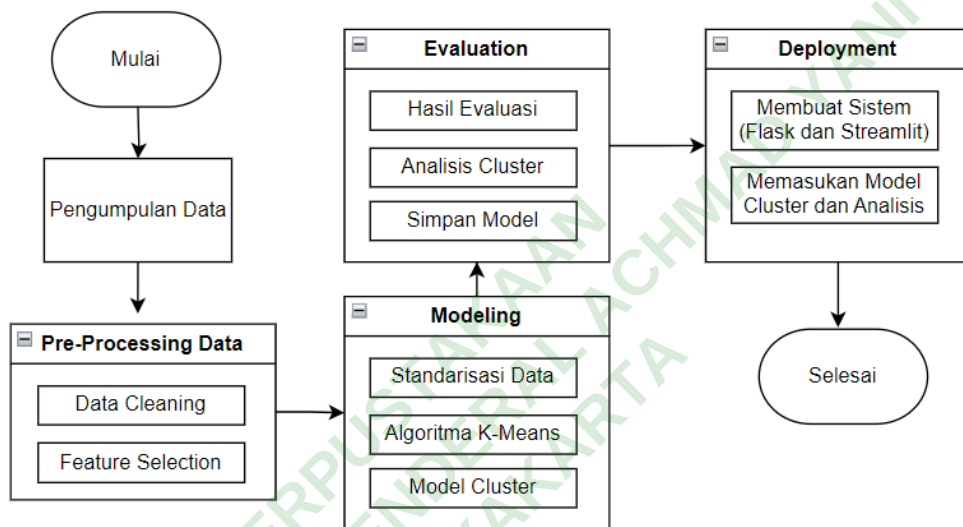


BAB 3

METODE PENELITIAN

Penelitian ini menggunakan metode *clustering* dengan algoritma K-Means yang tahapannya dijabarkan pada Gambar 3.1 yang sudah dimodifikasi dari penelitian Wongso et al., (2023).



Gambar 3.1 Tahapan Metode Penelitian

Berdasarkan tahapan metode penelitian diatas, berikut adalah penjabaran dari setiap tahapannya:

1. Pengumpulan Data

Tahap awal dimulai dengan mengumpulkan data dari Biro Kerjasama, Promosi dan Admisi (KPA) yang berasal dari database SICAMA Unjaya.

2. *Pre-processing Data*

Pada tahap *pre-processing* ini diawali dengan dengan *data cleaning*, dan *feature selection*. Tahap ini dilakukan untuk menentukan atribut atau mereduksi data, menghilangkan nilai yang hilang untuk atribut dengan data kosong (*null*) sehingga data yang dihitung akan lebih akurat karena nilai yang hilang sudah dihapus, dan dilakukan pemilihan variabel yang akan digunakan untuk model.

3. *Modeling*

Pada tahap ini dilakukan standarisasi data menggunakan *standar scaler* yang bertujuan untuk menempatkan data ke dalam rentang data yang tidak terlalu jauh dan mencegah terjadinya *outlier* data yang dapat mengganggu saat proses klasterisasi (Mayoana Fitri et al., 2023). Selanjutnya, tahap *modeling* dilakukan dengan cara pemodelan pada *dataset* yang sudah diproses dengan algoritma *clustering K-Means* untuk pengolahan data. *K-Means* adalah algoritma untuk mengintegrasikan data yang bertujuan dalam menggabungkan titik-titik data berdasarkan karakteristik dan kedekatannya dengan pusat cluster (Asnawi et al., 2023).

4. *Evaluation*

Pada tahap ini dilakukan dengan mengevaluasi hasil model, apakah sudah sesuai dengan tujuan penelitian atau melihat pada pertanyaan penelitian apakah sudah terjawab. Evaluasi menggunakan *silhouette score* yaitu untuk menentukan seberapa baik suatu titik data yang berada dalam *cluster* dan jaraknya dari *cluster* lainnya. *Silhouette score* dengan nilai skor yang mendekati 1 mengindikasikan bahwa titik data yang terkelompok sudah baik (Shutaywi & Kachouie, 2021). Selain itu, setelah hasil sesuai maka dilakukan proses penyimpanan model dengan *library pickle*.

5. *Deployment*

Terakhir tahap *Deployment* yaitu menerapkan hasil proses data yang sudah diolah kedalam program menggunakan *framework flask* yang dijalankan pada *local host*. Pada tahap deployment juga menggunakan *library streamlit* untuk membangun dashboard hasil analisis *K-Means*.

3.1 BAHAN PENELITIAN

Penelitian ini menggunakan data yang diperoleh dari hasil ekspor Biro KPA Unjaya yang berisi data mahasiswa Prodi Sistem Informasi Unjaya yang tersimpan pada database SICAMA PMB Unjaya. Data yang diperoleh yakni data profil pendaftar angkatan tahun 2019, 2020, 2021, dan 2022.

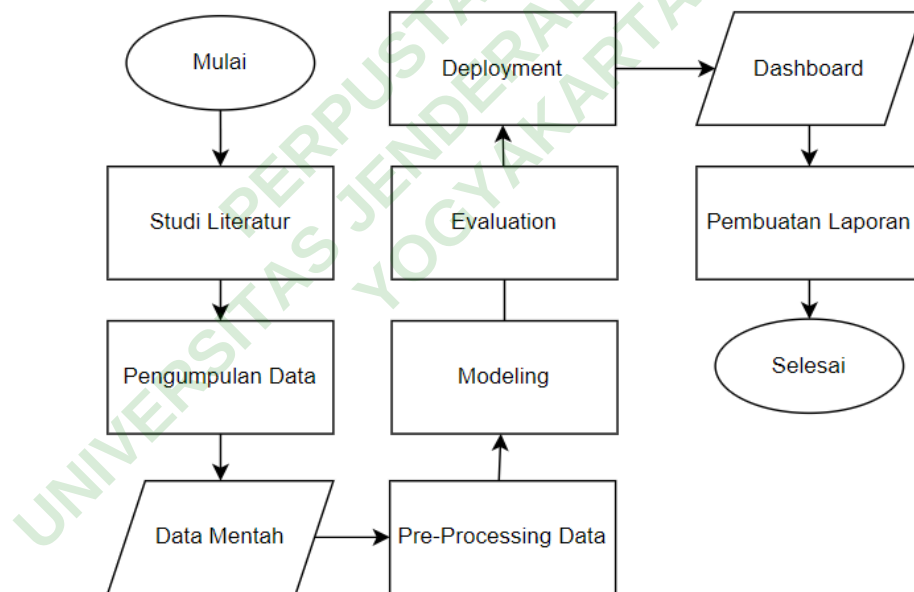
3.2 ALAT PENELITIAN

Penelitian ini memerlukan perangkat lunak dan sistem operasi tertentu, serta dukungan jaringan internet, untuk menjalankan proses penelitian. Berikut adalah beberapa program aplikasi dan sistem operasi yang digunakan dalam penelitian:

1. Laptop Lenovo Thinkpad Yoga 380 dengan *Operating System* Windows 11
2. Microsoft Office tahun 2021
3. Google Collaboratory
4. Visual Studio Code

3.3 JALAN PENELITIAN

Jalan penelitian ini dilakukan dengan menerapkan pendekatan CRISP-DM. Pada gambar 3.2 terdiri dari enam (6) tahapan (Plotnikova et al., 2022) yaitu *Pre-Processing Data, Modeling, Evaluation, Deployment*.



Gambar 3.2 Jalannya Penelitian

Berdasarkan jalan penelitian yang sudah dipaparkan, berikut adalah penjelasan dari setiap tahapan:

1. Studi Literatur

Tahap awal penelitian dilakukan dengan studi literatur dengan membaca berbagai referensi jurnal untuk dijadikan acuan dalam menulis.

2. Proposal Penelitian

Penulisan proposal penelitian berdasarkan masalah yang ada pada Unjaya dan juga menggunakan referensi jurnal yang sudah dikumpulkan dijadikan sitasi pada setiap paragraf yang berkaitan.

3. Pengumpulan Data

Pada tahap pengumpulan data dikumpulkan dari Biro Kerjasama, Promosi dan Admisi (KPA) yang berasal dari database SICAMA Unjaya. Data yang dikumpulkan adalah profil mahasiswa angkatan tahun 2019, 2020, 2021, dan 2022. Di dalam data terdapat no, tahun, nama, ID, prodi pilihan 1, prodi pilihan 2, asal provinsi, dan media promosi yang digunakan oleh mahasiswa pada saat mengetahui sumber informasi mengenal kampus.

4. Pre-Processing Data

Kemudian pada tahap *pre-processing data* dilakukan *cleaning data*, dan *feature selection*. Tahap *preparation* dilakukan untuk *missing value* sehingga data yang dihitung akan lebih akurat karena *missing value* telah dihapus, penghapusan atribut yang tidak digunakan dan menentukan atribut variabel yang akan digunakan untuk *modeling*.

5. Modeling

Modeling dilakukan dengan menentukan titik optimal kluster dengan metode elbow untuk menentukan jumlah cluster yaitu penentuan k sebagai jumlah cluster dengan menerapkan *elbow method*, dilanjutkan dengan perhitungan *centroid* awal pada cluster, dimana jarak tiap objek data dengan centroid akan dihitung dengan metode *Euclidian Distance*. Setiap objek dialokasikan kedalam centroid yang terdekat menggunakan *Euclidian Distance* (Kharisma & Aesy, 2023). Selanjutnya proses standarisasi data dengan *standar scaler* untuk menempatkan data ke dalam rentang data yang sesuai, dan membuat model dari data siap olah dengan menggunakan algoritma *K-Means* (`sklearn.cluster import KMeans`) dalam pengolahan data dengan mengklusterkan wilayah berdasarkan jumlah pendaftar. Setelah selesai mengklusterkan wilayah, selanjutnya dilakukan pengklasteran media promosi di wilayah cluster 1 dan 2.

6. Evaluation

Evaluasi dilakukan dengan menggunakan *silhouette score* untuk menentukan seberapa baik suatu titik data yang berada dalam *cluster* serta jaraknya dari *cluster* lainnya. *Silhouette score* dengan nilai skor yang mendekati 1 mengindikasikan bahwa titik data yang terkelompok sudah baik (Shutaywi & Kachouie, 2021). Setelah dilakukan evaluasi, dilanjutkan menganalisis cluster dengan membuat *bar chart* sesuai informasi yang didapat. Terakhir, model cluster yang terbentuk disimpan menjadi file excel agar bisa lanjut ke tahap berikutnya.

7. Deployment

Pada tahap ini membuat dashboard menggunakan *framework flask* dan library *streamlit* dengan memasukkan model yang sudah disimpan. Setelah dashboard jadi dilakukan pengecekan, apakah data yang dianalisis sudah sesuai dengan tujuan awal.

8. Pembuatan Laporan

Terakhir dilakukan penyusunan laporan untuk memenuhi syarat tugas akhir dengan membuat dokumen yang terstruktur, lengkap, dan informatif.