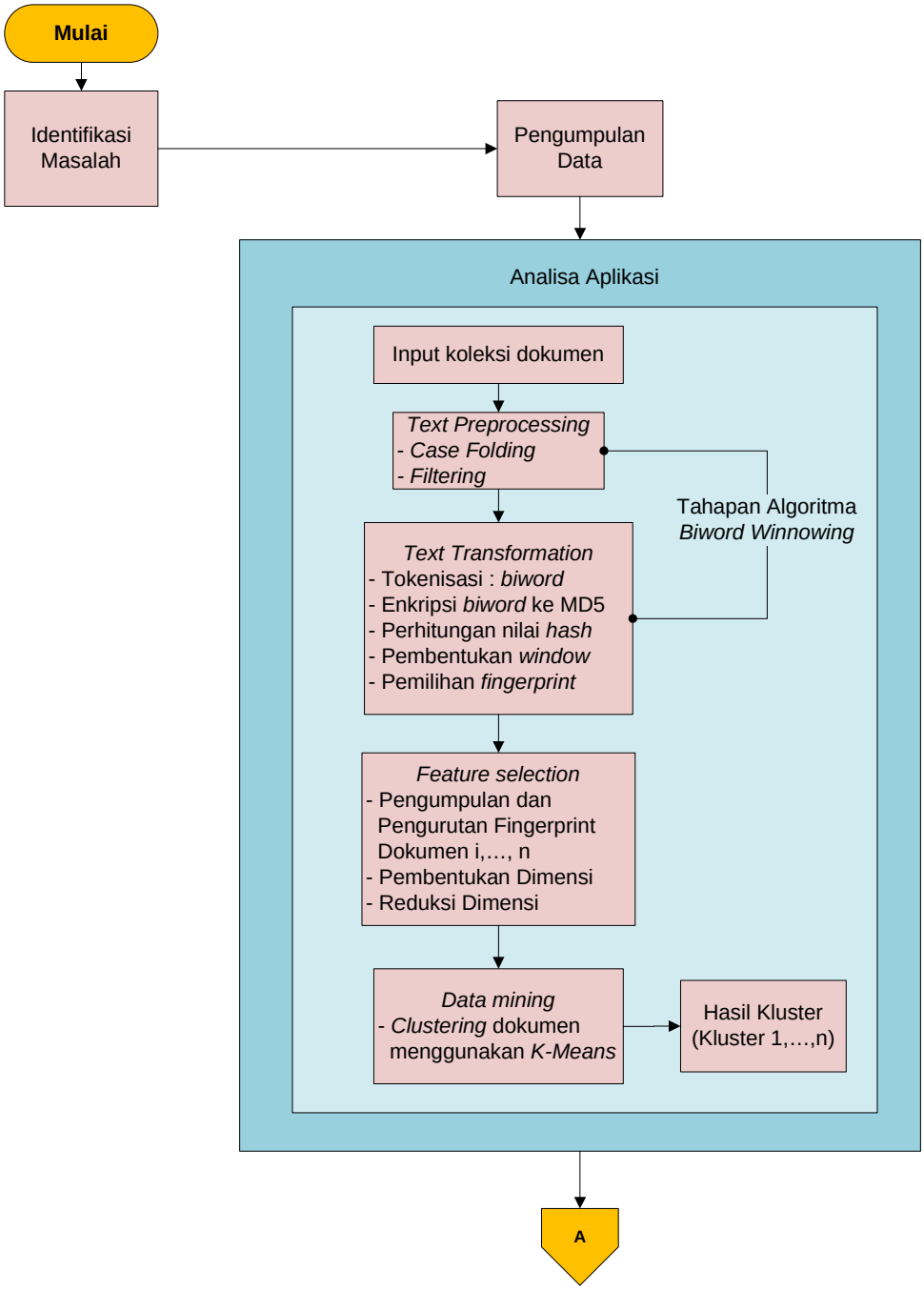
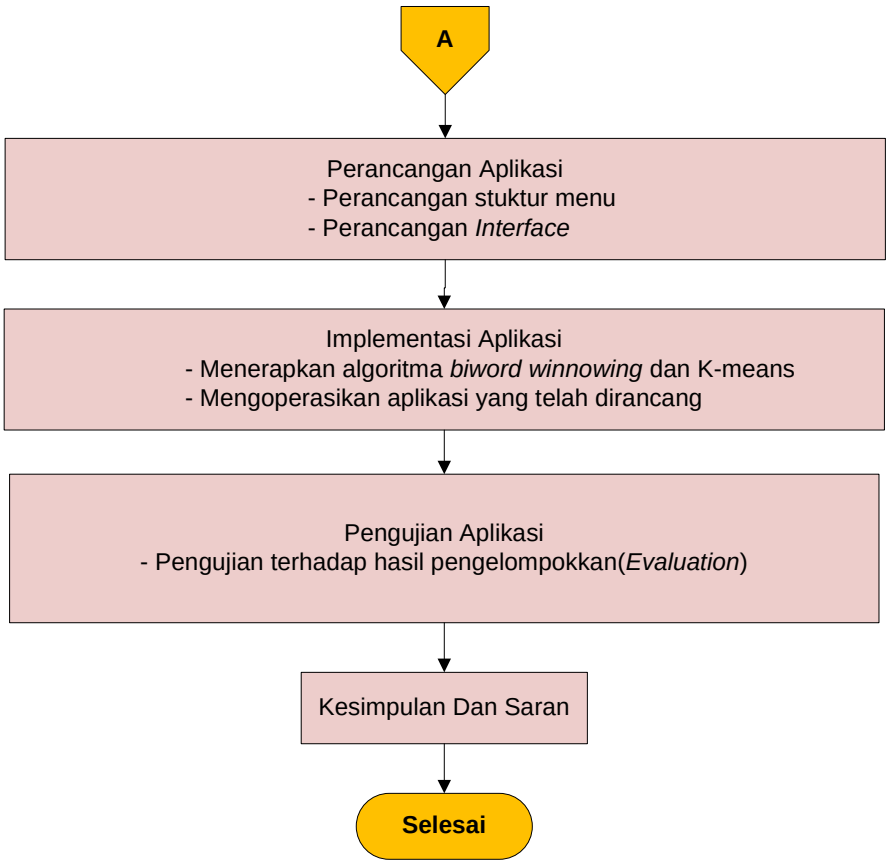
BAB III

METODOLOGI PENELITIAN

Pada penelitian ini ada beberapa tahapan penelitian yang akan dilakukan seperti terlihat pada gambar 3.1 berikut :





Gambar 3.1 Tahapan Penelitian

3.1 Identifikasi Masalah

Pada tahap ini penulis melakukan identifikasi terhadap masalah yang akan diteliti, yakni dengan cara mencari isu yang sedang marak dibahas. Kemudian penulis menetapkan penyelesaian untuk masalah yang ditemukan. Isu yang penulis pilih yakni mengenai *clustering* dokumen teks.

3.2 Pengumpulan Data

Setelah masalah yang akan diteliti didapat, penulis melakukan pengumpulan data dengan cara mencari referensi-referensi terkait yang dibutuhkan untuk penelitian. Referensi tersebut dapat berupa buku-buku, jurnal-jurnal, tulisan penelitian dan juga artikel-artikel dari internet yang memiliki kaitan dengan kasus mengenai *clustering* dokumen.

3.3 Analisa Aplikasi

Setelah melakukan hipotesa, langkah selanjutnya adalah melakukan analisa terhadap aplikasi yang akan dibangun dengan cara menganalisa hal-hal yang berhubungan dengan *clustering* dokumen teks berdasarkan *fingerprint biword winnowing* dengan metode *K-Means*. Beberapa tahapan dalam analisa yang akan dilakukan adalah :

1. *Text Preprocessing*
   a. *Case Folding*
      Pengubahan setiap huruf besar menjadi huruf kecil.
   b. *Filtering*
      penghilangan karakter atau simbol yang tidak relevan.
2. *Text Transformation*
   a. Tokenisasi/*Tokenizing*
      Membagi dokumen menjadi token berbentuk *biword*.
   a. Eksport menjadi nilai MD5
      Pengubahan token *biword* yang telah terbentuk menjadi nilai MD5 sepanjang 32 karakter, sehingga setiap token memiliki panjang karakter yang sama.
   b. Perhitungan Nilai *Hash*
      Perhitungan nilai *hash* dari masing-masing token *biword* yang telah berbentuk nilai MD5 menggunakan persamaan *rolling hash*.
   c. Pembentukan *window-window*
      Pembagian nilai-nilai *hash* yang diperoleh ke dalam beberapa *window* dengan ukuran w.
   d. Memilih *Fingerprint*
      Pemilihan nilai *hash* terkecil dari tiap-tiap *window*.
3. *Feature Selection* (Analisa Tahapan Pembentukan vektor)
   a. Pengumpulan dan pengurutan data *fingerprint* dari masing-masing dokumen.
   b. Perhitungan frekuensi *fingerprint* dari masing-masing dokumen.
   c. Pengambilan nilai *frekuensi* dari masing-masing dokumen sebagai representasi dimensi dokumen.

d. Melakukan proses reduksi terhadap dimensi dokumen yang telah terbentuk.

4. *Data Mining*
   a. Penentuan jumlah kluster sebanyak k kluster.
   b. Pembentukan *centroid* secara acak untuk masing-masing kluster. Pada penelitian ini penulis menggunakan 2 cara pembentukan *centroid*. Pada cara pertama, pembentukan *centroid* aplikasi *clustering* berdasarkan dokumen yang telah ada yang dipilih secara acak. Sedangkan pada cara kedua, pembentukan *centroid* dilakukan berdasarkan *range* nilai dimensi dokumen terbesar dan terkecil dari keseluruhan dokumen, lalu menentukan nilai dimensi dokumen *centroid* secara acak berdasarkan *range* tersebut.
   c. Hitung jarak antara *centroid* dengan masing-masing dokumen menggunakan rumus jarak *cosine similarity*.
   d. Tentukan kembali *centroid* untuk masing-masing kluster. Jika nilai *centroid* berubah, ulang ke langkah c. Jika nilai *centroid* tetap, hentikan proses.

3.4 Perancangan Aplikasi

Pada tahap ini akan dilakukan perancangan aplikasi *clustering*. Tahap ini bertujuan untuk memberikan gambaran kepada pengguna terhadap aplikasi *clustering* dokumen yang akan dibangun. Tahapan yang akan dilakukan adalah :

a. Perancangan Struktur Menu

   Merancang struktur menu pada aplikasi yang memiliki fungsi masing-masing sesuai tujuan yang akan dicapai. Adapun menu-menu yang terdapat dalam aplikasi adalah :

   - Menu *Home* : halaman yang akan memperkenalkan aplikasi kepada pengguna
   - Menu *Form* Kluster : halaman yang akan digunakan untuk memulai proses pengelompokkan dokumen.
   - Menu *Text Preprocessing* : halaman yang akan menampilkan hasil dari proses *preprocessing* dokumen yang telah dilakukan.

- Menu *Text Transformation* : halaman yang akan menampilkan hasil dari proses *text transformation* yang telah dilakukan.
- Menu *Feature Selection* : halaman yang akan menampilkan hasil dari proses *feature selection* yang telah dilakukan.
- Menu *Data Mining* : halaman yang akan menampilkan hasil dari proses *data mining* yang telah dilakukan, yaitu hasil pengelompokkan.
- Menu *Help* : halaman yang akan memberikan informasi atau petunjuk cara penggunaan aplikasi kepada pengguna.

b. Perancangan *Interface* Aplikasi

Merancang atau mendesain tampilan antar muka aplikasi dengan pengguna. *Interface* yang akan dibangun adalah *interface input* dan *output*. Dengan demikian akan terlihat *interface* dari aplikasi dan dapat memberikan gambaran terhadap aplikasi yang akan dibangun.

c. Perancangan *pseudo code* aplikasi, agar pembaca dapat memahami algoritma yang terdapat dalam perancangan aplikasi *clustering* tahap demi tahap.

3.5 Implementasi

Setelah dilakukan perancangan aplikasi, maka akan dilakukan tahap implementasi. Implementasi merupakan tahap di mana aplikasi siap untuk dioperasikan sesuai dari hasil analisis dan perancangan yang telah dilakukan, sehingga akan diketahui apakah aplikasi yang dirancang benar-benar dapat menghasilkan tujuan yang ingin dicapai.

Adapun spesifikasi *hardware* dan *software* yang akan digunakan pada implementasi ini adalah :

a. Perangkat Lunak (*Software*)

Sistem Operasi	: *Windows 7 Home Premium 32-bit OS*
Bahasa Pemrograman	: *Hypertext Prepocessor* (PHP)
Tool	: *Notepad++*

b. Perangkat Keras (*Hardware*)

Processor	: *Intel Pentium Dual Core* 2.20 GHz
Memori (RAM)	: 1 GB
Harddisk	: 160 GB

3.6 Pengujian

Setelah tahap implementasi, selanjutnya akan dilakukan pengujian terhadap aplikasi yang telah dibangun. Dengan demikian dapat diketahui tingkat keberhasilan aplikasi, apakah telah mencapai tujuan yang diharapkan.

Pengujian terhadap aplikasi *clustering* dokumen teks dilakukan dengan mengevaluasi hasil pengelompokkan (Menggunakan pembentukan *centroid* cara I dan cara II) dengan menggunakan beberapa dokumen uji, lalu menginputkan jumlah kelompok. Sehingga dapat dianalisa mengenai efektifitas metode *clustering k-means* berdasarkan *fingerprint byword winnowing*. Apakah dokumen-dokumen yang berada pada satu kluster benar-benar memiliki tingkat kesamaan yang tinggi dan dokumen pada kluster yang berlainan memiliki tingkat kesamaan yang rendah. Dalam tahapan *text mining*, tahapan ini masuk ke dalam tahap *evaluation*.

3.7 Kesimpulan Dan Saran

Tahapan ini meliputi tentang kesimpulan yang dapat ditarik dari penelitian yang telah dilakukan. Selanjutnya pada bagian saran berisi tentang saran-saran yang dapat diberikan demi penelitian yang lebih baik mengenai tema yang sama ataupun berkaitan dengan penelitian yang penulis lakukan.