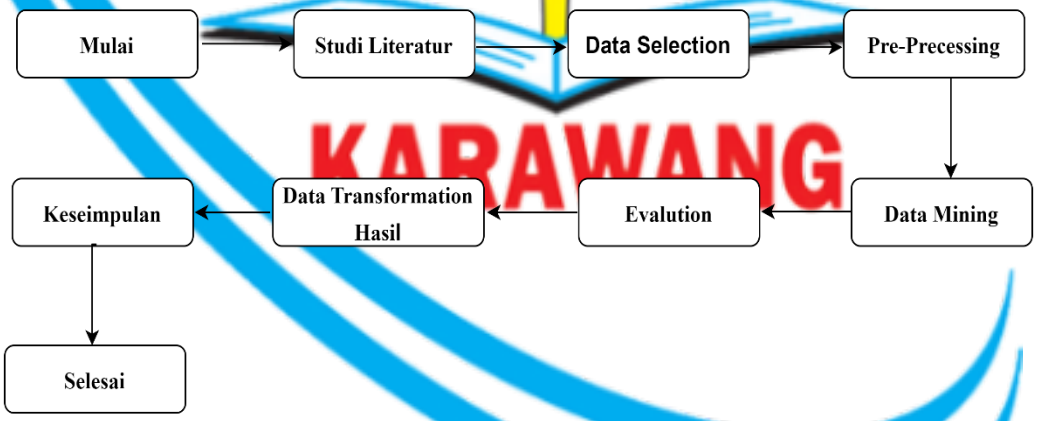
BAB III
METODE PENELITIAN

3.1. Objek Penelitian

Analisis data yang digunakan pada penelitian ini adalah *Knowledge Discovery Database* (KDD). Adapun menjelaskan langkah-langkah yang akan digunakan serta perancangan dalam melakukan jumlah tenaga kesehatan di dinas kesehatan kabupaten Karawang metode *K-Medoids*tahapan dari metode *Knowledge Discovery in Database* yang digunakan pada penelitian ini clusteringuntuk pengelompokan data jumlah tenaga kesehatan di dearah kabupaten Karawang.

3.2. Prosedur Penelitian

Dalam penelitian ini terdapat tahapan-tahapan yang harus dipahami dan dilakukan. Terdapat pada gambar 3.2 Alur penelitian



Gambar 3.1 Alur Penelitian

1.2.1. Studi Literatur

Peneliti melakukan studi literatur dengan mempelajari buku dan jurnal karya ilmiah yang berkaitan dengan teknik data mining dengan konsep clustering menggunakan algoritma *K-Medoids* yang bersumber di internet dan perpustakaan.

1.2.2. Data Selection

Pengumpulan data adalah tahapan yang bertujuan untuk memperolehinformasi maupun data yang berkaitan dengan penelitian. Bahan atau data yang akan digunakan dalam penelitian ini yaitu *dataset* yang diperoleh dari *website UCI Machine Learning Repository*, yang merupakan data Jumlah Tenaga Kesehatan Di Dinas Kesehatan Kabupaten Karawang. dalam menggunakan format excel dengan jumlah total atribut 19 dari 72 data, sehingga dapat mendukung untuk proses klasifikasi jumlah tenaga kesehatan di dinas kesehatan kabupaten karawang.

1.2.3. Pre-Processing

Pre-Processing bertujuan untuk memahami data agar dapat di seleksi dan di kelola, tentunya data tersebut dikategorikan menjadi dua yaitu data *training* 80% dan data *testing* 20%. Dari data jumlah tenaga kesehatan yang ada di Kecamatan Kabupaten Karawang berjumlah sekitar 13.477 orang, maka data *training* dapat di akumulasikan sebesar 80% sehingga perhitungannya yaitu $13.477 \times 80\% = \frac{1.078.160}{100} = 10.781$. Sedangkan data *testing* dapat di akumulasikan sebesar 20% sehingga perhitungannya yaitu $13.477 \times 20\% = \frac{269.540}{100} = 2.695$. Sehingga memudahkan dalam menentukan kesalahan atau eror dalam mengakurasikan algoritma Medoids terhadap klasifikasi Jumlah Tenaga Kesehatan Di Dinas Kesehatan Kabupaten Karawang.

1.2.4. Data Transformation Hasil

Data Transformation Hasil adalah proses pembersihan data dari atribut-atribut yang tidak diperlukan dalam pengolahan data *mining*. seleksi data dilakukan untuk mengeliminasi data yang memiliki *missing value*.

1.2.5. Data Mining

Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan isualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Sutan Fasial.01:2019). data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisa terhadap data mining (big data) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini di pahami dan berguna bagi pemilik data tersebut(Bramer.dkk.180:2007).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. data mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. jadi data mining dapat diartikanData mining merupakan suatu proses otomatis untuk menganalisa data data yang akan diproses merupakan data yang sangat besar, sehingga sulit untuk menproses dengan manual, Tujuan data mining adalah mendapatkan hubungan atau pola/ Teknik yang memberikan indikasi yang bermanfaat. Data mining bukanlah hal yang baru, satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang dulu sudah mapan terlebih dulu. data mining memiliki akar yang Panjang dari bidang ilmu yang berbeda seperti kecerdasan buatan (artificial ntelligent).

Machine learning, statistik, database, dan juga information retrieval. Database yang tersimpan di media penyimpanan jarang sekali dimanfaatkan oleh sebagian besar penggunanya dan bahkan dalam jangka waktu tertentu data-data tersebut dihapus karena

dianggap sampah dan hanya memenuhi media penyimpanan saja. Anggapan tersebut tidak sepenuhnya benar, karena sesungguhnya database dalam ukuran yang besar dapat memberikan informasi yang dibutuhkan untuk berbagai kepentingan, baik untuk kepentingan bisnis dalam mengambil keputusan maupun untuk ilmu pengetahuan dan penelitian.

1.2.6. Evaluation

Pada tahap evaluation ini diambil dari hasil akurasi algoritma *Medoids* dan *clustering* kemudian algoritma yang mempunyai nilai akurasi yang tinggi merupakan algoritma terbaik pada proses klasifikasi ini. Untuk mendapatkan nilai akurasi dalam klasifikasi kanker serviks, digunakan persamaan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} * 100\% \quad (7)$$

Keterangan : 6

False Positive (FP), yaitu data negatif tapi terprediksi sebagai data positif.

False Negative (FN), yaitu data positif yang terprediksi sebagai data negatif.

True Positive (TP), yaitu data positif yang terprediksi benar.

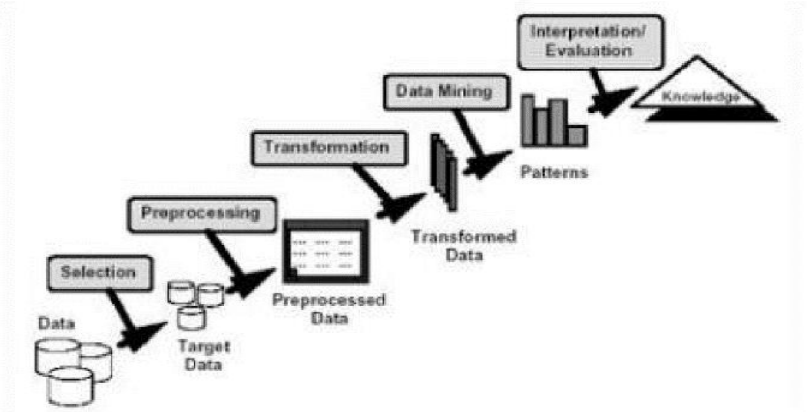
True Negative (TN), yaitu data negatif yang terprediksi dengan benar.

1.3. Knowledge Discovery in Database (KDD)

Istilah data miningdan *knowledge discovery in database* (KDD) 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. Salah satu tahapan dalam keseluruhan proses KDD adalah data mining

(Aishwarya Batra, dkk, 2011:05). Data Selection, pemilihan (seleksi) data dari sekumpulan dataoperasional perlu dilakukan sebelum tahappenggalian informasi dalam KDD dimulai. Data hasilseleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah daribasis data operasional. Pre-processing/Cleaning sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yangmenjadi fokus KDD. Proses pembersihan mencakupantara lain membuang *duplikasi data*, memeriksa data yang *inkosisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi)*. Transformation *Coding* adalah transformasi pada data yang telah di pilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis ataupola informasi yang akan dicari dalam basis data.

Data mining, data mining adalah proses mencari pola atauinformasi menarik dalam data terpilih denganmenggunakan teknik atau metode tertentu..Teknik, metode, atau algoritma dalam *data mining* sangatbervariasi. Pemilihan metode atau algoritma yangtepat sangat bergantung pada tujuan dan proses KDDsecara keseluruhan. *Interpretation / Evaluation*, pola informasi yang *dihasilkan* dari proses *dataminingperlu* ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahapini merupakan bagian dari proses KDD yang disebut*interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya



Gambar 3.2. Proses Data Mining Dalam Penemuan Pengetahuan Dalam Databas

1.4. Lokasi dan Waktu Penelitian

Penelitian Tugas Akhir dilaksanakan di Dinas Kesehatan Kabupaten Karawang jalan Parahyangan No.39 Kelurahan Adiarsa kecamatan Karawang Timurkabupaten Karawang. KegiatanPenelitian Propasal dilaksanakan selama 3 bulan, dilaksanakan mulai dari tanggal 29 November 2022 sampai dengan 28 Febuari 2023. Adapun perincian kegiatan Penelitian ditunjukkan pada Tabel 3.1.

Tabel 3.1. Pelaksanaan Penelitian Proposal

NO	Prosedur Penelitian	BULAN															
		NOVEMBER 2022				DESEMBER 2022				JANUARI 2023				FEBRUARI 2023			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Studi Literatur																
2	Data Selection																
3	Pre-Precessing																
4	Data Mining																
5	Evalution																
6	Data Transformation Hasil																
7	Keseimpulan																