

BAB I

PENDAHULUAN

1.1. Latar Belakang

Persaingan antar penyedia layanan internet semakin hari semakin banyak. Masing- masing penyedia layanan internet berlomba-lomba untuk mendapatkan pelanggan sebanyak-banyaknya. Semakin banyak pengguna tentunya akan menguntungkan bagi pihak penyedia layanan internet. Namun, kualitas penyedia layanan internet tidak dapat ditentukan dari jumlah penggunanya saja. Tetapi seberapa luas infrastruktur yang di miliki oleh penyedia layanan internet, semakin luas jangkauan maka probabilitas pelanggan semakin tinggi.

Dalam industri yang menyediakan layanan tentunya tidak menjamin semua pelanggan puas dengan layanan yang diberikan, pasti ada opini dari pelanggan yang bermacam-macam. Di jaman yang moderen ini pelanggan biasanya meluapkan opini positif maupun negatif di sosial media mereka, karena mereka mempunyai kontrol penuh terhadap sosial media sendiri.

Dengan itu opini-opini yang muncul bisa memberikan nilai positif terhadap pelayanan dari Telkom dan Biznet untuk meningkatkan pelayanan, strategi pemasaran, dan peningkatan produk agar lebih baik. Salah satu media sosial yang digunakan tempat beropini yaitu twitter. Dari twitter dapat dilakukan *opinion mining* atau yang disebut analisis sentimen. Untuk mendapatkan data guna melakukan twitter mining peneliti akan memanfaatkan *API Twitter* untuk mengambil data.

Penulis bermaksud untuk membuat sebuah aplikasi analisis sentimen pengguna twitter terhadap pelayanan Telkom dan Biznet menggunakan algoritma *Support Vector Machine* yang digunakan oleh Telkom dan Biznet, dengan tujuan mengetahui presentase sentimen negatif dan positif dari Telkom dan Biznet sehingga kedua perusahaan tersebut bisa meningkatkan pelayanan.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang permasalahan di atas, maka penulis mengidentifikasi masalah sebagai berikut :

1. Bagaimana melakukan analisis sentimen yang diberikan oleh pengguna Telkom dan Biznet dari *tweet* yang ada di twitter menggunakan algoritma *Support Vector Machine*?
2. Bagaimana mengukur nilai akurasi dengan *K-Fold Cross Validation* dan *Confusion Matrix* dari algoritma *Support Vector Machine* pada klasifikasi sentimen terhadap pelayanan Telkom dan Biznet?

1.3. Batasan Masalah

Dari rumusan masalah diatas maka dibuatlah batasan masalah agar penulisan skripsi ini tidak menyimpang dari tujuan yang direncanakan :

1. Bahasa yang diambil dari twitter adalah Bahasa Indonesia.
2. Penelitian hanya dilakukan pada *tweet* yang memiliki hastag yang berhubungan dengan Telkom dan Biznet seperti 'telkom', 'telkomindonesia', 'indihome', 'biznet', 'biznetnetworks', dan juga 'biznethome'.
3. Bahasa pemrograman yang digunakan adalah Python dengan versi 3.7.

4. Analisa sentimen menggunakan library yang sudah tersedia di Python dan menggunakan metode *Support Vector Machine*.
5. Data sentimen diklasifikasikan kedalam 2 kelas yaitu positif dan negatif.
6. Data yang digunakan untuk pengujian dalam penelitian ini menggunakan data *tweet* dari twitter.
7. Jumlah data *training* 500 data, 250 biznet, dan 250 telkom.
8. Perhitungan akurasi dari algoritma *Support Vector Machine* menggunakan *K-Fold Validation (10-fold)* dan *Confusion Matrix*.

1.4. Tujuan Penelitian

Tujuan dari pembuatan skripsi ini adalah:

1. Menerapkan metode *Support Vector Machine* untuk analisis sentimen terhadap pelayanan Telkom dan Biznet.
2. Menemukan sentimen yang berhubungan dengan Telkom dan Biznet di twitter menggunakan algoritma *Support Vector Machine*.
3. Mencari nilai akurasi algoritma *Support Vector Machine* dalam klasifikasi sentimen terhadap pelayanan Telkom dan Biznet dengan menggunakan *K-Fold Validation (10-fold)* dan *Confusion Matrix*.

1.5. Manfaat Penelitian

- a. Bagi Perusahaan Telkom dan Biznet

Mendapatkan presentase sentimen publik dari data twitter terhadap pelayanan yang diberikan oleh Telkom dan Biznet guna meningkatkan kualitas pelayanan dan mutu dari Telkom dan Biznet.

b. Bagi ilmu pengetahuan

1. Mengetahui cara memperoleh hingga mengolah data untuk analisis sentimen.
2. Mengetahui bagaimana implementasi algoritma *Support Vector Machine* untuk analisis sentimen menggunakan pemrograman Python.

1.6. Metode Penelitian

Metode penelitian yang dilakukan untuk mendapatkan informasi dan data mengenai objek permasalahan ini adalah :

1.6.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan sebagai berikut :

1.6.1.1. Studi Pustaka

Mencari literatur atau jurnal yang bersangkutan dengan analisis sentimen, *opinion mining*, algoritma *Support Vector Machine*, *K-Fold Cross Validation*, dan *Confusion Matrix*.

1.6.1.2. Wawancara

Perlunya pengetahuan mengenai langkah-langkah yang digunakan dalam analisis sentimen, sehingga dibutuhkannya wawancara bersama orang yang ahli atau sudah pernah melakukan analisis sentimen, selain itu dibutuhkan juga informasi mengenai cara kerja dari metode-metode yang biasa digunakan dalam analisis sentimen.

1.6.2. Metode Analisis

Dalam metode ini terdapat beberapa tahapan, yaitu sebagai berikut :

1. *Collecting Data*

Data diperoleh dari *crawling* data twitter menggunakan *API Twitter* selama 4 kali dalam rentang waktu 4 minggu.

2. *Preprocessing Data*

Merupakan tahap awal pengolahan data dalam data mining, yang tahapannya terdiri dari :

- a. *Data Cleaning*, adalah tahap menghapus kata yang tidak dipakai seperti *hashtag* dan *username*, dan juga mengurangi *noise* saat menganalisis sentimen.
- b. *Case folding* merupakan cara untuk membuat keseluruhan data menjadi *lowercase* (huruf kecil).
- c. *Tokenization* merupakan tahap memisahkan kalimat menjadi kata.
- d. *Stopword* merupakan tahap yang berfungsi menghapus kata yang tidak memiliki arti yang deskriptif.

3. *Feature Weighting*

Feature Wighting merupakan proses pembobotan setiap kata pada setiap dokumen, pada penelitian ini menggunakan teknik *TF-IDF*.

4. *Consine Similarity*

Merupakan fitur yang digunakan untuk mencari tingkat kesamaan data untuk meningkatkan nilai akurasi.

5. Klasifikasi *Support Vector Machine*

Pada tahap ini data akan dilatih menggunakan Algoritma *Support Vector Machine* dengan menentukan suatu kelas berdasarkan pola yang terbentuk dari hasil proses *training*. Klasifikasi dilakukan dengan memisahkan antara opini positif dan opini negatif.

6. Menghitung akurasi model

Untuk menguji keakuratan model dari *Support Vector Machine* maka dilakukan pengujian menggunakan *confusion matrix* agar menghasilkan performa model berupa nilai akurasi, *precision*, *recall*, dan *f1-score* dengan *10-Fold Cross Validation* sebagai metode untuk membagi data *training* dan data *testing*, dengan hasil akhir yaitu nilai rata-rata seluruh *fold*.

1.6.3. Metode Perancangan

Perancangan yang perlu dilakukan adalah perancangan untuk mengimplementasikan algoritma ke dalam aplikasi menggunakan *flowchart*. Penerapan algoritma *Support Vector Machine* untuk suatu system diperlukan perancangan system dari tahap *preprocessing*, *training*, hingga *testing*, serta alur dari *Support Vector Machine* bekerja. Selain itu untuk memudahkan penggunaan aplikasi perlu dirancangnya antarmuka pengguna.

1.6.4. Metode Implementasi

Proses implementasi pembuatan sistem di lakukan menggunakan bahasa pemrograman Python dengan memanfaatkan *library* dari *sklearn*, *numpy*, *pandas*, *sastrawi*, dan *nltk*.

1.6.5. Metode Pengujian

Pengujian dilakukan dengan menggunakan metode *K-fold Cross Validation* yang merupakan teknik untuk melakukan evaluasi kinerja model. Pengujian membagi kedalam data *training* dan data *testing* menggunakan metode *10-Fold Cross Validation*. Masing-masing *fold* akan dihitung performanya menggunakan metode *confusion matrix* untuk memberikan hasil akhir performa model.

1.7. Sistematika Penulisan

Laporan ini akan disusun secara sistematis dalam bentuk bab dan masing-masing bagian akan diuraikan ke dalam beberapa sub-bab agar yang akan diurutkan sebagai berikut :

BAB I PENDAHULUAN

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

BAB II LANDASAN TEORI

Pada bab ini menjelaskan mengenai teori yang digunakan sebagai dasar dalam analisis penelitian. Mencangkup tinjauan pustaka meliputi pengertian *data mining*,

Support Vector Machine (SVM), analisis sentimen, Pre-Processing, Term Frequency - Inverse Document Frequency, Cosine Similarity, K-Fold Cross Validation dan Confusion Matrix.

BAB III ANALISIS DAN PERANCANGAN

Pada bab ini akan menjelaskan mengenai perancangan model, cara kerja yang akan dilakukan di dalam analisis sentimen twitter Telkom dan Biznet.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Pada bagian bab ini akan menjelaskan kode program mulai dari *preprocessing*, hingga mendapatkan model algoritma *Support Vector Machine*, dan *testing*. Serta menerapkan dari model algoritma *Support Vector Machine* ke dalam aplikasi.

BAB V PENUTUP

Pada bab ini berisi kesimpulan yang di dapat dari hasil analisis, perancangan, dan implementasi yang telah dibahas pada bab-bab sebelumnya serta saran untuk pengembangan penelitian yang lebih baik.