

BAB III METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Dalam penelitian analisis sentimen ulasan aplikasi M-Paspor tahun 2023 pada *google play store* untuk merancang dan membangun sistem memiliki beberapa tahapan yaitu:

1. Studi Literatur

Studi literatur merupakan tahapan awal dalam penelitian. Tahapan ini bertujuan untuk mempelajari dan memahami informasi dari topik yang diteliti. Informasi yang dikumpulkan berasal dari jurnal, artikel, karya tulis, buku dan website. Teori-teori yang relevan dalam penelitian ini antara lain; Analisis sentimen, *Multinomial Naive Bayes*, TF-IDF, M-Paspor dan *Confusion Matrix*.

2. Pengumpulan Data

Data dikumpulkan dengan menggunakan metode scraping data melalui *library google-play-scraper*. *library google-play-scraper* merupakan *library* yang digunakan dalam mengambil data aplikasi yang terdapat *google play store*. Pada penelitian ini data yang akan diambil merupakan ulasan pada aplikasi M-Paspor sepanjang tahun 2023 dimulai dari tanggal 01 januari 2023 hingga saat dilakukan penelitian pada tanggal 18 oktober 2023 dan didapatkan sebanyak 6443 data ulasan aplikasi M-Paspor.

3. Perancangan

Tahapan Perancangan sistem dimulai dengan membuat alur penelitian atau *flowchart*, melakukan pelabelan dengan *lexicon based vader*, melakukan proses *text pre-processing* pada data, Melakukan fitur ekstraksi TF-IDF dan melakukan pembagian data latih dan data tes.

4. Implementasi Algoritma

Tahapan implementasi algoritma dilakukan setelah tahapan perancangan berhasil dilakukan. Tahapan Implementasi Algoritma adalah melakukan mengimplementasikan algoritma *Multinomial Naive Bayes* kedalam sistem dengan membuat pemodelan dan melakukan pembelajaran mesin terhadap pemodelan yang telah dibuat dengan data train.

5. Pengujian

Tahapan pengujian merupakan tahapan yang memiliki fungsi untuk melakukan pengujian terhadap code yang telah dibangun untuk mencari apakah ada kesalahan dan bug dari sistem yang telah dirancang bangun dan juga akan melakukan pengujian performa pemodelan yang telah diterapkan ke dalam sistem. Hal ini dilakukan untuk memastikan apakah sistem yang dibuat dapat berjalan dengan baik. Pada proses pengujian, data latih dan data uji dibagi menjadi 60:40, 70:30 dan 80:20. Hingga menghitung tingkat *accuracy*, *precision*, *recall* dan *f1-score* yang terdapat pada *confusion matrix* dengan membandingkan label klasifikasi sebelum pemodelan dan membandingkan label klasifikasi setelah dilakukan pemodelan.

6. Dokumentasi

Menokumentasikan semua proses mulai dari awal proses penelitian, desain, dan pemodelan pemodelan klasifikasi. Hal ini dilakukan secara bertahap mulai dari pendahuluan hingga simpulan dan saran.

7. Konsultasi

Tahapan konsultasi merupakan tahapan untuk berkonsultasi dan mendiskusikan penelitiannya dengan dosen pembimbing yang bertujuan untuk memperoleh saran dan pendapat terhadap proses dan laporan yang telah dilakukan dalam penelitian.

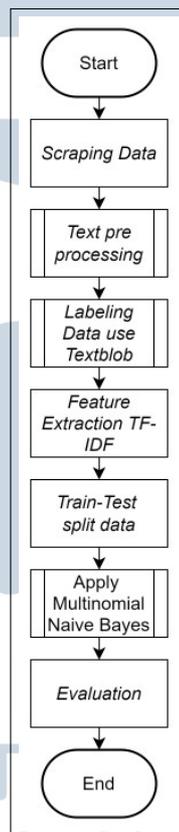


3.2 Alur Pembangunan Sistem

Pada tahap pembangunan sistem menggunakan diagram alur *flowchart*. Diagram alur dalam penelitian ini antara lain *flowchart* gambaran umum dari sistem yang akan dibangun, *Text-Preporcessing*, *Labelling data* dan *Apply Multinomial Naive Bayes*.

3.2.1 Flowchart Gambaran Umum Sistem

Terlihat pada Gambar 3.1 merupakan *flowchart* gambaran umum dari sistem yang akan dibangun dalam penelitian. Alur dari *flowchart* dalam penelitian dimulai dari pengumpulan data menggunakan metode *scraping data*, *Text-preprocessing*, *Feature Extraction TF-IDF*, *Labelling Data Use Textblob*, *Split data*, *Apply MultinomialNB* dan *Evaluation*.



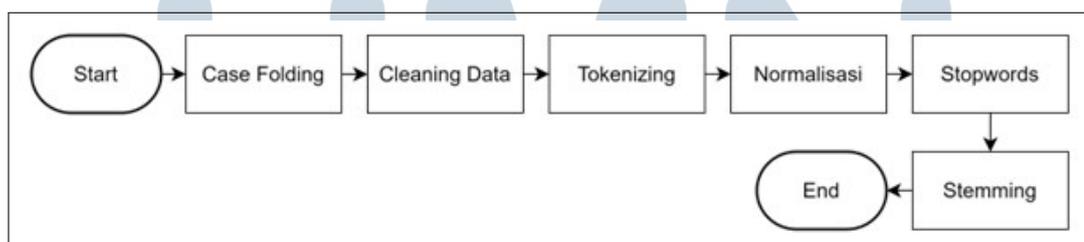
(Gambar 3.1. Flowchart Gambaran Umum Sistem)
(Sumber: Dokumentasi Pribadi)

3.2.2 Scraping Data

Scraping data merupakan proses mengumpulkan dan menyimpan data dari halaman web secara otomatis. Pada proses *scraping data* menggunakan alat atau program untuk mendapatkan informasi yang terdapat dalam web dan menyimpannya dalam format yang digunakan seperti csv [24]. Pada penelitian ini pengumpulan data dilakukan dengan *scraping data* dari review aplikasi M-Paspor di *google play store* menggunakan *library google-play-scraper*. Data yang diambil dalam analisis sentimen M-paspor di *google play store* berasal dari Indonesia beserta bahasa yang diambil adalah bahasa Indonesia dengan menggunakan fungsi *country* dan *language* yang disediakan oleh *library google-play-scraper*. Data yang diambil sepanjang tahun 2023 dimulai dari tanggal 01 januari 2023 hingga saat dilakukan penelitian pada tanggal 18 oktober 2023 dan didapatkan sebanyak 6443 data ulasan aplikasi M-Paspor.

3.2.3 Text-preprocessing

Text-preprocessing adalah tahapan yang berguna untuk menyeleksi teks agar menjadi terstruktur. Gambar 3.2 merupakan alur tahapan *text-preprocessing* dimulai dari *Case Folding*, *Cleaning Data*, *Tokenizing*, *Stemming*, *Normalisasi*, *Stopwords* dan *stemming*.



(Gambar 3.2. Flowchart Text Pre-processing)

(Sumber: Dokumentasi Pribadi)

1. Case Folding

Tahapan *Case folding* dalam *text-preprocessing* adalah tahapan yang memiliki fungsi untuk mengubah teks menjadi huruf kecil atau *lowercase*.

2. Cleaning Data

Menyeleksi dan membersihkan data teks dari simbol, emoji, link, *single char*, angka, *whites space* dan tanda baca yang tidak memiliki sentimen dalam suatu teks.

3. Tokenizing

Pada tahapan *tokenizing* akan menggunakan *library nltk* untuk memecah kalimat menjadi daftar kata.

4. Normalisasi

Normalisasi dilakukan dengan membuat file csv yang terdiri dari kata yang salah dalam penulisan beserta dengan perbaikannya.

5. Stopwords

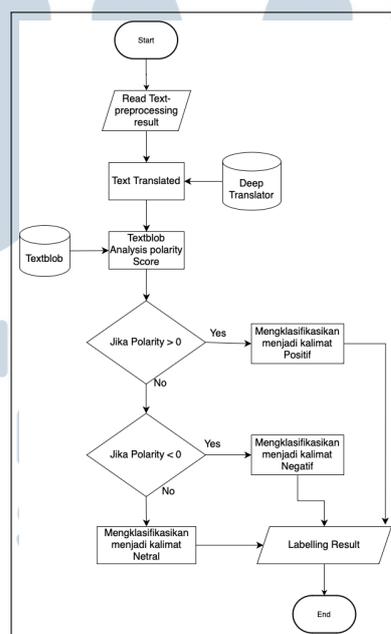
Pada tahapan *stopwords* akan menghapus kata-kata yang tidak memiliki makna khusus contohnya adalah kata konjungsi.

6. Stemming

Tahapan *Stemming* akan menggunakan *library sastrawi* yang akan mengubah atau mengekstrak kata yang memiliki imbuhan menjadi dasar kata.

3.2.4 Labelling Data use Textblob

Labelling data merupakan tahapan yang berguna sebagai pemberian label klasifikasi sentimen pada data teks yang didapatkan melalui proses *scraping data* dan telah melalui tahapan *text-preprocessing*. Pada penelitian ini tahap *labelling* akan dilakukan secara otomatis menggunakan *Library Textblob*.



(Gambar 3.3. Flowchart Labelling Data Use Textblob)

(Sumber: Dokumentasi Pribadi)

Gambar 3.3 merupakan alur dari pelabelan menggunakan *Textblob* dimulai dari membaca data kemudian menterjemahkan bahasa asal yaitu bahasa Indonesia ke dalam bahasa Inggris. Penerjemahan ini dilakukan karena *Textblob* merupakan kamus yang berisi daftar kata yang menggunakan bahasa Inggris. Oleh karena itu, *textblob* hanya membaca data teks yang menggunakan bahasa Inggris.

Pada proses penerjemahan bahasa *library deep-translator* digunakan sebagai *library* yang membantu dalam penerjemahan bahasa. setelah data telah berhasil diterjemahkan langkah selanjutnya, menghitung nilai polaritas menggunakan *library textblob* dan mengklasifikasikannya ke dalam kelas positif, negatif atau netral.

3.2.5 Feature Extraction TF-IDF

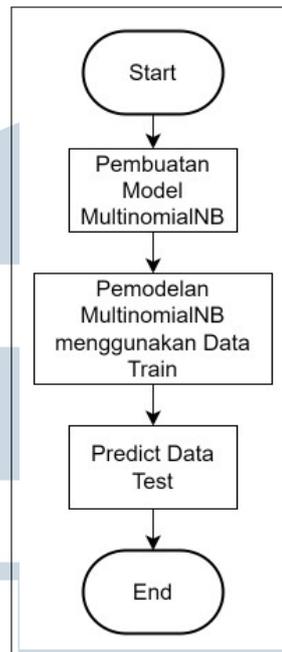
Feature Extraction TF-IDF merupakan tahapan yang memiliki berfungsi untuk memberikan nilai bobot pada data teks yang terdapat dalam kalimat. Pembobotan ini dilakukan dengan tujuan supaya algoritma *multinomial naive bayes* dapat memahami teks yang diinput.

3.2.6 Train-Test Split Data

Setelah *Feature Extraction* dilakukan selanjutnya melakukan tahap pembagian data menggunakan *train-test split data*. Pada tahap *train-test split data* akan membagi kumpulan data menjadi dua bagian yaitu, *data train* dan *data test*. *data train* berguna sebagai data untuk melatih pemodelan yang dibuat menggunakan algoritma *multinomial naive bayes*. Sedangkan, *data test* berguna sebagai data untuk menguji pemodelan yang berhasil dibuat. Hal ini, dilakukan untuk menilai dan mengetahui kinerja dalam pemodelan.

3.2.7 Flowchart Apply Multinomial Naive Bayes

Setelah data telah dibagi menjadi dua langkah selanjutnya adalah mengimplmentasikan algoritma *multinomial naive bayes* dan mengklasifikasikan data. Gambar 3.4 merupakan *flowchart* dari alur pemodelan algoritma *multinomial naive bayes* dimulai dari pembuatan model klasifikasi *multinomial naive bayes* kemudian melakukan pemodelan dengan *data train* yang telah dilakukan *feature extraction* dan hasil dari pemodelan yang telah dibuat dengan menggunakan algoritma *multinomial naive bayes* akan digunakan untuk memprediksi data tes.



(Gambar 3.4. *Flowchart Apply Multinomial Naive Bayes*)
(Sumber: Dokumentasi Pribadi)

3.2.8 *Evaluation*

Tahapan *evaluation* akan mengukur kinerja dari pemodelan yang telah berhasil dibuat menggunakan algoritma *Multinomial Naive Bayes*. Pada penelitian ini, *evaluation* dilakukan menggunakan *confusion matrix* dengan tiga perbandingan label yang terdiri dari label positif, negatif dan netral. Selanjutnya, hasil yang didapatkan dari *confusion matrix* tersebut akan digunakan untuk menghitung *Accuracy*, *Precision*, *Recall* dan *F1-score* guna untuk mengetahui performa dan kinerja dari pemodelan *multinomial naive bayes*.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A