

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis sentimen adalah bidang studi yang bertujuan untuk menganalisis opini, emosi, sikap dan sentimen dari teks bahasa alami menggunakan metode komputasi [1]. Analisis sentimen dapat digunakan sebagai tolak ukur penilaian terhadap suatu tokoh, brand, produk, kebijakan, serta berbagai hal yang dituangkan ke dalam media sosial sebagai opini seperti saat masa pandemi Covid-19, Twitter menjadi tempat masyarakat untuk beropini mengenai kebijakan-kebijakan yang dikeluarkan oleh pemerintah seperti pemberlakuan pembatasan kegiatan masyarakat (PPKM) Nataru. Dalam mengeluarkan opini di Twitter, banyak digunakan kata yang tidak baku dan kesalahan pada ejaan.

Pada penelitian yang berkaitan dengan perbaikan kata tidak baku dan *typo-correction* mengenai analisis sentimen oleh Lulu Luthfiana [2] pada kasus *feedback* aplikasi didapatkan hasil akurasi sebesar 77%. Penelitian tersebut menyebutkan perlunya melakukan pembenaran ejaan atau *typo correction* untuk meningkatkan hasil akurasi dalam pengklasifikasiannya. Kemudian, penelitian yang dilakukan oleh Prananda Antinasari [3] pada kasus opini film pada Twitter, penggunaan normalisasi *Levenshtein Distance* dalam penelitian ini menunjukkan hasil peningkatan akurasi dengan nilai akurasi sebesar 98.33% dari 91,67%. Selanjutnya, penelitian yang dilakukan oleh Ferly Gunawan [4] mengenai kasus ulasan aplikasi BCA mobile memiliki hasil klasifikasi dengan nilai akurasi sebesar 96,9% pada penggunaan *Levenshtein Distance* dan 94,4% tanpa *Levenshtein Distance*. Dengan menggunakan metode normalisasi *Levenshtein Distance* pada penelitian sebelumnya, keduanya menyimpulkan bahwa metode tersebut dapat digunakan dalam *typo-correction* dan menghasilkan peningkatan akurasi. Penelitian sebelumnya yang berkaitan dengan algoritma *Support Vector Machine* oleh Hennie Tuhuteru [5] penggunaan algoritma SVM mendapatkan hasil lebih baik

sebesar 81.67% dibandingkan dengan Naïve Bayes yang mendapatkan hasil 67.20%. Selanjutnya, pada penelitian oleh Mulkan Azhari [6] menghasilkan kesimpulan bahwa algoritma SVM menghasilkan akurasi yang lebih baik sebesar 95% dibandingkan dengan C4.5 yang hanya 86.67%, Random Forest 83.33% dan Naïve Bayes 86.67%. Berdasarkan kedua penelitian sebelumnya yang menggunakan metode SVM, disimpulkan bahwa metode SVM memiliki tingkat akurasi yang lebih tinggi dibandingkan metode C4.5, Random Forest, dan Naïve Bayes.

Berdasarkan latar belakang dan beberapa penelitian terdahulu, maka pada penelitian ini akan dilakukan analisis sentimen menggunakan metode *Support Vector Machine* dengan studi kasus PPKM Nataru untuk mengklasifikasikan kalimat kedalam kata positif, dan negatif. Serta, penggunaan *word normalization* untuk perbaikan kata tidak baku dan metode *Levenshtein Distance* untuk melakukan *typo-correction* atau perbaikan dalam ejaan kata sehingga dapat diketahui seberapa besar pengaruh yang diberikan perbaikan kata tidak baku dan *typo-correction* terhadap analisis sentimen.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka ditemukan masalah yang ada yaitu, sering ditemukannya teks dalam bentuk tidak lengkap atau tidak konsisten seperti adanya penggunaan kata tidak baku, dan kesalahan dalam ejaan kata yang menyebabkan kesalahan dalam penelitian dan mengurangi hasil nilai akurasi.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah mengimplementasikan metode *Support Vector Machine* dan normalisasi *Levenshtein Distance* untuk *typo-correction* dan memperbaiki kata tidak baku dalam melakukan analisis sentimen.

Adapun tujuan pada penelitian ini adalah mengukur pengaruh *preprocessing* perbaikan kata tidak baku dan *typo correction* terhadap tingkat akurasi dalam melakukan analisis sentimen pada media sosial Twitter.

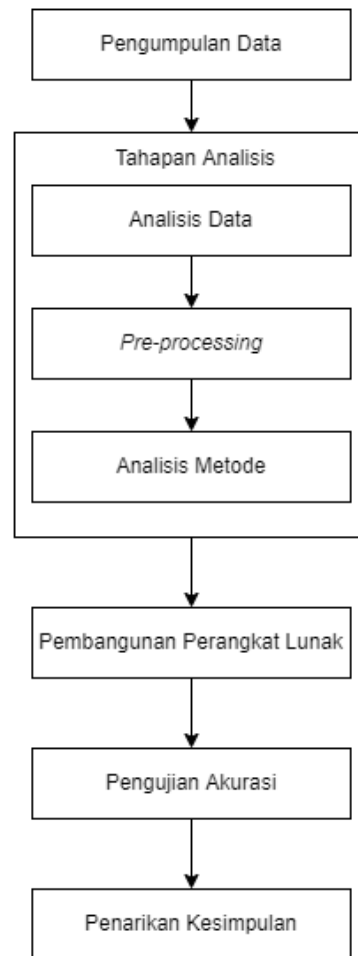
1.4 Batasan Masalah

Berdasarkan masalah yang telah didapatkan, maka batasan yang dibuat ialah sebagai berikut :

1. Tweet yang diambil hanya tweet Bahasa Indonesia.
2. Data diambil dari media sosial Twitter yang berhubungan dengan PPKM Nataru 2021 dengan menggunakan kata kunci PPKM Nataru pada tanggal 1 Desember 2021 sampai 1 Maret 2022.
3. Metode yang dipakai untuk melakukan klasifikasi adalah metode *Support Vector Machine* dan *TF-IDF* digunakan dalam pembobotan kata.
4. Perbaikan kata tidak baku diselesaikan dalam proses *word normalization* dengan menggunakan kamus_singkatan.csv.
5. *Typo-correction* dilakukan dengan membandingkan kata terhadap kamus_indonesia.txt oleh Damar Zaky[7] dan menggunakan metode *Levenshtein Distance*.
6. Tweet dikelompokkan kedalam positif, dan negatif.
7. Pelabelan kata dilakukan secara manual.

1.5 Metode Penelitian

Metode penelitian yang akan dilakukan pada penelitian dapat dilihat pada gambar 1.1.



Gambar 1.1 Metode Penelitian

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut :

1. Studi Literatur

Studi Literatur merupakan teknik pengumpulan data dengan cara mengumpulkan jurnal, paper, atau buku yang berkaitan dengan analisis sentimen dan pengaruh normalisasi yang diterbitkan dari tahun 2017 sampai 2022.

2. Pengumpulan Dataset

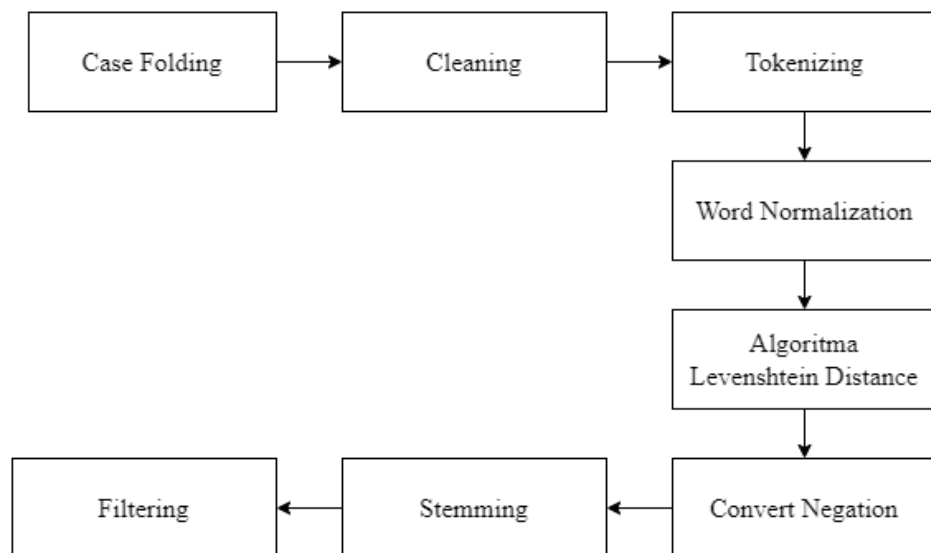
Dataset yang digunakan pada penelitian ini berupa unggahan dari pengguna twitter dengan kata kunci PPKM Nataru yang didapatkan dengan

melakukan *crawling data* menggunakan *API Key Twitter* dengan bahasa pemrograman python, *library* tweepy, pandas dan sncrape. Data yang diambil dimulai dari tanggal 1 Desember 2021 sampai 1 Maret 2022.

1.5.2 Metode Tahapan Analisis Data

Tahapan Analisis data yang digunakan dalam penelitian ini adalah sebagai berikut:

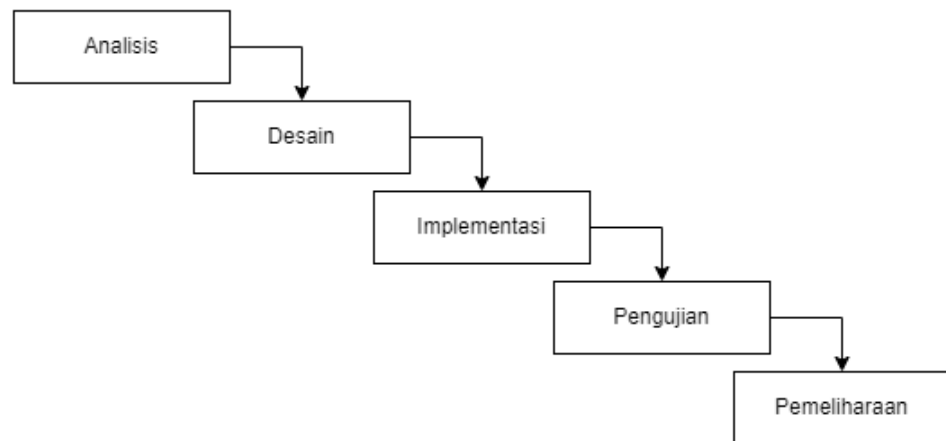
1. Analisis Data dengan menganalisis pola pada kalimat dari data yang telah didapat, data yang dianalisis merupakan unggahan dari pengguna twitter dengan kata kunci PPKM Nataru.
2. *Preprocessing*, yaitu tahap untuk memproses data sehingga menjadi lebih terstruktur dan mudah untuk diolah. Langkah yang dilakukan pada tahapan preprocessing terdiri dari *case folding*, *cleaning*, *tokenizing*, *word normalization*, algoritma *levenshtein distance*, *convert negation*, *stemming*, dan *filtering*. Dengan tahapan yang dapat dilihat pada gambar 1.2.



Gambar 1.2 *Pre-processing*

3. Perbaiki kata tidak baku dengan *word normalization* dan *Typo Correction* diperbaiki dengan menggunakan metode *Levenshtein Distance* dengan tiga operasi utama yaitu penambahan karakter, pengubahan karakter, dan penghapusan karakter.

4. Analisis Metode, pada tahap ini dilakukan klasifikasi menggunakan metode *Support Vector Machine* dengan pembobotan kata menggunakan TF-IDF (*Term Frequency — Inverse Document Frequency*) sehingga mendapatkan hasil sentimen positif, dan negatif, serta tingkat akurasi yang didapatkan dari metode yang digunakan.
5. Metode pembangunan perangkat lunak yang akan digunakan dalam penelitian ini adalah metode *waterfall*. Metode *waterfall* adalah sebuah metode pengembangan sistem dimana antar satu fase ke fase yang lain dilakukan secara berurutan[8]. Dengan tahapan yang dapat dilihat pada gambar 1.3.



Gambar 1.3 Metode Waterfall

1. Analisis dan Perencanaan

Pada tahap ini dilakukan analisis dan perencanaan mengenai kebutuhan tentang perangkat lunak yang akan dibuat dan sesuai dengan kebutuhan penelitian berdasarkan masalah yang ada.
2. Desain

Tahap desain dilakukan untuk pembuatan desain sistem secara keseluruhan dan menentukan alur dari perangkat lunak yang akan dibangun.
3. Implementasi

Pada tahap implementasi, data dan informasi yang terkumpul dan desain yang ada akan diimplementasikan menggunakan Bahasa pemrograman.

4. Pengujian

Pengujian dilakukan untk menguji bila terjadinya kesalahan dalam sistem yang dibuat.

5. Pemeliharaan

Perbaikan dari sistem yang telah diuji sebelumnya.

1.5.3 Metode Pengujian Akurasi

Dalam tahapan ini dilakukan pegujian akurasi metode menggunakan *Confusion Matrix* dan *K-fold Cross Validation* pada hasil sentimen dengan klasifikasi *Support Vector Machine* tanpa perbaikan kata tidak baku dan *typo-correction*, *Support Vector Machine* tanpa perbaikan kata tidak baku, *Support Vector Machine* tanpa *typo-correction*, dan *Support Vector Machine* dengan perbaikan kata tidak baku dan *typo-correction* terhadap unggahan pada media sosial twitter mengenai PPKM Nataru dalam mengukur pengaruh *preprocessing* analisis sentimen dengan perbaikan kata baku dan *typo correction*.

1.5.4 Penarikan Kesimpulan

Pada tahapan ini dilakukan penarikan kesimpulan berdasarkan hasil dari pengujian pengaruh pre-processing dalam penerapan metode *Support Vector Machine* tanpa perbaikan kata tidak baku dan *typo-correction*, *Support Vector Machine* tanpa perbaikan kata tidak baku, *Support Vector Machine* tanpa *typo-correction*, dan *Support Vector Machine* dengan perbaikan kata tidak baku dan *typo-correction* pada analisis sentimen.

1.6 Sistematika Penulisan

Sistematika penulisan proposal penelitian ini dibuat untuk menjelaskan secara umum tentang penelitian yang dilakukan. Dengan sistematika penulisan adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini membahas tentang pengertian dari twitter, analisis sentimen, normalisasi atau *preprocessing*, teknik dari *preprocessing*, dan masalah yang perlu diselesaikan seputar klasifikasi opini tweet yang ditunjukkan oleh masyarakat

tentang PPKM Nataru dengan menggunakan metode dalam melakukan klasifikasi tersebut dan mengukur pengaruh normalisasi terhadap analisis sentimen, metode penelitian yang akan digunakan dalam melakukan analisis sentimen, penarikan kesimpulan, dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Bab ini membahas tentang konsep dasar dan teori yang berkaitan dengan analisis sentimen, seperti kata tidak baku, *typo correction*, proses *pre-processing*, *Levenshtein Distance*, klasifikasi, metode *Support Vector Machine*, pembobotan dengan TF-IDF dan akurasi pengujian untuk dijadikan acuan dalam melakukan penelitian ini.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini membahas tentang analisa seperti analisis masalah, analisis data, gambaran umum sistem, analisis *preprocessing*, perbaikan kata tidak baku dan *typo correction* menggunakan *Levenshtein Distance*, analisis pembobotan kata dengan TF-IDF, klasifikasi dan analisis kebutuhan sistem yang akan dibangun.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas tentang implementasi sistem pada perangkat lunak, pengujian sistem, pengujian akurasi menggunakan *confusion matrix* dan *k-fold cross validation*, serta kesimpulan dari pengujian sistem dan disertai pembahasan pengujian dari sistem yang telah dibangun.

BAB 5 KESIMPULAN DAN SARAN

Bab ini membahas tentang kesimpulan yang didapatkan dari hasil penelitian serta saran yang dapat diimplementasikan untuk penelitian selanjutnya.