

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam dunia pendidikan, analisis sentimen sering digunakan sebagai dasar dalam pengambilan keputusan maupun evaluasi terhadap kinerja suatu sistem. Analisis sentimen merupakan penelitian yang banyak dikembangkan pada saat ini [1]. Analisis sentimen digunakan untuk mengekstrak pendapat atau perasaan dari teks yang akan dianalisis. Analisis sentimen adalah teknik yang digunakan untuk mengekstrak data teks untuk mendapatkan informasi tentang sentimen positif, netral, atau negatif [2]. Analisis sentimen juga dikenal sebagai penambangan komentar. Dalam proses analisa sentimen, metode analisis ini melibatkan pemrosesan bahasa alami (*Natural Language Processing/NLP*), analisis teks, pembelajaran mesin, linguistik komputasi, dan berbagai metode lainnya guna menganalisis, memproses, merasionalisasi, serta mendapatkan sentimen dari teks [2], [3].

Analisis sentimen pada layanan akademik di UIN Sunan Gunung Djati Bandung adalah topik utama dari penelitian ini. Metode analisis sentimen yang digunakan melibatkan metode *lexicon based* dan metode *feature extraction*, dengan penerapan algoritma *Support Vector Machine (SVM)*.

Metode *lexicon based* adalah sebuah pendekatan yang praktis dalam menganalisis sentimen pada teks dengan menggunakan kamus kata-kata yang memiliki nilai polaritas untuk mengidentifikasi sentimen dalam suatu konten [4]. Salah satu kelebihan metode ini terletak pada kemudahan implementasinya. Namun, tantangan muncul ketika menghadapi kata-kata yang tidak terdapat dalam kamus, mengakibatkan keterbatasan dalam menangani variasi dan konteks yang lebih kompleks [5].

Sementara itu, metode *feature extraction* melibatkan perubahan teks menjadi fitur numerik untuk diolah oleh algoritma pembelajaran mesin [11]. Meskipun mampu memperhatikan konteks teks dan mengatasi kata-kata ambigu, teknik ini dianggap lebih canggih. Namun, kehati-hatian diperlukan karena

metode *feature extraction* rentan terhadap *overfitting*, yang dapat mengurangi kegeneralan model terhadap data yang baru [12].

Metode *Lexicon based* dan *Metode Feature extraction* menggunakan Algoritma *Support Vector Machine* (SVM) dipilih untuk perbandingan akurasi dalam analisis sentimen layanan akademik UIN Sunan Gunung Djati Bandung karena memiliki karakteristik yang dapat memberikan wawasan yang mendalam terhadap perbandingan performa keduanya.

Pertama, metode *Lexicon based* dipilih karena kemudahan implementasinya serta kecenderungan untuk memberikan hasil yang konsisten dalam mengklasifikasikan sentimen. Dengan menggunakan kamus kata-kata yang memiliki nilai polaritas, metode ini dapat memberikan pemahaman yang baik terhadap sentimen positif, negatif, atau netral terkait layanan akademik.

Kedua, metode *Feature extraction* dipilih karena kemampuannya menerjemahkan teks menjadi fitur numerik untuk diproses oleh algoritma SVM. Hal ini memberikan fleksibilitas dalam menangani konteks teks yang lebih kompleks dan kata-kata yang mungkin tidak terdapat dalam kamus. Meskipun canggih, metode ini perlu diuji dalam konteks layanan akademik untuk melihat sejauh mana kemampuannya.

Penggunaan SVM sebagai algoritma klasifikasi dipilih dengan pertimbangan untuk memberikan hasil yang dapat dipercaya dalam mengklasifikasikan sentimen. SVM dikenal efektif dalam menangani masalah klasifikasi pada data berdimensi tinggi.

Dengan menggabungkan dua metode yang berbeda ini, penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang kinerja keduanya dalam konteks analisis sentimen layanan akademik UIN Sunan Gunung Djati Bandung. Keseluruhan, perbandingan ini diharapkan dapat memberikan wawasan yang bermanfaat untuk pengembangan dan peningkatan layanan akademik di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang penelitian yang telah diuraikan sebelumnya, berikut rumusan masalah dari penelitian yang akan dilakukan:

- a. Bagaimana algoritma *Support Vector Machine* (SVM) dapat melakukan klasifikasi teks dalam analisis sentimen yang dilakukan?
- b. Bagaimana perbandingan akurasi dari metode *lexicon based* dan metode *feature extraction* dalam proses identifikasi sentimen pada data layanan akademik UIN Sunan Gunung Djati Bandung?

1.3 Tujuan

Tujuan dari penelitian ini adalah:

- a. Untuk memahami dan menjelaskan mekanisme serta kinerja algoritma *Support Vector Machine* (SVM) dalam melakukan klasifikasi teks pada konteks analisis sentimen terhadap layanan akademik di UIN Sunan Gunung Djati Bandung. Penelitian ini bertujuan untuk mendapatkan wawasan mendalam tentang bagaimana SVM dapat secara efektif memproses dan mengklasifikasikan teks dalam konteks analisis sentimen
- b. Mengevaluasi dan membandingkan akurasi dari kedua metode yaitu metode *lexicon based* dan metode *feature extraction* yang digabungkan dengan algoritma *Support Vector Machine* (SVM).

1.4 Batasan Masalah

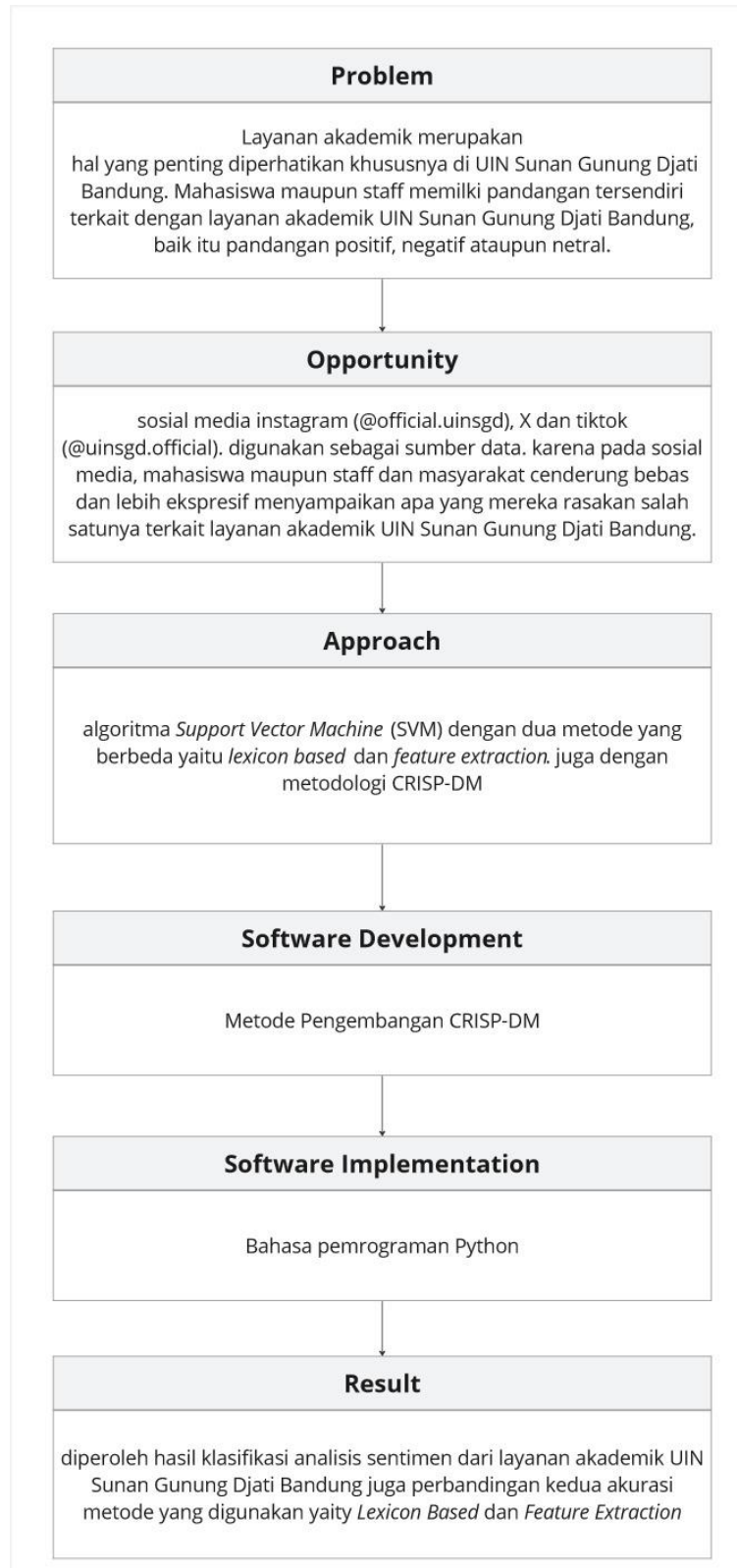
- a. Objek yang digunakan dalam penelitian ini adalah data teks berupa data layanan akademik yang didapatkan dari media sosial UIN Sunan Gunung Djati Bandung yaitu X, Instagram (@uinsgd.official), TikTok (@uinsgd.official).
- b. Pengumpulan data dilakukan selama periode satu tahun untuk media sosial Instagram (@uinsgd.official) dan TikTok (@uinsgd.official), yaitu selama tahun 2023. Adapun, pengumpulan data untuk media sosial X dilaksanakan dari bulan Maret hingga Mei 2023.
- c. Data yang digunakan adalah data yang berbahasa Indonesia saja.
- d. Algoritma yang digunakan adalah algoritma *Support Vector Machine* (SVM).
- e. Metode yang digunakan adalah metode *lexicon based* dan metode *feature extraction*.
- f. *Systems Development Life Cycle* (SDLC) yang digunakan adalah metode *waterfall*.

- g. Aplikasi yang dibuat hanya menggunakan cmd sebagai media keluaran (output).



1.5 Kerangka Pemikiran

Langkah-langkah kerangka penelitian dijelaskan pada Gambar 1.1



Gambar 1. 1 Kerangka Pemikiran

1.6 Metodologi Penelitian

Algoritma *Support Vector Machine* (SVM) adalah tipe mesin pembelajaran *supervised learning* yang bisa digunakan untuk menangani banyak kasus atau masalah diantaranya *binary class* dan *multi-class* [4]. Dalam kasus *binary-class* data akan dikelompokkan pada tiga kelas. dalam penelitian ini, kelas akan dikelompokkan menjadi 3 kelas dengan range label positif, netral dan negatif, dengan nilai yang lebih tinggi sesuai dengan emosi yang lebih positif, nilai yang lebih rendah sesuai dengan emosi negatif, dan nilai normal dengan emosi netral.

Penelitian ini dimulai dengan mengumpulkan data yang diambil dari media sosial UIN Sunan Gunung Djati Bandung. Data pada dataset yang diperoleh kemudian dilabeli dengan label positif, netral dan negatif. Kemudian dilakukan pendekatan berbasis leksikon atau *lexicon based* untuk memberikan nilai sentimen pada masing-masing data. Pendekatan leksikal ini banyak digunakan untuk pemrosesan bahasa alami atau NLP dan perhitungan seperti klasifikasi teks, analisis sentimen, dll. Penelitian menggunakan kamus data leksikon yang berasal dari inSet (Indonesia Sentiment Lexicon) yang sudah diberi label dengan *weight range* -5 s.d 5 dengan masing-masing 3.609 kata positif dan 6.609 kata negatif.

Setelah itu, data tersebut akan di *cleaning* untuk menghilangkan data yang tidak valid atau tidak relevan dengan data yang dibutuhkan. Beberapa cara untuk membersihkan data atau *cleaning* ini meliputi *case folding*, normalisasi data (menghilangkan kata gaul), *stopword removing* (kata hubung atau kata imbuhan), *stemming* (mengubah ke kata dasar).

Data yang sudah bersih akan masuk ke proses *labelling* dengan menggunakan pencocokan berdasarkan kamus *lexicon*. Pada tahap ini kalimat di dalam teks akan diiterasi setiap kata per kata untuk menentukan polarisasi dari kalimat tersebut. Dan hasil akhirnya adalah berupa label sentimen berupa label positif, negatif dan netral.

Kemudian dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) Klasifikasi ini merupakan proses pengklasifikasian data

berdasarkan label yang telah ditentukan sebelumnya. Data dibagi menjadi dua bagian yaitu data uji dan data latih, dengan lima kali pengujian dengan rasio yang berbeda.

Tahapan selanjutnya yaitu metode ekstraksi fitur, Metode ekstraksi digunakan untuk menemukan sentimen pada teks opini layanan akademik UIN Sunan Gunung Djati Bandung. Proses ekstraksi fitur ini meliputi *Term Presence*, BoW (*Bag of Words*) dan TF-IDF (*Term Frequency-Inverse Document Frequency*). TF-IDF adalah metode yang digunakan untuk mengukur seberapa penting suatu kata berada dalam suatu kalimat dengan cara menghitung jumlah kemunculan kata tersebut. BoW adalah metode dengan cara mengubah teks menjadi vektor dengan cara menghitung jumlah kemunculan setiap kata dalam suatu kalimat atau teks. Kemudian dilakukan evaluasi untuk mengukur kualitas hasil klasifikasi menggunakan evaluasi multimetrik yang meliputi skor presisi, akurasi, recall, dan F1-skor.

Tahapan terakhir adalah analisis terhadap hal apa saja dan metode mana yang cocok digunakan dalam proses klasifikasi sentimen ini yang tentunya akan memberikan rekomendasi metode apa yang tepat digunakan dalam klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) tentunya dari hasil penelitian yang sudah dilakukan.

1.6 Sistematika Penulisan

Sistematika penulisan laporan tugas akhir dirancang dengan pembagian menjadi beberapa bab, di mana masing-masing bab akan membahas beberapa topik tertentu. Oleh karena itu, berikut adalah rincian sistematika penulisan laporan tugas akhir:

BAB I : PENDAHULUAN

Bab I ini berisi latar belakang permasalahan topik yang diteliti, rumusan masalah, tujuan penelitian, batasan masalah dan kerangka pemikiran.

BAB II : KAJIAN LITERATUR

Bagian ini berisi rangkuman dari sumber literatur yang ditinjau terkait dengan topik penelitian yang dilakukan. Tujuan bab tinjauan literatur adalah untuk menyediakan kerangka dan konteks yang diperlukan untuk penelitian

yang akan dilakukan dan untuk membantu pembaca memahami landasan teori penelitian. Bab ini juga sering disebut dengan *state of the art* dari teori yang sedang dikaji.

BAB III : METODOLOGI

Bab ini memaparkan secara sistematis dan berurutan mengenai langkah-langkah dan teknik yang digunakan dalam penelitian, dengan menekankan pada penggunaan metode penelitian yang relevan dengan karakteristik dan tujuan penelitian. Umumnya, metode penelitian yang digunakan merujuk pada model-model pengembangan perangkat lunak yang ada atau model-model lain yang sesuai dengan kebutuhan penelitian yang dilakukan.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini membahas secara lebih rinci hasil dari penelitian yang sudah dilakukan berikut dengan pembahasan mengenai faktor apa yang mempengaruhi hasil tersebut. Bab ini merupakan interpretasi terhadap data yang telah diperoleh untuk menjelaskan hasil yang telah dicapai. Pembahasan harus menjawab rumusan masalah yang telah dijelaskan sebelumnya.

BAB V : SIMPULAN DAN SARAN

Bab ini berisi tentang dua hal utama yang disampaikan secara padat dan jelas. Bagian simpulan pada bab ini diharapkan dapat memberikan jawaban atas pertanyaan peneliti atau rumusan masalah. Dalam memberikan saran atau rekomendasi untuk penelitian selanjutnya, disarankan untuk memfokuskan pada dua atau tiga hal utama yang ditemukan oleh peneliti.