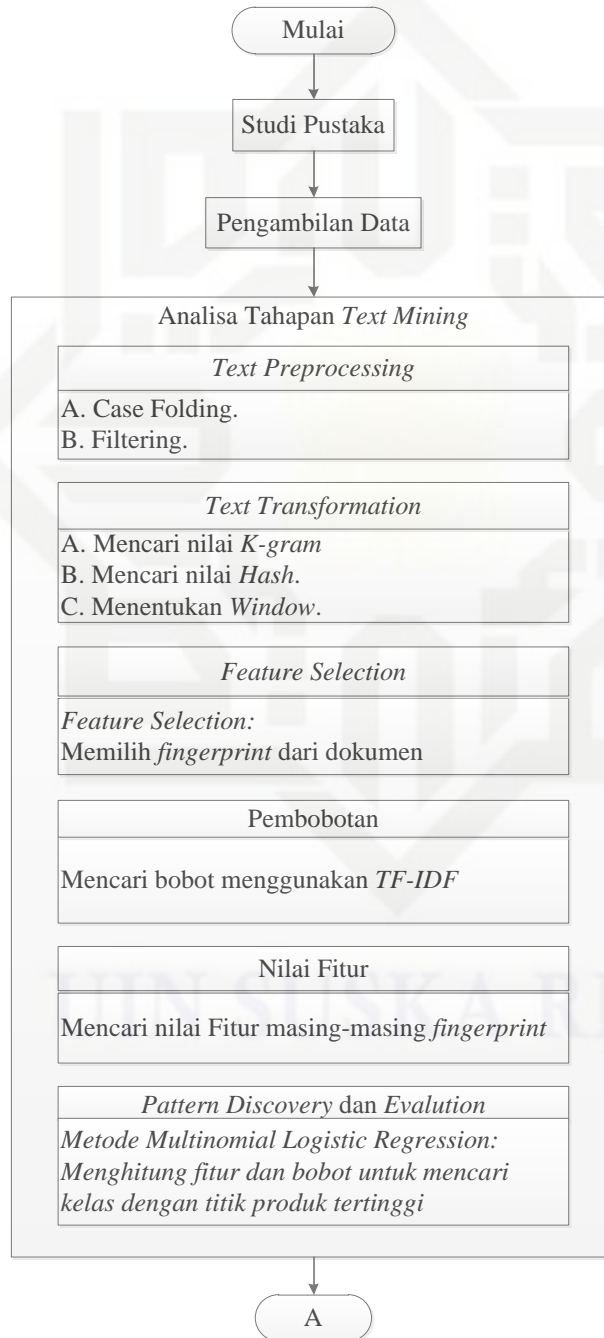


Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III METODOLOGI PENELITIAN

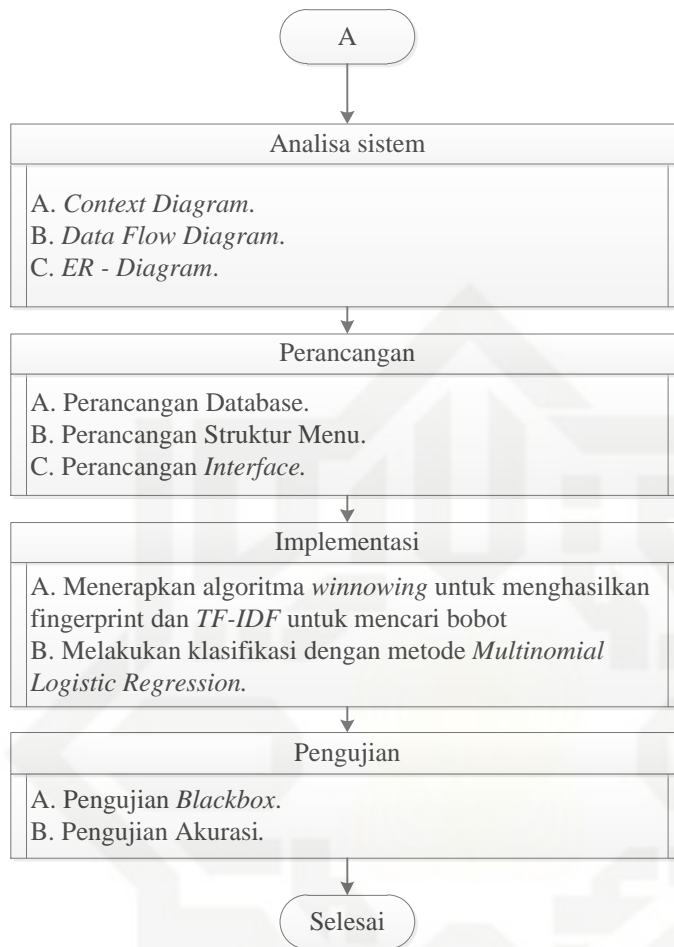
Pada penelitian ini ada beberapa tahapan penelitian yang akan dilakukan sebagai berikut :



Gambar 3.1 Tahapan Penelitian

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 3.2 Tahapan Penelitian Lanjutan

3.1 Studi Pustaka

Mencari referensi terkait yang dibutuhkan. Referensi dapat berupa jurnal, tulisan penelitian sebelumnya, buku dan artikel dari internet yang memiliki kaitan dengan kasus yang sedang dilakukan dalam penelitian yaitu *text mining*, *algoritma winnowing*, *TF-IDF* dan *Multinomial logistic regression*.

3.2 Pengumpulan Data

Pengumpulan data berasal dari data sekunder dari penelitian Radili tahun 2016 yang berasal dari situs *Digital repository* ITS (Institut Teknologi Sepuluh Nopember) Surabaya. Data yang di gunakan berupa dokumen text digital yang mempunyai format *.pdf* dan *.txt*. Kelompok yang digunakan terdiri dari 8

kelompok, yaitu : Komputasi Cerdas dan Visualisasi (KCV), Komputasi Berbasis Jaringan (KBJ), Rekayasa Perangkat Lunak (RPL), Interaksi Grafika dan Seni (IGS), Arsitektur dan Jaringan Komputer (AJK), Algoritma dan Pemrograman (AP), Dasar dan Terapan Komputasi (DTK) dan Manajemen Informasi (MI).

Jumlah data yang digunakan (Radili, 2016) berjumlah 1050 dokumen. Perbandingan data latih dan data uji pada penelitian ini adalah 90 : 10 dengan mengambil kutipan penelitian (Raharjo & Winarko, 2014) yang menyatakan bahwa data latih dan uji dengan perbandingan 90 : 10 secara umum cukup banyak yang menggunakan.

3.3 Analisa Tahapan *Text Mining*

Merupakan tahap yang dilakukan setelah mengumpulkan data yang diperlukan. Beberapa tahapan analisa yang dilakukan, adalah:

1. *Text Preprocessing*

Pada tahap ini dilakukan penghapusan karakter yang tidak relevan, seperti spasi, angka, tanda baca yaitu ~, !, @, #, \$, %, ^, &, *, (,), _, +, *, /, ?, ;, :,], }, \, |, ', ", <, >, =, +, -, _, dan simbol yaitu j, ¢, £, ¤, ¥, §, ¨, ©, ª, «, ¬, ®, ¯, °, ±, ², ³, µ, ¶ serta :

- 1) Case Folding, yaitu mengubah huruf besar menjadi huruf kecil.
- 2) Filtering, yaitu mengambil kata penting menggunakan *stopwords* untuk menghilangkan kata penghubung. Yaitu: dan, atau, tetapi, sedangkan, serta, melainkan, padahal, sesudah, setelah, sejak, selesai, ketika, sebelum, seandainya, seumpamanya, andaikan, sekiranya, jika, kalau, agar, supaya, meskipun, walaupun, seolah-olah, seakan-akan, sehingga, sampai, karena, sebab, bahwa, dengan, biarpun begitu, sesudah itu, sebaliknya, sesungguhnya, bahkan, akan tetapi.

2. *Text Transformation*

Tahap ini berguna untuk mentransformasi teks :

- 1) Menerapkan metode *k-gram*
- 2) Menerapkan teknik *hashing*
- 3) Membentuk *window*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4) Memilih *fingerprint*

3. *Feature Selection*

Pada tahapan ini dilakukan pengumpulan semua *fingerprint* yang ada.

4. Pembobotan

Pada tahapan ini dilakukan perhitungan mencari bobot pada kata menggunakan *TF-IDF*.

5. *Pattern Discovery* dan *Evaluation*

Pada tahapan ini digunakan metode klasifikasi *Multinomial Logistic Regression* untuk mencari nilai kemiripan dokumen dengan melihat titik produk tertinggi berdasarkan bobot dan fitur.

Kemudian menganalisa sistem dengan menggunakan *context* diagram, *data flow* diagram, dan *ER* - diagram.

3.4 Perancangan Sistem

Merupakan tahapan untuk merancang sistem yang telah dianalisa dengan tujuan untuk memberikan kemudahan dan menyederhanakan suatu proses aliran data, perancangan terhadap model, dan merancang bangun sistem, tahap yang dilakukan yaitu :

1. Perancangan *Database*.
2. Perancangan Struktur Menu.
3. Perancangan Antar Muka atau *Interface*.

3.5 Implementasi

Tahap ini akan dilakukan pembuatan modul yang telah dirancang dan dianalisa selanjutnya diimplementasikan pada bahasa pemrograman dan dilakukan pengujian untuk mengetahui tingkat keberhasilan aplikasi yang telah ada. Modul berisi penerapan algoritma *winnowing* untuk menghasilkan *fingerprint* dan pembobotan dengan menggunakan *TF-IDF*. Selanjutnya adalah melakukan klasifikasi dengan metode *Multinomial Logistic Regression*.

Adapun pada tahap implementasi ini, spesifikasi lingkungan implementasi perangkat keras dan perangkat lunak adalah sebagai berikut:

1. Perangkat Keras

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- a. *Processor* : AMD FX(tm)-6300 Six-Core Processor 3.50 GHz
- b. *Memory* : RAM 8 GB DDR3
2. Perangkat Lunak
 - a. Sistem Operasi : *Windows 7 Ultimate*
 - b. Bahasa Pemrograman : PHP

3.6 Pengujian

Pengujian *blackbox* dilakukan dengan menguji sistem yang telah dibuat apakah sesuai dengan harapan atau tidak. Pengujian *confusion matrix* dilakukan dengan menguji apakah hasil klasifikasi dokumen sesuai dengan prediksi serta mencari tingkat akurasi.

3.7 Kesimpulan dan Saran

Tahapan kesimpulan dan saran merupakan tahap akhir dari penelitian tugas akhir. Tahapan ini membahas hasil evaluasi dari seluruh kegiatan yang dilakukan dalam melakukan penelitian serta saran dari penulis berikan untuk pengembangan penelitian selanjutnya.