

BAB 3

METODOLOGI PENELITIAN

Berikut adalah metodologi penelitian yang dilakukan dalam penelitian ini.

3.1 Studi Literatur

Dalam melakukan sebuah penelitian, tahap ini adalah tahap dimana dilakukan pencarian teori-teori yang akan digunakan melalui jurnal, prosiding, atau buku. Selain mencari teori-teori yang akan digunakan dalam penelitian, pada tahap studi literatur dilakukan pencarian informasi-informasi terhadap penelitian terdahulu yang dapat dijadikan sebagai referensi.

3.2 Pengumpulan Data

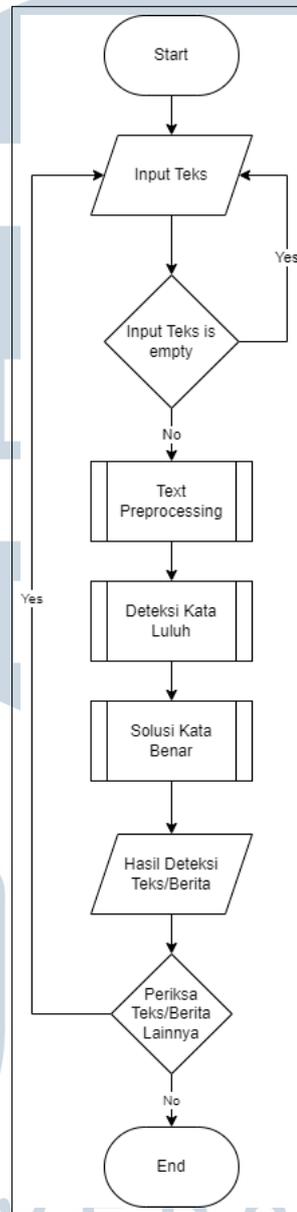
Dalam penelitian ini, data-data yang dibutuhkan harus dikumpulkan terlebih dahulu. Data-data tersebut adalah *dataset* kata dasar, *dataset* kata benar, dan berita dari portal berita Tribunnews. *Dataset* kata dasar dan *dataset* kata benar dikumpulkan dengan cara memasukkan kata-kata yang telah diperiksa kebenarannya pada aplikasi KBBI V secara manual ke dalam dua file excel yang berbeda. Sedangkan untuk *dataset* berita portal berita Tribunnews didapatkan melalui kumpulan berita yang telah diberikan oleh pihak Tribunnews dalam google drive.

3.3 Perancangan Sistem

3.3.1 Diagram Alir

Bagian ini berisi diagram alir dari fungsi yang telah dibuat pada sistem.

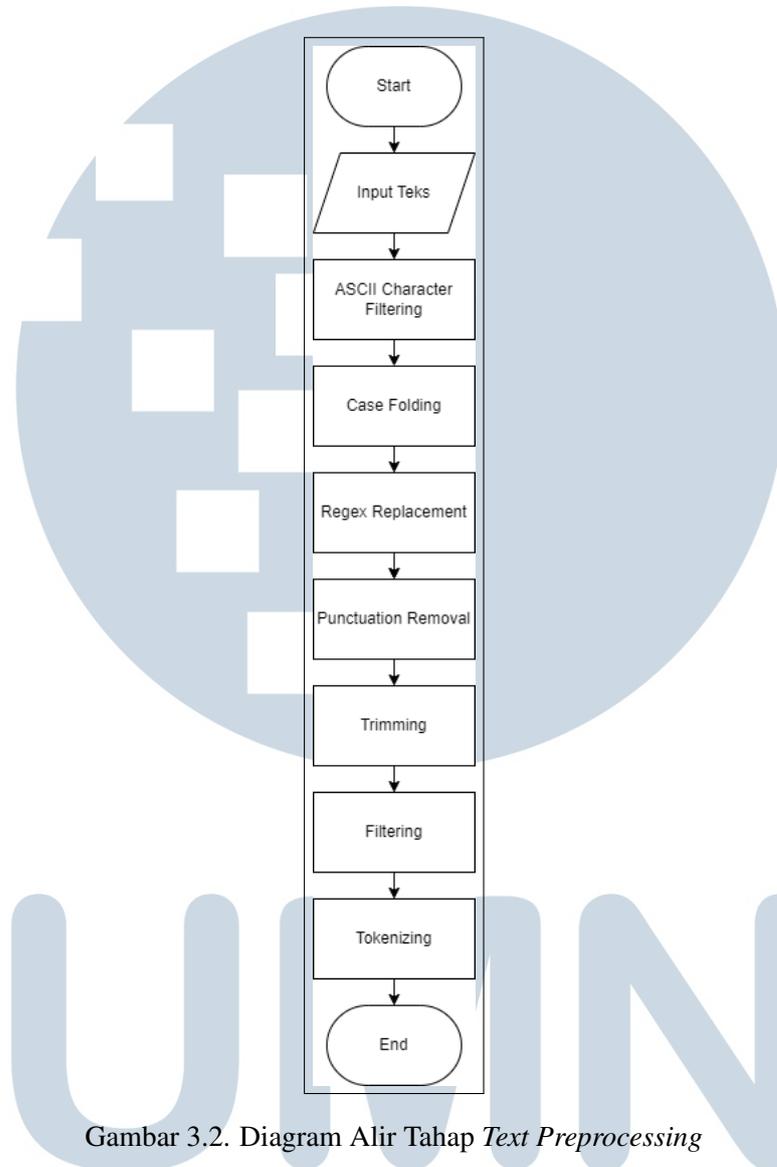
A Diagram Alir Utama



Gambar 3.1. Diagram Alir Utama

Setelah data-data yang dibutuhkan telah dikumpulkan, proses pendeteksian kesalahan ejaan kata luluh sistem akan dilakukan berdasarkan gambar 3.1. Langkah pertama yang dilakukan sistem adalah menerima sebuah *input* teks. Setelah sistem menerima sebuah *input* teks, tahap *text preprocessing* akan dilakukan.

B Diagram Alir Text Preprocessing

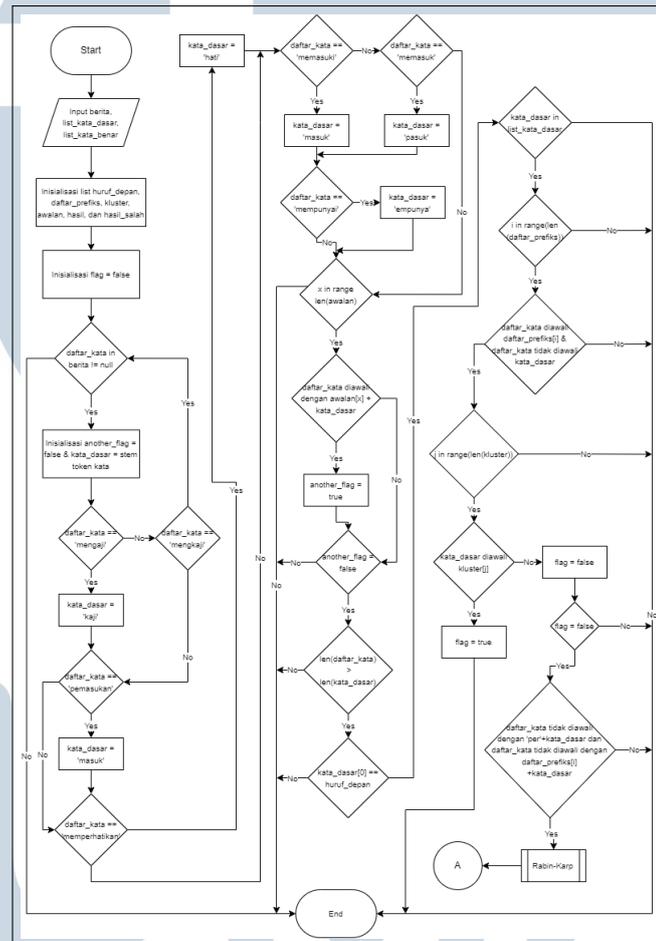


Gambar 3.2. Diagram Alir Tahap *Text Preprocessing*

Secara umum, tahap *Text Preprocessing* dilakukan melalui 4 langkah yaitu, *case folding*, *tokenizing*, *filtering*, dan *stemming*. Namun, dalam penelitian ini, dilakukan proses tambahan seperti *ASCII character filtering*, *regex replacement*, *punctuation removal*, dan *trimming* untuk mempersempit data yang digunakan dalam penelitian ini. Dengan penambahan beberapa tahap tersebut, proses *text preprocessing* yang dilakukan pada penelitian ini sesuai dengan gambar 3.2. Selain itu, untuk proses *stemming* penelitian ini, digunakan *stemmer* sastrawi.

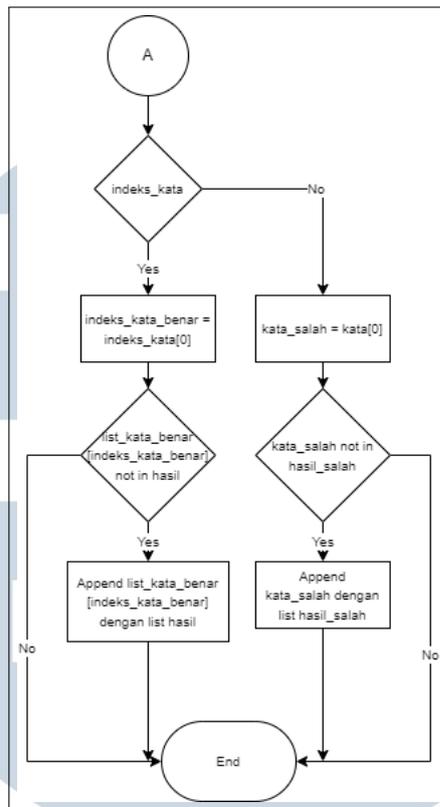
C Diagram Alir Deteksi Kesalahan Ejaan Kata Luluh

Pada bagian ini, teks/berita yang sudah diproses pada tahap *text preprocessing* dideteksi menggunakan algoritma Rabin-Karp untuk diperiksa apakah terdapat kesalahan ejaan kata luluh. Berikut adalah diagram alir pendeteksian kesalahan ejaan kata luluh yang telah dibuat.



Gambar 3.3. Diagram Alir Deteksi Kesalahan Ejaan Kata Luluh

UNIVERSITAS
MULTIMEDIA
NUSANTARA

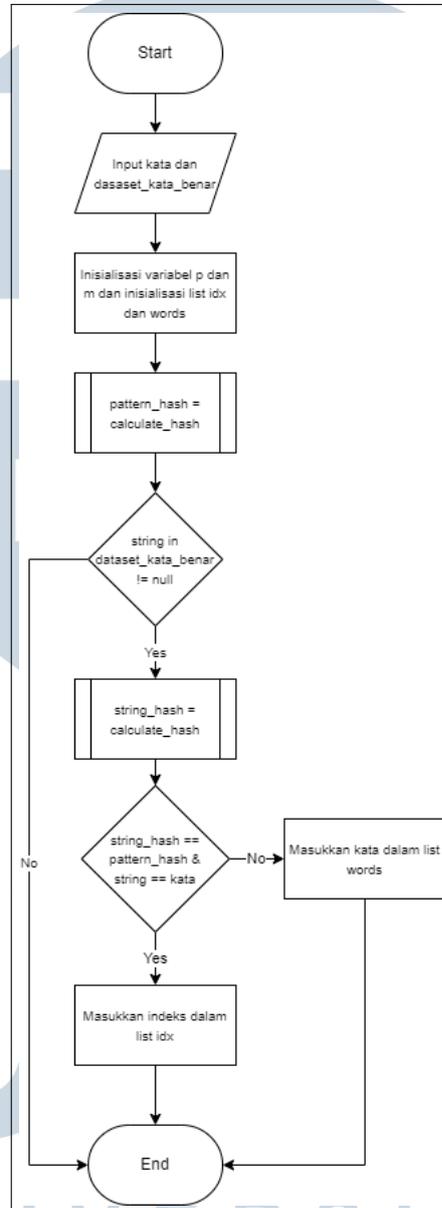


Gambar 3.3. Diagram Alir Deteksi Kesalahan Ejaan Kata Luluh (lanjutan)

Dalam gambar 3.3, proses pertama yang dilakukan untuk mendeteksi kesalahan ejaan kata luluh adalah dengan membuat *filter* agar hanya kata-kata luluh yang terdeteksi. Ciri-ciri kata luluh adalah katanya tidak memiliki 2 prefiks (prefiks me- dan pe-), katanya berimbuhan me- dan pe-, kata dasar dari kata luluh memiliki huruf awal s, p, k, atau t, ditambahkan juga baris untuk mengecek apakah kata dasarnya terdapat dalam *dataset* kata dasar untuk memastikan bahwa kata tersebut adalah kata yang benar ada. Selanjutnya, ciri-ciri kata luluh adalah kata yang berawalan me- atau pe- yang tidak diikuti dengan kata dasarnya. Kata dasar kata luluh juga tidak berawalan kluster. Kluster adalah gabungan dari 2 huruf konsonan. Terakhir, ciri-ciri kata luluh adalah tidak berawalan 'per' yang diikuti kata dasarnya.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

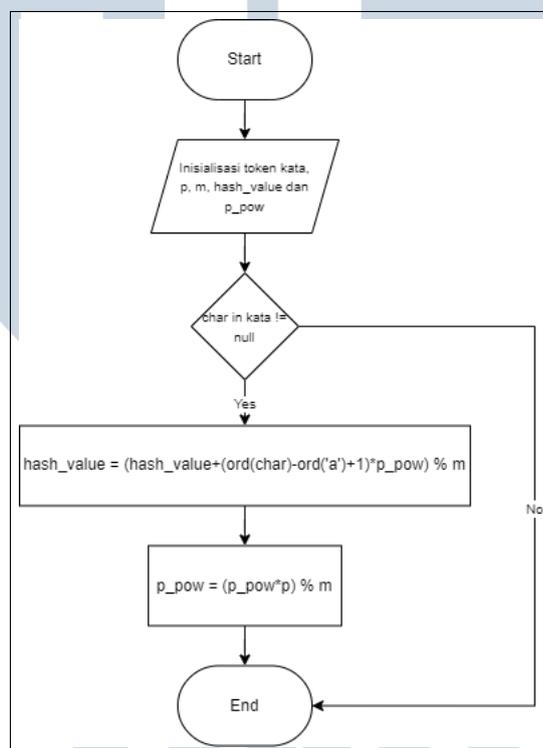
D Diagram Alir Algoritma Rabin Karp



Gambar 3.4. Diagram Alir Algoritma Rabin Karp

Dengan mengetahui ciri-ciri dari kata luluh, penulisan kode untuk mendapatkan kata luluh dalam suatu teks/berita dapat dilakukan. Setelah mendapatkan kumpulan kata luluh tersebut, pemeriksaan dilakukan menggunakan algoritma Rabin-Karp yang ditunjukkan pada gambar 3.4. Pada diagram alir tersebut, kata luluh diperiksa berdasarkan nilai dari *hash* kata luluh tersebut yang dibandingkan dengan nilai *hash* dari kata yang ada dalam *dataset* kata benar. Jika

nilai *hash* dari kata luluh sama dengan nilai *hash* dari salah satu kata dalam *dataset* kata benar dan kata luluh tersebut sama dengan kata yang ada pada *dataset* maka, indeks kata yang ada dalam *dataset* kata benar dimasukkan ke dalam *list* *idx*. Namun, jika nilai *hash* dari kata luluh tidak sama dengan nilai *hash* dari salah satu kata dalam *dataset* kata benar atau kata luluh tersebut tidak sama dengan kata yang ada pada *dataset* maka, kata yang diperiksa dimasukkan ke dalam *list* *words*. Berikut adalah perhitungan nilai *hash* dalam algoritma Rabin-Karp yang ditunjukkan pada gambar 3.5.

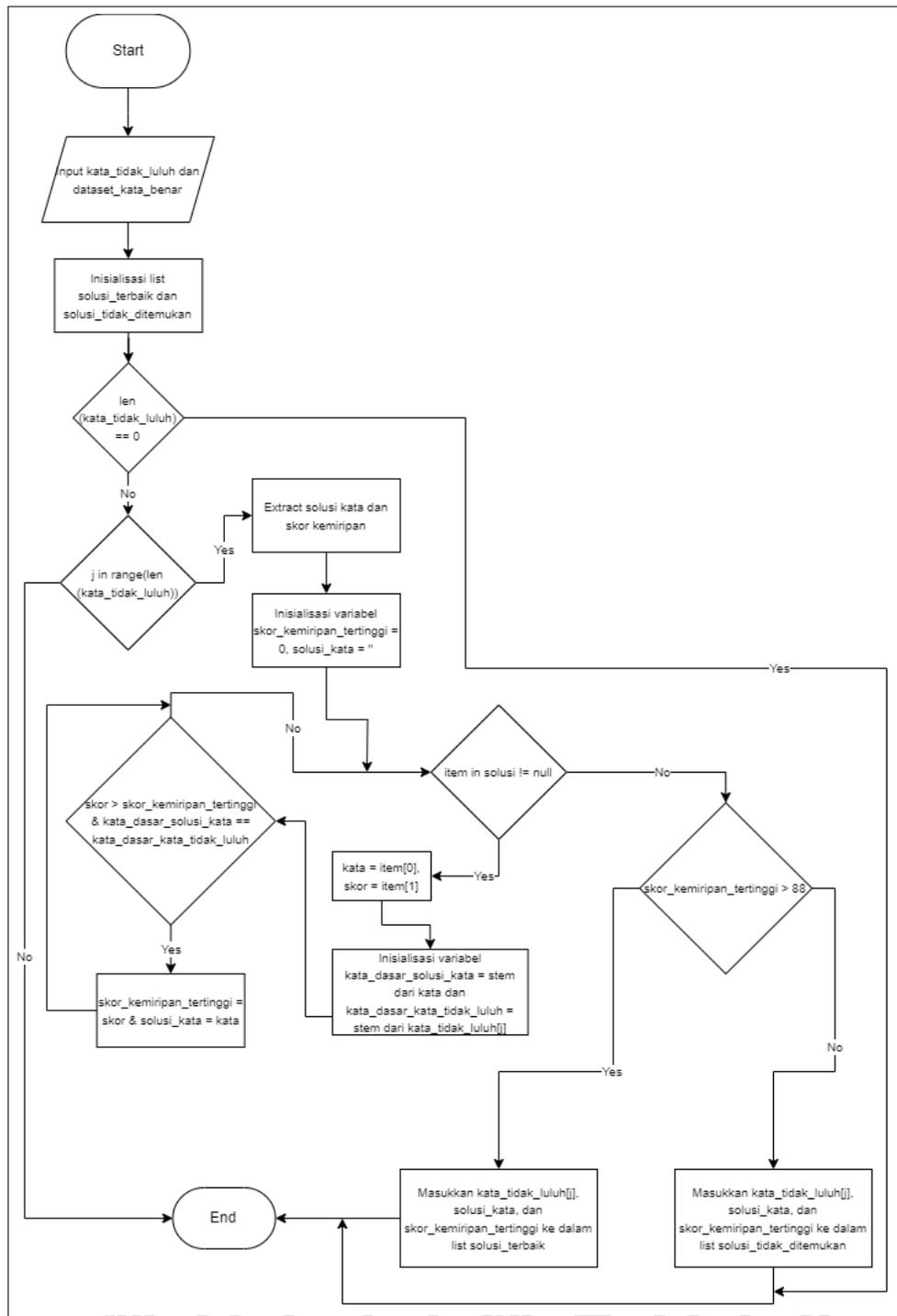


Gambar 3.5. Diagram Alir Perhitungan Nilai *Hash*

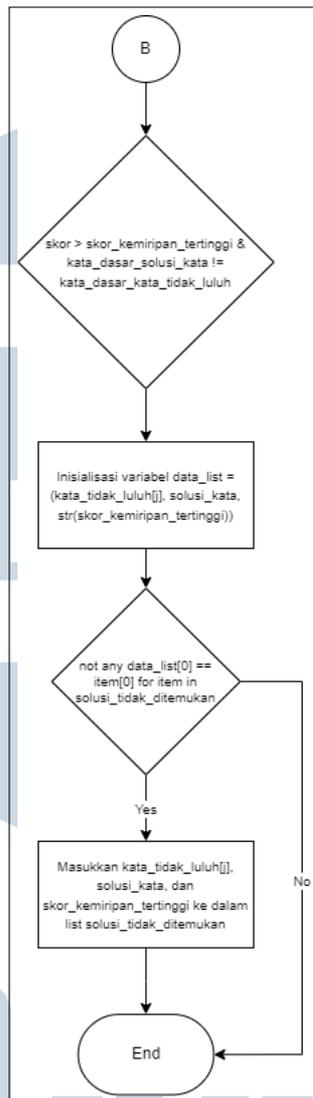
Perhitungan nilai *hash* pada gambar 3.5 menggunakan metode *polynomial rolling hash* yang mengembalikan nilai dari *hash value* dari kata yang dimasukkan.

E Diagram Alir Pencarian Solusi Kata Benar

Setelah dapat mendeteksi kata luluh yang mengalami kesalahan ejaan, sistem akan memberikan solusi kata benar untuk kesalahan kata luluh tersebut. Gambar 3.6 merupakan alur bagaimana sistem dapat mencari solusi kata yang terbaik untuk kesalahan ejaan kata luluh.



Gambar 3.6. Diagram Alir Pencarian Solusi Kata Benar



Gambar 3.6. Diagram Alir Pencarian Solusi Kata Benar (lanjutan)

Pencarian solusi kata benar dilakukan menggunakan *library fuzzy-wuzzy* dimana *library fuzzy-fuzzy* melakukan pencocokan kata berbasis jarak dengan menggunakan algoritma Levenshtein Distance.

3.4 Implementasi

Setelah perancangan sistem selesai dilakukan, sistem akan dibuat menggunakan jupyter notebook dengan menerapkan algoritma Rabin-Karp. Sistem yang telah selesai dibuat pada jupyter notebook kemudian akan dipindahkan ke dalam visual studio code karena dalam melakukan pengujiannya, akan dibuat tampilan UI sederhana dari sistem dalam bentuk *website* menggunakan *Flask*

Framework.

3.5 Pengujian dan Evaluasi Sistem

Setelah sistem selesai dibangun, pengujian dilakukan untuk memeriksa apakah sistem dapat mendeteksi kesalahan ejaan kata luluh dan memberikan solusi kata benarnya atau tidak. Pengujian dilakukan dengan berbagai parameter, yaitu berdasarkan jumlah berita, nilai *threshold/cutoff* yang digunakan, dan *kluster*. Evaluasi dilakukan jika pada tahap pengujian, sistem tidak bekerja sesuai dengan yang diinginkan. Sistem diperbaiki agar hasil dari pendeteksian yang dilakukan dapat sesuai dengan yang diinginkan. Langkah terakhir adalah melakukan perhitungan *confusion matrix* untuk mengetahui performa dari sistem yang telah dibuat.

3.6 Penulisan Laporan

Penulisan laporan dilakukan sebagai salah satu bentuk dokumentasi terhadap seluruh proses pengerjaan dalam penelitian ini. Laporan yang dibuat akan menjelaskan bagaimana proses dari penelitian yang dilakukan mulai dari pencarian teori-teori dalam penelitian ini hingga kesimpulan yang didapatkan berdasarkan penelitian yang telah dilakukan.

