

## **BAB III. METODELOGI PENELITIAN**

### **3.1 Data**

Data testing yang digunakan dalam penelitian ini adalah tweet pada sosial media Twitter yakni sentiment terhadap kinerja presiden joko widodo disaat pandemi corona berlangsung. Data tweet diambil pada saat virus corona mulai masuk indonesia pada tanggal 2 maret sampai 14 juni 2020, jumlah data testing algoritma naïve bayes yang digunakan penulis yaitu 90 data tweet dan jumlah data testing Opinion Detection yaitu 120.

Data training yang digunakan penulis di bagi menjadi 2 yaitu data training untuk algoritma opinion detection dan data training untuk algoritma naïve bayes. Data training untuk opinion detection yang digunakan penulis di bagi menjadi dua yaitu dataset POSTagging Viterbi sebanyak 259.368 kata dan dataset rule opinion sebanyak 15 rule. Data training untuk algoritma naïve bayes yang digunakan penulis yaitu 300 data tweet.

### **3.2 Metode Pengambilan Data**

Dalam penelitian ini data didapatkan dari media sosial Twitter dengan metode pengambilan data yaitu metode scraping.

Pengambilan data dilakukan secara berkala dengan target 8 kata kunci, yaitu bantuan corona jokowi, kartu prakerja jokowi, psbb jokowi, BLT jokowi, cicilan jokowi, Bansos jokowi, jokowi asimilasi narapidana, jokowi new normal. Pengambilan data secara periodik ini dilakukan dengan metode scraping menggunakan library scrapy pada bahasa pemrograman python, setelah data berhasil di scraping, data tersebut akan disimpan di database.

### **3.3 Metode Pengolahan Data**

Tahap analisa data dilakukan dengan melalui beberapa proses diantaranya proses preprocessing, pembobotan, postagging dan klasifikasi sentiment. Preprocessing adalah pengolahan data yang bertujuan untuk meminimalisir data / kata yang tidak memiliki nilai dalam object penelitian dan data yang tidak konsisten. Tahap preprocessing dibagi menjadi beberapa proses yaitu cleaning, casefolding, tokenizing, normalisasi, filtering dan stemming.

- Case Folding

Case Folding adalah mengubah semua teks dalam dokumen menjadi konsisten, bisa berupa lowercase atau uppercase.

- Cleaning

Cleaning adalah proses penghapusan url, mention, hastag, tanda baca dan angka.

- Tokenizing

Tokenizing adalah proses pemotongan kalimat menjadi kata.

- Normalisasi

Normalisasi adalah proses mengubah kata gaul menjadi kata KBBI.

- Filtering

Filtering adalah proses hasil tokenizing dilakukan penghapusan kata-kata yang kurang memiliki arti dalam tweet yang ada dalam list stopwords.

- Stemming

Stemming adalah proses pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya.

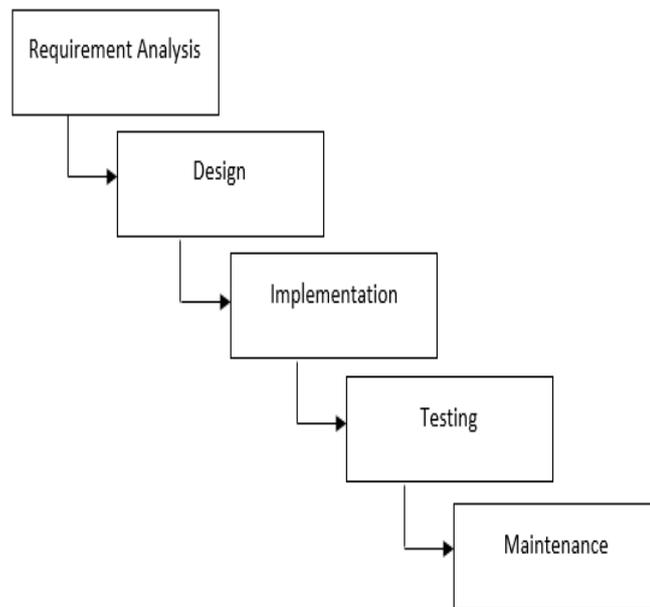
Hasil dari preprocessing kemudian dilakukan proses penilaian probabilitas dengan metode laplace correction. Laplace correction adalah suatu cara untuk menangani nilai probabilitas 0 (nol). Dari sekian banyak data di training set, pada setiap perhitungan datanya ditambah 1 dan tidak akan membuat perbedaan yang berarti pada estimasi probabilitas sehingga bisa menghindari kasus nilai probabilitas 0 (nol).

Setelah proses pembobotan atau proses laplace correction berhasil, maka akan di lanjutkan ke proses opinion detection (penggabungan algoritma postagging dan rule opini) untuk memisahkan antara sentiment opini dan sentimen bukan opini.

Setelah proses laplace correction berhasil, selanjutnya adalah proses naïve bayes dimana metode naïve bayes berguna untuk memisahkan antara opini positif dan opini negatif.

### **3.4 Metode Pengembangan Perangkat Lunak**

Metode dalam penelitian ini menggunakan metode waterfall. Dalam metode ini terdapat 6 tahap, langkah-langkahnya adalah sebagai berikut:



Gambar 3.1 Diagram Alur *Waterfall*

a. *Requirement Analysis*

Pada tahapan ini dilakukan pengumpulan bahan-bahan yang dibutuhkan untuk mengelola sistem yang akan dibuat dan didapat dari hasil observasi dan studi literatur. Seperti fungsi-fungsi yang akan digunakan, metode yang akan digunakan, kinerja sistem tersebut, dan juga output yang dihasilkan. Setelah itu, akan didefinisikan kebutuhan tentang sistem yang akan dibuat.

b. *Design*

Proses desain akan menerjemahkan syarat kebutuhan ke sebuah perancangan perangkat lunak yang dapat diperkirakan sebelum dibuat koding. Mempelajari kebutuhan dari tahap sebelumnya, yang kemudian menyiapkan design dari sistem untuk membantu mendefinisikan arsitektur secara keseluruhan

c. *Implementation*

Dalam tahap ini dilakukan pembuatan aplikasi. Pembuatan perangkat lunak dipecah menjadi modul-modul kecil yang nantinya akan digabungkan dalam tahap berikutnya. Selain itu dalam tahap ini juga dilakukan pemeriksaan terhadap modul yang dibuat, apakah sudah memenuhi fungsi yang diinginkan.

d. *Testing*

Di tahap ini dilakukan uji coba pada aplikasi yang telah dibuat untuk mengetahui apakah perangkat lunak yang dibuat telah sesuai dengan desainnya dan apakah masih terdapat kesalahan.

e. *Maintenance*

Tahap ini dilakukan jika perangkat lunak sudah jadi dan mengalami kerusakan. Maka akan dilakukan pemeliharaan. Pemeliharaan termasuk dalam memperbaiki kesalahan yang tidak ditemukan pada langkah sebelumnya.

### 3.5 Metode Pengujian Data

Pengujian yang dilakukan untuk menguji sistem dilakukan dengan 2 langkah, yaitu pengujian fungsionalitas dan pengujian performa sistem. Pengujian bertujuan untuk memastikan bahwa sistem yang dibangun telah berjalan sesuai dengan yang diharapkan.

a. Pengujian Fungsionalitas

Uji coba fungsionalitas dikatakan berhasil apabila fungsi yang ada pada sistem sesuai dengan yang diharapkan. Uji fungsionalitas yang diterapkan pada penelitian ini ialah dengan mencoba menjalankan setiap fitur yang ada dan yang disediakan pada sistem ini

b. Pengujian Akurasi Sistem

Dalam pengujian ini, dilakukan untuk mengetahui keakuratan aplikasi dalam menampilkan data hasil pencarian oleh pengguna, yang diantara lainnya yaitu :

1. *Accuracy*

Accuracy merupakan tingkat ketepatan antara data atau informasi yang ditemukan oleh sistem dibandingkan dengan jumlah data sebenarnya. Nilai Accuracy tertinggi adalah 1, yang berarti seluruh dokumen yang ditemukan adalah relevan.

$$Accuracy = \frac{\sum \text{benar sistem}}{\sum \text{data sebenarnya}}$$

2. *Precision*

Precision merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Nilai Precision tertinggi adalah 1, yang berarti seluruh dokumen yang ditemukan adalah relevan.

$$Precision = \frac{\sum \text{ditemukan sistem}}{\sum \text{data sistem}}$$

### 3. *Recal*

Recall merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Nilai Recall tertinggi adalah 1, yang berarti seluruh dokumen dalam koleksi berhasil ditemukan.

$$Recall = \frac{\sum \text{ditemukan sistem}}{\sum \text{data sebenarnya}}$$