

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Sentimen analisis adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [1]. Sentimen analisis biasa dilakukan pada *public figure*, produk atau perusahaan yang mendapat perhatian dari masyarakat luas. Contoh *public figure* dan perusahaan yang pernah dianalisis seperti calon presiden [2], gubernur [3], PT. Pos Indonesia [4] dan Samsung [5]. Sentimen ini dilakukan untuk mengetahui opini masyarakat mengenai subjek yang dianalisis. Sentimen ini biasa memiliki beberapa tahapan penelitian yaitu tahap pra proses, lalu tahap analisis sentimen dengan berbagai macam algoritma. Tiap tahapan memiliki peranan penting dalam penelitian ini, tahap pra proses memiliki peran untuk mengatur dataset menjadi format yang sesuai agar bisa diproses oleh algoritma yang akan digunakan untuk menganalisis sentimen. Tahapan pra proses memiliki beraneka ragam metode untuk mengubah dataset menjadi format yang beraneka ragam pula. Tahapan-tahapan pra proses yang biasa digunakan adalah tokenizing, normalisasi fitur, *case folding* dan *stopword removal*. Setelah pra proses, tahap selanjutnya adalah tahap klasifikasi dengan algoritma klasifikasi. Algoritma yang sering digunakan untuk sentimen analisis adalah Naive Bayes dan SVM. Tahap klasifikasi ini dilakukan dengan tujuan memperoleh sentimen dari tiap data.

Banyaknya macam metode pada pra proses membuat tiap peneliti menggunakan kombinasi metode pra proses yang berbeda pula. Perbedaan kombinasi pada tahap pra proses ini disesuaikan dengan kebutuhan peneliti. Beranekaragamnya jumlah tahapan yang digunakan oleh peneliti dipengaruhi oleh banyaknya metode pra proses yang dapat digunakan. Peneliti dapat menggunakan 2 tahapan, 3 tahapan, 4 tahapan bahkan sampai 7 tahapan pra proses tergantung

pada kebutuhan penelitian. Berdasarkan analisis sebelumnya terhadap 50 jurnal mengenai sentimen analisis, meskipun setiap penelitian memiliki kombinasi tahapan pra proses yang berbeda tetapi ada tahapan-tahapan yang cukup sering digunakan pada penelitian sentimen analisis.

Kombinasi metode pra proses yang berbeda juga akan menghasilkan persentase keberhasilan yang berbeda juga. Berdasarkan penelitian dari Amelia Mustika, menghasilkan kesimpulan bahwa menggunakan algoritma SVM sebagai algoritma klasifikasi pada analisis sentimen yang menggunakan 5 tahapan pra proses menghasilkan akurasi sebesar 72,5% [6]. Sedangkan berdasarkan penelitian dari Valonia Inge Santoso, menghasilkan kesimpulan bahwa menggunakan algoritma SVM pada analisis sentimen yang menggunakan 4 tahapan pra proses menghasilkan akurasi sebesar 67,83% [7]. Berdasarkan 2 penelitian yang telah disebutkan, dapat disimpulkan bahwa meskipun menggunakan algoritma yang sama akan menghasilkan akurasi yang berbeda.

Berdasarkan penelitian dari Atika Rahmawati menghasilkan kesimpulan bahwa algoritma SVM menghasilkan akurasi lebih baik daripada algoritma K-means yaitu 98% untuk SVM dan 82% untuk K-Means [8]. Berdasarkan penelitian dari Andi Nurul Hidayat menghasilkan kesimpulan bahwa algoritma SVM menghasilkan akurasi lebih baik daripada algoritma Naive Bayes yaitu 90,50% untuk SVM dan 59,98 untuk Naive Bayes [9]. Berdasarkan dua penelitian yang telah disebutkan sebelumnya, maka dilakukan penelitian “Pengaruh Pra Proses Pada Analisis Sentimen Dalam Teks Berbahasa Indonesia” agar dapat mengetahui pengaruh pra proses pada akurasi dan mengetahui nilai akurasi dari setiap kombinasi pra proses.

1.2. Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan maka dapat diidentifikasi sebuah masalah yaitu penggunaan algoritma klasifikasi yang sama tetapi menggunakan kombinasi pra proses yang berbeda bisa menghasilkan akurasi yang berbeda.

1.3. Maksud dan Tujuan

Berdasarkan masalah yang telah diuraikan dalam latar belakang dan perumusan masalah, maka maksud dari penelitian ini adalah untuk membangun aplikasi untuk mengetahui pengaruh tahapan pra proses pada analisis sentimen. Sedangkan tujuan dari penelitian ini adalah untuk mengetahui kombinasi tahapan pra proses dengan nilai akurasi tertinggi.

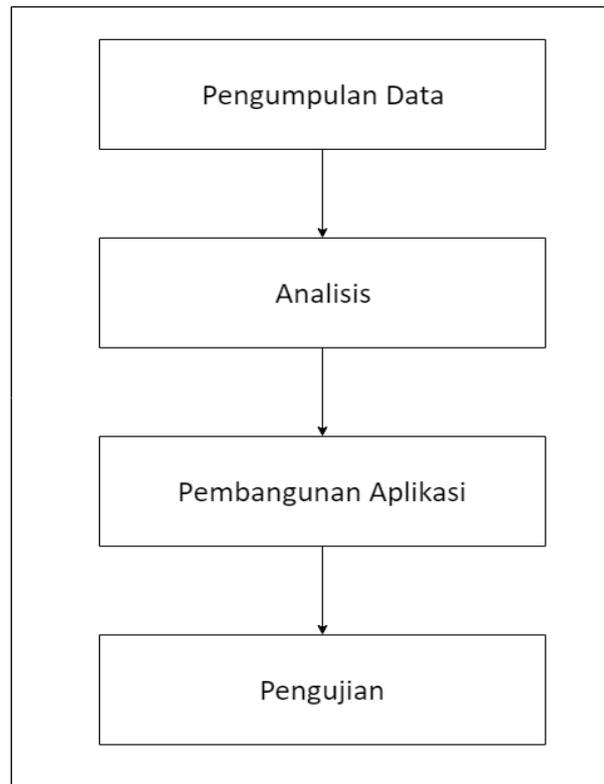
1.4. Batasan Masalah

Agar penelitian yang dilakukan lebih terarah dan dapat mencapai tujuan, maka diperlukan suatu pembatasan masalah atau ruang lingkup kajian yaitu sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini adalah dataset yang didapat dari media sosial twitter.
2. Tahapan pra proses yang digunakan adalah hasil dari analisis tahapan pra proses pada 50 jurnal sentimen analisis
3. Metode klasifikasi yang digunakan adalah *Support Vector Machine*

1.5. Metode Penelitian

Penelitian ini menggunakan beberapa tahapan yaitu pengumpulan data, analisis, pembangunan aplikasi, pengujian, dan penarikan kesimpulan dari penelitian yang telah dilakukan. Gambaran metode yang akan dilakukan adalah seperti berikut:



Gambar 1. 1. Metode Penelitian

1.5.1. Pengumpulan Data

Pada tahap ini melakukan pengumpulan data, pengumpulan data pada tahap ini dibagi menjadi 2 proses yaitu pengumpulan dataset dan pengumpulan 50 jurnal sentimen analisis

1.5.1.1. Pengumpulan Dataset

Pada proses ini dilakukan pengumpulan dataset dari twitter dengan cara mengambil tweet masyarakat. Tweet yang diambil adalah tweet yang dicuitkan selama bulan September dan Desember.

1.5.1.2. Pengumpulan 50 Jurnal

Pada proses ini dilakukan pengumpulan 50 jurnal. Jurnal yang dikumpulkan adalah jurnal yang mengangkat topik mengenai sentimen analisis yang diterbitkan dari tahun 2013 sampai 2018. Jurnal diambil dari web <https://scholar.google.com>.

1.5.2. Analisis

Pada tahap ini terdapat 2 proses analisis yaitu proses analisis pra proses 50 jurnal dan analisis sentimen

1.5.2.1. Analisis Pra Proses 50 Jurnal

Pada tahap ini melakukan analisis tahapan pra proses pada 50 jurnal mengenai sentimen analisis. Tiap tahapan pra proses yang digunakan oleh 50 jurnal tersebut didata dan dianalisis sebestumnya menyamakan pengertian pra proses yang digunakan lalu dikumpulkan perkomposisi tahap pra proses. Jika ada yang menggunakan komposisi pra proses yang sama maka komposisi pra proses tersebut hanya dihitung satu kali. Setelah itu data komposisi pra proses tersebut diurutkan berdasarkan penggunaan pra proses terbanyak.

1.5.2.2. Analisis Sentimen

Pada tahap analisis sentimen ini dilakukan beberapa proses yaitu penerapan pra proses, pembobotan Tf/Idf, dan klasifikasi menggunakan SVM. Dataset yang sudah dikumpulkan pada tahap pertama diproses menggunakan pra proses yang telah dianalisa dari 50 jurnal sentimen analisa. Setelah melalui tahap pra proses, dilakukan pembobotan menggunakan Tf/idf. Setelah dilakukan pembobotan, lalu dilakukan klasifikasi menggunakan SVM dengan kernel linier hingga diketahui sentimen pada dataset tersebut.

1.5.3. Pembangunan Aplikasi

Pada tahap ini dilakukan pembangunan aplikasi untuk mengetahui akurasi pada sentimen analisis dengan beberapa komposisi pra proses yang berbeda. Aplikasi ini dibangun berbasis web dengan bahasa pemograman html dan php.

1.5.4. Pengujian

Pada tahap ini dilakukan pengujian pada sentimen yang didapat dari simulator. Pengujian ini dilakukan dengan metode *conclusion matrix* pada semua hasil sentimen analisis dengan beberapa komposisi pra proses yang berbeda. Sentimen yang didapat akan dibandingkan dengan sentimen sesungguhnya yang telah dianalisis secara manual di luar simulator. Pengujian dilakukan untuk

mendapatkan akurasi dari sentimen analisis dengan tahapan pra proses yang berbeda.

1.6. Sistematika Penulisan

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

BAB 1 PENDAHULUAN

Bab 1 ini membahas mengenai latar belakang dilakukannya penelitian, perumusan masalah berdasarkan latar belakang, menentukan maksud dan tujuan penelitian, menentukan batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab 2 ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang diangkat dan hal-hal yang berguna yang dijadikan acuan penelitian ini serta tinjauan dari penelitian-penelitian serupa yang pernah dilakukan sebelumnya.

BAB 3 ANALISIS DAN PERANCANGAN

Bab 3 ini membahas mengenai analisis sentimen, analisis tahapan pra proses, serta implementasi *Support Vector Machine* sebagai algoritma klasifikasi.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab 4 ini berisi mengenai implementasi dari hasil analisis dan perancangan yang telah dibuat serta pengujian sistem yang telah dibangun untuk menentukan apakah sistem telah sesuai dengan kebutuhan.

BAB 5 KESIMPULAN DAN SARAN

Bab 5 ini berisi tentang kesimpulan dari penelitian yang telah dilakukan serta saran-saran yang dapat diberikan untuk pengembangan kedepannya.