses, yaitu proses pendeteksian lokasi pusat tiap cluster dan proses pencarian anggota dari tiap- tiap cluster. Cara kerja algoritma *K-Means*:

1. Tentukan *k* sebagai jumlah *cluster* yang ingin dibentuk.
2. Bangkitkan *k centroid* (titik pusat *cluster*) awal secara random.
3. Hitung jarak setiap data ke masing-masing *centroid*.
4. Setiap data memilih *centroid* yang terdekat.
5. Tentukan posisi *centroid* yang baru dengan cara menghitung nilai rata rata dari data yang terletak pada *centroid* yang sama.

6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* yang lama tidak sama.

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karateristik yang berbeda di kelompokan ke dalam cluster yang lain. Secara umum algoritma dasar dari K-Means Clustering adalah sebagai berikut :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung *centroi*d/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke *centroid*/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada datayang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan

Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satu anya yaitu *Euclidean Distance Space. Euclidean distance space* sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang diperhitungkan. Adapun persamaannya adalah sebagai berikut:

$$d(i,j) = \left[\sum_{k=1}^{p} |xi_k - yj_k| \quad ^2 \right] \quad ^2$$

Dimana

D_{ij} = Jarak objek antara objek *i* dan *j*

P = Dimensi Data

X_{ik} = Koordinat dari objek *I* pada dimensi *k*

X_{jk} = Koordinat dari objek *i*

2.4.1 Clustering

Clustering merupakan salah satu teknik data mining yang menggunakan untuk mendapatkan kelompok-kelompok dari objek yang mempunyai karakteristik yang umum didata yang cukup besar tujuannya untuk menemukan cluster yang berkualitas dalam waktu yang layak. Clustering dalam data mining berguna untuk menemukan pola distribusi didalam sebuah data set yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi.

2.5 Algoritma *Support Vector Machine*

Pengertian *support vector machine* (SVM) yaitu sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi – fungsi linier dalam sebuah fitur yang berdimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi (Puspitasari, Ratnawati, & Widodo, 2018). *Support vector machine* (SVM) diperkenalkan oleh Vapnik, Boser dan Guyon pada tahun 1992 (Fridayanthie, 2015) sebagai rangkaian dari beberapa konsep–

konsep unggulan dalam bidang *pattern recognition* (Susilowati, Sabariah, & Gozali, 2015). Tingkat akurasi pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan (Parapat & Furqon, 2018). Berdasarkan dari karakteristiknya, metode SVM dibagi menjadi dua, yaitu SVM Linier dan SVM Non-Linier. SVM linier merupakan data yang dipisahkan secara linier, yaitu memisahkan kedua class pada hyperplane dengan soft margin. Sedangkan SVM Non-Linier yaitu menerapkan fungsi dari kernel trick terhadap ruang berdimensi tinggi (Rachman & Purnami, 2012). Pada dokumentasi Matlab, SVM untuk prediksi disebut dengan SVM Regression terdiri dari fungsi linear dan nonlinear dengan primal formula dan dual formula) (Nurmasani, Utami, & Al Fatta, 2017). Metode SVM Regression disebut sebagai suatu teknik nonparametric karena bergantung pada fungsi kernel. Adapun rumus persamaannya sebagai berikut:

Persamaan garis lurus dengan model matematika seperti berikut

$$y = mx + b \quad \text{... (1)}$$

Dengan m disebut dengan gradient, persamaan berikut

$$y = 2x + 1 \quad \text{... (2)}$$

2.6 Pemodelan Sistem

Perancangan sistem yang baru dimulai dengan perancangan database, yang dimulai dengan pembuatan UML yang akan dilanjut dengan perancangan aplikasinya. *Unified Modeling Language* (UML) adalah sebuah bahasa yang telah menjadi standar dalam industry yang digunakan untuk visualisasi, merancang dan

mendokumentasikan sistem peranti lunak.

2.6.1 *Unfield Modeling Language* (UML)

Menurut Hasdiana & Hasanul Fahmi (2018), *Unified Modeling Language* merupakan satu kumpulan konvensi pemodelan yang digunakan untuk menentukan atau menggambarkan sebuah sistem software yang terkait dengan objek. *Unified Modeling Language* atau yang lebih dikenal dengan UML merupakan salah satu materi ajar yang penting dalam matakuliah Analisis dan Perancangan Sistem Informasi. UML digunakan sebagai salah satu alat untuk melakukan perancangan atau memodelkan sistem. UML sering digunakan karena penggunaanya yang tidak terpengaruh pada perangkat lunak, perangkat keras, sistem operasi, jaringan, basis data dan bahasa pemrograman yang digunakan. Pada UML versi 2 terdiri atas tiga kategori dan memiliki 13 jenis diagram yaitu:

1. Struktur Diagram, terdiri dari *Class diagram, Object diagram, Component diagram, Deployment diagram, Composite structure diagram, Package diagram*
2. *Behavior Diagram,* terdiri dari *Use case diagram, Activity diagram, State Machine diagram* (State chart diagram in version 1.x)
3. *Interaction diagram,* terdiri dari *Communication diagram, Interaction Overview diagram, Sequence diagram, Timing diagram.*

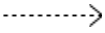
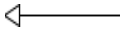
   Adapun penjabaran dari masing-masing UML adalah sebagai berikut:

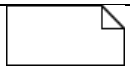
   a. *Use Case Diagram*

      Menurut Isa dan Hartawan (2017), *Use Case diagram* digunakan untuk menggambarkan sistem dari sudut pandang pengguna sistem tersebut

(*user*). sehingga pembuatan *use case diagram* lebih dititik beratkan pada fungsionalitas yang ada pada sistem, bukan berdasarkan alur atau urutan kejadian. Sebuah *use case diagram* mempresentasikan sebuah interaksi antara aktor dengan sistem.

Tabel 2.1 Simbol *Use Case Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		*Actor*	Menspesifikasikan himpuan peran yang pengguna mainkan ketika berinteraksi dengan *use case*.
2		*Dependency*	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri *(independent)* akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri (*independent*).
3		*Generalization*	Hubungan dimana objek anak (*descendent*) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (*ancestor*).
4		*Include*	Menspesifikasikan bahwa *use case* sumber secara *eksplisit*.
5		*Extend*	Menspesifikasikan bahwa *use case* target memperluas perilaku dari *use case* sumber pada suatu titik yang diberikan.
6		*Association*	Apa yang menghubungkan antara objek satu dengan objek lainnya.

NO	GAMBAR	NAMA	KETERANGAN
7		*System*	Menspesifikasikan paket yang menampilkan sistem secara terbatas.
8		*Use Case*	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor
9		*Collaboration*	Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan prilaku yang lebih besar dari jumlah dan elemen-elemennya (sinergi).
10		*Note*	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi

sumber : Winda Aprianti dan Umi Maliha,2016.

b. Activity Diagram

Menurut Isa dan Hartawan (2017), *Activity Diagram* Menggambarkan rangkaian aliran dari aktivitas, digunakan untuk mendeskripsikan aktifitas yang dibentuk dalam suatu operasi sehingga dapat juga digunakan untuk aktifitas lainnya. Diagram ini sangat mirip dengan *flowchart* karena memodelkan *work flow* dari suatu aktifitas ke aktifitas yang lainnya, atau dari aktifitas ke status. *Activity diagram* juga digunakan untuk menggambarkan interaksi antara beberapa *use case*.

Tabel 2.2 Simbol *Activity Diagram*

NO	GAMBAR	NAMA	KETERANGAN

1		*Actifity*	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain
2		*Action*	State dari sistem yang mencerminkan eksekusi dari suatu aksi
3		*Initial Node*	Bagaimana objek dibentuk atau diawali.
4		*Actifity Final Node*	Bagaimana objek dibentuk dan dihancurkan
5		*Fork Node*	Satu aliran yang pada tahap tertentu berubah menjadi beberapa aliran

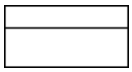
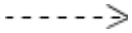
sumber : Winda Aprianti dan Umi Maliha,2016.

c. Class Diagram

Menurut Isa dan Hartawan (2017), *Class Diagram* adalah spesifikasi yang akan menghasilkan objek dan merupakan inti dari pengembangan dan desain berorientasi objek. *Class* menggambarkan keadaan (atribut atau properti) suatu sistem, sekaligus menawarkan layanan untuk memanipulasi keadaan tersebut (metode atau fungsi).

Tabel 2.3 Simbol *Class Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		*Generalization*	Hubungan dimana objek anak *(descendent)* berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (*ancestor*).
2		*Nary Association*	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.

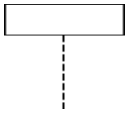
NO	GAMBAR	NAMA	KETERANGAN
3		*Class*	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
4		*Collaboration*	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu actor
5		*Realization*	Operasi yang benar-benar dilakukan oleh suatu objek.
6		*Dependency*	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri *(independent)* akan mempegaruhi elemen yang bergantung padanya elemen yang tidak mandiri
7		*Association*	Apa yang menghubungkan antara objek satu dengan objek lainnya

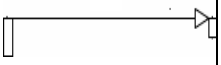
sumber : Winda Aprianti dan Umi Maliha,2016.

d. Squence Diagram

Menurut Isa dan Hartawan (2017), *Squence Diagram* menggambarkan interaksi antara sejumlah objek dalam urutan waktu. Kegunananya untuk menunjukan rangkaian pesan yang dikirim antara objek juga interaksi antar objek yang terjadi pada titik tertentu dalam eksekusi sistem.

Tabel 2.4 Simbol *Squence Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		*LifeLine*	Objek *entity*, antarmuka yang saling berinteraksi.

NO	GAMBAR	NAMA	KETERANGAN
2		*Message*	Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktifitas yang terjadi
3		*Message*	Spesifikasi dari komunikasi antar objek yang memuat informasi-informasi tentang aktifitas yang terjadi

sumber : Winda Aprianti dan Umi Maliha,2016

Penelitian Terkait

Tabel 2.5 Hasil Penelitian

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
1.	Normah, Siti Nurajizah, Arinda Salbinda (2021)	Penerapan Data Mining Metode *K-Means Clustering* Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten	Penerapan metode K-Means pada toko Helai, yaitu dengan cara mengelompokan data stok baju. Kemudian memilih 3 cluster secara acak sebagai centroid awal. Setelah data pada setiap cluster tidak berubah-ubah, maka dapat diketahui hasil akhirnya yaitu yang sangat laris ada 11 artikel, yang laris ada 55 artikel dan 34 artikel untuk yang kurang laris.	-	Dapat diketahui hasil akhirnya yaitu yang sangat laris ada 11 artikel, yang laris ada 55 artikel dan 34 artikel untuk yang kurang laris.
2.	Muhammad Rizki, Fira Fania, Agus P Windarto (2020)	Implementasi *K-Means Clustering* Dalam Mengelompokkan Jumlah Penjualan Ikan Laut Di Tpi Menurut	Terdapat 20 Provinsi di Indonesia di dalam penelitian ini.Peneliti menggunakan algoritma K-Means clustering dalam penelitian ini,Peneliti membuat 2 buah cluster di dalamnya yakni cluster tingkat tinggi (C1) dan cluster tingkat rendah (C2).Peneliti memanfaatkan software RapidMiner dalam proses	-	Terdapat 2 Provinsi yang menduduki posisi cluster tinggi yakni

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
		Wilayah	pengujian penelitian ini.Hasil dari penelitian ini menghasilkan 2 Provinsi memnduduiki posisi cluster tingkat tinggi (C1) dan 18 Provinsi menduduki cluster tingkat rendah (C2).		DKI Jakarta dan Jawa Timur. Sedangkan 18 Provinsi lainnya menduduki cluster rendah. Penelitian ini diharapkan dapat menjadi acuan pemerintah wilayah dalam menstabilkan harga ikan terutama untuk provinsi yang menduduki cluster tingkat

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
					tinggi
3.	Suhandio Handoko, Fauziah ,Endah Tri Esti Handayani (2020)	Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode *K-Means Clustering*	Pada penelitian ini data penjualan dikelompokkan menjadi 3 yaitu data penjualan rendah, data penjualan sedang dan data penjualan tinggi. Pengujian clustering dengan algoritma K-Means pada aplikasi terhadap data transaksi penjualan paket telkomsel diperoleh persentase kesesuaian yaitu 100% dibandingkan dengan clustering manual.	-	Diperoleh persentase kesesuaian yaitu 100% dibandingkan dengan clustering manual.
4.	Ridho Ananda, Achmad Zaki Yamani (2021)	Penentuan Centroid Awal K-means pada proses Clustering Data Evaluasi Pengajaran Dosen	Hasil-hasil clustering dengan preprocessing tersebut selanjutnya dibandingkan untuk memperoleh clustering optimal. Hasil perbandingan menunjukan bahwa k-means menggunakan kriteria Elbow memberikan hasil clustering terbaik pada data nilai evaluasi dosen ITTP tahun 2018 dengan terbentuk empat cluster dan nilai index Silhouette 0.6772. Disimpulkan juga bahwa hasil k-means dengan kriteria Elbow memberikan hasil yang lebih baik dibandingkan k-means tanpa preprocessing dengan peluang 0.778 pada data yang digunakan. Interpretasi hasil clustering kmeans dengan kriteria Elbow menunjukkan,	-	Hasil k-means dengan kriteria Elbow memberikan hasil yang lebih baik dibandingkan k-means tanpa preprocessing dengan peluang

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
			ada 3 dosen dengan kinerja pengajaran yang rendah yaitu N16, N25, dan N84		0.778 pada data yang digunakan.
5.	Helena Nurramdhani Irmanda, Ria Astriratma (2021)	Klasifikasi Jenis Pantun dengan Metode *Support Vector Machines* (SVM)	-	Hasil klasifikasi kemudian dievaluasi untuk mendapatkan nilai akurasi dan akan dianalisis apakah model klasifikasi yang dibuat layak digunakan. Hasil penelitian menunjukkan bahwa SVM	Hasil klasifikasi akurasi sebesar 81,91%.

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
				dapat dengan baik mengklasifikasi jenis pantun dengan akurasi sebesar 81,91%.	
6.	Emy Haryatmi, Sheila Pramita Hervianti (2019)	Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu	-	Dengan menerapkan algoritma SVM dari data training dan data testing dan evaluasi untuk memvalidasi dan mengukur keakuratan model. Hasil	Algoritma SVM memberikan nilai akurasi yang sangat baik yaitu 94,4%

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
				pengujian kelompok pertama dengan jumlah data training sebanyak 90% dan data testing sebanyak 10% menunjukkan bahwa algoritma SVM memberikan nilai akurasi yang sangat baik yaitu 94,4%	
7.	Fithri Selva Jumeilah (2017)	Penerapan Support Vector Machine (SVM)	-	Dari hasil penelitian	Hasil pengujian yang

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
		untuk Pengkategorian Penelitian		diperoleh bahwa pengkategorian penelitian yang dihasilkan oleh SVM sudah sangat baik. Hal ini dibuktikan oleh hasil pengujian yang menghasilkan tingkat akurasi 90%.	menghasilkan tingkat akurasi 90%.
8.	A A Aldino, D Darwis, A T Prastowo, and C Sujana (2021)	*Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency*	*Data that have the same characteristics are grouped in one cluster and the remaining is grouped into another cluster so that the data that is in one cluster has a small degree of variation. So the authors tried to apply the K-Means clustering method from*	-	*The region with the highest number of corn harvest is Penengahan*

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
			the corn crop data of the last 2 years to produce feasibility information from each sub-district.		*with 79448.30257 centroid 2 value. The region with the lowest number of corn harvest is Candipuro with 1424.036868 centroid 2 value.*
9.	Rozzi Kesuma Dinata,Novia Hasdyna, Sujacka Retno, Muhammad Nurfahmi (2021)	*K-means algorithm for clustering system of certain plant seeds specialization areas in east aceh*	*This test stops at the 8th iterations. The result of the last iteration will be used for the clustering parameters. The author determines which cluster members are included in the cluster that is high interest, decent interest, and low interest based on the average value of the last centroid. The average value of C1 = 6133,778364, C2 = 13149,225, C3 = 31217,514, the members of C1 are described as low interest, members of C2 are decent interest, and members of C3 are high interest. As for*	-	*The average value of C1 = 6133,778364, C2 = 13149,225, C3 = 31217,514, the members of C1 are described as low interest, members of C2 are decent*

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
			the calculation of other commodities, it is the same as the calculation of the rice commodity.		*interest, and members of C3 are high interest.*
10.	Babacar Gaye , Dezheng Zhang, and Aziguli Wulamu (2021)	*Improvement of Support Vector Machine Algorithm in Big Data Background*	-	*The algorithm speed can be improved by transforming the solution of the dua lproblemi nto the classification surface of the original space. Conclusion. By improving the calculation rate of traditional machine learning algorithms, it is concluded that the accuracy of the fitting prediction*	*it is concluded that the accuracy of the fitting prediction between the predicted data and the actual value is as high as 98%,which can make the traditiona lmachine learning algorithm meett here quire ments of the big data era.It can be widely used in the context of big data.*

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
				between the predicted data and the actual value is as high as 98%,which can make the traditiona lmachine learning algorithm meett here quire ments of the big data era.It can be widely used in the context of big data.	
11.	Dakhaz Mustafa Abdullah, Adnan Mohsin Abdulaz (2021)	*Machine Learning Applications based on SVM Classification: A Review*	-	*the highest accuracy obtained was 97.47% and the same for and diseases diagnostics where the best-*	*the highest accuracy obtained was 97.47% and the same for and diseases diagnostics where the best-*

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
				obtained accuracy was 97% in cancer identification. However, for Sentiment analysis the highest obtained accuracy was 77% which is not as high as other previous applications.	*obtained accuracy was 97% in cancer identification. However, for Sentiment analysis the highest obtained accuracy was 77% which is not as high as other previous applications.*
12.	Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Naoureen Hameed (2018)	*Sentiment Analysis using SVM: A Systematic Literature Review*	-	*This study has provided a compact and comprehensive review of latest research by focusing on SVM technique of sentiment analysis. This study has followed a systematic*	-

No.	Penulis	Judul Penelitian	Klasterisasi	Klasifikasi	Akurasi
				framework for review and provided the answers of identified research questions after critical review of selected papers. For future work it is suggested to perform a comparative analysis of the customized techniques with same dataset.	