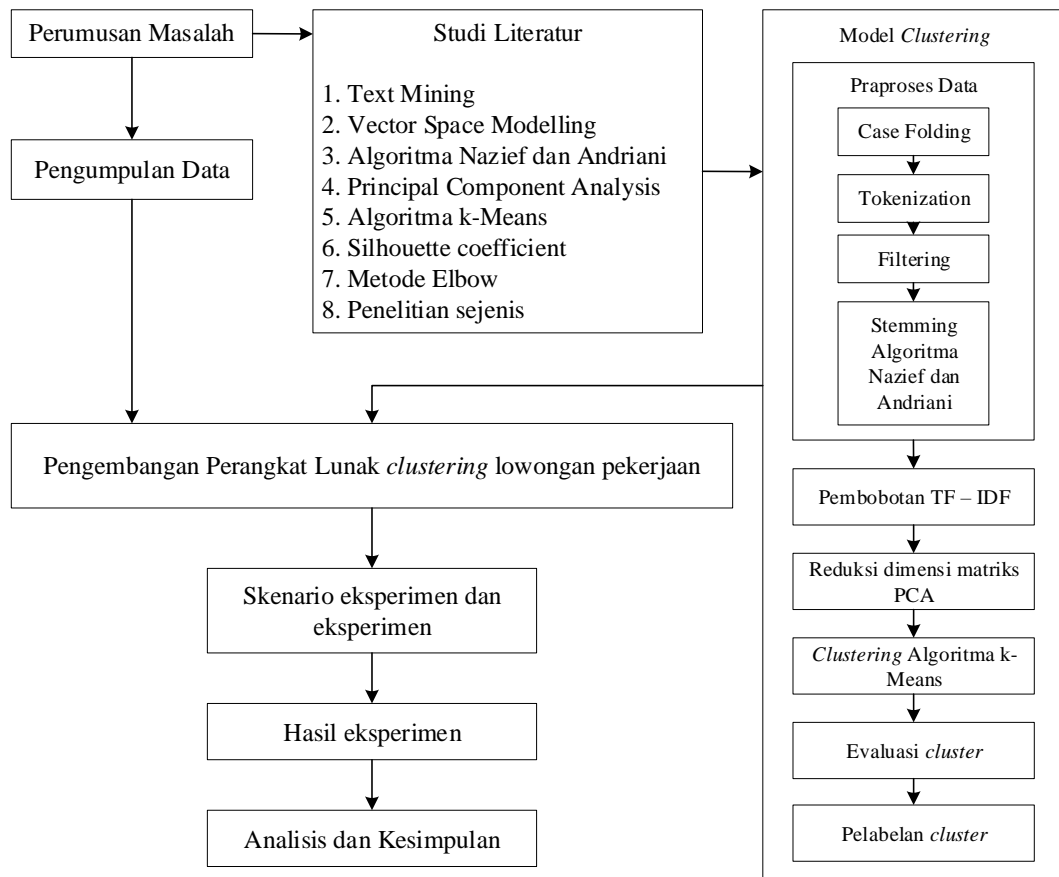


# BAB III

## METODOLOGI PENELITIAN

### 3.1 Desain Penelitian

Desain penelitian adalah kerangka kerja yang digunakan untuk melakukan penelitian. Pada bagian ini penulis akan memaparkan kerangka kerja dari mulai penelitian sampai selesai.



**Gambar 3.1** Desain Penelitian

Pada Gambar 3.1 diatas menggambarkan garis besar kerangka kerja yang akan dilakukan oleh penulis pada penelitian ini. Penjelasan lebih lengkap akan dijabarkan sebagai berikut.

### 3.1.1 Perumusan Masalah

Pada tahap ini penulis memulai penelitian dengan berkonsultasi dengan pembimbing 1 Ibu Dr. Rani Megasari, M.T.. Beliau menjelaskan mengapa topik ini perlu didiskusikan dan menggali masalah terkait. Beberapa permasalahan yang muncul mengenai lowongan pekerjaan diharapkan dapat terselesaikan dengan dilakukannya topik ini. Akses untuk mendapatkan data diberikan oleh beliau. Setelah perumusan selesai, maka penelitian dapat dimulai dengan mempelajari beberapa literatur terkait dan memulai untuk mengumpulkan data.

### 3.1.2 Pengumpulan Data

Pengumpulan data dimulai setelah perumusan masalah, karena pengumpulan data dapat dimulai secara *parallel* dengan studi literatur. Data didapatkan melalui *platform* sosial media yang digunakan oleh alumni dari Departemen Pendidikan Ilmu Komputer, salah satunya adalah grup *Facebook* alumni. Pengumpulan data dilakukan secara manual sesuai data yang masuk dalam grup tersebut.

### 3.1.3 Studi Literatur

Dalam tahap ini, dilakukan pengumpulan dan pemahaman berbagai teori serta pembahasan terkait penelitian yang akan dilakukan dari berbagai sumber yang valid untuk dijadikan referensi. Sumber literatur yang didapatkan berasal dari berbagai media seperti, jurnal, buku, *website*, dan video. Adapun subjek teori yang telah didapat untuk dipahami mengenai *text mining*, *principal component analysis*, TF-IDF, perhitungan distance *euclidean*, *clustering* algoritma *k-Means*, *silhouette coefficient*, *sum of squared error* serta penelitian serupa yang telah dilakukan sebelumnya.

### 3.1.4 Model *clustering*

Sebelum melakukan implementasi ke dalam bentuk *coding*, perlu dibuatkan sebuah model sebagai kerangka kerja pada saat tahap implementasi. Model yang dirancang berdasarkan studi literatur yang telah dipelajari sebelumnya. Dalam perancangan model *clustering*, alur dimulai dengan melakukan praproses data, lalu melakukan pembobotan kata TF-IDF, pereduksi dimensi PCA, pengelompokan *k-Means*, evaluasi *cluster* menggunakan *silhouette* dan SSE, serta terakhir pelabelan

*cluster* yang telah terbentuk. Untuk lebih jelasnya, berikut penjelasan dari sub proses yang terdapat pada model *clustering*.

#### 1. Praproses Data

Sebelum data dikelompokkan, data mengalami tahap praproses terlebih dahulu. Langkah-langkah dalam praproses meliputi penyeragaman huruf kapital menjadi huruf kecil, dan penghilangan karakter selain huruf, tahap ini disebut *case folding*. Setelah *case folding* data mengalami pemisahan kalimat menjadi per-kata yang disebut dengan *tokenizing*. Lalu data mengalami pengurangan kata *stop word* yang disebut *filtering*. Terakhir data akan mengalami pemenggalan kata dasar yang disebut *stemming*.

#### 2. Pembobotan TF-IDF

Data yang telah dilakukan praproses akan diproses dengan melakukan pembobotan. Proses ini mengubah data teks tidak terstruktur menjadi matriks jarang (*sparse matrix*). Yang disebut matriks jarang adalah isi dari matriks ini sebagian besar berupa 0. Sehingga apabila matriks ini diproses pada pengelompokan akan membuat hasil pengelompokan menjadi kurang bagus.

#### 3. Pengurangan Dimensi Matriks PCA

Hasil matriks dari pembobotan sebagian besar adalah 0 dan berdimensi besar. Jika ini dipaksakan untuk dilakukan *clustering*, tetap dapat dikelompokkan namun hasilnya tidak akan bagus. Maka dari itu, perlu diubah menjadi data baru yang tetap merepresentasikan data itu namun tidak akan 100% sama. Metode pengurangan dimensi PCA merupakan solusinya.

#### 4. *Clustering* Algoritma *k-Means*

Setelah data mengalami pengurangan dimensi dan mendapatkan dimensi baru dari PCA, langkah berikutnya adalah pengelompokan menggunakan algoritma *k-Means*. Pengelompokan dilakukan dari jumlah *cluster* sebanyak 2 hingga 10 *cluster*. *Stopping criteria* dari algoritma *k-Means* pada penelitian ini ketika anggota

*cluster* tidak berubah, maksimum iterasi pada 100 iterasi, dan anggota *cluster* tidak berubah sebanyak 10 kali (ini untuk mengatasi *local optimum*).

#### 5. Evaluasi

Dari *cluster* yang telah terbentuk, akan dilakukan pengujian berdasarkan metrik yang telah ditentukan, yaitu SSE dan *silhouette*. Tujuan dari evaluasi ini merupakan untuk dianalisis dan mendapatkan jumlah *cluster* paling optimal.

#### 6. Pelabelan *cluster*

Setelah pengelompokan berhasil dilakukan, tahap berikutnya adalah memberikan *cluster* label agar dapat dimengerti oleh pengguna. Pelabelan diambil dari kata yang sering muncul pada anggota yang terkumpul.

### 3.1.5 Pengembangan Perangkat Lunak *clustering* lowongan pekerjaan

Setelah dilakukan studi literatur dan pengumpulan data, maka pengembangan perangkat lunak baru dapat dilakukan. Pengembangan didasarkan pada alur yang telah dirancang sebelumnya dan menggunakan model *waterfall* (Sommerville, 2010) sebagai metode pengembangan perangkat lunak. Pengembangan menggunakan bahasa pemrograman *Python 3* dan *web application Jupyter notebook*.

### 3.1.6 Skenario Eksperimen dan Eksperimen

Pada tahap ini perangkat lunak sudah selesai dikembangkan dan akan dibuat sebuah skenario eksperimen. Tujuannya merupakan untuk menguji data yang telah dikumpulkan serta untuk mendapatkan hasil yang beragam. Skenario yang akan dilakukan terbagi ke dalam 2 skenario, yang pertama dengan memberikan batas minimal kemunculan kata dengan tidak memberikan batas minimal kemunculan kata. Skenario ini diaplikasikan dalam *parameter* pada fungsi TF-IDF. Setelah skenario dibuat maka langsung diaplikasikan ke dalam perangkat lunak.

### 3.1.7 Hasil Eksperimen

Hasil eksperimen pada kedua skenario yang telah dilakukan akan dicatat dan dikumpulkan. Hasil ini akan dilakukan analisis pada tahap berikutnya.

### 3.1.8 Analisis dan Kesimpulan

Pada tahap ini hasil dari eksperimen dilakukan analisis. Mengenai hasil dari perbedaan kedua skenario dibahas pada tahap ini. Hasil evaluasi ditampilkan ke dalam sebuah grafik untuk mempermudah analisis. Dari beberapa analisis yang dilakukan, akan dibuat menjadi kesimpulan.

## 3.2 Metode Penelitian

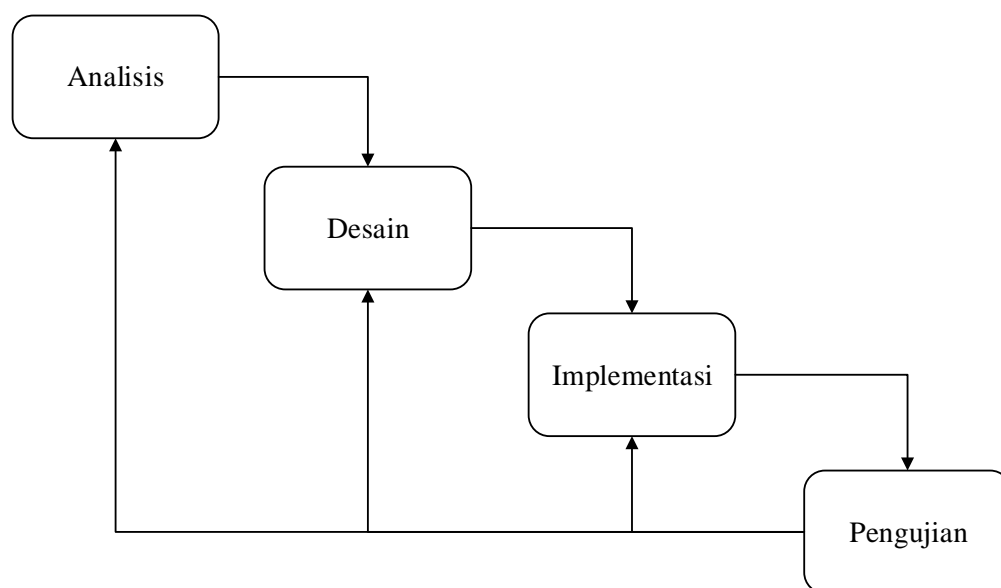
Metode penelitian yang dilakukan dalam melakukan penelitian ini terbagi ke dalam dua bagian, meliputi metode yang digunakan dalam proses pengumpulan data serta metode yang digunakan dalam proses pengembangan perangkat lunak. Penjelasan mengenai kedua tahapan metode tersebut akan dijelaskan lebih lanjut sebagai berikut.

### 3.2.1 Metode Pengumpulan Data

Metode pengumpulan data adalah dengan cara menyalin secara manual pada grup Alumni Departemen Pendidikan Ilmu Komputer yang terdapat pada *platform Facebook*. Sebuah lowongan pekerjaan disalin dan ditempel menjadi satu buah data berupa teks panjang satu baris. Setiap data yang diambil hanya data yang terdapat dalam grup tersebut, dan tidak berasal dari luar grup, sehingga data benar-benar valid dari alumni. Data yang diambil berupa teks dan tentunya tidak terstruktur.

### 3.2.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan dalam pembangunan penelitian ini adalah dengan pendekatan model *waterfall*, yang ditunjukkan pada Gambar 3.9.



**Gambar 3.2** Model Pengembangan Perangkat Lunak *Waterfall*

Metode pengembangan perangkat lunak dilakukan dengan metode waterfall. Model SDLC air terjun (*waterfall*) sering juga disebut model sekuensial linier (*sequential linier*). Model *waterfall* menyediakan pendekatan alur hidup perangkat lunak secara sekuensial atau urut dimulai dari analisis, desain, pengkodean, pengujian dan tahap support (Sukamto & Shalahuddin, 2011). Penulis menggunakan metode modern waterfall seperti pada Gambar 3.2 agar jika suatu saat ada kesalahan pada salah satu tahap, bisa dikembalikan ke tahap sebelumnya. Berikut pengertian dari tahap-tahap pada model waterfall pada Gambar 3.2 menurut (Sommerville, 2010) :

1. *Requirements Analysis and Definition* (Analisis)

Analisis adalah tahap menentukan aplikasi atau *software* seperti apakah yang akan dibuat. Analisis merupakan tahapan penetapan fitur, kendala dan tujuan sistem melalui konsultasi dengan pengguna sistem. Semua hal tersebut akan ditetapkan secara rinci dan berfungsi sebagai spesifikasi sistem. Analisis ini terdiri atas analisis kebutuhan dan analisis pembuatan sistem.

2. *System and Software Design* (Desain)

Dalam tahapan ini akan dibentuk suatu arsitektur sistem berdasarkan persyaratan yang telah ditetapkan. Dan juga mengidentifikasi dan menggambarkan abstraksi dasar sistem perangkat lunak dan hubungan-hubungannya. Desain terdiri atas desain database, desain arsitektur system, dan desain antarmuka (*user interface*).

### 3. *Implementation and Unit Testing* (Implementasi)

*Coding* adalah tahap proses implementasi dari desain, dalam tahapan ini, hasil dari desain perangkat lunak akan direalisasikan sebagai satu set program atau unit program. Setiap unit akan diuji apakah sudah memenuhi spesifikasinya.

### 4. *Integration and System Testing* (Pengujian)

Proses testing atau pengujian dilakukan pada logika internal untuk memastikan semua pernyataan sudah diuji. Dalam tahapan ini, setiap unit program akan diintegrasikan satu sama lain dan diuji sebagai satu sistem yang utuh untuk memastikan sistem sudah memenuhi persyaratan yang ada. Setelah itu sistem akan dikirim ke pengguna sistem.

## 3.3 Perangkat dan Data Penelitian

Dalam menunjang keberlangsungan penelitian, penulis memanfaatkan alat dan bahan yang digunakan selama kegiatan ini berlangsung. Adapun alat dan bahan ini meliputi perangkat keras, perangkat lunak, bahan kajian beserta sejumlah data yang akan dijelaskan sebagai berikut.

### 3.3.1 Perangkat Penelitian

Perangkat yang digunakan dalam penelitian ini meliputi perangkat keras dan perangkat lunak. Berikut daftar alat yang digunakan selama penelitian ini berlangsung.

Perangkat Keras (*Hardware*) yaitu komputer dengan spesifikasi:

- *Processor Intel® Core™ i3 CPU M390 @2.67GHz (4 CPUs), ~2.7GHz*
- Memori 8192MB RAM
- Intel® HD Graphics

- HDD 500 GB
- SSD 250 GB

Perangkat Lunak (*software*) sebagai berikut:

- Sistem Operasi Windows 10 Pro 64bit (10.0, Build 17134)
- Browser Google Chrome versi 61.0.3163.100 (Official Build) (64-bit)
- Python 3.7.0
- Jupyter Notebook
- Anaconda

### **3.3.2 Data Penelitian**

Data penelitian yang digunakan dalam kegiatan penelitian ini terdiri atas data lowongan pekerjaan yang dibagikan oleh alumni Departemen Pendidikan Ilmu Komputer melalui *platform* sosial media *Facebook*. Data yang diambil memiliki jangka waktu 2014 hingga 2019 saja, di luar itu tidak akan penulis ambil sebagai data.