BAB II
TINJAUAN PUSTAKA

2.1 Teori Penyakit Stroke

Istilah stroke telah dikenal oleh masyarakat luas baik di luar negeri maupun di Indonesia. Menurut World Health Organisation (WHO), stroke adalah penyakit yang disebabkan adanya kelainan pembuluh darah otak yang menimbulkan gangguan fungsi otak baik fokal maupun global dan berlangsung selama lebih dari 24 jam (Kasner et al., 2013). Stroke merupakan penyakit nomor dua di dunia yang menyebabkan kematian, dan nomor tiga yang menyebabkan disabilitas. Menurut survei WHO, negara berkembang dengan tingkat pemasukan rendah hingga menengah, angka kejadian stroke bertambah hingga 2 kali lipat dibanding 40 tahun sebelumnya (Johnson et al., 2016). Menurut Riset Kesehatan Dasar (Riskedas) tahun 2018, jumlah penderita stroke di Indonesia meningkat seiring bertambahnya usia. Kelompok usia 75 tahun ke atas jumlahnya mencapai 50.2%, sedangkan yang terendah adalah pada kelompok usia 15-24 tahun, yaitu sebesar 0,6%. Di Indonesia, populasi penderita stroke lebih tinggi di daerah perkotaan dibandingkan di pedesaan (12,6% vs 8,8%) (Riskesdas 2018). Hal ini dipengaruhi oleh gaya hidup akibat globalisasi seperti mengkonsumsi makanan cepat saji yang dapat menyebabkan kadar kolesterol tinggi (Kusuima et al., 2009). Terjadinya stroke dapat disebabkan oleh kurangnya pengetahuan mengenai penyakit tersebut. Sosialisasi dengan cara memberikan informasi dan pemahaman kepada masyarakat sebagai suatu kegiatan pengabdian masyarakat tentang stroke sangat diperlukan. Kegiatan penyuluhan ini bertujuan untuk memberikan informasi dan meningkatkan pemahaman tentang stroke agar dapat mencegah terjadinya stroke.[4]

2.2.1 Pengertian *Data Mining*

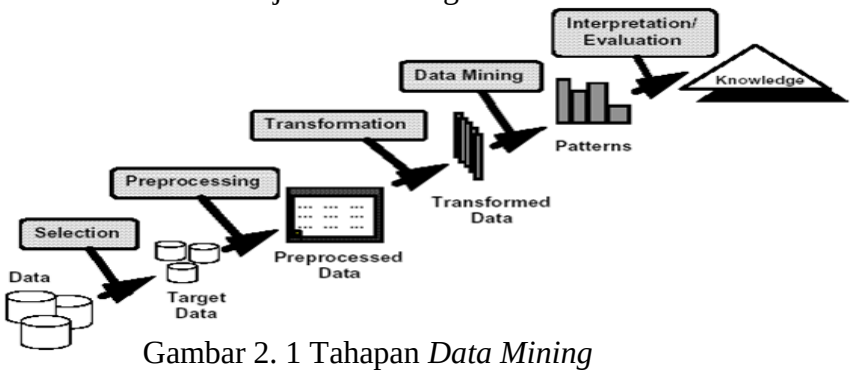
Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dari dalam database. *Data Mining* adalah

proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Pengetahuan itu bisa dijadikan acuan untuk mengambil keputusan dalam bisnis perusahaan.

Data Mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola. *Data Mining* dibagi menjadi beberapa fungsi,yaitu : Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklasteran, dan Asosiasi. Di Pada penelitian ini akan dilakukan proses asosiasi. Tugas asosiasi dalam *Data Mining* adalah menemukan atribut yang muncul dalam satu waktu. Tugas ini, alam dunia marketing lebih dikenal dengan analisis keranjang pasar *(market based analysis)*. [5]

2.2.2 Tahapan *Data Mining*

Tahapan yang dilakukan pada proses *Data Mining* diawali dari seleksi data dari data sumber ke data target, tahap *Pre-processing* untuk memperbaiki kualitas data, *Transformasi*, *Data Mining* serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Secara detail dijelaskan sebagai berikut :



Gambar 2. 1 Tahapan *Data Mining*

Berikut ini adalah penjelasan tahapan *Data Mining* berdasarkan gambar:

a. Pemilihan Data

   Dari sekumpulan data operasional yang perlu dilakukan sebelum penggalian informasi dalam KDD dimulai. Data dari hasil seleksi digunakan untuk proses *Data Mining*, disimpan pada suatu berkas, terpisah dari basis data operasional.

b. *Pre-processing/Cleaning*

   Tahap cleaning ini mencakup seperti membuang data duplikasi, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan dalam menulis (tipografi).

c. *Transformation*

   Tahap ini dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Data diubah atau digabung ke dalam format khusus sebelum bisa diaplikasikan. Sebagai contoh, beberapa metode standar seperti analisis asosiasi dan cluterung hanya bisa menerima input data kategorikal. Karenanya data berupa numerik yang berlanjut perlu dibagi menjadbeberapa interval, seperti pada atribut data umur dapat ditranformasikan ke dalam rentang umur.

d. *Data Mining*

   Merupakan proses mencari pola atau informasi yang menarik dalam data yang terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangatbergantung pada tujuan dan proses KDD.

e. *Interpretation*/Evaluasi

   Pola informasi yang dihasilkan dari proses *Data Mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah

pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya. [6]

2.3 Clustering

Clustering adalah suatu pekerjaan mengelompokkan sekumpulan objek datasehingga objek-objek dalam suatu kelompok memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan objek-objek dalam kelompok lain. Proses pada *Clustering* akan mempartisi sekumpulan objek data menjadi subset yang dapat dimanfaatkan untuk mengatur hasil pencarian ke dalam kelompok-kelompok dan menyajikan hasilnya secara ringkas dan mudah diakses. *Clustering* umumnya digunakan di berbagai bidang dengan berbagai aplikasi yang sangat penting termasuk riset pasar, sistem rekomendasi, sistem keamanan, dan mesin pencari.[7]

2.4 Algoritma *K-Means*

Algoritma K-Means Clustering merupakan teknik Data Mining yang dapat mengelompokkan objek berdasarkan kesamaan karakteristiknya. Algoritma K-Means Clustering bisa meng-cluster banyak data dalam waktu relatif singkat dan efisiensi yang tinggi. Penggunaan algoritma ini pilihan yang tepat dalam melakukan pengelompokan sebuah data [1].Teknik K-Means Clustering berguna untuk mengatur dan menyederhanakan berbagai macam data terkait cluster karena mudah digunakan dan diimplementasikan. Data yang didapatkan pada tahap pengumpulan data, akan diproses dan diolah mengimplementasikan algoritma K-Means. Tahapan untuk perhitungan algoritma K-Means dijelaskan di bawah ini:

a. Tetapkan banyaknya jumlah cluster (k)
b. Pilih random titik pusat untuk cluster(centroid)
c. Pakai rumus Euclidean Distance untuk mendapatkan jarak tiap data terhadap centroid, rumusnya yakni:

$d(x,y) = \sqrt{\sum}^{n} = 1(xi - yi)^2$

Keterangan:

$(x,)$: jarak data x dan y

x: titik data objek

y: titik data *centroid*

i: banyaknya objek

d. Kelompokkan data yang sudah dikalkulasikan menurut jarak terkecil (minimum) antara data tersebut dengan pusat *cluster* atau data *centroid* dan mendapatkan *cluster* baru.

e. Lakukan perhitungan kembali berdasarkan data yang mengikuti *cluster* masing-masing untuk pusat *cluster* (*centroid*) baru. Nilai *centroid* baru didapatkan dari hasil perhitungan rata-rata data terhadap setiap *cluster*. Rumus sebagai Berikut:

$$C1 = \frac{X1+X2+X3+\cdots+Xn}{\sum x}$$

Keterangan:

CI : *centroid* baru

x1 : nilai *cluster* ke-1

xn : nilai *cluster* ke-n

x : jumlah data

f. Setelah dapat *centroid* baru, maka lakukan iterasi selanjutnya atau ulangi langkah c samapai e sampai tidak ditemukan data yang berpindah-pindah dari *cluster*. [8]

2.5 Analisa Metode Elbow

Metode elbow adalah metode di mana pada suatu titik tertentu terjadi penurunan yang signifikan dalam grafik, berbentuk lengkungan yang tajam. Nilainya kemudian akan menjadi nilai k atau banyaknya cluster yang baik. Mencari nilai k optimal dapat dilakukan dengan membandingkan nilai Sum of Square Error(SEE) yang disajikan dalam bentuk grafik. Tujuan dari metode elbow yaitu memilih nilai k yang terkecil dan mempunyai nilai internal yang rendah. Penentuan jumlah cluster yang optimal diidentifikasi dengan mempertimbangkan perbandingan perhitungan SEE pada setiap nilai cluster, peningkatan jumlah cluster akan membentuk siku, sehingga semakin besar nilai k, nilai SEE akan semakin kecil.[9]

2.6 Rapidminer

RapidMiner adalah sebuah aplikasi data mining yang bersifat open source atau sumber terbuka. Lisensi yang digunakan untuk RapidMiner adalah AGPL (GNU Affero General Public License) versi 3. Penelitian dan pengembangan alat ini dimulai oleh beberapa individu, termasuk Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer dari departemen kecerdasan buatan Universitas Dortmund pada tahun 001, dan kemudian diambil alih oleh SourceForge. Pada tahun 010-011, dalam survei KD nuggets, sebuah media data mining, Pencapaian RapidMiner menduduki peringkat pertama sebagai alat data mining untuk proyek nyata adalah prestasi yang signifikan. Ini menunjukkan bahwa RapidMiner telah terbukti sangat efektif dan relevan dalam dunia praktis analisis data dan data mining. Dalam penerapannya, Rapid Miner memberikan metode penambangan informasi dan pembelajaran mesin yang menggabungkan ETL (Extricate, Change, Stack), preprocessing informasi, visualisasi, pemodelan dan penilaian. Persiapan penambangan informasi di Rapid Miner digambarkan dalam format XML yang dapat diatur melalui antarmuka grafis (GUI). Hal ini membuat penggunaan dan pengawasan alur kerja penambangan

informasi menjadi lebih mudah. Instrumen Rapid Miner disusun dalam dialek pemrograman Java dan berkoordinasi dengan usaha dan wawasan penambangan informasi Weka.[10]

2.7 Penelitian Terdahulu

Dalam melakukan penelitian ini terdapat penelitian yang telah dilakukan sebelumnya yang menjadi bahan referensi bagi peneliti. Berikut adalah penelitian terdahulu yang digunakan sebagai referensi, penelitian terdahulu dapat dilihat pada tabel 2.1.

No	Peneliti	Judul	Metode	Kesimpulan
1.	Nisar 1, Wasilah 2, Haris Kusuma jaya 3	Pemanfaatan K Means Clustering dalam Pengelompokan Judul Skripsi	Metode *K-Means Clustering*	Hasil percobaan yang dilakukan pada penelitian ini menunjukkan bahwa algoritma *Clustering K-Means* melakukan pengelompokan skripsi dengan membagi data menjadi sejumlah k *cluster* yang ditentukan, dan menggunakan perhitungan jarak untuk mengukur kemiripan antar data. *Data Mining* menggunakan k-mean mampu melakukan pengelompokan data dalam jumlah besar dengan sangat cepat

				sehingga dapat membantu mempercepat proses pengelompokan.[7]
2.	Aviv Fitria Yulia, Handoyo Widi Nugroho	Implementasi Algoritma K-Means Classifier Sebagai Pendukung Keputusan Penerima Dana Bantuan Siswa Miskin (Studi Kasus : SMKN Sukoharjo)	Metode K-Means *Clustering*	Berdasarkan hasil penelitian yang dilakukan oleh penulis, dapat diambil kesimpulan sebagai berikut : - Penerapan algoritma K-Means membagi dataset menjadi tiga kelompok yaitu layak menerima bantuan siswa miskin ,dapat di pertimbangkan menerima bantuan siswa miskin dan tidak layak menerima bantuan siswa miskin. - Hasil pengujian mendapatkan nilai devies bouldin indeks sebesar 0,262 yang memiliki arti kesamaan antar anggota *cluster* yang cukup baik. - Hasil

				dari pengujian ini sangat layak untuk dijadikan rekomendasi penerima bantuan siswa miskin di sekolah (SMKN SUKOHARJO).[11]
3.	Melda Agarina, Sutedi, Arman Suryadi Karim, Erlinda Ratna Sari	Strategi Marketing Promosi Penerimaan Mahasiswa Baru Menggunakan K-Means Clustering	Penelitian ini menggunakan Metode K-Means Clustering	Penelitian ini menunjukkan bahwa penggunaan algoritma K-Means Clustering efektif dalam merancang strategi pemasaran untuk penerimaan mahasiswa baru secara efisien. Melalui analisis karakteristik calon mahasiswa, institusi pendidikan tinggi dapat mengembangkan strategi pemasaran yang lebih spesifik, yang pada gilirannya dapat meningkatkan jumlah pendaftar dan mahasiswa yang diterima. Diharapkan

				bahwa temuan penelitian ini dapat berkontribusi pada pengembangan strategi pemasaran penerimaan mahasiswa baru yang lebih efektif dan efisien di perguruan tinggi.[12]
4.	Nurjoko, Defi Dwirohayati, Novi Herawadi Sudibyo	Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means *Clustering*	Penelitian ini menggunakan Metode K-Means *Clustering*	*Clustering* menggunakan K-Means bertujuan untuk mengelompokkan data yang berkarakteristik sama dalam satu *cluster* dan data yang berkarakteristik berbeda ke dalam *cluster* yang lain. Data yang digunakan dalam penelitian ini adalah data crime indeks yang ada di POLRESTA Bandar Lampung berdasarkan jenis kasusnya yaitu pencurian, penganiayaan,

				penipuan, pemerkosaan, pembunuhan, perjudian dan narkoba. Tabel 3.1 adalah data yang digunakan untuk percobaan *Clustering*[13]
5.	Bani Nurhakim, Intan septiani,Khaerul Anam,Denni Pratama	Penarapan Algoritma K-Means Clustering Dalam Menganalisis Resiko Penyakit Stroke	Penelitian ini menggunakan Metode K-Means *Clustering*	Dari permasalahan penelitian ini dapat disimpulkan bahwa pelaksanaan analisis dengan menggunakan algoritma K-means clustering dapat digunakan untuk mengidentifikasi pola risiko penyebab stroke. Hasil analisis menggunakan algoritma K-means clustering dibuat dua kelompok berupa Cluster0 dengan hasil 4338 elemen dan Cluster1 dengan hasil 772 elemen. Hasil pengujian menggunakan aplikasi Rapid Miner

				menghasilkan nilai K optimal 2 dengan nilai DBI (Davies Bouldin Index) dengan hasil cluster 0.082. [2]
6.	Putri Syifa, Sarah Oktia Putri, Putri Almunawarah, Munirul Ula	Identifikasi Tingkat Resiko Kesehatan Jantung Dengan Menggunakan Algoritma K-Means Clustering	Metode K-Means *Clustering*	Berdasarkan percobaan yang telah diselesaikan dalam studi ini, disimpulkan bahwa metode klasterisasi dengan Algoritma K-Means memberikan wawasan baru terkait pengelompokan tingkat risiko penyakit jantung berdasarkan usia menjadi dua cluster. Cluster 0 mengelompokkan kategori usia dengan risiko penyakit jantung yang tergolong sedang atau rendah, sedangkan cluster 1 mengelompokkan kategori usia dengan risiko penyakit

				jantung yang tinggi. Melalui pendekatan yang berbeda ini memberi kesempatan kepada penyedia layanan kesehatan dalam memberikan perawtan yang lebih efisien, yang sesuai dengan tingkatan cluster yang alami oleh masing masing pasien serta meningkatkan kualitas hidup pasien melalui pengelolaan penyakit yang lebih individual dan efektif. Diperlukan uji klinis tambahan untuk memastikan bahwa pendekatan ini dapat diimplementasikan secara luas dan memberikan manfaat yang optimal bagi pasien.[14]
7.	Elsa Safutri, Ade Irma	Pengelompokan Tekanan Darah Lansia	Pada penelitian ini menggunakan metode K-Means	Penerapan algoritma K-meansuntuk pengelompokkan tekanan darah lansia

	Purnamasari, Agus Bahtiar, Edi Wahyudin	Dengan Algoritmak-Meansdi Kp.Lebak Jero		di Posbindu Kp. Lebak Jero berhasil mencapai hasil optimal pada k=2 dengan nilai Davies-Bouldin Index(DBI)sebesar 0,881 setelah 10 kali percobaan. Hasil pengelompokan membagi data menjadi dua klaster, yaitu Cluster_0 yang terdiri dari 75 lansia dengan risiko hipertensi lebih rendah, mencakup kategori normal dan pra-hipertensi, serta Cluster_1 yang terdiri dari 71 lansia dengan risiko hipertensi lebih tinggi, mencakup hipertensi Tingkat 1 dan Tingkat 2. Model ini menunjukkan efektivitas dalam mengelompokkan lansia berdasarkan tingkat tekanan darah sistolik, meskipun

				variabel lain seperti jenis kelamin, umur, dan berat badan hanya memberikan sedikit perbedaan.[15]
8.	Wawan Gunawan, Arief Wibow	Pemanfaatan Algoritma K-Means dalam Klasterisasi Gempa Sulawesi	Metode K-Means *Clustering*	Hasil penelitian ini memiliki dampak signifikan dalam memahami pola gempa bumi dan dapat dijadikan landasan untuk pengembangan strategi mitigasi risiko bencana gempa. Melalui pemrosesan data yang cermat dan penggunaan algoritma K-Means telah berhasil mengelompokkan pola gempa ke dalam klaster yang memiliki karakteristik berbeda.[16]
9.	Dwi Astuti, Relita Buaton, Magdalen	Pengelompokan Data Kasus Keracunan Makanan Biologis	Pada penelitian ini menggunakan Metode *Clustering*	Berdasarkan penelitian yang dilakukandengan menggunakan metode algoritma k-means maka dapat diketahui bahwasannya hasil

	a Simanjuntak	Berdasarkan Faktor Penyebab Menggunakan Metode Clustering		terbanyak terdapat pada cluster 2 dengan data kasus keracunan biologis terdapat 11 data dengan titik centroid usia (x) 2 yaitu 12-16 Tahun, titik centroid pada jenis keracunan (y) 6.36 yaitu roti isi, dan titik centroid pada factor penyebab (z) 2.9 yaitu Bakteri Gram-negatif berbentuk batang yang biasanya ditemukan di usus manusia dan hewan berdarah panas.REFERENSIAdawiah, N., Suliatiyowati, N., & Jajuli, M. (2021). Klasterisasi kasus kekerasan terhadap anak dan perempuan berdasarkan algoritma K-Means. Generation Journal, 5(2), 69–80.Alfikri, T. R., Fakhriza, M., & Ikhwan, A. (2024). Penerapan algoritma

				K-Means dalam mengelompokkan penyebaran penyakit di Kecamatan Sei Balai. Jurnal Teknik Informatika Kaputama (JTIK), 8(1), 1–7.Arhami, M., & Nasir, M. (2020). Data mining(R. Indah Utami, Ed.; 1st ed.). CV Andi Offset.Dewi, R. (2023). Aplikasi Matlab untuk simulasi pengolahan sinyal(A. Prijono, Ed.; 1st ed.). Zahir Publishing.Efori Buulolo, S. Kom., & M. Kom. (2020). Data mining(1st ed.). CV Budi Utama.Fitriana, N. F. (2021). Gambaran pengetahuan pertolongan pertama keracunan makanan. Jurnal Kesehatan Tambusai, 2(3), 173–178.Laksana, F., Hidayat, R., & Dewi, Y. (2023).

				Pengelompokan jumlah kekerasan terhadap anak berdasarkan kecamatan di Kabupaten Banyumas menggunakan penerapan algoritma K-Means. INDEXIA: Informatic and Computational Intelligent Journal, 5(2), 123–135.[17]
10.	Abdi Subayu	Penerapan Metode K-Means Untuk Analisis Stunting Gizi Pada Balita: Systematic Review	Penulis ini menggunakan Metode K-Means pada Penelitian tersebut	Banyak negara telah melakukan penelitian mengenai masalah Stunting Gizi pada balita, dan hasilnya adalah masih banyak balita mengalami Stunting atau Gizi buruk, Indonesia termasuk ke dalam 13 negara dengan masalah Gizi buruk. Dari penelitian dengan Systematic Review menjelaskan bahwa faktor yang menyebabkan Gizi buruk adalah,

				kurangnya pemahaman orang tua terhadap Gizi seimbang, faktor lingkungan yang belum menerapkan pola kebersihan, faktor Pola hidup yang tidak Sehat, faktor Ekonomi, faktor kurangnya kesadaran orangtua akan pentingnya Posyandu dan banyak faktor-faktor lain. Dari banyak faktor di atas dibuat sebuah makalah dengan Systematic Review yang berfokus pada satu metode penelitian untuk mencari kesimpulan apakah metode yang dipilih mampu menyelesaikan masalah persoalan Clustering Stunting Gizipada Balita. Dari 6 jurnal diatas penulis menarik kesimpulan mengenai studi literatur penggunaan metode K-Means untuk

				Clustering Stunting Gizi pada balita bahwa hasil akurasi yang diperoleh sangat rendah hingga kurang efektif untuk digunakan.[18]

Tabel 2. 1 Penelitian Terdahulu