

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Seiring dengan semakin berkembang pesat terdapat banyaknya unggahan yang ada di media sosial sehingga butuh adanya analisis sentimen, yang berfungsi untuk mengklasifikasi data yang terstruktur maupun yang tidak terstruktur. Analisis sentimen pada media sosial dapat memberikan informasi yang bermanfaat bagi entitas yang diberikan opini. Analisis sentimen dapat mengklasifikasikan berupa opini kalimat positif, negatif, dan netral. Sehingga dapat menghasilkan sebuah informasi bagi entitas baik itu perusahaan maupun instansi[1].

Pada proses analisis sentimen beberapa masyarakat sering menggunakan sarkasme yaitu opini positif yang mengandung makna negatif. Berdasarkan hasil penelitian yang dilakukan oleh E. Inderasari[2] dalam penelitian ini menyatakan bahwa komentar pada pengguna sosial media dalam kolom komentar akun Lambe Turah mayoritas didominasi oleh pesan dengan sentimen negatif yang disampaikan dalam bentuk sarkasme. Adanya sarkasme pada data dalam proses analisis sentimen mengakibatkan kesalahan pada hasil klasifikasi karena ambiguitas akan mempengaruhi pada polaritas sentimen.

Lalu pada penelitian yang dilakukan oleh Lunando[3] menyatakan bahwa kalimat sarkasme pada topik makanan, kehidupan, dan kesehatan jarang sekali ditemukan kalimat mengandung sarkasme. Akan tetapi pada topik pemerintahan, selebritis, atau tokoh politik penggunaan kalimat sarkasme sering ditemukan.

Penelitian mengenai analisis sentimen telah banyak dilakukan, diantaranya adalah penelitian yang dilakukan oleh E. Indrayuni[4] melakukan perbandingan dengan membandingkan algoritma *Naïve Bayes* dan *Support Vector Machine*, hasil pengujian algoritma *Naïve Bayes* menghasilkan nilai akurasi hingga 84.50% sedangkan untuk algoritma *Support Vector Machine* menghasilkan nilai akurasi lebih tinggi dan akurat dengan nilai akurasi hingga 90.00%. Selanjutnya, penelitian yang dilakukan oleh S. Fanissa, M. A. Fauzi, S. Adinugroho[5] penelitian yang

dilakukan memperhatikan data terdapat pola kalimat sarkastik yang bermakna negatif dan pada penelitian ini menggunakan metode *Multinomial Naïve Bayes* yang menghasilkan akurasi tertinggi sebesar 86.6%, penulis menyarankan untuk memperhatikan kalimat sarkastik agar hasil klasifikasinya lebih sempurna.

Salah satu penelitian analisis sentimen sarkasme yang dilakukan oleh Dwi A. P. Rahayu, S. Kuntur, N. Hayatin[6] dikatakan bahwa analisis sentimen sarkasme dengan menguji dua kombinasi metode yaitu *K - Nearest Neighbors* menghasilkan *f-measure sarcasm* 82% dan *Naïve Bayes* menghasilkan *f-measure sarcasm* 50%, dan pada kombinasi *K - Nearest Neighbors* signifikan lebih tinggi dari kombinasi *Naïve Bayes*.

Dari beberapa penelitian diatas dapat disimpulkan bahwa pendeteksian sarkasme pada analisis sentimen memiliki kebutuhan penting, kemudian dari penelitian sebelumnya disimpulkan bahwa metode *Support Vector Machine* memiliki tingkat hasil akurasi yang terbaik dengan nilai tertinggi diantara metode – metode lain, dan juga penggunaan metode *K - Nearest Neighbors* menghasilkan *f-measure sarcasm* lebih baik dari kombinasi metode *Naïve Bayes*.

Berdasarkan penelitian sebelumnya yang sudah diuraikan dan terdapat permasalahan yang terjadi, maka pada penelitian ini dilakukan pendeteksian sarkasme pada analisis sentimen terhadap selebgram menggunakan *Support Vector Machine* dan *K-Nearest Neighbors*.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan maka dapat diidentifikasi masalah yaitu :

1. Terdapat polaritas pada opini publik di sosial media Twitter terhadap selebgram.
2. Terdapat pola kalimat positif yang mengandung makna negatif (sarkasme) yang dapat mempengaruhi penentuan polaritas pada analisis sentimen.

1.3 Maksud dan Tujuan

Maksud dilakukannya penelitian ini yaitu untuk melakukan pendeteksian sarkasme pada analisis sentimen terhadap selebgram dengan metode *Support Vector Machine* untuk analisis sentimen tanpa deteksi sarkasme dan *K-Nearest Neighbor* untuk melakukan analisis sentimen dengan deteksi sarkasme. Adapun tujuan pada penelitian ini adalah untuk mengukur performansi dari kombinasi metode *Support Vector Machine* dan *K-Nearest Neighbor* pada kasus analisis sentimen sarkasme terhadap selebgram.

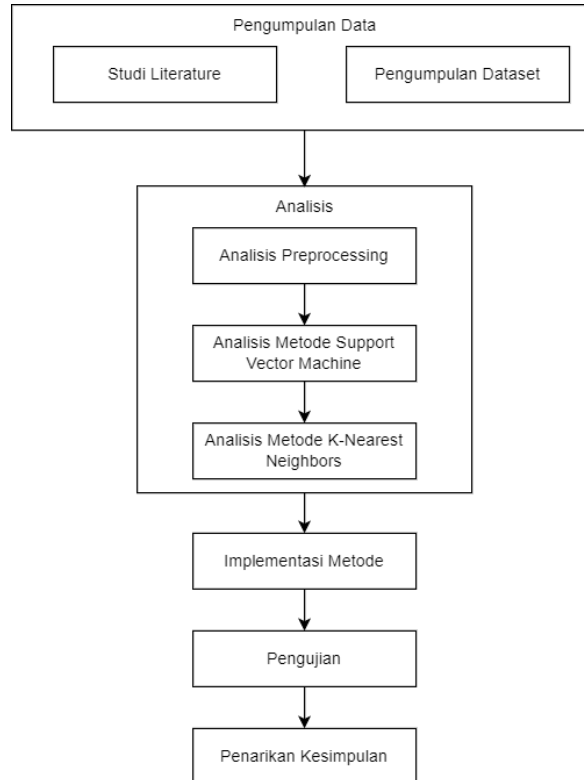
1.4 Batasan Penelitian

Permasalahan dalam penelitian ini penulis batasi agar ruang lingkup penelitian ini menjadi lebih jelas. Adapun batasan masalah dalam penelitian ini sebagai berikut :

1. Metode klasifikasi data menggunakan *Support Vector Machine*.
2. Sumber data yang akan diperoleh adalah opini publik terhadap selebgram dengan menggunakan teknik *crawling* dari Twitter yang mengandung kata atau kalimat para selebgram di Indonesia dengan *keyword* *lutfiagizal*, *deddycorbuizer*, *meldi*, *xrickolim*, *isyana*, *meyden*.
3. Sentimen twitter yang diambil hanya sentimen yang terkait dengan selebgram.
4. Sumber data diambil pada bulan Januari – November 2022.
5. Kelas untuk pelabelan pada analisis sentimen dibagi menjadi dua kelas positif dan negatif.
6. Bahasa yang digunakan pada penelitian ini yaitu data berbahasa Indonesia.
7. Kamus yang digunakan <https://raw.githubusercontent.com/haryoa/ind-collex/main/dict/inforformal-formal-Indonesian-dictionary.tsv> untuk kamus pada proses *spelling normalization*.
8. Format file untuk data latih dan data uji adalah .csv.
9. Pembobotan kata menggunakan TF-IDF untuk klasifikasi sentimen dan deteksi sarkasme.

1.5 Metode Penelitian

Berikut alur penelitian yang akan dilakukan seperti pada Gambar 1.1.



Gambar 1.1 Metode Penelitian

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini yaitu :

1. Studi Literatur

Studi literatur dilakukan untuk mencari sumber terpercaya jurnal paper, dan situs internet resmi dengan topik yang sesuai dengan penelitian ini yaitu berkaitan dengan analisis sentimen.

2. Pengambilan Data

Pengumpulan data yang dilakukan menggunakan sumber data primer yang diperoleh secara langsung menggunakan teknik *scraping* dari twitter menggunakan bahasa pemrograman python dengan 6 selebgram di Indonesia selama bulan Januari – November 2022.

1.5.2 Metode Analisis

Analisis yang dilakukan pada penelitian ini dibagi menjadi 3 yaitu :

1. Analisis *Preprocessing*

Pada bagian ini dilakukan analisis terhadap tahapan praproses data masukan yang akan diolah sebelum masuk kedalam tahap klasifikasi. Adapun tahapan praproses yang dilakukan yaitu sesuai dengan kebutuhan dari data mentah yang diperoleh dari tahap pengumpulan data.

2. Analisis Metode *Support Vector Machine*

Pada bagian ini dilakukan analisis terhadap metode klasifikasi yang akan digunakan yaitu metode *Support Vector Machine* untuk analisis sentimen positif dan negatif pada opini publik.

3. Analisis Metode *K - Nearest Neighbor*

Pada bagian ini dilakukan analisis terhadap metode klasifikasi yang akan digunakan yaitu metode *K - Nearest Neighbor* untuk analisis sentimen positif yang memiliki maksud sarkasme.

1.5.3 Implementasi Metode

Pada tahapan implementasi akan dilakukan pembuatan sistem berupa simulator, adapun hal yang bisa dilakukan oleh simulator yaitu memasukkan dan membaca data tweet yang telah dilabeli secara manual, kemudian melakukan implementasi preprocessing yaitu: *case folding*, *cleaning*, *tokenization*, *normalization*, *convert negation* dan *stopword removal*, setelah data dimasukkan telah melewati praproses dilanjutkan pada tahap pembobotan kata dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) menggunakan *library sklearn*. Lalu data dari pembobotan kata akan dilakukan pembagian data, yaitu data latih dan data uji. Lalu implementasi algoritma dengan menggunakan SVM untuk sentimen tanpa deteksi sarkas dan KNN untuk sentimen dengan deteksi sarkas.

1.5.4 Pengujian

Pada tahap ini akan dilakukan pengujian untuk menghitung tingkat akurasi dari metode *Support Vector Machine* dengan menggunakan parameter C (1.5, 1.75, 2) dan *K-Nearest Neighbor* menggunakan kernel Euclidean dan parameter K (2, 4, 8, 10, 15, 20, 25, 30, 35, dan 40) yang akan diimplementasikan dalam analisis sentimen pada opini publik terhadap selebgram.

1.5.5 Penarikan Kesimpulan

Pada tahap yang terakhir ini akan menyimpulkan dari hasil performansi pengujian metode *Support Vector Machine* analisis sentimen tanpa sarkas, performansi pengujian *K-Nearest Neighbor* deteksi sarkas, dan performansi sentimen dengan deteksi sarkas yang sesuai dengan tujuan penelitian.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan penelitian ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Pada Bab 1 berisi pembahasan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, manfaat, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada Bab 2 akan dijelaskan tentang pembahasan berbagai konsep dasar mengenai analisis sentimen, SVM, sarkasme, *K-Nearest Neighbor*, *preprocessing*, twitter, dan teori-teori pendukung lainnya yang berkaitan dengan topik.

BAB 3 ANALISIS

Pada Bab 3 ini berisi tentang analisis dan perancangan sistem yang dibangun meliputi analisis masalah, analisis data, analisis sistem, analisis *preprocessing*,

analisis algoritma SVM, analisis algoritma K - *Nearest Neighbors*, analisis non fungsional dan fungsional.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada Bab 4 ini berisi tentang gambaran dari hasil implementasi dan pengujian pada sistem yang mulai pada proses pengklasifikasian dengan metode SVM, proses deteksi sarkasme dengan metode K - *Nearest Neighbor* dan melakukan tahapan pengujian dengan metode pengujian yang akan menampilkan hasil keluaran seperti nilai *recall*, *precision* dan *f1- score*, menguji akurasi dari analisis sentimen dan pendeteksian sarkasme.

BAB 5 KESIMPULAN DAN SARAN

Pada Bab 5 berisi mengenai kesimpulan yang sudah diperoleh dari hasil penulisan tugas akhir dan agar dapat dilakukan pengembangan untuk penelitian selanjutnya.