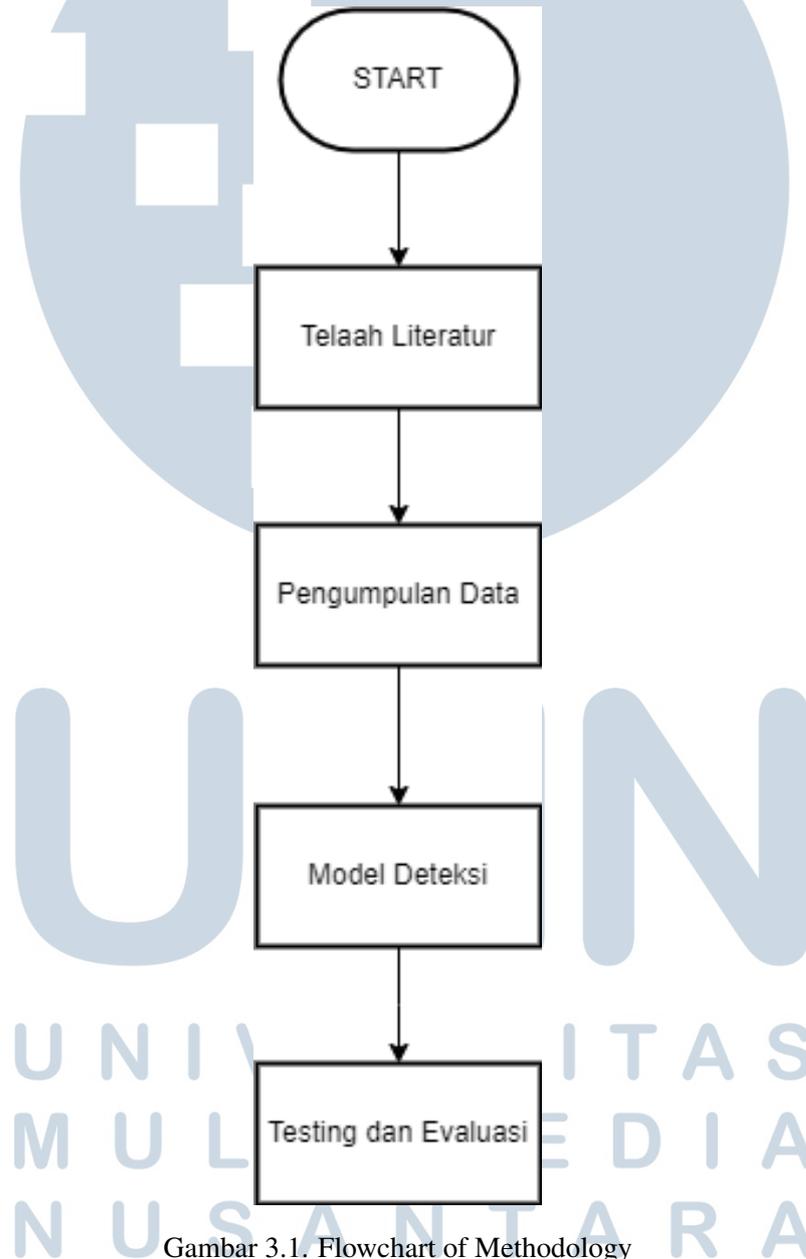


BAB 3 METODOLOGI PENELITIAN

Gambar 3.1 merupakan diagram alur dari penelitian yang dilakukan. Berikut penjelasan masing-masing dari alur penelitian.



Gambar 3.1. Flowchart of Methodology

3.1 Telaah Literatur

Telaah literatur merupakan tahap pertama dalam penelitian ini. Literatur yang ditelaah merupakan literatur dari jurnal ataupun *proceedings*. Telaah literatur menjadi tahap dalam penunjang teori untuk pemaksimalan penelitian. Literatur yang ditelaah adalah literatur-literatur seputar kesalahan ejaan, kata majemuk, *library fuzzywuzzy* dan NLP-Id.

3.2 Pengumpulan Data

Pengumpulan data menjadi tahap selanjutnya. Dalam penelitian ini, data yang digunakan adalah data mengenai kata majemuk. Kata majemuk yang dikumpulkan didapat dari berita milik Tribunnews. Dari berita milik Tribunnews yang sudah tersedia di Google Drive, didapatkan 1846 kata majemuk. 1846 Kata majemuk yang didapat berupa kata majemuk yang terpisah(non-senyawa) maupun kata majemuk yang digabung(senyawa). Kata majemuk yang didapatkan ini digunakan sebagai patokan ataupun saran kata majemuk yang benar untuk kata majemuk yang terdeteksi salah.

Selain data kata majemuk, terdapat data lain yang dikumpulkan, yaitu data kata benar. Data kata benar ini didapatkan dari peneliti U-Tapis sebelumnya. Data kata benar pada penelitian ini digunakan sebagai filtrasi untuk kata pada artikel yang masuk pada saat mendeteksi kata majemuk.

3.3 Model Deteksi

Setelah telaah literatur sudah dilakukan dan data-data yang diperlukan pada penelitian ini sudah dikumpulkan, selanjutnya adalah membuat model atau program untuk mendeteksi serta memberikan saran untuk kata majemuk yang salah.

Model deteksi dimulai dengan mendapatkan teks artikel yang di-*input* oleh pengguna atau jurnalis. Dari teks artikel yang masuk, dilakukan *preprocess* teks untuk melakukan proses-proses awal sebelum mendeteksi kesalahan kata majemuk. *Preprocess* tersebut seperti *case folding*, menghapus spasi-spasi yang berlebih, kemudian pengaplikasian *PosTag* pada teks artikel untuk setiap kata yang bertujuan untuk mendeteksi kata majemuk berdasarkan pola pembentukan kata majemuk.

Dari terkumpulnya kata-kata yang terdeteksi berdasarkan pola pembentuk kata majemuk, dilakukanlah proses pendeteksian kesalahan penulisan kata majemuk menggunakan *library fuzzywuzzy* yang berbasis algoritma *Levenshtein*

Distance. Jika kata terdeteksi benar, kata tersebut nantinya akan dilabeli dengan warna hijau, sedangkan jika kata terdeteksi salah, maka akan dilabeli dengan warna merah pada saat ditampilkan di website. Selain itu, kata yang salah akan dibetulkan dengan kata majemuk yang benar berdasarkan hasil deteksi algoritma *Levenshtein Distance*.

3.4 Testing dan Evaluasi

Tahap *testing* dan evaluasi merupakan tahap terakhir untuk mengetahui apakah hasil dari penelitian ini bagus atau tidak. Cara mendapatkan nilai tes akurasi yang pertama adalah dengan menggunakan fungsi *evaluate* berdasarkan model yang dibuat. Selanjutnya adalah dengan mencoba menggunakan teks artikel yang asli pada program yang telah dibangun. Dari percobaan tersebut, akan dihitung *confusion matrix* untuk mendapatkan nilai akurasi, presisi, *recall*, dan *F1-score*.

