

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode yang digunakan dalam penelitian ini yaitu menerapkan metode K-Means Clustering pada data hasil survey kesehatan mental pada Mahasiswa yang diperoleh dari repository Kaggle untuk dianalisa dan dikelompokkan berdasarkan faktor akademik, faktor kesehatan fisik, faktor dukungan sosial dan faktor faktor lain yang relevan.

3.2 Perangkat Penelitian

Dalam melakukan penelitian ini, dibutuhkan peralatan perangkat keras (Hardware) dan perangkat lunak (Software). Untuk lebih lengkapnya di jelaskan berikut ini:

3.2.1 Perangkat Keras

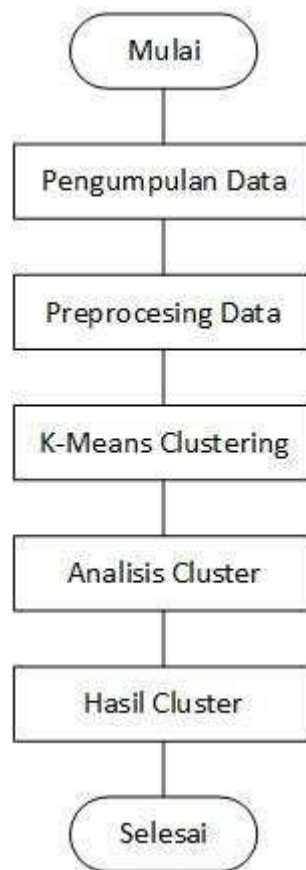
Perangkat keras (hardware) yang digunakan dalam penelitian ini yaitu menggunakan 1 buah laptop yang digunakan selama penelitian. Spesifikasi laptop sebagai berikut: MyBook14H, Intel(R) Celeron(R) N4020 CPU @ 1.10GHz 1.10 GHz, Ram 8,00 GB (7,82 GB usable), Windows 10 Pro 64bit.

3.2.2 Perangkat Lunak

Perangkat lunak memiliki peran penting pada penelitian ini karena hasil dari analisis data dapat diketahui dari pengolahan menggunakan perangkat lunak dalam mengetahui hasilnya. Pada penelitian ini peneliti menggunakan Aplikasi Rapidminer untuk melakukan proses clustering.

3.3 Tahap – Tahap Penelitian

Dalam penelitian ini, peneliti menggunakan salah satu metode dalam Data Mining yaitu metode K-Means Clustering. Dalam metode ini memiliki beberapa tahapan. Tahapan yang dilakukan adalah:



Gambar 3. 1 Bagan Tahap – tahap Penelitian

3.3.1 Pengumpulan Data

Untuk melakukan proses K-Means Clustering tentu membutuhkan sebuah data yang cukup banyak dan sesuai dengan yang dibutuhkan, di dalam penelitian ini peneliti menggunakan data set survey kesehatan mental pada Mahasiswa dari repository Kaggle.

3.3.2 Preprocessing Data

Tahap pemrosesan data adalah tahap dimana data yang sudah didapatkan, dipilah, dan dipisahkan agar mendapatkan data yang dibutuhkan untuk proses selanjutnya. Tahapan ini mempunyai beberapa proses dimana setiap proses tersebut saling berhubungan satu sama lainnya. Proses dalam tahapan pemrosesan data adalah sebagai berikut:

1. Cleaning Data

Cleaning Data adalah proses dalam tahapan pemrosesan untuk mengisi data kosong atau blank apabila memungkinkan, duplikasi

data, memperbaiki data yang tidak sesuai dengan ketentuan atau salah ketik seperti kurang huruf dan kelebihan huruf, mengubah dan memodifikasi data agar data yang akan diolah adalah data yang konsisten, mengatur data yang kurang rapi dalam penulisan huruf besar dan kecil, dan mengganti format penulisan angka dan huruf sesuai dengan yang dibutuhkan. Adapun dalam penelitian ini terdapat 27 data kotor yang terdiri dari 15 data substance use dan 12 data CGPA.

2. Transformasi Data

Karena metode K-Means Clustering adalah metode yang bisa dilakukan apabila data yang dipakai adalah data berupa angka, maka proses transformasi ini sangat dibutuhkan. Proses transformasi adalah tahap untuk mengubah data atribut yang selain angka ke dalam nilai angka agar data tersebut dapat diolah menggunakan algoritma K-Means Clustering. Pada analisa data set Kesehatan mental mahasiswa, saya menggunakan algoritma nominal to numerical untuk melakukan transformasi beberapa attribute yang memiliki type data kategorikal dan nominal.

3.3.3 K-Means Clustering

K-Means Clustering merupakan suatu pendekatan analisis data yang bertujuan untuk mengelompokkan entitas data secara otomatis berdasarkan kemiripan karakteristiknya, tanpa adanya label kelas yang telah ditentukan sebelumnya. Clustering, seringkali diidentifikasi sebagai metode unsupervised learning, memberikan kemampuan untuk mengklasifikasikan data yang belum memiliki label kelas yang jelas atau terdefinisi. Metode ini menjadi esensial dalam konteks analisis data kompleks, di mana informasi tentang struktur kelas tidak diketahui atau sulit ditentukan. Proses clustering melibatkan identifikasi pola-pola intrinsik dalam data, dengan tujuan membentuk kelompok-kelompok homogen yang memiliki tingkat kemiripan tinggi di antara anggotanya.

3.3.4 Analisis Cluster

Analisis kluster merupakan salah satu teknik pada data mining yang digunakan untuk menemukan kluster-kluster dari himpunan data secara otomatis atau semi-otomatis. Berdasarkan pendekatan dan/atau konsep yang diadopsi, teknik clustering dapat dikategorikan ke dalam metoda yang berbasis partisi, hirarki, dan densitas/kerapatan, grid, model dan konstrain. Adapun pada penelitian ini saya menggunakan simulasi model untuk menentukan kluster terbaik menggunakan Elbow Method.

3.3.5 Hasil Cluster

Hasil cluster dapat dieksplorasi melalui berbagai metode, termasuk visualisasi data, analisis statistik, dan penelitian lebih lanjut tentang sifat-sifat individu dalam setiap cluster. Tujuan utama dari memahami hasil cluster adalah untuk mendapatkan wawasan yang berharga yang dapat digunakan untuk pengambilan keputusan, perencanaan strategis, atau penelitian lanjutan dalam berbagai bidang seperti ilmu sosial, bisnis, biologi, dan lainnya.