

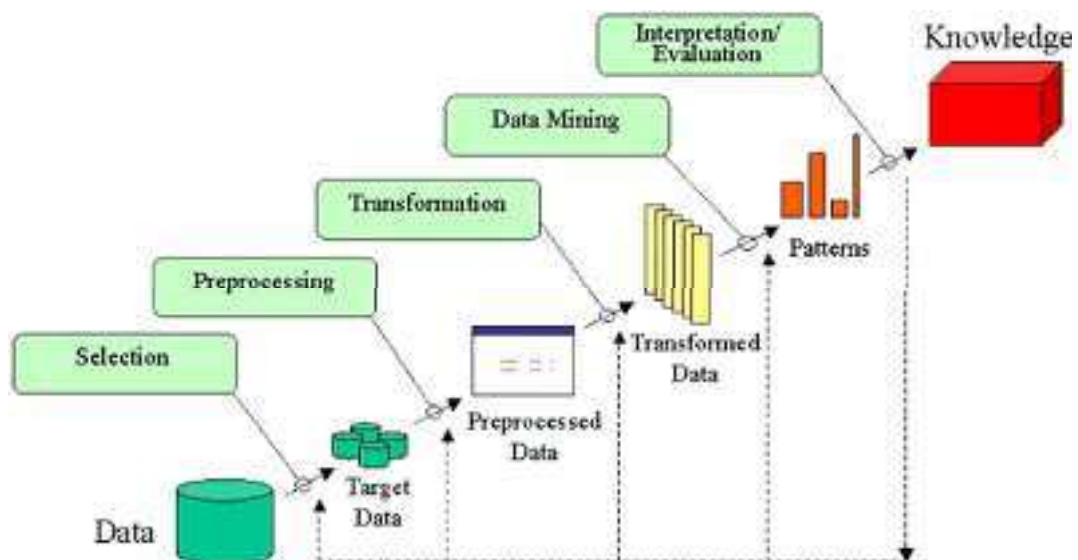
BAB II

TINJAUAN PUSTAKA

2.1. Teori Umum

2.1.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan proses yang terdiri dari pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar *Knowledge Discovery in Database* (KDD) didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data. Proses *Knowlegde Discovery in Database* melibatkan hasil proses data mining (proses pengekstrak kecenderungan suatu pola data) kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami (P. P. Putra & Chan, 2018).



Gambar 2.1 Knowledge Discovery in Database
Sumber: (Data Penelitian, 2023)

Berikut Langkah proses data mining pada *knowledge discovery in database* (KDD) adalah sebagai berikut (R. R. Putra & Wadisman, 2018):

1. Penyeleksian Data

Tidak semua data dalam database dilakukan untuk proses data mining, sehingga diperlukan proses seleksi yang sesuai dengan analisis yang dilakukan dan hasil pemilihan kemudian dipakai cara selanjutnya.

2. *Preprocessing/Cleaning*

Saat sebelum proses data mining bisa dicoba, butuh dilakukan cara membersihkan kepada informasi yang jadi fokus KDD. Cara membersihkan mencakup penghapusan data duplikasi, pengecekan data yang tidak konsisten, serta pemeriksaan kekeliruan pada data tersebut, semacam kekeliruan cap (tipografi).

3. Transformasi Data

Transformasi data adalah dilakukan dengan memilih data yang cocok buat cara penambangan. Cara pengkodean KDD merupakan cara inovatif serta sungguh terkait pada tipe ataupun bentuk informasi yang didapat dari database.

4. Data mining

Data mining merupakan sistem pencarian pola ataupun data yang menarik dari informasi tertentu dengan memakai metode ataupun tata cara khusus. Metode, tata cara ataupun algoritma penambangan informasi bermacam - macam. Memilih tata cara ataupun algoritma yang pas sungguh tergantung pada tujuan serta totalitas sistem KDD.

5. Interpretasi dan Evaluasi

Interpretasi dan evaluasi merupakan bentuk data yang diperoleh dari metode data mining yang mesti disajikan dalam wujud yang gampang dimengerti oleh pihak yang bersangkutan. Tahap ini menggambarkan bagian dari sistem KDD yang disebut rendering. Pada langkah ini kita memeriksa apakah pola ataupun data yang ditemui sebaliknya dengan kebenaran ataupun dugaan yang terdapat pada fakta.

2.1.2. Data Mining

Menurut (Nabila et al., 2021) data mining merupakan teknik bantu pengambilan keputusan pada penulis mencari pola dalam data. Ini dapat dilakukan secara manual oleh pengguna (melalui kueri, misalnya) atau secara otomatis oleh perangkat lunak yang menanyakan basis data untuk mencari model data. Penemuan mengacu pada proses mencari informasi baru. Pemilihan tugas data mining, pemilihan goal dari proses KDD misalnya klasifikasi, regresi, *clustering*, dll.

Pemilihan algoritma data mining untuk pengelompokkan (*clustering*). Proses data mining yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan (P. P. Putra & Chan, 2018).

2.1.3. Clustering

Menurut (Aditya et al., 2020) *Clustering* data mining adalah nama lain untuk

metode *segmentation*. Tujuan dari segmentasi pada metode data mining adalah mengelompokkan suatu class ke dalam beberapa segmen berdasarkan atribut yang ditentukan. Penentuan atribut harus sesuai kesamaan yang dimiliki beberapa *class* tadi. *K-means* merupakan algoritma *clustering*. *K-means Clustering* adalah salah satu “*unsupervised machine learning algorithms*” yang paling sederhana dan populer. *K-Means Clustering* adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistemn partisi. *Cluster* adalah suatu kumpulan dari entitas yang hampir sama

Clustering adalah metode penganalisaan data yang sering dimasukkan sebagai salah satu metode data mining yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Ada beberapa pendekatan yang digunakan dalam mengembangkan metode *clustering*. Dua pendekatan utama adalah *clustering* dengan pendekatan partisi dan *clustering* dengan pendekatan hirarki. *Clustering* dengan pendekatan partisi atau sering disebut dengan *partition-based clustering* mengelompokkan data dengan memilah-milah data yang dianalisa ke dalam *cluster-cluster* yang ada (Nabila et al., 2021).

Clustering dengan pendekatan hirarki atau sering disebut dengan *hierarchical clustering* mengelompokkan data dengan membuat suatu hirarki berupa dendogram dimana data yang mirip akan ditempatkan pada hirarki yang berdekatan dan yang tidak pada hirarki yang berjauhan. Di samping kedua pendekatan tersebut, ada juga *clustering* dengan pendekatan *automatic mapping*.

2.1.4. *K-Means*

Menurut (Nabila et al., 2021) metode *K-means* adalah metode pengelompokan data dengan mengambil parameter sejumlah k *cluster* dan mempartisi data kedalam *cluster* tersebut, dengan berpatokan pada kemiripan antara data dalam suatu *cluster* dan ketidakmiripan di antara *cluster* berbeda, pusat dari *cluster* adalah rata-rata dari nilai anggota *cluster* yang disebut dengan *centroid* atau *center of gravity*. Selain itu *K-means* melakukan pengelompokan dengan meminimalkan jumlah kuadrat dari jarak (*distance*) antara data dengan *centroid cluster* yang cocok. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

Adapun dalam melakukan perhitungan untuk mengetahui jarak data terhadap pusat data adalah dengan menggunakan rumus *Euclidean* yaitu:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Rumus 2.1 Perhitungan *Euclidean*
Sumber: (Data Penelitian, 2023)

Keterangan:

De = *Euclidean Distance*

I = Banyaknya objek

(x, y) = Koordinat objek

(s, t) = Koordinat centroid

Menurut (Aditya et al., 2020) Algoritma *k-means* dimulai dengan pembentukan prototipe *cluster* di awal kemudian secara iteratif prototipe *cluster* ini diperbaiki hingga konvergen (tidak terjadi perubahan yang signifikan pada prototipe *cluster*). Perubahan ini diukur menggunakan fungsi objektif yang

umumnya didefinisikan sebagai jumlah atau rata-rata jarak tiap item data dengan pusat massa kelompoknya.

2.1.5. RapidMiner



Gambar 2.2 Rapidminer
Sumber: (Data Penelitian, 2023)

Penerapan teknik data mining dengan algoritma *K-Means* pada penelitian ini yaitu untuk menentukan pola pelaksanaan tilang dalam proses razia yang membutuhkan software pendukung data mining yang bersifat *open source* yaitu *RapidMiner*. *RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah sebuah solusi untuk melakukan analisis terhadap data mining, *text mining* dan analisis prediksi. *RapidMiner* menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator data mining termasuk operator untuk input, output, data preprocessing dan visualisasi (Nabila et al., 2021).

RapidMiner merupakan perangkat lunak yang bersifat sumber terbuka (*Open Source*). *RapidMiner* adalah teknologi untuk memeriksa penambangan data, penambangan teks dan analitik prediktif. Sebagai alat analisis data dan mesin

penambangan data, *RapidMiner* dapat digunakan sendiri atau dimasukkan ke dalam produk yang sudah ada. Karena dikembangkan di Java, *RapidMiner* dapat digunakan dengan sistem komputer apa pun. Pengguna bisa mendapatkan wawasan berharga dari berbagai metode deskriptif dan prediktif *RapidMiner* yang memungkinkan mereka untuk membuat pilihan yang lebih tepat. *RapidMiner* sebelumnya bernama YALE (*Yet Another Learning Environment*) dimana versi awalnya mulai di kembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*. *RapidMiner* distribusikan di bawah lisensi AGPL (*GNU Affero General Public License*) versi tiga. Hingga saat ini telah ribuan aplikasi yang dikembangkan menggunakan *RapidMiner* di lebih dari 40 negara.

Menurut (Zaki Muhammad et al., 2018) *RapidMiner* adalah platform perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, yang menyediakan lingkungan terpadu untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*) dan analisis prediktif (*predictive analytics*).

Rapidminer memiliki beberapa sifat sebagai berikut :

1. Ditulis dengan bahasa pemrograman Java sehingga dapat dijalankan di berbagai sistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator *trees*.
3. Representasi XML internal untuk memastikan format standar pertukaran data.

4. Bahasa *scripting* memungkinkan untuk eksperimen skala besar dan otomatisasi eksperimen.
5. Konsep *multi-layer* untuk menjamin tampilan data yang efisien dan menjamin penanganan data.
6. Memiliki GUI, *command line mode* dan Java API yang dapat dipanggil dari program lain.

Beberapa Fitur dari Rapidminer, antara lain:

1. Banyaknya algoritma data mining, seperti *decision tree* dan *self-organization map*.
2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, *tree chart* dan *3D Scatter plots*.
3. Banyaknya variasi *plugin*, seperti *text plugin* untuk melakukan analisis teks.
4. Menyediakan prosedur data mining dan *machine learning* termasuk ETL (*extraction, transformation, loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi.
5. Proses data mining tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML dan dibuat dengan GUI.
6. Mengintegrasikan proyek data mining Weka dan statistika R.

2.2. Teori Khusus

2.2.1. Pengertian Lalu Lintas

Lalu lintas diartikan sebagai gerak bolak-balik manusia atau barang dari suatu tempat ke tempat lainnya dengan menggunakan sarana jalan (Rahayu Nurfauziah & Hetty Krisnan, 2021). Menurut Pasal 1 Undang-Undang Nomor 22 Tahun 2009

Tentang Lalu Lintas dan Angkutan Jalan mendefinisikan bahwa lalu lintas adalah gerak kendaraan dan orang di ruang lalu lintas jalan. Lalu lintas menurut Poerwodarminto (1993:55) yaitu:

1. Perjalanan bolak-balik
2. Perihal perjalanan di jalan dan sebagainya
3. Berhubungan antara sebuah tempat.

Definisi-definisi tersebut dapat diartikan bahwa lalu lintas adalah segala sesuatu hal yang berhubungan langsung dengan sarana jalan yang menjadi sarana utamanya untuk dapat mencapai satu tujuan yang dituju baik disertai maupun tidak disertai oleh alat angkut. Jadi, di dalam lalu lintas ada tiga komponen penyusunnya yaitu manusia, kendaraan, dan jalan yang saling berinteraksi dalam proses pergerakan.



Gambar 2.3 Sistem Komponen Lalu Lintas
Sumber: (Data Penelitian, 2023)

a. Manusia

Dalam komponen lalu lintas manusia berperan sebagai pengendara atau penumpang atau pejalan kaki dan mempunyai keadaan yang berbeda beda.

b. Kendaraan

Dalam komponen lalu lintas kendaraan merupakan suatu sarana angkut penumpang maupun barang yang terdiri atas kendaraan bermotor dan kendaraan tidak bermotor. Dalam Undang - Undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, jenis kendaraan bermotor dibagi menjadi beberapa jenis yaitu sepeda motor, mobil penumpang, mobil bus mobil barang dan mobil khusus.

c. Jalan

Dalam komponen lalu lintas jalan merupakan lintasan yang direncanakan dan digunakan kepada pengguna kendaraan bermotor maupun tidak bermotor, jalan juga digunakan untuk mengalirkan aliran lalu lintas dengan lancar, mendukung beban muatan kendaraan.

2.2.2. Sepeda Motor

Sepeda Motor atau angkutan bermotor menurut Nasution (1996) adalah roda dua transportasi yang menggunakan kendaraan bermotor sebagai fasilitasnya yang bergerak di jalan raya. Salah satu angkutan bermotor itu adalah sepeda motor, sepeda motor pertama kali dirancang oleh *Ernest Michaud* pada Tahun 1868 dengan menggunakan mesin berjenis uap. Tapi pada saat itu proyek sepeda motor dengan mesin uap gagal, lalu pada Tahun 1885 *Edward Butler* memperbaiki kendaraan tersebut dengan menggunakan mesin berjenis mesin pembakaran dalam dan pada tahun tersebut juga *Gottlieb Daimler* dan *Wilhelm Maybach* sebagai ahli mesin di Jerman menjadi seorang perakit motor pertama kali di dunia. Sejak saat itu lah banyak penemuan lainnya dalam perkembangan jenis sepeda motor ini.

Selain itu sepeda motor juga dapat membantu masyarakat untuk melakukan aktivitas dan dapat pula menghemat perjalanan untuk sampai tujuan. Selain menghemat perjalanan sepeda motor juga dapat menghindari macet, mempermudah aktivitas, menghemat waktu dan dapat meminimalisir pengeluaran bahan bakar minyak (Vivin Tannia & Nyoman Yulianthini, 2021).

2.2.3. Pelanggaran Lalu Lintas

Pelanggaran Lalu Lintas merupakan perbuatan melanggar sesuatu dan berhubungan dengan hukum. Pelanggaran lalu lintas itu adalah satu perbuatan atau tindakan pengendara yang bertentangan dengan ketentuan dan peraturan Undang-Undang Lalu Lintas. Jika ketentuan tersebut dilanggar oleh pengendara maka kecelakaan dalam berkendara kemungkinan dapat terjadi. Jadi dari definisi pelanggaran lalu lintas diatas dapat diketahui bahwa pelanggaran lalu lintas itu adalah suatu perbuatan yang bertentangan dengan hukum dan akan menimbulkan akibat dari perbuatan itu (Nurfauziah, 2021).

2.2.4. Faktor Pelanggaran Lalu Lintas

Menurut (Suryandari et al., 2022) faktor Pelanggaran lalu lintas saat ini terdapat banyak faktor yang mempengaruhi terjadinya pelanggaran lalu lintas. Faktor-faktor tersebut antara lain yaitu:

a. Faktor manusia

Menurut Suwardjoko mengatakan bahwa hampir semua bentuk pelanggaran lalu lintas dan kecelakaan lalu lintas yang ada disebabkan oleh pengendara. Hal ini dipertegas juga oleh Hoobs mengatakan bahwa manusia adalah

penyebab paling banyak dalam pelanggaran dan kecelakaan yang ada. Faktor manusia ini mencakup psikologi dan sistem indra seperti penglihatan dan pendengaran dan pengetahuan akan tata cara lalu lintas. Ada beberapa indikator yang membentuk sikap dan perilaku manusia dalam berkendara, diantaranya adalah:

1) Pengetahuan

Pemerintah telah membuat peraturan lalu lintas yang ditujukan kepada setiap pengguna jalan demi menciptakan keamanan, keselamatan, ketertiban dan kelancaran lalu lintas dalam bentuk Undang-Undang, Perpu, Perda, dan aturan lainnya. Oleh karena itu setiap pengguna jalan wajib memahami dan menjalani setiap aturan yang telah dibuat sehingga terbentuk satu persepsi dalam pola pikir dan tindakan dalam berinteraksi di jalan raya. Pola pikir yang terbentuk dapat terjadinya perbedaan tingkat pemahaman dan pengetahuan antara pengguna jalannya terhadap peraturan yang ada sehingga berpotensi munculnya masalah dalam berlalu lintas, baik antara pengguna jalan sendiri maupun antara pengguna jalan dengan aparat yang bertugas. Selain pemahaman terhadap peraturan yang ada, pengendara juga harus memiliki pemahaman tentang karakteristik kendaraannya. Setiap kendaraan memiliki karakteristik yang berbeda-beda. Pemahaman terhadap karakteristik kendaraan sangat berpengaruh terhadap operasional kendaraan di jalan raya yang berpengaruh juga terhadap situasi lalu lintas.

2) Mental

Salah satu yang menjadi faktor utama terhadap situasi lalu lintas adalah

mental pengendara tersebut. Untuk menciptakan sebuah interaksi dengan hasil seperti keamanan, keselamatan, kelancaran lalu lintas pengendara harus bisa menjaga etika, sopan-santun, toleransi antar pengguna jalan dan mengendalikan emosi. Jika pengendara tidak bisa menjaga hal itu maka dampak negatif yang diperoleh seperti menimbulkan kemacetan, pelanggaran lalu lintas dan bahkan dapat menyebabkan kecelakaan lalu lintas.

3) Keterampilan

Demi menciptakan keamanan, keselamatan, ketertiban, dan kelancaran lalu lintas pengendara harus memiliki keterampilan dalam mengendalikan kendaraannya, karena hal ini akan berpengaruh besar terhadap situasi lalu lintas yang ada. Setiap pengendara harus memiliki lisensi terhadap kemampuan dalam mengendalikan kendaraan yang diwujudkan secara formal melalui Surat Izin Mengemudi (SIM). Di antara faktor-faktor yang ada, faktor manusia menjadi penyebab yang paling tinggi dalam pelanggaran lalu lintas karena untuk faktor manusia berkaitan erat dengan tingkah laku, etika dan tata cara berkendara di jalan.

2.2.5. Dampak Pelanggaran Lalu Lintas

Dampak yang akan terjadi akibat pelanggaran-pelanggaran lalu lintas yang ada antara lain seperti:

1. Terjadinya peningkatan angka kecelakaan di jalan baik pada perempatan maupun tidak.
2. Rawan terjadi kecelakaan tunggal maupun beruntun.
3. Dapat merugikan diri sendiri maupun orang lain.

4. Tindakan melanggar rambu lalu lintas dapat membuat meningkatnya angka kecelakaan.
5. Kurangnya kesadaran pengendara dalam mematuhi peraturan yang ada dapat membuat kemacetan semakin parah.
6. Tindakan melanggar peraturan lalu lintas akan menciptakan suatu kebiasaan melanggar lalu lintas yang terjadi secara terus menerus.

2.3. Penelitian Terdahulu

Berikut ini adalah beberapa penelitian terdahulu sebagai acuan dalam penelitian ini:

1. Berdasarkan jurnal penelitian (Joseph & Thanakumar, 2019) dengan judul *“Survey Of Data mining Algorithms For Intelligent Computing System”*. *The Intelligent computing system, described to be a collection of the connected device working in mutual understanding to attain a particular purpose, is an incorporation of artificial intelligence and the computational intelligence, and are employed in variety of applications. The paper presents the survey on the data mining algorithms and the techniques that could be employed with the intelligent computing system, presenting a basic conception of the data mining along with the prominent algorithms of the data mining and the classification of its techniques, further the survey concludes with the challenges included in the overview of the survey done along with the future enhancement in the research that analyses the data mining techniques in the intelligent computing applications.*
2. Berdasarkan jurnal penelitian (Jalal & Ali, 2021) dengan judul *“Text*

documents clustering using data mining techniques". Increasing progress in numerous research fields and information technologies, led to an increase in the publication of research papers. Therefore, researchers take a lot of time to find interesting research papers that are close to their field of specialization. Consequently, in this paper we have proposed documents classification approach that can cluster the text documents of research papers into the meaningful categories in which contain a similar scientific field. Our presented approach based on essential focus and scopes of the target categories, where each of these categories includes many topics. Accordingly, we extract word tokens from these topics that relate to a specific category, separately. The frequency of word tokens in documents impacts on weight of document that calculated by using a numerical statistic of term frequency-inverse document frequency (TF-IDF). The proposed approach uses title, abstract, and keywords of the paper, in addition to the categories topics to perform the classification process. Subsequently, documents are classified and clustered into the primary categories based on the highest measure of cosine similarity between category weight and documents weights.

3. Berdasarkan jurnal penelitian (Simanjuntak et al., 2021) dengan judul "*Data mining Rekomendasi Pemakaian Skincare*". Pada proses pengolahan data dengan menggunakan algoritma *naïve bayes* yang merupakan salah satu metode klasifikasi *data mining* untuk menentukan pola pembelian konsumen terhadap produk yang ditawarkan oleh Toko *Skincare* dengan menggunakan data *training* dari transaksi selama ini dengan mengambil 4 variabel yang

sangat mempengaruhi di mana untuk membuat penjualan mereka laris di pasaran harus memperhatikan produk, jenis kulit, manfaat dan harga. Penerapan teknik data ini sangat efisien dan efektif untuk memprediksi bisnis ke depannya hal ini dapat dilihat dari hasil probabilitas dari penjualan mereka selama ini terhadap pembelian memiliki 72 % prediksi yang benar dan 28% prediksi yang salah.

4. Berdasarkan jurnal penelitian (Indraputra et al., 2020) dengan judul “*K-Means Clustering Data COVID-19*”. Tujuan dari penelitian ini adalah untuk memperoleh dan mengolah data *COVID-19* yang terdapat pada *Kaggle* menggunakan metode *Data mining* yaitu *K-Means Clustering* Untuk *K-Means Clustering* pada penelitian ini akan digunakan tiga metode untuk mengolah data yaitu pengolahan menggunakan *software. Microsoft Excel* dan *software Data mining* yaitu *Weka* dan *KNIME*. Dari hasil pengolahan data, diperoleh dua *cluster* data, dimana *cluster 2* memiliki jumlah terjangkit dan meninggal yang lebih tinggi dibandingkan dengan *cluster 1*, maka daerah - daerah *cluster* tersebut perlu diprioritaskan penanganannya. Sampel data ini dapat dibagi menjadi 2 *cluster*, dimana pada *cluster 1* terdapat 29 data dan pada *cluster 2* terdapat 19 data. Dengan informasi ini, dapat diketahui *cluster* daerah-daerah yang memerlukan penanganan COVID-19 yang lebih darurat, di mana pada sampel ini merupakan *cluster 2* dengan jumlah terjangkit dan meninggal yang lebih tinggi dibandingkan dengan *cluster 1*.
5. Berdasarkan jurnal penelitian (Daeli et al., 2022) dengan judul “Implementasi *Data mining* Untuk Mengelompokkan Lokasi Berdasarkan Tingkat

Kejahatan Pada Kabupaten Nias Barat Menggunakan Metode *K-Means Clustering*". Kabupaten Nias Barat terdiri dari 115 lokasi atau desa yang memiliki tingkat kejahatan berbeda-beda. Pihak kepolisian pada Kabupaten Nias Barat akan membagi tugas untuk menjaga setiap lokasi di mana jumlah petugas pada setiap desa disesuaikan dengan tingkat kejahatan pada lokasi tersebut. Sulitnya untuk mengetahui atau mengelompokkan lokasi berdasarkan tingkat kejahatan pada Kabupaten Nias Barat menimbulkan permasalahan yang mempersulit pembagian tugas oleh pihak kepolisian untuk menjaga keamanan setiap lokasi. Atas dasar masalah tersebut, maka dibutuhkan sebuah aplikasi *data mining* yang mampu mengelompokkan lokasi berdasarkan tingkat kejahatan pada Kabupaten Nias Barat. Di mana setiap data yang diperoleh dari kepolisian Kabupaten Nias Barat akan dihitung menggunakan salah satu metode *data mining* yang mampu dalam mengelompokkan lokasi kejahatan yaitu metode *K-means Clustering*. Hasil dari penelitian adalah sebuah aplikasi *data mining* yang mengadopsi metode *K-Means Clustering* yang mampu menjawab permasalahan yang ada pada Kabupaten Nias Barat terkait pengelompokan lokasi berdasarkan tingkat kejahatan.

2.4. Kerangka Pemikiran

Kerangka pemikiran adalah alur utama penelitian, dalam urutan masalah, peluang, pendekatan, identifikasi dan pemetaan, pemodelan, evaluasi dan hasil. Pada dasarnya, penelitian adalah kerangka hubungan antar konsep yang ingin diamati atau diukur melalui penelitian yang ingin dilakukan. Karena standar

kualitas pendaftar berada dalam kisaran yang cukup luas, studi ini tidak konsisten berdasarkan jurusan yang dapat diterima (Annisa Ekasetya & Jananto, 2020).



Gambar 2.4 Kerangka Pemikiran
Sumber: (Data Penelitian, 2023)

Kerangka Pemikiran merupakan tahapan sebelum pelaksanaan penelitian.

Berikut langkah-langkah kerangka pemikiran yang akan dilaksanakan, yaitu :

1. Melaksanakan studi literatur dengan cara membaca jurnal untuk memahami mengenai data mining dan algoritma *K-Means*.
2. Melakukan observasi, wawancara dan pengumpulan data yang akan diperlukan.
3. Mengemukakan masalah serta Batasan masalah yang akan diteliti.
4. Melakukan pengolahan data tilang kendaraan roda dua untuk mempermudah

clustering.

5. Menerapkan metode *K-Means* untuk mendapatkan hasil *cluster* dan melakukan pengujian dengan software *RapidMiner*.
6. Menyimpulkan hasil akhir dari pengolahan data yang telah dilakukan.