

BAB I

PENDAHULUAN

1.1. Latar Belakang

Knowledge Discovery Database (KDD) merupakan kegiatan yang melibatkan penggalian informasi dan kontribusi yang berguna untuk mengeksplorasi pengetahuan melalui kumpulan data yang besar (Palacios, 2021). *Knowledge Discovery from Data* (KDD) memungkinkan kita untuk menganalisis kumpulan data besar untuk memecahkan masalah. *Data mining* adalah proses menemukan pola data yang berguna dan pengetahuan yang dibutuhkan KDD (Abdulah & Fang, 2021).

Data Mining digunakan untuk menemukan pola hubungan dalam kumpulan data berjumlah besar menggunakan komposisi berbagai metode *machine-learning*, manipulasi database dan statistik. Proses ini melibatkan komputer dan disiplin statistika menjadi satu bidang yang digunakan untuk mengekstraksi pola data dari dataset (Thakkar dkk, 2021).

Clustering atau klusterisasi merupakan salah satu metode analisis data mining yang digunakan di berbagai bidang (Harahap dkk, 2021). *Clustering* adalah alat yang ampuh dalam analisis data. Digunakan untuk menemukan struktur *cluster* dalam kumpulan data dengan kesamaan terbesar dalam *cluster* yang sama (Sinaga dkk, 2021).

Telah banyak penelitian sebelumnya membahas metode *clustering*. Salah satunya adalah dengan menggunakan algoritma *K-Means*. *K-Means* adalah metode *clustering* data *non-hierarchical* yang dapat membagi data menjadi dua atau lebih kelompok (Abdullah, 2020). Metode *K-Means* merupakan metode analisa data yang digunakan untuk mengelompokkan data yang sudah ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama dengan lainnya

dan data yang ada di dalam kelompok lain mempunyai karakteristik yang berbeda (Suryani dkk, 2021), jadi *K-Means clustering* dapat mengelompokkan data menjadi beberapa kelompok sesuai dengan karakteristik datanya (Rahmadani dkk, 2021). Algoritma ini dimulai dengan nilai pusat (*centroid*) *cluster* yang dipilih secara acak. Setiap titik data dialokasikan ke pusat terdekat dari kelompok berdasarkan kesamaan. Kesamaan titik data dihitung dengan menggunakan nilai jarak (Revathi, 2021).

Seperti pada penelitian yang dilakukan oleh (Said dkk, 2021) menggunakan metode *K-Means* dalam klasterisasi dana bantuan pada Program Keluarga Harapan. Penelitian ini bertujuan untuk mempermudah tim penyeleksi memberikan bantuan sesuai dengan kriteria yang telah ditentukan apakah layak atau tidak layak menerima bantuan tersebut. Data yang digunakan merupakan data penyaluran dana bantuan pada 10 (sepuluh) Desa yang didapat dari Dinas Sosial Kabupaten Indragiri Hilir dan setelah diproses dengan menggunakan metode *K-Means* maka didapatkan hasil 3 *cluster* yaitu *Cluster 1* (C1) Rumah Tangga Hampir Miskin (RTHM), *Cluster 2* (C2) Rumah Tangga Miskin (RTM), *Cluster 3* (C3) Rumah Tangga Sangat Miskin (RTSM), Sehingga penelitian ini sangat membantu dalam merelokasikan bantuan yang tepat sasaran sesuai klaster keluarga harapan.

Selanjutnya dalam bidang Pendidikan (Elda dkk, 2021) menggunakan metode *K-Means* untuk mengklasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas. Penelitian ini bertujuan untuk melakukan pengelompokan siswa untuk mendapatkan komposisi kelas yang seimbang guna meningkatkan mutu dan hasil belajar siswa yang dilihat dari meningkatnya nilai rata-rata kelas. Data yang diolah pada penelitian ini bersumber dari data pokok sekolah sebanyak 90 siswa Kompetensi Keahlian Teknik Komputer Jaringan kelas XI SMKN Negeri 2 Padang Panjang tahun pelajaran 2020/2021. Perhitungan klasterisasi siswa dengan menggunakan metode *K-Means* ini berhasil mengelompokkan 90 siswa menjadi 3 klaster dimana cluster 1 berjumlah 47 siswa, cluster 2 berjumlah 10 siswa dan cluster 3 berjumlah 33 siswa. Masing-masing anggota klaster akan dibagi merata kedalam 3 kelompok belajar untuk mendapatkan komposisi kelas yang seimbang. Riset ini bisa dijadikan dasar pengambilan keputusan oleh sekolah dalam melakukan klasterisasi penempatan siswa untuk meningkatkan hasil belajar.

Metode *K-Means* juga dapat diterapkan di bidang Kesehatan yaitu digunakan untuk mengklasterisasi Virus Covid-19 di Wilayah Kabupaten Lamongan sehingga

didapatkan zonasi atau wilayah cluster persebaran virus Covid-19 yang terus meningkat di Wilayah Kabupaten Lamongan dengan aplikasi berbasis website. Data yang digunakan adalah data kasus Covid-19 tahun 2020 di Kabupaten Lamongan dan jumlah data yang digunakan sebanyak 27 Kecamatan. Setelah diproses dengan metode *K-Means* diperoleh klaster dengan 3 zona yaitu zona merah, kuning, dan hijau, zona merah diartikan sebagai daerah bahaya dengan jumlah terjangkit virus covid-19 cukup banyak, sedangkan pada zona kuning menandakan zona itu masuk di daerah zona siaga, dimana daerah tersebut termasuk penyebaran virus covid-19 dengan jumlah yang sama tidak banyak dan tidak sedikit. Yang ketiga adalah zona hijau yang berarti zona aman, maksudnya di zona ini adalah daerah yang terjangkit virus covid-19 datanya paling sedikit terjangkit. Tujuan dilakukannya penelitian ini adalah mengklaster persebaran Virus Corona di Wilayah Kabupaten Lamongan berdasarkan parameter jumlah ODP, PDP, kasus Positif, pasien sembuh dan pasien meninggal, penelitian ini berperan penting dalam mengklasifikasi zona yang tersebar di tingkat Kecamatan serta dapat mempermudah Gugus Tugas Percepatan Penanganan Covid-19 dalam menangani dan menanggulangi penyebaran Covid-19 berdasarkan tingkat cluster (Susilo, 2021).

Metode *K-Means* juga dapat diterapkan di bidang pertanian yaitu melakukan klasterisasi hasil panen jagung, diperlukan pengelompokan daerah penghasil jagung potensial untuk mengetahui daerah mana yang menghasilkan jagung dalam jumlah besar atau kecil. Pembagian hasil panen biasanya dilakukan berdasarkan nama Kecamatan penghasil jagung. Data yang digunakan yaitu hasil panen jagung tahun 2018 dan 2019 pada 17 (tujuh belas) Kecamatan di Kabupaten Lampung Selatan. Setelah diproses dengan metode *K-Means* diperoleh 2 (dua) *cluster* yaitu wilayah dengan jumlah panen jagung terbanyak dan wilayah dengan jumlah panen jagung terendah dengan hasil wilayah dengan jumlah panen jagung terbanyak adalah Penengahan dengan 79448.30257 nilai *centroid* 2 dan wilayah dengan jumlah panen jagung terendah adalah Candipuro dengan sentroid 1424.036868 nilai *centroid* 2 (Aldino dkk, 2021).

Metode *K-Means* telah digunakan untuk mengklasterisasi kualitas udara di Uttarakhand-India selama periode *lockdown* pandemi Covid-19 (Sunori dkk, 2021). Masa *lockdown* pandemi Covid-19 telah mengakibatkan penurunan drastis polusi udara, hal ini menyebabkan komponen udara yang menunjukkan perbaikan indeks

kualitas udara. Penelitian ini menggunakan data kualitas udara dari dua tahun yaitu kualitas udara tahun 2019 (periode normal) dan kualitas udara tahun 2020 (periode *lockdown* Covid-19). Pengelompokan dibagi menjadi dua *cluster* yaitu udara tercemar dan udara bersih menggunakan algoritma *K-Means clustering* di Matlab. Pengujian dilakukan dengan dua metode yang berbeda yaitu yang pertama jarak *Euclidean* dan metode kedua adalah jarak *cosinus* kemudian hasilnya dibandingkan. Hasil pengujian menunjukkan bahwa untuk set data yang diberikan pendekatan jarak *Euclidian* memberikan hasil *clustering* yang jauh lebih baik daripada *cosinus*.

Penelitian selanjutnya menggunakan metode *K-Means* dalam proses pengolahan citra Pap *smear*. Tujuan penelitian ini adalah melakukan segmentasi, ekstraksi fitur dan klasifikasi pada citra Pap *smear*. Data yang digunakan pada penelitian ini yaitu 180 citra Pap *smear* RepoMedUNM. Metode yang digunakan untuk melakukan identifikasi citra Pap *smear* diawali dengan *preprocessing*, yaitu mengubah warna pada citra menjadi warna L^*a^*b . Segmentasi menggunakan metode *K-means*, ekstraksi 6 fitur yaitu *metric*, *eccentricity*, *contras*, *correlation*, *energy* dan *homogeneity* dan selanjutnya diidentifikasi dengan menghitung jarak terdekat antara fitur data latih dengan fitur data uji dengan persamaan *euclidean distance*. Hasil identifikasi citra Pap *smear ThinPrep* pada 3 kelas mencapai rata-rata akurasi 93,33%, sedangkan citra Pap *smear Non-ThinPrep* pada 2 kelas mencapai rata-rata akurasi 90% dan rata-rata akurasi keseluruhan citra pada 4 kelas mencapai 92%. Hasil ini menunjukkan bahwa metode usulan dapat mengidentifikasi citra Pap *smear* dengan baik (Riana dkk, 2022).

Algoritma *K-Means clustering* juga digunakan untuk pemetaan kerusakan jalan di Kabupaten Malang yaitu dengan mengelompokkan tingkat kerusakan jalan di Kecamatan Lawang, Kecamatan Singosari, Kecamatan Karangploso, dan Kecamatan Dau. Data yang digunakan pada penelitian ini yaitu 310 data kerusakan jalan di Kabupaten Malang. Pengelompokan dibagi menjadi tiga *cluster* yaitu C1 (Ringan), C2 (Sedang) dan C3 (Berat). Kemudian setelah diproses dengan metode *K-Means* didapatkan jumlah C1 (Ringan) sebanyak 221 data, C2 (Sedang) sebanyak 24, dan C3 (Berat) sebanyak 65 data, sehingga pemerintah mengetahui tingkatan rusak berat dapat ditangani terlebih dahulu, penelitian ini juga menampilkan pemetaan kerusakan jalan dengan Sistem Informasi Geografis (Suryani dkk, 2021).

Berdasarkan Peraturan Bupati Kerinci Nomor 50 Tahun 2018 tentang Petunjuk Pelaksanaan Pemungutan Retribusi Pengendalian Menara Telekomunikasi di Kabupaten Kerinci, disebutkan pada pasal 1 bahwa instansi pemungut adalah Bagian Komunikasi dan Informatika Sekretariat Daerah Kerinci yang saat ini telah berubah menjadi Dinas Komunikasi dan Informatika Kabupaten Kerinci, dalam rangka pemungutan retribusi pengendalian menara telekomunikasi maka Dinas Komunikasi dan Informatika Kabupaten Kerinci melakukan pengawasan, dan pengendalian menara telekomunikasi ke seluruh lokasi Menara Telekomunikasi atau disebut juga dengan *Base Transceiver Station* (BTS) dengan jumlah kunjungan per BTS 2 (dua) kali setahun dan kapasitas kunjungan 1 (satu) BTS per hari per tim, dengan 1 (satu) tim terdiri dari 3 (tiga) orang. Permasalahan yang dihadapi oleh Tim Pengawas dalam melakukan pengendalian dalam pengawasan ke lokasi BTS yaitu mengalami kendala dalam perencanaan pengawasan BTS agar dalam penganggaran, penjadwalan dan pembagian jumlah tim menjadi lebih tepat dan efisien. Perencanaan ini terkait dengan jumlah Tim Pengawas yang turun berdasarkan kedekatan jarak antar BTS dengan patokan jarak dari Kantor Diskominfo Kerinci, untuk itu diperlukan metode pengelompokan atau klasterisasi agar didapatkan kluster BTS berdasarkan titik koordinat yang berdekatan untuk mempermudah Tim Pengawas melakukan pengawasan. Data yang akan digunakan pada penelitian ini yaitu data BTS tahun 2021 yang bersumber dari Dinas Komunikasi dan Informatika Kabupaten Kerinci.

Berdasarkan permasalahan yang telah dijelaskan maka penelitian ini nantinya akan menghasilkan informasi mengenai pengelompokan BTS berdasarkan jarak yang selanjutnya akan dibuat visualisasi hasil *cluster* BTS berdasarkan *Google Map*. Berdasarkan latar belakang tersebut, maka penulis melakukan penelitian dengan judul **“Metode K-Means dalam Visualisasi Berbasis *Google Map* terhadap Klasterisasi Koordinat BTS (*Base Transceiver Station*)”**, sehingga diharapkan penelitian ini dapat memberi kontribusi dalam perencanaan dan pelaksanaan kegiatan pengawasan BTS.

1.2. Perumusan Masalah

Berdasarkan permasalahan yang telah diuraikan pada latar belakang, maka perlu dilakukan pengidentifikasian rumusan masalah pada penelitian ini yaitu antara lain sebagai berikut:

1. Bagaimana mengklasterisasi titik koordinat BTS berdasarkan kedekatan jarak antar BTS dengan menggunakan metode *K-Means Clustering*?
2. Bagaimana penerapan metode *K-Means Clustering* dalam pengklasteran titik koordinat BTS dengan menggunakan *RapidMiner*?

1.3. Batasan Masalah

Mengingat luasnya permasalahan dan ruang lingkup terkait klasterisasi, maka agar penelitian ini lebih terarah dan mencapai tujuan yang diharapkan, maka ditetapkan batasan-batasan masalah sebagai berikut:

1. Data yang digunakan merupakan data titik koordinat BTS tahun 2021 yang bersumber dari Dinas Komunikasi dan Informatika Kabupaten Kerinci.
2. Metode yang digunakan dalam proses klasterisasi titik koordinat BTS adalah *K-Means Clustering*.
3. Pengujian data hasil klasterisasi menggunakan aplikasi *RapidMiner*.

1.4 Tujuan Penelitian

Sesuai dengan latar belakang dan rumusan masalah yang telah diuraikan di atas, adapun tujuan yang ingin dicapai dari penelitian terkait klasterisasi ini adalah sebagai berikut:

1. Mengklasterisasi titik koordinat BTS dengan menggunakan metode *K-Means Clustering*.
2. Mengimplementasikan metode *K-Means* ke dalam aplikasi berbasis web dan melakukan pengujian menggunakan *software RapidMiner*.

3. Menampilkan hasil klasterisasi titik koordinat dengan visualisasi berbasis Google Map.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat positif dalam penelitian bidang klasterisasi atau pengelompokan khususnya menggunakan metode *K-Means clustering*, adapun manfaat khusus dari penelitian ini adalah :

1. Memberi kemudahan bagi Tim Pengawas BTS dalam melakukan monitoring dan pengawasan ke seluruh lokasi BTS.
2. Pengelompokan ini diharapkan dapat memberikan gambaran lokasi yang tepat untuk penambahan BTS baru.

1.6 Sistematika Penulisan

Sistematika penulisan dibuat untuk mempermudah melihat dan mengetahui pembahasan yang ada dalam penelitian ini secara keseluruhan, maka penulis menyusun laporan penelitian ini sebagaimana uraian berikut ini :

BAB I : PENDAHULUAN

Berisi Latar Belakang, Perumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penelitian

BAB II : LANDASAN TEORI

Bab ini menguraikan tentang teori dan penerapan metode yang berhubungan dengan konsep dasar, pengertian dan contoh kasus.

BAB III : METODOLOGI PENELITIAN

Bab ini membahas tentang penjabaran perancangan berupa literatur kerangka kerja antara lain metode pengumpulan data termasuk penentuan kriteria dan metode Data Mining.

BAB IV : ANALISA DAN PERANCANGAN

Bab ini berisi tentang penguraian hasil analisa sistem, kebutuhan sistem seperti perangkat lunak (*software*) dan kerangka perancangan sistem pada klaster.

BAB V : IMPLEMENTASI DAN HASIL

Bab ini berisi tentang penjabaran hasil Analisa dan pengujian dari sistem yang digunakan.

BAB VI : KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari penelitian dan saran yang dapat dilakukan setelah penelitian.