

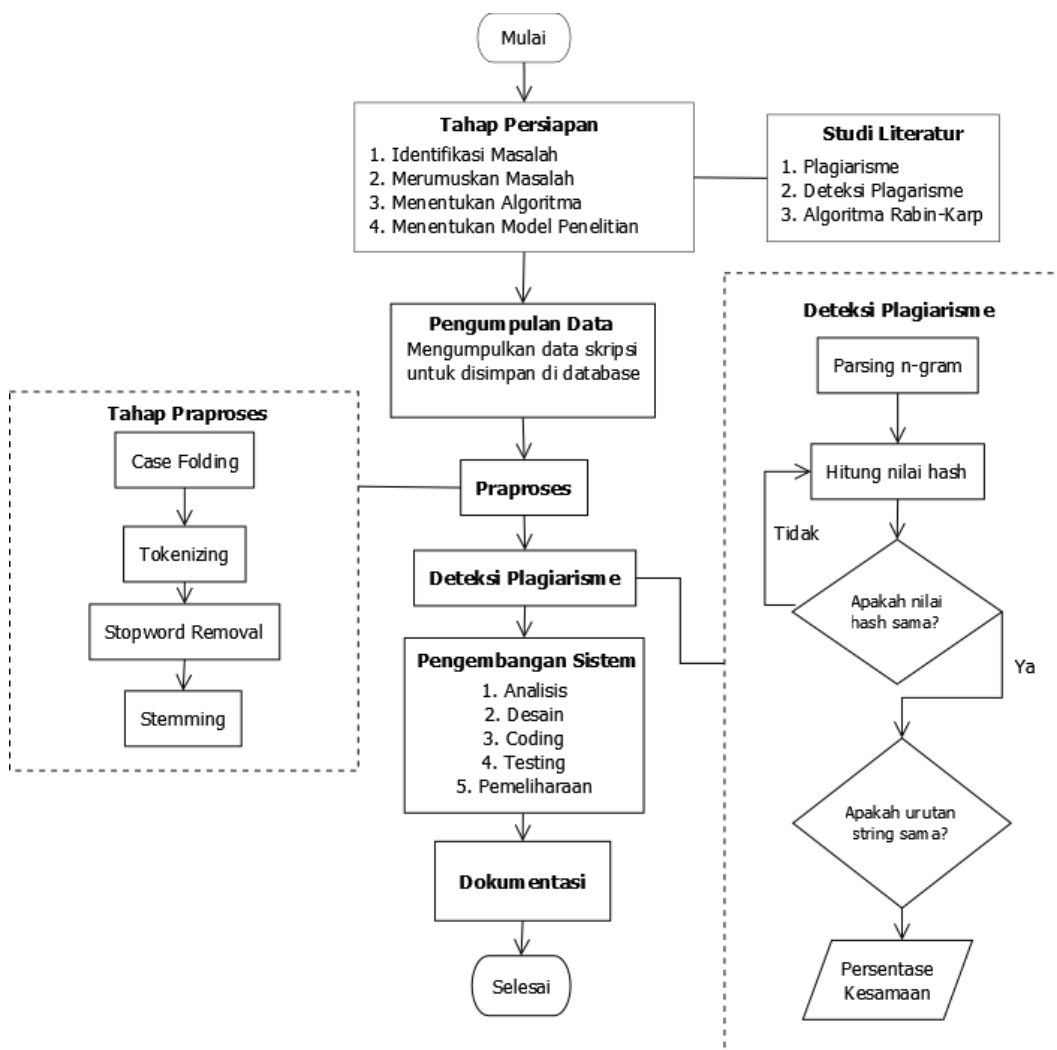
## BAB III

### METODOLOGI PENELITIAN

Bab ini memaparkan tentang metodologi penelitian yang terdiri dari desain penelitian, alat dan bahan penelitian, serta metode penelitian yang digunakan.

#### 3.1 Desain Penelitian

Desain penelitian merupakan kerangka kerja penelitian yang dilakukan. Berikut ini akan dipaparkan kerangka kerja dari penelitian ini. Desain penelitian tersebut digambarkan pada Gambar 3.1.



Gambar 3. 1 Desain Penelitian

1. Tahap persiapan adalah tahap awal dari penelitian, tahap ini diawali dengan mengidentifikasi masalah, kemudian merumuskan masalah, lalu mencari metode atau algoritma yang sesuai untuk menyelesaikan masalah tersebut, kemudian membuat model desain penelitian.
2. Studi literatur merupakan tahap persiapan sebelum melakukan penelitian. Pada bagian ini penulis mempelajari dan memahami terlebih dahulu teori yang akan digunakan berkaitan dengan penelitian. Teori yang harus dipahami dalam melakukan penelitian ini diantaranya memahami sekilas mengenai plagiarisme, memahami deteksi plagiarisme, memahami perhitungan algoritma Rabin-Karp. Teori-teori tersebut didapat dari literatur berupa buku, jurnal, *paper*, undang-undang, serta artikel yang terkait dengan penelitian.
3. Pengumpulan data dilakukan dengan mengumpulkan dokumen skripsi yang kemudian dimasukkan ke dalam *database* aplikasi. Bagian bab skripsi yang akan diambil dan digunakan untuk penelitian ini adalah seluruh bab yang terdiri dari judul, abstrak, bab satu hingga bab lima.
4. Setelah data terkumpul, data tersebut akan memasuki tahap praproses atau pembersihan data sebelum dilakukan perbandingan atau pendeteksian plagiarisme. Tahap praproses yang akan dilakukan pada penelitian ini diantaranya :
  - a. *Case Folding*

*Case Folding* merupakan salah satu tahap dalam praproses untuk mengubah semua karakter huruf dalam dokumen menjadi huruf kecil. Karakter atau huruf yang diterima oleh sistem ini adalah huruf dari a sampai z, sedangkan karakter lain akan dihilangkan.
  - b. *Tokenizing*

Proses *Tokenizing* merupakan proses pemisahan kata yang terdapat pada suatu dokumen. Umumnya setiap kata terpisahkan oleh spasi, sehingga proses *tokenizing* mengandalkan karakter spasi pada dokumen untuk melakukan pemisahan kata.

c. *Stopword Removal*

*Stopword Removal* merupakan proses menghapus kata yang tidak diperlukan seperti kata penghubung, kata baku, dan sebagainya.

d. *Stemming*

*Stemming* merupakan suatu proses transformasi kata yang terdapat dalam suatu dokumen menjadi kata dasarnya dengan menghapus imbuhan kata menggunakan aturan-aturan tertentu.

Kemudian setelah data bersih, data siap untuk dibandingkan dan dianalisis persentase kesamaannya.

5. Tahap deteksi plagiarisme adalah tahap dilakukannya seluruh proses algoritma Rabin-Karp. Data yang telah melalui praproses selanjutnya di *parsing* menjadi n-gram dan akan membentuk beberapa *window*, setiap *window* akan dihitung nilai hashnya kemudian nilai hash tersebut dibandingkan dan dicari persentase *similarity* nya.
6. Tahap pengembangan sistem yaitu pembuatan perangkat lunak (*software*) penelitian ini. Pengembangan sistem dilakukan sesuai model pengembangan perangkat lunak *waterfall*. Tahap pertama yang dilakukan yaitu analisis, pada tahap ini dilakukan analisis mengenai bagaimana *software* akan dibuat. Setelah itu masuk tahap desain, tahap ini mencakup desain aplikasi, desain *database* serta desain *interface*. Setelah itu masuk tahap implementasi, tahap implementasi ini yaitu tahap pembuatan *software* atau biasa disebut dengan tahap *coding*, setelah itu dilakukan tahap *testing* atau pengujian *software*.
7. Tahapan ini merupakan tahapan terakhir dalam penelitian yang dilakukan. Pada tahap ini, proses-proses yang dilakukan selama melaksanakan penelitian dan juga hasil dari penelitian yang telah dilakukan akan didokumentasikan. Dokumentasi tersebut akan dibuat dalam bentuk laporan skripsi.

### 3.2 Alat dan Bahan Penelitian

Berikut ini adalah pembahasan lebih detail mengenai alat dan bahan penelitian yang digunakan pada penelitian ini.

### 3.2.1 Alat Penelitian

- a. Perangkat keras yang digunakan berupa laptop dengan spesifikasi
  - Processor Intel Core i3-5005U CPU 2.00 GHz
  - Random Access Memory (RAM) 4 GB
  - Harddisk Drive 500 GB
- b. Perangkat lunak yang digunakan untuk perancangan aplikasi berupa
  - Sublime Text 3
  - Sistem Operasi Windows 8.1 Pro 64 bit
  - Python 2.7.12
  - Python Tkinter GUI

### 3.2.2 Bahan Penelitian

Bahan yang digunakan pada penelitian ini yaitu dokumen skripsi mahasiswa berupa judul, abstrak, bab satu sampai bab lima.

### 3.3 Metode Penelitian

Adapun metode yang dilakukan dalam penelitian ini dibagi kedalam dua bagian, yaitu metode pengumpulan data dan metode pengembangan perangkat lunak.

#### 3.3.1 Metode Pengumpulan Data

Untuk menunjang penelitian ini penulis melakukan metode pengumpulan data seperti berikut ini:

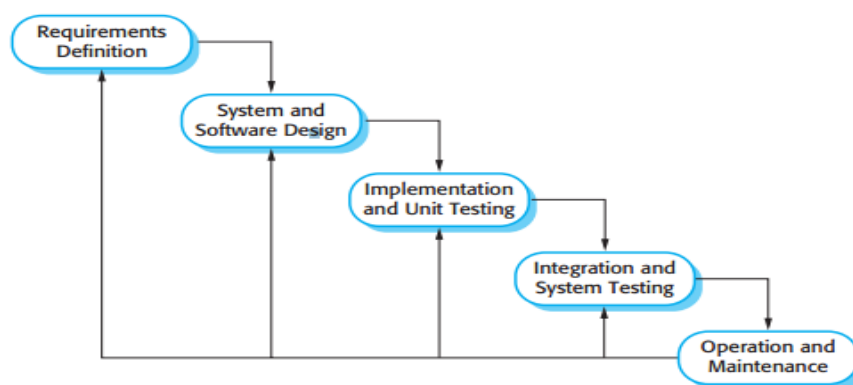
1. Studi Literatur

Studi literatur dilakukan dengan mempelajari dan memahami teori dan konsep yang menjadi pendukung untuk penelitian ini. Teori-teori tersebut diantaranya plagiarisme, deteksi pagiarisme dan algoritma Rabin-Karp melalui buku, jurnal, *paper*, undang-undang, serta artikel yang berhubungan dengan penelitian ini.

## 2. Pengumpulan Data

Pengumpulan data yang dimaksud yaitu dokumen skripsi mahasiswa berupa judul, abstrak, bab sampai bab lima. Kemudian dokumen-dokumen tersebut akan memasuki tahap praproses untuk tahap pembersihan data sebelum dilakukan deteksi plagiarisme.

### 3.3.2 Metode Pengembangan Perangkat Lunak di metode penelitian



Gambar 3. 2 Model *Waterfall* (Sommerville, 2011)

Metode pengembangan perangkat lunak dilakukan dengan metode *waterfall*. Model SDLC air terjun (*waterfall*) sering juga disebut model sekuensial linier (sequential linier). Penulis menggunakan metode *modern waterfall* 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 Ian Sommerville (2011) :

#### 1. *Requirements Analysis and Definition* (Analisis)

Analisis adalah tahap menentukan aplikasi atau *software* seperti apakah yang akan dibuat. Analisis merupakan tahapan penetapan fitur, ruang lingkup, tujuan dan batasan yang dimiliki oleh perangkat lunak. Semua hal tersebut akan ditetapkan secara rinci dan berfungsi sebagai spesifikasi perangkat lunak. Dalam penelitian ini proses analisis yang dilakukan mencakup analisis terhadap data yang digunakan dan batasannya yaitu perangkat lunak hanya

dapat memproses data berbentuk pdf dan keluaran perangkat lunak yaitu persentase kesamaan dokumen *input* dengan dokumen pada *database*.

## 2. *System and Software Design* (Perancangan)

Dalam tahapan ini dilakukan perancangan lebih detail yang dikembangkan dari hasil analisis. Pada penelitian ini, dilakukan perancangan antarmuka atau *mock-up* dari perangkat lunak. Pada tahapan ini juga dibahas gambaran umum mengenai alur dan proses perangkat lunak yang dibuat.

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

*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* (Testing)

Perangkat lunak yang dibuat akan memasuki tahap pengujian atau *testing*. Pengujian yang dilakukan meliputi pengujian fungsi-fungsi yang terdapat pada perangkat lunak dan juga pengujian terhadap data yang digunakan, yaitu data dokumen yang telah di plagiat.

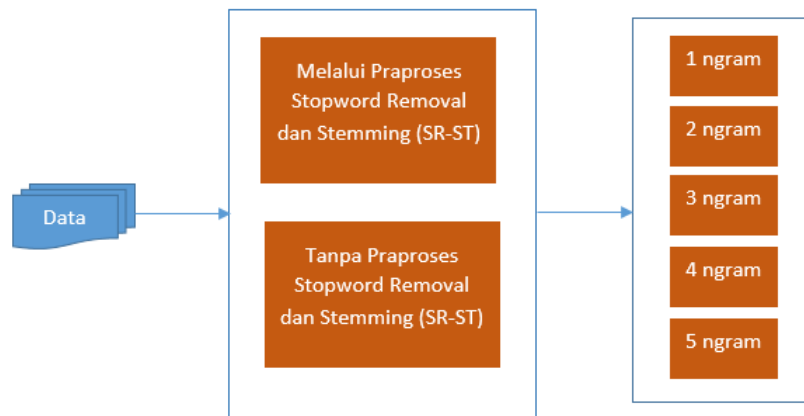
## 5. *Operation and Maintenance* (Pemeliharaan)

Dalam tahapan ini, perangkat lunak mulai digunakan. Selain itu juga memperbaiki *error* yang tidak ditemukan pada tahap pembuatan. Dalam tahap ini juga dilakukan pengembangan perangkat lunak seperti penambahan fitur dan fungsi baru.

### 3.4 Skenario Eksperimen

Eksperimen dilakukan dengan melakukan pendeteksian plagiarisme dalam beberapa skenario. Skenario eksperimen ditentukan dengan mengatur praproses yang digunakan dan besar n-gram yang digunakan.

Dilihat pada Gambar 3.3, terdapat dua parameter dalam skenario yang digunakan yaitu praproses dan n-gram. Berikut adalah penjelasan masing-masing parameter yang digunakan pada eksperimen.



Gambar 3. 3 Skenario Eksperimen

### 1. Praproses

Sebelum memasuki proses deteksi plagiarisme dilakukan empat tahap praproses, yaitu *case folding*, *tokenize*, *stopword removal* dan *stemming*. Untuk proses deteksi plagiarisme sebenarnya cukup dengan praproses *case folding* dan *tokenize*, sehingga praproses *stopword removal* dan *stemming* hanya menjadi praproses tambahan agar mencapai hasil yang optimal. Pada eksperimen ini akan dibandingkan skenario eksperimen tanpa menggunakan praproses *stopword removal* dan *stemming* (SR-ST) dan skenario eksperimen dengan menggunakan praproses *stopword removal* dan *stemming* (SR-ST) untuk mengetahui pengaruh penambahan praproses *stopword removal* dan *stemming* pada hasil eksperimen.

### 2. N-gram

Pada algoritma Rabin-Karp, n-gram adalah banyak kata yang diparsing dan akan membentuk kumpulan window yang akan dibandingkan *similarity* nya. Semakin besar n-gram maka semakin banyak kata yang akan dikelompokkan menjadi sebuah window. Semakin besar n-gram maka semakin banyak kata yang tercampur menjadi satu window, sehingga semakin sulit untuk mencari *window* yang similar jika plagiator banyak melakukan modifikasi pada teks yang ditirunya. Maka dari itu perlu ditentukan besar n-gram yang digunakan untuk mendapatkan hasil yang optimal.

Dengan berdasarkan parameter tersebut dilakukan beberapa skenario pada eksperimen ini agar dapat mengetahui pengaruh dari kombinasi parameter yang digunakan. Pada eksperimen ini besar n-gram yang digunakan yaitu 1, 2, 3, 4 dan 5 gram.

Selain itu, dilihat pula dampak dari enam metode plagiat yang diterapkan pada data eksperimen. Yaitu diantaranya metode hapus kata, tambah kata, modifikasi kata, hapus modifikasi dan tambah modifikasi. Pada setiap metode plagiat penulis menguji coba dengan melakukan perubahan perubahan pada depan kalimat, tengah kalimat, dan akhir kalimat pada dokumen untuk mengetahui hasil dari setiap perubahan tersebut.

Hasil yang ditampilkan yaitu berupa nilai *error* yang dimiliki dari masing-masing n-gram yang digunakan pada eksperimen ini, yaitu 1 sampai 5 gram. Nilai error ditentukan dengan metode *Root Mean Square Error* (RMSE).

RMSE merupakan parameter untuk mengevaluasi nilai hasil pengukuran terhadap nilai sebenarnya atau nilai yang dianggap benar. Rumus RMSE dapat dilihat pada persamaan (4.1). Perhitungan RMSE didasarkan pada akar kuadrat selisih nilai observasi ( $y$ ) dengan nilai sebenarnya ( $f$ ) dan dibagi dengan banyaknya data ( $n$ ).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (3.1)$$

Nilai observasi pada eksperimen ini adalah nilai yang dihasilkan oleh perangkat lunak DPS, sedangkan nilai sebenarnya adalah persentase plagiat data eksperimen yang berdasarkan *human judgement*.