

BAB II

LANDASAN TEORI

2.1 Data Mining

Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Namun tidak semua pekerjaan pencarian informasi dapat dinyatakan sebagai data mining [4]. Data Mining serta KDD memiliki tujuan yang sama yakni menggunakan data yang tersedia pada basis data kemudian memproses data untuk mendapatkan sebuah informasi baru yang berguna.

Data mining adalah proses mengekstraksi/menggali/ mencari informasi yang tersembunyi dari dalam database yang besar serta bermanfaat bagi pihak yang berkepentingan. Data mining dapat membantu pemegang data untuk menganalisis dan menemukan hubungan yang tidak terduga antar data, pada akhirnya akan membantu pengambilan keputusan[2]. Dalam data mining sendiri memiliki beberapa cara penerapan sesuai dengan penelitian ini menggunakan metode K-Means Clustering adalah teknik pengelompokan data non-hirarki yang memisahkan data ke dalam cluster, mengelompokkan data dengan fitur yang sama bersama-sama dan mengelompokkan data dengan karakteristik yang berbeda ke dalam kelompok yang berbeda.

2.1.1 Proses Data Mining

Proses data mining melibatkan beberapa tahapan yang bertujuan untuk menghasilkan informasi yang relevan dan dapat diterapkan dalam pengambilan keputusan. langkah-langkah dalam proses data mining meliputi

2.1.1.1 Pemahaman Masalah

Tahap awal ini melibatkan pemahaman mendalam mengenai:

Mengidentifikasi jenis-jenis pola asuh (otoriter, demokratis, permisif) dan faktor-faktor yang relevan. Mengidentifikasi indikator kesejahteraan emosional yang akan dianalisis (misalnya, tingkat stres, kebahagiaan, atau hubungan sosial). Mengungkap hubungan antara pola asuh dengan kesejahteraan emosional mahasiswa Fakultas Sains dan Teknologi Universitas Labuhanbatu.

2.1.1.2 Pengumpulan Data

Pengumpulan data dilakukan melalui metode seperti :

- a. Kuesioner ini dirancang untuk mengumpulkan informasi terkait pola asuh orang tua dan kesejahteraan emosional mahasiswa Fakultas Sains dan Teknologi Universitas Labuhanbatu. Kuesioner terdiri dari dua bagian utama. Bagian pertama berfokus pada pola asuh orang tua, dengan pertanyaan yang menggali sejauh mana orang tua menerapkan batasan, memberikan dukungan, serta melibatkan anak dalam pengambilan keputusan.
- b. wawancara atau observasi sering digunakan untuk memperkuat data yang telah diperoleh dari metode utama, seperti kuesioner.

1. Persiapan Data

Pada tahap ini, data yang telah dikumpulkan akan diproses menjadi format yang dapat digunakan untuk analisis.

2. Pemrosesan dengan Metode *K-Means Clustering*

Mengelompokkan data berdasarkan pola asuh orang tua dan Tentukan jumlah clustering.

3. Pemrosesan dengan metode K- Nearest Neighbor (KNN)

Memprediksi kesejahteraan emosional mahasiswa berdasarkan cluster pola asuh yang telah dihasilkan dan pilih atribut kesejahteraan emosional sebagai variabel target.

4. Evaluasi model yang dihasilkan untuk memastikan bahwa pola asuh orang tua atau kesejahteraan emosional mahasiswa sesuai.

5. Pelaporan Hasil

Tahap akhir adalah menyusun laporan hasil analisis

2.2.2 Manfaat Data Mining

Manfaat dari data mining ini adalah untuk menemukan pola yang sebelumnya tidak diketahui. Apabila pola sudah ditemukan, selanjutnya digunakan untuk membuat suatu keputusan untuk keperluan tertentu bagi yang berkepentingan. Data mining adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan, dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika[2]. Secara khusus, data mining memiliki metode tersendiri berdasarkan tujuan dari pemanfaatan himpunan data yaitu

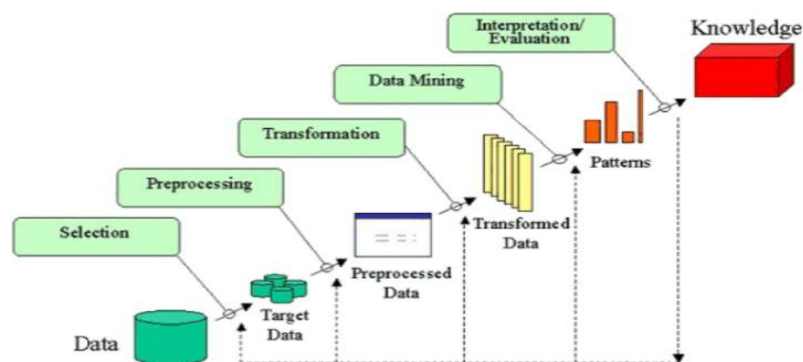
estimasi, prediksi, klasifikasi, klasterisasi dan asosiasi[8]. Data mining adalah proses ekstraksi pengetahuan tersembunyi dalam data secara otomatis atau semi-otomatis [9]. Proses data mining merujuk kepada semua aktivitas yang biasanya muncul dalam proses data mining; penggalian data (mencari dan menemukan pola-pola data) hanya satu bagian dari proses tersebut[10].

2.2 Knowledge Discovery in Database

Knowledge Discovery In Database (KDD) merupakan metode untuk memperoleh pengetahuan dari database yang ada. Dalam database terdapat tabel - tabel yang saling berhubungan / berelasi. Hasil pengetahuan yang diperoleh dalam proses tersebut dapat digunakan sebagai basis pengetahuan (*knowledge base*) untuk keperluan pengambilan keputusan. Istilah *Knowledge Discovery in Database* (KDD) dan data mining sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut[1].

Knowledge Discovery in Database (KDD) adalah proses yang tidak sepele yang digunakan untuk mengidentifikasi validitas data, potensi, guna, dan pada akhirnya menghasilkan pola data yang dapat dimengerti . KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah kumpulan data. Ada pendekatan kuantitatif, semacam pendekatan probabilistic semacam logika induktif, pencarian pola, dan analisis tumbuhan keputusan. Pendekatan yang lain meliputi deviasi, analisis

kecenderungan, algoritma genetik, jaringan saraf tiruan, dan pendekatan campuran 2 maupun lebih dari sebagian pendekatan yang ada. Pada dasarnya ada 6 elemen yang sangat esensial dalam tata carapencaria ninformasi/ pengetahuan dalam KDD yakni:mengerjakan sebagian besar data, diperlukan efesiensi berkaitan dengan volume data, mengutamakan ketetapan/ keakuratan, membutuhkan pemakaian bahasa tingkatan besar, menggunakan sebagian wujud dari pendidikan otomatis, menghasilkan hasil yang menarik[2]. Data mining sendiri adalah bagian dari tahapan proses KDD, Terdapat beberapa tahapan di dalam proses KDD diantaranya yaitu *selection, preprocessing, transformation, data mining*, dan *evaluation* yang terdapat pada gambar 2.1



Gambar 2.1 Tahapan Proses KDD

Sumber : Haris Kurniawan, SarjonDefit, Sumijan. [3]

Berikut ini penjelasan setiap tahapan dalam proses KDD :

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

2. *Pre-processing / Cleaning* Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus *Knowledge Discovery in Database* (KDD). Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk *Knowledge Discovery in Database* (KDD), seperti data atau informasi eksternal lainnya yang diperlukan.

3. *Transformation Coding* adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam *Knowledge Discovery in Database* (KDD) merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan.

5. *Interpretation / Evaluation* Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses *Knowledge Discovery in Database* (KDD) yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.3 Clustering

Clustering adalah metode pengelompokan data yang sering digunakan sebagai salah satu metode data mining atau penggalian data. *Clustering* adalah proses partisi satu set objek data kedalam himpunan bagian yang disebut dengan cluster. Oleh karena itu, metode *clustering* ini sangat berguna untuk menemukan kelompok yang tidak dikenal dalam data[11]. *Clustering* merupakan proses pengelompokan sekumpulan objek ke dalam kelas/pola objek yang serupa. *Cluster* merupakan teknik yang sangat penting dalam analisis data. *Cluster* merupakan membagi data menjadi kelompok-kelompok atau *cluster* berdasarkan suatu kemiripan atribut-atribut diantara sekumpulan data, atribut-atribut yang mirip tersebut dipresentasikan sebagai titik-titik dalam ruang multidimensi[2]. *Clustering* merupakan salah satu metode Data Mining yang bersifat tanpa arahan (*unsupervised*) Ada dua jenis data clustering yang sering dipergunakan dalam proses pengelompokan data yaitu hierarchical (hirarki) data *clustering* dan *non-hierarchical* (non hirarki) data *clustering*. *Clustering* adalah suatu metode pengelompokan berdasarkan ukuran kedekatan. Perbedaan *Clustering* dengan grup, kalau grup berarti kelompok yang sama kondisinya kalau tidak ya pasti bukan kelompoknya[10]. Teknik data mining clustering akan mengelompokkan data dengan karakteristik yang sama dan mengelompokkan data dengan karakteristik yang berbeda ke kelompok lain[12].

2.4 Pola Asuh

Pola asuh adalah pola perilaku yang semua orang terapkan kepada anak dan sifatnya relative konsisten. Ada beberapa macam pola asuh orang tua yaitu demokratis, otoriter, dan permisif. Masing-masing pola asuh memiliki keuntungan dan kelemahan tersendiri. Sudah menjadi tugas dan kewajiban orang tua untuk memperlakukan, mendisiplinkan, dan mendidik serta melindungi anak dalam mencapai proses kedewasaan[13]. Pola asuh adalah suatu cara yang digunakan oleh orang dalam mencoba berbagai strategi untuk mendorong anak – anaknya mencapai tujuan yang diinginkan, dimana tujuan tersebut antara lain pengetahuan, nilai moral dan standart perilaku yang harus dimiliki anak bila dewasa nanti Pola asuh merupakan pola interaksi antara orang tua dan anaknya yaitu bagaimana sikap antara perilaku orangtua saat berinteraksi dengan anak, termasuk caranya menerapkan aturan, mengajarkan nilai / atau norma, memberikan perhatian dan kasih sayang serta menunjukkan sikap dan perilaku yang baik sehingga dijadikan contoh atau panutan bagi anaknya[6].

Pola asuh merujuk pada cara orang tua memenuhi kebutuhan fisik, emosional, serta juga sosial anak-anak mereka, serta cara mereka memberikan bimbingan serta juga batasan didalam menghadapi perilaku anak. Berdasarkan teori serta juga temuan, terdapat beberapa jenis pola asuh utama, seperti otoriter, otoritatif, permisif, serta juga penelantaran Pola asuh otoriter ditandai dengan adanya kontrol yang tinggi serta juga kaku, sedangkan pola asuh permisif cenderung kurangnya kontrol serta juga keterlibatan yang terlalu sedikit. Pola asuh orang tua juga menjadi salah satu pengaruh dalam proses perubahan-perubahan

yang dialami remaja serta bagaimana remaja bersikap serta berperilaku di masyarakat[15]. Sementara itu, pola asuh otoritatif dianggap sebagai pola asuh yang optimal dikarenakan mencakup kontrol yang tepat serta juga dukungan emosional yang kuat[16]. Pola asuh harus pula berkaitan menggunakan tanggungjawab dan kewajiban orang tua terhadap anak.

Menurut Undang-Undang Republik Indonesia Nomor 35 Tahun 2014, Pasal 26, orangtua dalam keluarga berkewajiban dan bertanggungjawab[17]. Peran pola asuh orangtua yang menerapkan jenis pola asuh demokratis mampu meningkatkan motivasi belajar terhadap anak seperti selalu membiasakan anak untuk disiplin belajar dan beribadah, orangtua juga memberikan pendampingan belajar yang maksimal terhadap anak, hal tersebut mampu memberikan dampak positif bagi peningkatan motivasi belajar anak. Tidak hanya itu pola asuh orang tua juga memiliki keterkaitan dalam pendidikan karakter yang meliputi : religius, disiplin, toleransi, bersahabat dan mandiri[18].

2.5 Orang Tua

Orang tua adalah ayah, ibu, Orang tua, juga dikenal sebagai keluarga, atau yang identik dengan pengasuh anak dalam keluarga. Menurut Hamalik (2007: 33) peran adalah pola tingkah laku tertentu yang merupakan ciri-ciri khas semua petugas dari pekerjaan atau jabatan tertentu[19]. Meskipun orang-orang tua dibagi menjadi tiga, yaitu orang tua, orang tua asuh, dan tua tiri. Orang tua merupakan anggota keluarga. yang terdiri dari ayah dan ibu dan berfungsi sebagai hasil dari ikatan perkawinan legal yang dapat menjadi keluarga. Orang tua memiliki kewajiban untuk mengasuh, membimbing, dan mendidik anak-anaknya untuk

mencapai tingkat tertentu yang membuat anak siap dalam hidup dalam masyarakat. Anak-anak dapat mulai belajar dari kedua orang tua atau mulai dari masa kandungan, ayunan, berdiri, berjalan, dan seterusnya. Perkembangan hidup manusia memiliki fase dimulai dari anak-anak, remaja, dewasa, lanjut usia. Empat fase tersebut memiliki karakter masing-masing[20].

Orang tua adalah orang pertama yang bertanggung jawab untuk mendidik anak. Interaksi sehari-hari antara orang tua dan anak, serta pola asuh orang tua, membentuk karakter anak. Orang tua biasanya harus menjaga potensi psikomotor, kognitif, dan efektif anak, serta memberikan nutrisi dan kehidupan yang layak. Beberapa orang tua memilih untuk memberikan aturan yang ketat dan disiplin yang tinggi, sementara yang lain lebih memilih untuk melibatkan anak dalam pengambilan keputusan dan memberikan kebebasan dalam batasan yang aman[21]. Oleh karena itu, peran orang tua sangat penting dalam membentuk karakter dan tingkah laku anak. Pendidikan yang diberikan orang tua kepada anak saat mereka masih kecil akan berdampak pada perilaku mereka saat mereka menjadi remaja dan dewasa[8].

2.6 Emosional

Mental emosional adalah suatu kondisi dimana seseorang mengalami distress psikologik, terjadi perubahan psikologis pada keadaan tertentu tetapi bisa kembali pulih seperti semula, akan tetapi masalah mental emosional ini apabila tidak ditangani secara tepat akan menimbulkan dampak yang buruk bagi proses perkembangan remaja Masalah mental emosional merupakan suatu keadaan yang

di alami oleh individu ditandai dengan perubahan emosional dan apabila berkelanjutan akan berkembang menjadi keadaan patologis.

World Health Organization (WHO, 2018), menyatakan Prevalensi orang dengan gangguan mental emosional di dunia dalam rentang usia 10-19 tahun kondisi kesehatan mental mencakup 16% dari beban penyakit dan cedera global. Berbicara tentang perkembangan sosial emosional, sebagai salah satu aspek perkembangan anak sejatinya tidak dapat dipisahkan satu sama lain. Dengan kata lain, membahas perkembangan emosi harus bersinggungan dengan perkembangan sosial anak[24]Setengah dari semua kondisi kesehatan mental dimulai pada usia 14 tahun tetapi kasus tidak terdeteksi dan tidak diobati karena sejumlah alasan, seperti kurangnya pengetahuan atau kesadaran tentang kesehatan mental diantara petugas kesehatan, atau stigma yang mencegah remaja mencari bantuan, hal ini bisa meningkatkan kemungkinan pengambilan perilaku beresiko lebih lanjut dan dapat mempengaruhi kesejahteraan kesehatan mental dan emosi pada remaja[9].Perilaku sosial emosional anak dipengaruhi oleh proses perlakuan, pengasuhan atau bimbingan dari orangtua terhadap anak dalam mengenalkan berbagai aspek kehidupan sosial, atau norma-norma kehidupan bermasyarakat serta mendorong dan memberikan contoh kepada anaknya bagaimana menerapkan norma-norma tersebut dalam kehidupan sehari-hari. perkembangan emosional anak berlangsung dalam interaksi dinamis antara hubungan orangtua dengan anak dan lingkungan mereka [26].

2.7 Remaja

Masa remaja merupakan masa dimana seseorang mengalami perkembangan diri karena pada masa ini remaja bertransisi dari anak-anak menuju dewasa. Perkembangan yang terjadi pada masa remaja ini akan menuntunnya kepada perubahan-perubahan dalam diri individu baik yang terlihat maupun tidak terlihat. Remaja yang tidak mampu menyesuaikan diri dengan peran barunya tersebut dapat membuat dirinya labil dan emosional bahkan dapat membuat frustrasi dan depresi[27]Aspek perkembangan mental dan emosi remaja yang merupakan generasi penerus bangsa perlu diperhatikan. Kematangan mental emosional, sosial, dan fisik merupakan arti luas dari usia remaja mengatakan bahwa masa remaja merupakan masa peralihan perkembangan antara kanak-kanak dengan masa dewasa dimana pada masa ini akan lekat dengan beberapa perubahan, seperti perubahan biologis, kognitif, dan emosional yang nantinya akan memiliki masalahnya sendiri pada setiap periode perubahannya.

Masa remaja adalah masa di mana seseorang mengalami perubahan besar dalam dirinya, sehingga memiliki kerumitan yang cukup tinggi dalam dirinya. sifat peralihan ditunjukkan pada masa remaja karena mereka belum mendapatkan status dewasa, namun sudah terlepas dari status anak. Ada banyak remaja yang sudah merasa cukup dewasa dan cenderung tidak ingin diperlakukan seperti anak-anak. [6].Remaja merupakan masa peralihan dari masa kanak-kanak ke masa dewasa. Remaja merupakan periode transisi perkembangan individu antara masa kanak-kanak dan masa dewasa, yang melibatkan berbagai perubahan antara lain perubahan biologis, kognitif dan sosio-emosional. Rentan usia remaja dimulai sekitar usia 10-

13 tahun dan berakhir pada usia 18-22 tahun. Remaja memiliki tuntutan – tuntutan tugas seperti tuntutan perkembangan dimana remaja dituntut untuk menyesuaikan diri seperti mengembangkan hati Nurani, tanggung jawab, moralitas dan nilai – nilai yang sesuai dengan lingkungan dan kebudayaannya[10].

2.8 Metode K-Nearest Neighbor

K-Nearest Neighbor merupakan metode klasifikasi terhadap sekumpulan data berdasarkan pembelajaran data yang sudah terklasifikasikan sebelumnya[29]. *K-Nearest Neighbor* adalah metode untuk mengklasifikasi objek berdasarkan data latih yang paling dekat dengan objek tersebut[30]. K-Nearest Neighbor (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan (K) tetangga terdekatnya[35]. Algoritma *K-Nearest Neighbor* (KNN) adalah sekumpulan data berdasarkan pembelajaran data yang terklasifikasikan sebelumnya. Algoritma *K-Nearest Neighbor* (K-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek baru berdasarkan (K) tetangga terdekatnya. *K-Nearest Neighbor* (K-NN) termasuk algoritma Supervised Learning, dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari label class atau kategori pada K-NN.

Algoritma *K-Nearest Neighbor* adalah algoritma machine learning yang digunakan untuk klasifikasi data. Algoritma ini digunakan untuk mengklasifikasi objek baru berdasarkan data pembelajaran yang jarak tetangganya paling dekat atau memiliki nilai selisih yang kecil dengan objek tersebut. KNN dapat digunakan untuk estimasi dan prediksi data serta sering digunakan dalam teknik data mining[31] Kelas yang paling banyak muncul, yang akan menjadi kelas hasil

klasifikasi. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *Euclidian Distance*. *Euclidian Distance* sering digunakan untuk menghitung jarak. *Euclidian Distance* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua objek, di bawah ini merupakan rumus *Euclidian Distance*:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Dimana:

$d(p, q)$: jarak antara titik p dan q .

p_i : Nilai koordinat ke $-i$ pada vector p .

q_i : Nilai koordinat ke- i pada vector q .

n : jumlah dimensi atau fitur data.

Jika hasil nilai dari rumus di atas besar maka akan semakin jauh tingkat keserupaan antara kedua objek dan sebaliknya jika hasil nilainya semakin kecil maka akan semakin dekat tingkat keserupaan antar objek tersebut. Objek yang dimaksud adalah training data dan testing data. Dalam algoritma ini, nilai K yang terbaik itu tergantung pada jumlah data. Ukuran nilai K yang besar belum tentu menjadi nilai K yang terbaik begitupun juga sebaliknya. Langkah-langkah untuk menghitung algoritma K-NN:

1. Menentukan nilai K (nilai K dipilih secara manual).
2. Menghitung jarak antara data training dan data testing dengan menggunakan rumus *Euclidian Distance*.

3. Kemudian mengurutkan objek-objek (data training) tersebut ke dalam kelompok yang berdasarkan jarak terkecil.
4. Menetapkan kelas, dimana kelas yang dipilih adalah kelas dengan jumlah nilai K terbanyak pada data testing[13].

2.9 Metode K-Means Clustering

Metode *K-Means Clustering* termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke k daerah bagian yang terpisah. Pada *K-Means*, setiap objek harus masuk dalam kelompok tertentu, tetapi dalam satu tahapan proses tertentu, objek yang sudah masuk dalam satu kelompok, pada satu tahapan berikutnya objek akan berpindah ke kelompok lain[2]. K-Means merupakan metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric[33].

K-Means merupakan teknik *clustering* yang diperoleh dari sebuah dataset dengan cara menghitung jarak dari setiap titik ke pusat cluster secara *iterative*.

Proses untuk *clustering* menggunakan algoritma *k-means* adalah sebagai berikut:

1. Tentukan nilai k dalam total *cluster* yang ingin dibentuk.
2. Pilih nilai *centroid cluster* (*centroid*) berawal dari k.
3. Hitung jarak Setiap data input dan setiap centroid menggunakan rumus jarak *Euclidean* untuk menemukan jarak terdekat antara setiap titik data dan centroid.
4. Mengklasifikasikan/mengelompokkan setiap item data berdasarkan jarak (jarak minimum) ke centroid.
5. Perbarui medianya. Nilai pusat baru.

6. Mengulangi langkah ke 3 sampai 5 anggota setiap *cluster* tetap tidak berubah. Berikut ini merupakan rumus terhadap perhitungan algoritma *K-Means Clustering*.

$$de = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

x_i : Koordinat x dari titik pertama.

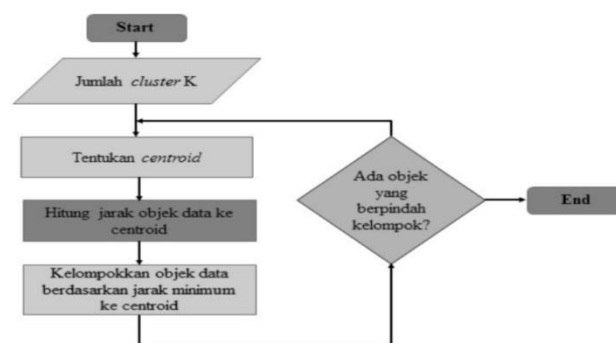
y_i : Koordinat y dari titik pertama.

s_i : Koordinat x dari titik kedua.

t_i : Koordinat y dari titik kedua.

d_e : Hasil jarak Euclidean antara kedua titik tersebut.

Pengelompokan data menggunakan algoritma *k-means clustering* ini dilakukan dengan langkah-langkah seperti pada Gambar 2.



Gambar 2.2 Diagram Alir Proses K-Means Clustering

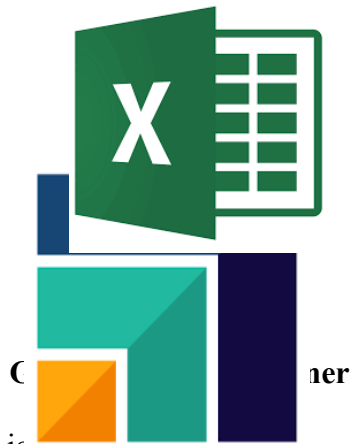
Sumber : Haris Kurniawan, SarjonDefit, Sumijan. [3]

Untuk memulai algoritma *K-Means Clustering* diawali dengan menentukan jumlah cluster, selanjutnya tentukan pusat cluster (centroid), kemudian dilanjutkan dengan menghitung jarak objek data ke centroid dan terakhir lakukan pengelompokan objek data berdasarkan jarak minimum terhadap centroid. Satu kali

siklus proses *K-Means Clustering* disebut dengan 1 kali iterasi. Selanjutnya hitung pusat cluster yang baru, setelah itu lanjutkan dengan menghitung jarak data ke pusat *cluster* baru yang samaprosesnya dengan itersi pertama (iterasi-1) tadi. Jika hasil pengelompokan iterasi ke-2 dan iterasi ke-1 belum sama lanjutkan ke iterasi berikutnya. Namun jika hasil pengelompokan iterasi tersebut sudah sama, maka proses k-means clustering di hentikan dan itulah hasil akhir dari pengelompokan akhir[2].

2.10 RapidMiner

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofman dari *Institute of Teknologi Blanchardstown* dan Ralf Klinkenberg (rapid-i.com). Rapid Miner dirancang mudah untuk digunakan, bahkan tanpa pengalaman pemrograman[34]. Rapid Miner adalah platform perangkat lunak yang kuat untuk ilmu data dan pembelajaran mesin. Ini menyediakan beragam alat untuk persiapan data, pemodelan, evaluasi, dan implementasi. Rapid Miner dirancang untuk mudah digunakan dan memungkinkan pengguna untuk dengan mudah membangun dan menguji berbagai model, bahkan tanpa pengalaman pemrograman. RapidMiner menawarkan antar muka drag-and-drop yang memungkinkan pengguna untuk membangun alur kerja untuk memproses dan menganalisis data. Ini mendukung beragam sumber data, termasuk file datar, basis data, dan platform big data seperti *Hadoop* dan *Spark*. Perangkat lunak ini juga mencakup beragam operator yang sudah dibangun, yang merupakan blok bangunan dari alur kerja, yang mencakup semua tahap proses data mining, seperti pembersihan data, pemilihan fitur, dan pemodelan[15].



Sumber : Webcatalog.ic

2.11 Microsoft Excel

Microsoft office excel adalah *Software* lembar kerja *Sphreadsheet* yang dibuat dan didistribusikan oleh Microsoft Corporation yang dapat dijalankan pada Microsoft Windows dan Mac OS. sejak diluncurkan pertama kali pada tahun 1985, Excel telah menjadi salah satu alat yang paling banyak digunakan untuk mengelola dan menganalisis data dalam berbagai bidang, termasuk keuangan, ilmu pengetahuan, pendidikan, dan bisnis. Excel memungkinkan penggunaanya untuk membuat tabel, menghitung angka, menganalisis data, serta membuat grafik dan diagram. Software ini memiliki banyak fitur yang membuat Microsoft Excel menjadi salah satu pogram komputer yang populer di seluruh dunia. Kelebihan Microsoft Excel diantaranya yaitu, *userfriendly*, tersedia fitur pembuatan grafik data, adanya fitur pemrograman VBA, dan kemampuan unutuk menyimpan data dalam jumlah besar.

Gambar 2.4 Microsoft Excel

Sumber : Wikipedia

2.12 Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini meliputi :

2.12.1 Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan melalui tiga pendekatan utama, yaitu studi pustaka, observasi, dan kuesioner. Studi pustaka dilakukan dengan mengumpulkan informasi dari berbagai sumber literatur, seperti buku, jurnal ilmiah, artikel, dan penelitian terdahulu. Tujuan dari metode ini adalah untuk membangun landasan teori yang kuat dan memahami konsep-konsep yang relevan dengan penelitian. Selanjutnya, metode observasi digunakan untuk mengamati secara langsung fenomena, objek, atau situasi yang menjadi fokus penelitian. Observasi dilakukan secara sistematis dengan mencatat data yang ditemukan selama pengamatan berlangsung. Pendekatan ini memberikan gambaran nyata dan aktual dari fenomena yang diteliti, sehingga menghasilkan data primer yang akurat.

Metode pengumpulan data lainnya adalah kuesioner, yang melibatkan penyebaran daftar pertanyaan kepada responden. Kuesioner dirancang secara terstruktur untuk mendapatkan informasi sesuai dengan tujuan penelitian. Dalam pelaksanaannya, responden mengisi kuesioner dengan menjawab pertanyaan yang

telah disusun, baik dalam bentuk tertutup maupun terbuka. Kuesioner ini memungkinkan pengumpulan data secara efisien dari banyak responden sekaligus.

2.12.2 Studi Pustaka

Studi pustaka dilakukan untuk memperoleh informasi dan teori yang relevan dengan topik penelitian. Peneliti mengumpulkan data sekunder dari buku, jurnal, artikel ilmiah, dan sumber referensi lain yang dapat memperkuat landasan teori. Hal ini bertujuan untuk memahami konsep dan teori yang mendasari penelitian serta memberikan konteks yang lebih luas.

2.12.3 Observasi

Observasi digunakan untuk mengamati secara langsung fenomena atau objek yang menjadi fokus penelitian. Metode ini memungkinkan peneliti untuk memperoleh data primer dari lingkungan atau situasi yang alami tanpa intervensi. Observasi yang dilakukan bersifat sistematis, terencana, dan bertujuan untuk memahami pola-pola tertentu dalam data yang diamati.

2.12.4 Kuesioner

Kuesioner disusun untuk mengumpulkan data secara langsung dari responden. Pertanyaan dalam kuesioner dirancang dengan jelas dan relevan dengan tujuan penelitian, mencakup dua bagian utama:

1. Menggali informasi terkait variabel pertama (misalnya, pola asuh, jenis, dan pengaruhnya).
2. Menganalisis variabel kedua (misalnya, kesejahteraan emosional) dengan indikator yang terukur, seperti tingkat kebahagiaan atau stres.

2.13 Penelitian Terdahulu

Tabel 2.2 Penelitian Terdahulu

Judul	Nama Peneliti	Tahun	Hasil
Analisis K-Means dan Naive Bayes Untuk Pengelompokan Rawan Bencana di Daerah Kabupaten Labuhanbatu	Nadira Jannah Adeni Lubis, Syaiful Zuhri Harahap, Irmayanti Ritonga.	2024	Dari pembahasan diatas Maka didapatkan hasil sebagai berikut: 1. Pada Algoritma K-Means didapatkan hasil Cluster 0 dengan jumlah bencana paling sedikit terdapat di daerah pangkatan, panai hulu, dan panai tengah. 2. Pada algoritma Naïve Bayes didapatkan hasil didapatkan hasil akurasi dari kedua yang sudah diuji data adalah sebesar 87.88%. 3. Dari hasil kedua algoritma diatas didapatkan perbedaan antar kedua algoritma tersebut yaitu algoritma K-Means

			digunakan untuk mengelompokkan bencana sesuai dengan cluster nya masing-masing.
Penerapan Data Mining Untuk Evaluasi Data Penjualan Menggunakan Metode Clustering Dan Algoritma Hirarki Divisive Studi Kasus Toko SembakoPujo.	Ade Eka Febriyanti, Syaiful Zuhri Harahap, Masrizal	2024	Dari analisa dan pengujian yang telah dilakukan, hasil dari pengujian algoritma hirarki divisive a. Cluster 1 terdiri dari 135 item yaitu dengan range pembelian 1-25. Customer yang berada didalam range 1-25. b. Cluster 2 terdiri dari 51 item yaitu dengan range pembelian 27-113. Customer yang berada di dalam range 27-113 c. Cluster 3 terdiri dari 18 item yaitu dengan range pembelian 117-358.

			Customer yang berada di dalam range 117-358. Setelah dilakukan pengujian secara manual menggunakan algoritma hirarki divisive dapat diketahui bahwa customer yang loyal berada di cluster 3 dengan range frekuensi pembelian 117-358.
Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell	Muhammad Rafi Nahjan, Nono Heryana, Apriade Voutama	2023	Dengan hasil penelitian ini maka dapat disimpulkan bahwa transaksi yang masuk dalam klaster 2 adalah transaksi yang diminati yaitu yang berisi transaksi isi pulsa dan pembelian paket data dengan nominal diatas Rp 50.000 dan memberi profit rata-rata Rp.2.729 per transaksi. Adapun hal yang

			peneliti ingin sarankan adalah sesuaikan lagi harga dan profit yang di ambil dalam transaksi yang termasuk klaster 1. dan juga membuka layanan online 24 jam seperti web atau bot sosial media seperti bot whatsapp yang memiliki fitur transaksi otomatis juga dapat membantu menarik minat dari pelanggan.
--	--	--	---