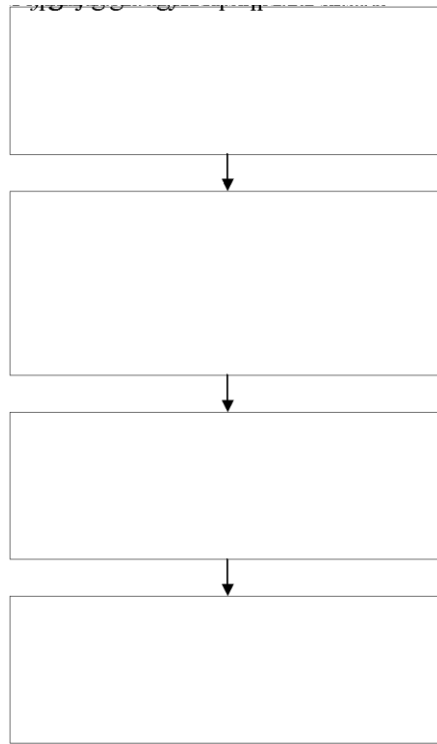


BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menerapkan metode penelitian kuantitatif, karena hasil akhir dari penelitian diubah menjadi data berbentuk angka untuk tujuan analisis statistik. Pemilihan algoritma *Naïve Bayes* dalam penelitian ini didasarkan pada sifat sederhana algoritma ini serta kebutuhan sedikitnya data latih, sehingga cocok sebagai algoritma yang digunakan dalam konteks penelitian ini. Dalam penelitian ini, penulis memanfaatkan sampel data sejumlah 10.000 data, dengan 2.000 data digunakan sebagai data latih dan 8.000 data lainnya digunakan sebagai data uji.



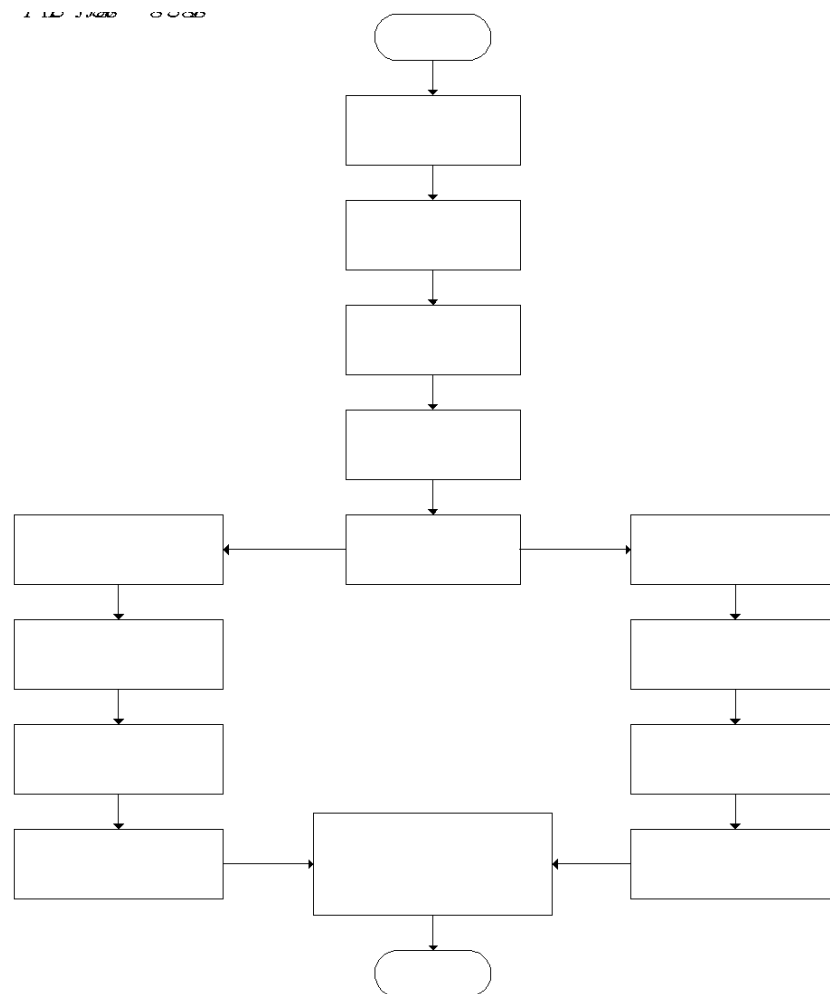
Gambar 3. SEQ Gambar_3. * ARABIC
1 Langkah Pengujian

Adapun langkah pengujian yang akan digunakan dalam penelitian ini terdapat pada Gambar 3.1. Gambar tersebut menjelaskan bahwa, penelitian ini secara umum melalui beberapa tahapan yaitu rancangan sistem penelitian, analisis persentase opini hasil rancangan

sistem analisis sentimen masyarakat pada Twitter, pengujian sistem analisis data, dan pembahasan hasil penelitian

3.1.1 Analisis Perancangan Penelitian

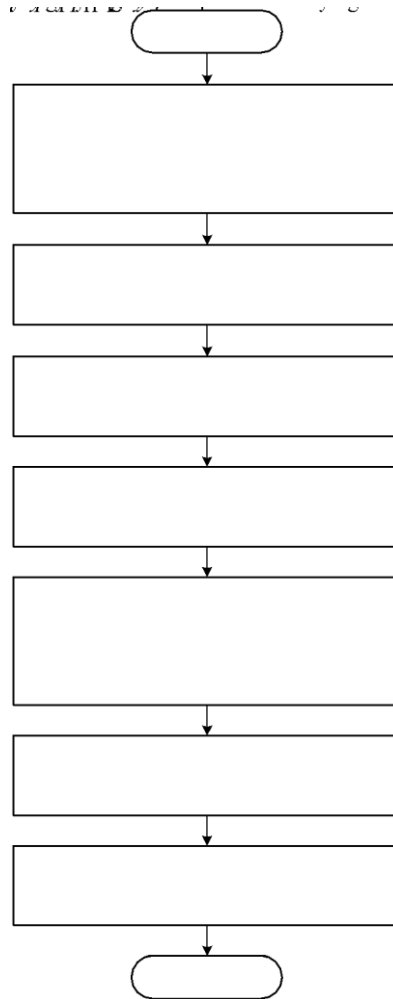
Pada pelaksanaan analisis perancangan sistem penulis akan menggambarkan bagaimana alur proses penyelesaian masalah, penulis juga akan membuat gambar alur proses agar mudah dipahami bagi pembaca. Alur proses penyelesaian masalah akan digambarkan seperti Gambar 3.2



Gambar 3.2 Alur Perancangan Sistem

Dari Gambar 3.2 dapat dijelaskan bahwa :

- a. *Crawling data* pada penelitian ini menggunakan *Google Colab* dan terdapat dua *library* yaitu *pandas* untuk *dataframe* dan *snscrepe* untuk pengambilan dataset. Alur melakukan *crawling data* dapat dilihat pada Gambar 3.3.



Gambar 3.3 Alur proses melakukan *crawling data*

Pada Gambar 3.3 terdapat alur proses dari *crawling data*, pertama penulis akan menentukan *keyword* yang merepresentasikan kekuatan sinyal dan kecepatan internet, kemudian melakukan *install library* yang dipakai yaitu *Pandas* dan *Snscrepe*, lalu memasukan data maksimal 10.000 data maka jika sudah mencapai maksimal *crawling* akan dihentikan, kemudian memasukan *keyword* yang akan dipakai, adapun *keyword* yang akan dipakai yaitu :

Kekuatan Sinyal:

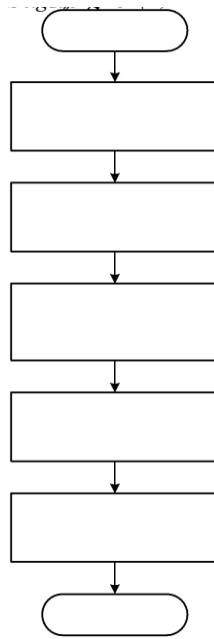
1. Sinyal Indosat.
2. Sinyal Indosat kuat.
3. Jaringan Indosat.
4. Koneksi jaringan Indosat.

Kecepatan Intenet :

1. Indosat lemot.
2. Indosat *error*.
3. Indosat lancar.

Tweet ini diambil dengan rentang waktu dari Maret 2020 hingga Februari 2023, setelah itu masuk ketahap selanjutnya itu menampilkan data dalam bentuk tabel dengan *DataFrame*, setelah *crawling data* selesai akan disimpan dalam bentuk csv.

- b. RapidMiner merupakan *tool* yang dipakai untuk melakukan pengolahan data dan pengujian akurasi dari sistem pengujian analisis sentimen.
- c. *Cleansing* yaitu untuk membersihkan atribut dari *userneame* yang tidak diperlukan seperti url, tagar, tanda baca dan simbol, dalam penelitian ini *cleansing data* pada *software* RapidMiner dengan menggunakan operator *replace*, *remove duplicate*.
- d. Normalisasi merupakan proses mengubah teks menjadi bentuk yang lebih standar atau seragam sebelum menerapkannya pada analisis sentimen.
- e. *Split Data*, pada penelitian ini, dikumpulkan data *sample* sebanyak 10.000 data kemudian dibagi dua. Dikarenakan penulis melakukan pelabelan secara manual maka penulis membagi data dengan jumlah 2.000 data sebagai data latih dan 8.000 data sebagai data ujicoba.
- f. *Preprocessing Tweet*, pada tahap ini melakukan rancangan rangkaian *filtering text*. Dalam alur proses *preprocessing* ini dapat dilihat pada gambar 3.4



Gambar 3.4 Alur proses *preprocessing*

Dapat dijelaskan dari gambar 3.4 bahwa operator dari *preprocessing* diantaranya:

1. *Tokenize* yaitu proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi bagian-bagian tertentu. Misalnya, “sinyal indosat jelek” menjadi “sinyal”, “indosat”, “jelek”.
2. *Transform Cases* yaitu proses untuk menyamakan kata menjadi huruf kecil atau tidak kapital.
3. *Filter Stopwords* yaitu mengurangi jumlah kata-kata yang tidak diperlukan biasanya kata umum yang dianggap tidak memiliki makna sehingga kata tersebut dihilangkan. Misalnya, “yg”, ”klo”, ”nih”, ”hehe” .
4. *Filter Tokenize By Length* yaitu proses pengambilan kata penting dari hasil token berdasarkan jumlah karakter, dimana pada penelitian ini token yang mempunyai kurang dari 3 karakter dan lebih dari 25 karakter akan dihapus.
5. *Stemming* yaitu mengubah kata yang mempunyai imbuhan menjadi kata dasar, misalnya “ter-gantung” output “gantung”, “di-pakai” output “pakai”.
- g. *Labeling*, pada proses ini didapatkan data latih kemudian melakukan pelabelan sentimen *positive* dan *negative*. Label *negative* untuk *tweet* yang mengandung unsur kata-kata kekecewaan dan label *positive* untuk *tweet* yang mengandung unsur kata-kata kepuasan atau kebanggaan. Hasil dari pelabelan data latih ini disebut dengan pemodelan.

- h. Algoritma *Naïve Bayes*, kemudian dilakukan pengujian sentimen dengan menggunakan algoritma *Naïve Bayes*. Algoritma *Naïve Bayes* akan mempelajari pola data yang telah dibuat oleh data latih sehingga menghasilkan model klasifikasi.
- i. *Machine Learning* ini merupakan *machine learning* dari sistem RapidMiner, data latih yang telah dilabeli secara manual akan dijadikan model untuk melatih algoritma *Naïve Bayes* sehingga menghasilkan *prediction* pada data uji.
- j. Hasil Akhir Pengujian, pada tahap ini mendapatkan hasil persentase sentimen *negative* dan *positive* serta menguji tingkat akurasi penggunaan *algoritma Naïve Bayes* pada penelitian ini.

3.1.2 Analisis Persentase Opini Hasil Rancangan Sistem Analisis Sentimen Masyarakat pada Twitter dengan *Software RapidMiner*

Setelah dataset melalui tahap *preprocessing*, kemudian masuk pada tahap *labeling*, lalu dibuatkan oleh penulis diagram *pie* untuk membantu pembaca lebih dapat memahami sentimen apa yang lebih mendominasi dalam penelitian ini.

3.1.3 Pengujian Sistem Analisis Data

Kemudian setelah tahapan *labeling*, dilakukannya pengujian sistem analisis guna melakukan uji validasi untuk melihat tingkat keakuratan penguji analisis sentiment menggunakan algoritma *Naïve Bayes*.

3.1.4 Pembahasan Hasil Penelitian

Pada tahap ini, berisi pembahasan mengenai perancangan sistem penelitian dan hasil pengujian sistem analisis data. Pada pembahasan perancangan sistem penelitian didalamnya terdapat hasil persentase sentiment negatif dan sentimen positif dan kata yang paling sering muncul yang akan divisualisasikan dalam bentuk *wordcloud*. Pada pembahasan mengenai hasil pengujian sistem analisis data berisi tingkat akurasi dari pengujian sistem kemudian terdapat *recall* dan *precision* masing-masing sentimen. Tahapan pembahasan ini dilakukan untuk mempermudah penulis dalam menyusun kesimpulan dan mempermudah pembaca untuk dalam memahami hasil dari penelitian ini.

3.2 Prosedur Penarikan Kesimpulan

Setelah dataset sampai pada tahapan akhir yaitu uji akurasi, kemudian dilakukan analisis secara keseluruhan sehingga dapat dilakukan penarikan kesimpulan. Kesimpulan dirumuskan berdasarkan analisis hasil pengujian sistem yang telah dilakukan. Kesimpulan ini dibuat dengan tujuan untuk memberikan informasi secara keseluruhan dengan singkat sehingga memudahkan pembaca mengetahui hasil dari penelitian.