



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB III

METODOLOGI PENELITIAN DAN PERANCANGAN SISTEM

3.1 Metodologi Penelitian

Penelitian akan dilakukan dengan beberapa tahap, yaitu, studi literatur, pengumpulan dataset, perancangan sistem, pemrograman sistem, pengujian sistem dan penulisan laporan. Penjabaran tiap tahap adalah sebagai berikut.

1. Studi literatur dilakukan untuk melakukan pembelajaran terhadap teori-teori yang berhubungan dengan perancangan dan pemrograman aplikasi. Pada penelitian kali ini, studi literatur dilakukan dengan mengumpulkan dan mempelajari jurnal tentang analisis sentimen, Multinomial Naïve Bayes.
2. Pengumpulan dataset dilakukan dengan melakukan *text mining* menggunakan API Twitter. Pencarian digunakan dengan menulis kata kunci @gojekindonesia pada fitur *search* dari API twitter. Setelah itu dataset akan diberikan label positif atau negatif secara *manual*.
3. Perancangan sistem dimulai dengan membuat *flowchart* untuk memetakan alur informasi pada proses *training* dan *testing* pada aplikasi sentimen analisis.
4. Pemrograman sistem dilakukan dengan melakukan *coding* berdasarkan rancangan *flowchart* dan antarmuka yang sudah dibuat. Selain yang dilakukan dalam pemrograman sistem diantaranya mengimplementasikan algoritma ke dalam *coding*, dan *training dataset*.
5. Pengujian sistem dilakukan untuk menguji apakah aplikasi berjalan dengan baik. Selain itu, pada tahap ini dilakukan juga pengujian terhadap *dataset* yang telah disiapkan untuk melakukan *testing*. Dilakukan juga pengujian

evaluasi peforma dengan menghitung *accuracy*, *precision*, *recall*, dan F1 *score* dari hasil pengklasifikasian teks.

6. Penulisan laporan dilakukan dengan mencatat dokumentasi dari perancangan hingga pemrograman aplikasi. Laporan dibuat secara bertahap dimulai dari pendahuluan hingga kesimpulan dan saran.

3.2 Perancangan Sistem

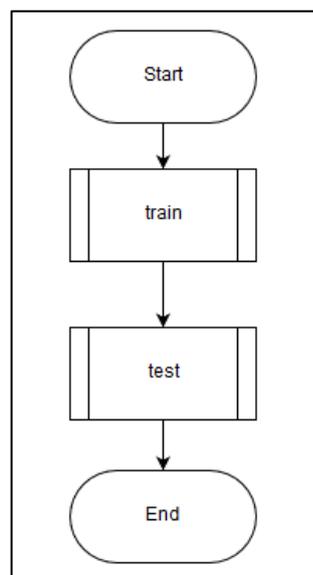
Perancangan sistem terdiri dari *flowchart* utama, *training*, *testing*.

A. Flowchart

Flowchart digunakan untuk merancang dan mengetahui proses antara satu proses dengan proses lainnya pada tahap testing, training.

A.1 Flowchart Utama

Flowchart pada Gambar 3.1 merupakan *flowchart* yang menggambarkan keseluruhan sistem sentimen analisis terkait jasa transportasi online *via* twitter.

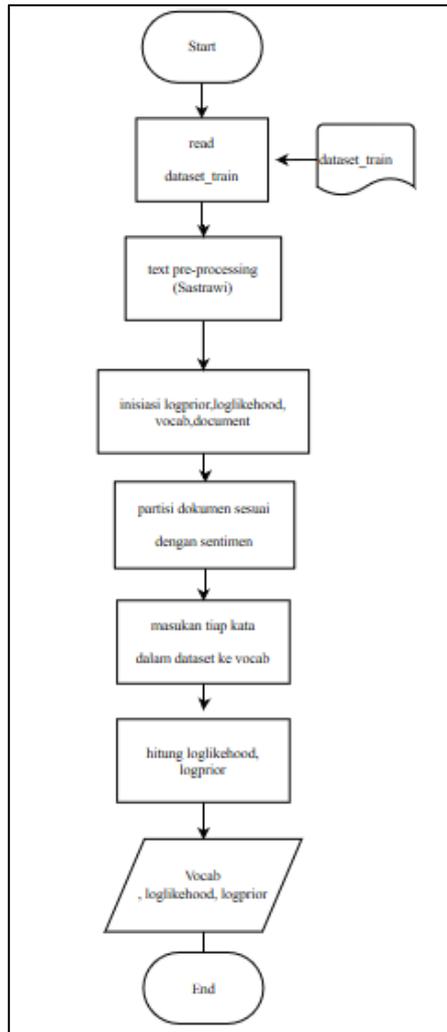


Gambar 3.1 Flowchart Utama

A.2 Flowchart Training

Flowchart pada Gambar 3.2 merupakan *flowchart* untuk training data menggunakan algoritma Multinomial Naïve Bayes. Dataset yang sudah diklasifikasi secara *manual* dibaca dari file berekstensi csv lalu dilakukan *text-preprocessing*. *Library* yang digunakan untuk melakukan *text-preprocessing* adalah Sastrawi. *Text-preprocessing* berjalan dengan tahapan *stopword removal*, *case folding*, *filtering*, *stemming* dan *tokenizing*. Setelah melalui proses *tokenizing*, dataset telah menjadi kumpulan kata dasar. Setelah itu, loglikelihood, logprior, vocab dan dokumen diinisiasi. Setelah penginisiasian *variable*, setiap dataset dipartisi sesuai dengan sentimennya yaitu 0 untuk negative dan 1 untuk positif.

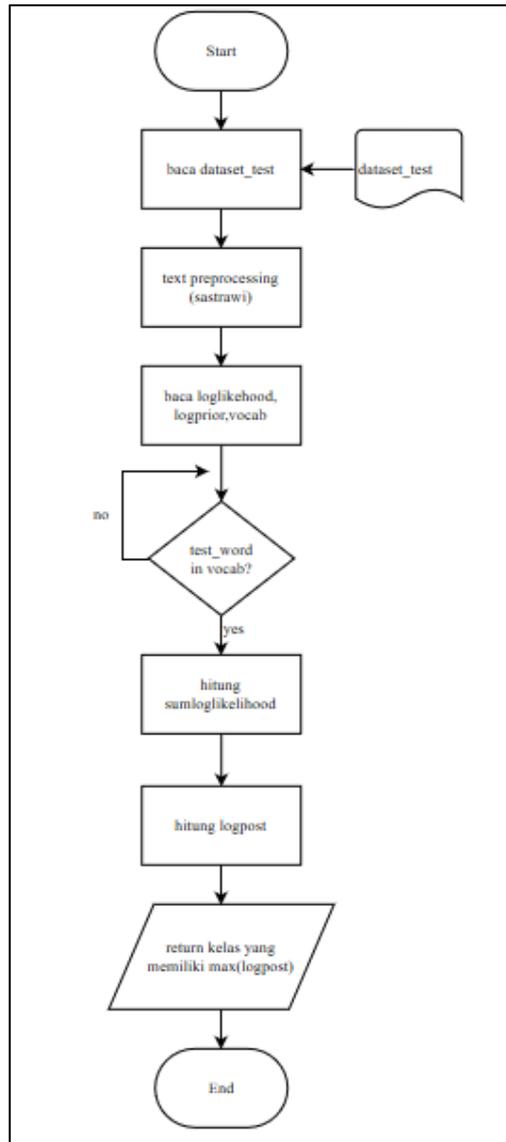
Setelah pemartisian dokumen, setiap kata yang ada didalam dokumen akan dimasukkan kedalam vocab. Langkah selanjutnya adalah melakukan perhitungan *likelihood* dan *prior probability*, rumus untuk perhitungan *likelihood* dan *prior probability* dapat dilihat pada rumus 2.2 dan 2.3. Pada perhitungan kali ini, *likelihood* dan *prior* akan dihitung menggunakan perhitungan log agar terhindar dari *floating point underflow*. Selanjutnya, *vocab*, *loglikelihood* dan *logprior* akan digunakan untuk melakukan testing.



Gambar 3.2 Flowchart Train

A.3 Flowchart Testing

Flowchart pada Gambar 3.3 merupakan *flowchart* untuk testing. Testing dimulai dengan melakukan *text preprocessing* terhadap dataset yang akan digunakan sebagai testing. Setelah itu loglikelihood, logprior dan vocab dibaca untuk dilakukan perhitungan untuk mendapatkan sentimen prediksi. Perhitungan dilakukan dengan melakukan perhitungan sumloglikelihood, yaitu menghitung total loglikelihood untuk setiap kalimat. Proses penambahan jumlah loglikelihood akan berjalan apabila kata yang ada di dataset testing, terdapat didalam vocabulary.



Gambar 3.3 Flowchart Test

Setelah mendapatkan sumloglikelihood, hasil dari sumloglikelihood akan ditambahkan dengan logprior untuk mendapatkan logpost. Logpost merupakan probabilitas masuknya suatu dokumen ke suatu kelas. Logpost mempunyai dua probabilitas, yang pertama probabilitas masuknya suatu dokumen ke dalam kelas positif, dan masuknya suatu dokumen ke dalam kelas negatif. Untuk mendapatkan prediksi, kita mengambil nilai probabilitas terbesar antara suatu dokumen tersebut termasuk kelas positif atau negatif.