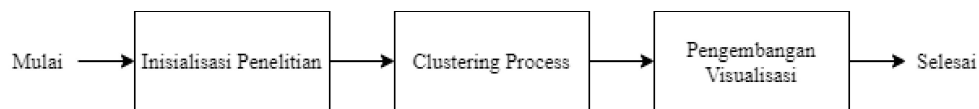


BAB III

METODOLOGI PENELITIAN

3.1. Metodologi Penelitian

Metodologi penelitian merupakan langkah – langkah yang dilakukan oleh peneliti dalam menyelesaikan penelitian. Beberapa langkah yang akan dilakukan peneliti dalam membangun *clustering* menggunakan algoritma *k-medoids*, secara umum dijelaskan pada diagram Gambar 3.1.



Gambar 3. 1. Metode Penelitian

Berdasarkan gambar diatas dapat dijelaskan lebih rinci mengenai langkah - langkah pada metode penelitian ini dibagi dalam 4 tahapan proses yaitu: Inisialisasi Penelitian, Clustering Process, Pengembangan Sistem, dan pengujian.

3.1.1. Inisialisasi Penelitian

Tahap ini merupakan tahap pertama dalam melakukan penelitian ini, yang terdiri dari mendefinisikan permasalahan, menentukan solusi, menentukan metode.

- a) Mendefinisikan masalah
 - Jumlah data yang dihasilkan oleh media online terus meningkat dan tidak dapat terkendali lagi jumlahnya.
 - Diperlukannya pengelompokan berita *online* berdasarkan *content* beritanya sehingga dapat mengetahui berita yang sedang diperbincangkan tanpa harus membaca semua sumber
- b) Menawarkan solusi

Membangun visualisasi yang dapat melakukan pengelompokan terhadap isi berita. Harapannya masyarakat tidak perlu membaca banyak berita dalam *cluster* yang sama, sehingga masyarakat dapat menerima informasi yang lebih banyak.
- c) Menentukan Metode

Metode yang digunakan dalam penelitian ini adalah *k-medoids* yang digunakan untuk melakukan pengelompokan berdasarkan isi berita yang ada.

3.1.2. Clustering Proses

Tahapan ini dilakukan Clustering proses yang terdiri dari pengumpulan data, *Pre-processing data*, *k-medoids clustering*, rincian setiap proses sebagai berikut:

a) *Data Collection*

Untuk dapat membentuk model *clustering* harian berita online maka diperlukan data yang cukup banyak dan selesai yang dibutuhkan, oleh sebab itu pada penelitian ini peneliti menggunakan data yang dimiliki oleh PT. Indonesia Indicator.

b) *Pre - processing data*

Tahap ini merupakan tahap yang bertujuan untuk mengolah data *Clustering* agar data sesuai dengan kebutuhan, dalam tahap *pre-processing* memiliki beberapa proses yaitu:

1. *Data Reduction*

Proses untuk mengurangi dimensi, seperti penghapusan field yang tidak digunakan dalam proses selanjutnya, di mana proses ini berguna untuk mendapatkan field dan sejumlah data yang akan digunakan.

2. *Data Cleansing*

Proses ini bertujuan untuk membersihkan data, agar data yang diolah nantinya minim *noise*, dan sangat berpengaruh terhadap hasil akhirnya. Contoh hal yang akan dibersihkan adalah penghapusan kata yang umum dibicarakan seperti kita, sama kami atau sering disebut juga *stop words*.

3. *Data Transformation*

Penelitian ini menggunakan metode *k-medoids*, di mana metode tersebut menggunakan data yang berupa angka, maka dalam proses ini data dari berita harian tersebut ditransformasikan ke dalam

bentung angka dengan menggunakan metode TF-IDF, di mana tf-idf dapat mentransformasikan isi berita yang ada ke dalam angka yang dapat dimengerti oleh metode k-medoids

4. *Data Integration*

Data integration merupakan proses yang digunakan untuk menggabungkan data dari beberapa sumber. Jika data yang dibutuhkan berasal dari berbagai sumber data. Langkah yang wajib dilakukan adalah mengintegrasikan skema, entitas, dan lain – lain.

c) *K-medoids clustering*

Setelah data yang didapat diolah melalui proses diatas, maka data tersebut sudah siap untuk diolah dengan menggunakan metode *k-medoids*. Metode tersebut adalah metode yang akan digunakan untuk mengelompokkan data ke dalam kumpulan data dengan titik pusat yang berbeda beda. Tahapan ini memiliki proses sebagai berikut:

1. Menentukan titik pusat cluster

Proses ini memiliki tujuan untuk menentukan titik awal sebagai patokan awal untuk setiap data mencari jarak antara data dan titik pusat cluster.

2. Menghitung jarak

Proses ini memiliki fungsi untuk menghitung jarak setiap data yang ada terhadap *centroid* atau titik awal *cluster*.

3. Alokasi data terhadap *cluster*

Pada proses ini data yang sudah dihitung jaraknya terhadap masing masing *centroid* akan melakukan proses *clustering*. Untuk parameter pengelompokan ini dilihat dari hasil perhitungan pada proses ke 2 dan diambil jarak terdekat.

4. Menentukan *centroid cluster* baru

Setelah setiap data mendapatkan mengelompokannya masing – masing selanjutnya melakukan perhitungan jarak terhadap data yang ada pada masing – masing *cluster*, dan memiliki kemungkinan membentuk *centroid* baru.

5. Validasi *clustering*

Dengan cara melakukan iterasi sejumlah tertentu jika dalam iterasi tersebut tidak mendapatkan perubahan atas *centroid* dan anggota *cluster* tersebut maka proses dari *clustering* sudah bisa dibilang selesai.

d) Pengujian

Pengujian ini dilakukan dengan cara melakukan pengukuran nilai *Silhouette* [13] dan *Purity Score* [14] pada *process clustering* yang dilakukan dengan menggunakan 2 dataset dengan jumlah yang berbeda yaitu 5 ribu data dan 10 ribu data dengan perbandingan *proses* , Hasil dari tes ini untuk menentukan model yang dipakai untuk melakukan prediksi data pada bagian *API*.

3.1.3. Pengembangan Visualisasi

Pengembangan visualisasi dapat dilakukan jika proses diatas sudah selesai dan sudah mendapatkan kan hasil berupa model yang dapat diambil dan digunakan lagi ke dalam *API* dan melakukan proses *clustering* berdasarkan data yang ada pada *database* untuk menghasilkan visualisasi.