

BAB I PENDAHULUAN

1.1 Latar Belakang

Pada saat ini internet telah menjadi hal yang penting dan menjadi kebutuhan bagi masyarakat untuk sekedar mencari informasi atau menghasilkan uang dari internet. Salah satu penyedia layanan internet di Indonesia adalah Biznet. Biznet merupakan operator telekomunikasi dan multimedia di Indonesia yang memberikan layanan jaringan, internet, pusat data, serta layanan hosting dan cloud computing. Biznet merupakan salah satu penyedia layanan internet dengan pengguna terbanyak di Indonesia. Peran pengguna Biznet dalam memberikan opini berupa kritis dan saran dapat berguna sebagai feedback untuk Perusahaan Biznet agar bisa lebih baik dalam memberikan pelayanannya.

Sebagian besar pengguna mengambil saran, ulasan atau pendapat dari opini orang lain sebelum mereka memutuskan untuk memilih penyedia layanan Internet yang akan digunakan. Media sosial adalah tempat dimana seseorang dapat mengemukakan sentimen, pendapat, ulasan tanpa ada yang membatasi. Dampak media sosial ini sangat berpengaruh baik untuk konsumen maupun produsen tanpa disadari.

Dengan demikian, dari ulasan-ulasan tersebut dapat memberikan nilai potensial yang dapat diimplementasikan untuk analisis sentimen yang bertujuan untuk meningkatkan strategi dalam pemasaran, peningkatan mutu produk atau pelayanan untuk menjadi lebih baik. Salah satu contoh media sosial yang sering digunakan untuk bertukar opini adalah Twitter. Dari twitter dapat dilakukan Opinion Mining atau biasa disebut Analisis Sentimen. Untuk melakukan Twitter Mining peneliti akan mengambil data yang digunakan dengan memanfaatkan API Twitter.

Analisis Sentimen merupakan suatu proses pengolahan data berupa teks untuk mengklasifikasi sentimen positif. Dalam proses pengolahan data sentimen akan melewati beberapa tahapan yaitu pengumpulan data, cleaning data, tokenization, stopword removal, stemming, feature list, pembobotan kemudian setelah itu dilakukan klasifikasi menggunakan algoritme Support Vector Machine (SVM). SVM merupakan salah satu metode klasifikasi dengan menggunakan machine learning (supervised learning) yang memprediksi kelas berdasarkan model atau pola dari hasil proses training. Klasifikasi dilakukan dengan mencari hyperplane atau garis pembatas (decision boundary) yang memisahkan antara suatu kelas dengan kelas lain [1], yang dalam kasus ini adalah opini

positif dan opini negatif. Dalam penelitian yang berjudul "Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques" yang dilakukan oleh Shweta Rana dan Archana Singh [2] dapat disimpulkan bahwa metode ini lebih baik dibandingkan metode lainnya. Dengan menggunakan metode Support Vector Machine, tingkat keakuratan mencapai 87,50%, dan lebih tinggi 3% hingga 7% dibandingkan dengan metode Naive Bayes. Oleh karena itu maka dalam penelitian ini akan menggunakan metode Support Vector Machine untuk mengukur nilai akurasi sentien yang diperoleh dari hasil crawling data pengguna twitter terhadap Biznet.

1.2 Rumusan Masalah

Perumusan masalah dalam skripsi ini adalah:

1. Bagaimana mengklasifikasikan sentimen pada data Twitter pengguna biznet menggunakan algoritme Support Vector Machine ?
2. Berapakah nilai akurasi algoritme Support Vector Machine pada klasifikasi sentimen terhadap pleyanan Biznet ?

1.3 Batasan Masalah

Penulis merumuskan batasan masalah yang hendak dibahas pada penelitian ini agar penulisan menjadi lebih terarah. Batasan masalah dalam pembuatan skripsi ini adalah :

1. Penelitian hanya dilakukan terhadap data dari twitter dengan kata kunci yang berhubungan dengan Biznet seperti '*Biznet*', '*Biznet Home*', '*Biznet Networks*', '*Internet Biznet*'
2. Bahasa yang diambil dari Twitter adalah bahasa Indonesia.
3. Analisis sentimen dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh sebanyak tiga kali dalam rentang waktu 2 minggu.
4. Analisa sentimen dengan metode SVM menggunakan tools yang tersedia di library python.
5. Data sentimen diklasifikasikan sebagai kelas positif, negatif dan netral.
6. Bahasa pemrograman yang digunakan adalah Python.
7. Python yang digunakan adalah python versi 3.8.1
8. Jumlah data training 500 data.

9. Analisis sentimen untuk data pengujian data testing dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh secara *realtime*.

1.4 Tujuan Penulisan

Tujuan dari pembuatan Skripsi ini adalah:

1. Menentukan sentiment yang di berikan oleh pengguna Biznet dari komentar di twitter menggunakan algoritme Support Vector Machine.
2. Menerapkan metode SVM pada analisis sentimen terhadap data twitter Biznet.
3. Mencari nilai akurasi dari algoritme Support Vector Machine dalam mengklasifikasi sentimen terhadap layanan Biznet.

1.5 Manfaat Penulisan

Manfaat yang diharapkan dari penulisan Skripsi ini antara lain:

- a. Bagi Biznet
Mendapatkan presentase sentimen publik dari pengolahan data twitter yang menjadi tolak ukur dan guna meningkatkan pelayanan dan mutu dari Biznet.
- b. Bagi Ilmu Pengetahuan
 1. Mengetahui bagaimana implementasi opinion mining dalam pemrograman python.
 2. Mengetahui cara memperoleh dan mengolah data untuk sentimen analisis dengan menggunakan algoritme Support Vector Machine

1.6 Metode Penellttian

Dalam penelitian ini akan menjelaskan tahapan – tahapan sentimen analisis sebagai berikut.

1.6.1 Metode Pengumpulan Data

1.6.1.1. Studi Pustaka

Mencari literatur atau jurnal yang bersangkutan dengan text mining, sentimen analisis, algoritme SVM dan K-Fold Cross Validation.

1.6.2 Metode Analisis

Dalam metode analisis ini terdapat beberapa tahapan yang akan dilakukan yaitu :

1. Collecting Data
Data yang dikumpulkan diperoleh dari crawling data twitter menggunakan Twitter API yang diunduh sebanyak 3 kali dalam rentang waktu 2 minggu.
2. Data Preprocessing
Tahap preprocessing adalah tahap awal pengolahan data teks dalam analisis sentimen. Terdapat beberapa metode yang digunakan pada tahap preprocessing diantaranya adalah:
 - a. Cleaning data merupakan tahapan untuk menghapus kata-kata yang tidak dipakai seperti username, hashtag, dan link.
 - b. Lower casing untuk membuat keseluruhan huruf di tweet menjadi huruf kecil.
 - c. Tokenization merupakan tahap untuk memisahkan data teks menjadi beberapa token dan menghilangkan kata yang tidak bermakna.
 - d. Stemming merupakan tahapan untuk merubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan awal atau akhir pada kata tersebut. Contohnya "Melelahkan" menjadi kata dasar yaitu "lelah".
3. Feature Weighting
 - a. Term Frequency (TF) merupakan frekuensi kemunculan kata pada kalimat.
 - b. Document Frequency (IDF) kemunculan berapa banyak kata dalam satu kalimat.
4. Klasifikasi Support Vector Machine
Pada tahapan ini data akan diklasifikasi menggunakan algoritme SVM dengan menentukan suatu kelas berdasarkan pola yang terbentuk dari hasil proses training. Data diklasifikasi menjadi 2 tahap yaitu dengan

tahap training berjumlah 500 data yang kemudian akan diberikan label secara manual dan untuk data pengujian diunduh secara realtime.

5. Menghitung akurasi

Pada tahapan ini akan dilakukan perhitungan akurasi dari penelitian yang dilakukan. Metode yang digunakan yaitu K-Fold Cross Validation. Dalam penelitian ini, tahap pengujian menggunakan 10-Fold sebagai parameter untuk melakukan Cross Validation yang diuji menggunakan tools pada python.

1.6.3 Metode Implementasi

1. Implementasi sistem di shell python menggunakan bahasa python.
2. Manual program.
3. Hasil implementasi.

1.6.4 Metode Testing

Metode yang digunakan untuk mengetahui rata-rata keberhasilan suatu sistem menggunakan K-Fold .

1.7 Sistematika Penulisan

Laporan penelitian ini akan disusun secara sistematis yang akan diurutkan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan menjelaskan mengenai latar belakang masalah, perumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang teori yang digunakan sebagai dasar dalam analisis penelitian. Mencakup tinjauan pustaka meliputi pengertian data mining, bahasa pemrograman Python, Support Vector Machine (SVM), sentiment analysis, pre-processing, Term Frequency-Inverse Document Frequency, K-Fold cross validation dan confusion matrix.

BAB III ANALISIS DAN PERANCANGAN

Bab ini menjelaskan mengenai perancangan dan model cara kerja yang akan dilakukan di dalam studio kasus sentimen analisis tweet pengguna Biznet.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini menjelaskan proses dan hasil analisis sentimen dari pre-processing data, tokenization, stemming, TF-IDF, SVM, dan Confusion Matrix.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa dan optimalisasi sistem berdasarkan yang telah diuraikan pada bab-bab sebelumnya.

