

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi, informasi, dan komunikasi di Indonesia semakin meningkat dikarenakan kebutuhan publik akan komunikasi dan akses informasi yang tinggi. Melalui internet pengguna dapat menemukan informasi dan data yang dibutuhkan dengan cepat dan mudah. Rata-rata penggunaan internet harian yang dilakukan oleh pengguna dengan umur 16 tahun hingga 64 tahun di Indonesia pada tahun 2020 adalah 7 jam 59 menit dimana 3 jam 26 menitnya dihabiskan untuk mengakses jejaring sosial[1]. Jejaring sosial merupakan situs dimana setiap orang saling terhubung dengan orang lain seperti teman atau masyarakat secara *online* sehingga terjalin interaksi komunikasi untuk berbagi informasi, berkomunikasi jarak jauh, berpendapat, memberikan pesan, dan masih banyak lagi[2].

Teknologi seluler di Indonesia terutama bidang industri telekomunikasi berkembang semakin pesat dan mengalami persaingan seiring munculnya layanan penyedia telekomunikasi baik lokal maupun asing. Hal tersebut menyebabkan para *provider* telekomunikasi saling gencar melakukan pemasaran guna menjaring pelanggan baru, salah satunya menggunakan strategi pembuatan akun sosial media perusahaan layanan di Twitter[3].

Twitter merupakan layanan *microblogging* dimana pengguna dapat saling berinteraksi melalui tulisan dengan batas 280 karakter. Melalui *platform* ini, perusahaan mampu melakukan interaksi dengan pelanggan sehingga tahap evaluasi layanan perusahaan dapat dilakukan sesuai dengan umpan balik pelanggan. Pengguna sosial media Twitter sebagai pengguna layanan provider

internet dapat memberikan keluhan, komentar, dan pertanyaan. Pada penelitian yang dilakukan oleh Setyohadi (2017), didapatkan hasil analisis berupa interaksi keluhan memiliki persentase terbesar sebanyak 42,20%, kemudian komentar 35,5%, dan pertanyaan 22,25%[4].

PT Telekomunikasi Seluler Tbk (Telkomsel) merupakan salah satu perusahaan penyedia layanan telekomunikasi dan operator telepon seluler pertama di Asia yang memperkenalkan jasa GSM prabayar pada November 1997, hingga saat ini menjadi operator seluler tingkat pertama yang memiliki pelanggan terbanyak di Indonesia[5]. Telkomsel menawarkan beberapa produk di bidang telekomunikasi berupa layanan pascabayar (kartu Halo), prabayar (simPATI, kartu As, Loop), layanan khusus modem USB/*mobile*, dan layanan digital *telco* (by.U)[6].

Pada bulan Oktober tahun 2019, Telkomsel meluncurkan produk baru sekaligus layanan seluler prabayar digital pertama di Indonesia yaitu by.U. Produk tersebut merupakan pembaruan dari produk sebelumnya dengan nama hüp! beta yang telah rilis pada Agustus 2019. Layanan baru ini ditujukan untuk generasi Z yang merupakan generasi dengan gaya hidup digital. Melalui by.U pengguna diberikan pengalaman digital dalam kebebasan untuk mempersonalisasikan layanan sesuai dengan kebutuhan mereka mulai dari pemilihan nomor telepon, pembelian kuota internet, hingga sistem pembayaran yang diinginkan melalui aplikasi by.U. Layanan by.U menyediakan akun resmi Twitter (@byu\_id), Instagram (@byu.id) dan Facebook. Melalui *platform* yang disediakan, pengguna layanan dapat memberikan opininya terhadap layanan yang diberikan oleh by.U, baik bersifat positif maupun negatif. Opini pengguna merupakan informasi penting

karena dapat digunakan untuk mengetahui tingkat kepuasan masyarakat terhadap layanan by.U melalui proses *text mining*.

*Text mining* adalah proses pengambilan informasi berupa data teks yang berkualitas sehingga dapat mengetahui permasalahan pada topik tertentu, salah satu tujuan dari *text mining* adalah analisis sentimen[7]. Analisis sentimen merupakan proses pengklasifikasian opini yang sebelumnya telah melalui beberapa tahapan proses pengolahan data sentimen ke dalam dua kelas yaitu kelas opini positif dan kelas opini negatif. Salah satu metode yang digunakan dalam analisis sentimen ini adalah Support Vector Machine (SVM). Metode ini bekerja dengan cara menemukan *hyperplane* terbaik yang memisahkan dua kelas dengan margin terbesar. Proses klasifikasi analisis sentimen sangat cocok dengan metode support vector machine karena akurasi yang dihasilkan tetap tinggi meskipun dataset yang diproses berjumlah besar dengan jumlah fitur dalam bentuk *term* yang sangat banyak[8].

Dalam penelitian dengan judul "*Feature Extraction Untuk Klasifikasi Pengenalan Wajah Menggunakan Support Vector Machine dan K-Nearest Neighbor*" menggunakan dataset berupa kumpulan foto wajah pria dan wanita didapatkan hasil nilai rata-rata akurasi klasifikasi dari Support Vector Machine sebesar 88.13% sementara K-Nearest Neighbour sebesar 84.40%[9].

Sementara pada referensi penelitian lain dengan judul "*Perbandingan Teknik Klasifikasi Neural Network, Support Vector Machine, dan Naïve Bayes dalam Mendeteksi Kanker Payudara*" dengan tujuan pemodelan klasifikasi kanker payudara yang berdasarkan akurasi algoritma, didapatkan algoritma

terbaik yaitu SVM dengan tingkat akurasi 98.9%, *neural network* dengan nilai 97.9% sementara *Naïve Bayes* mendapatkan nilai akurasi 96.1%[10].

Berdasarkan uraian latar belakang dan permasalahan diatas, pada penelitian ini penulis akan melakukan pengujian tingkat akurasi performa algoritma Support Vector Machine dalam proses pengklasifikasian analisis sentimen ke dalam dua kelas yaitu kelas positif dan kelas negatif terhadap layanan by.U pada sosial media Twitter.

## 1.2 Rumusan Masalah

Perumusan masalah dalam penelitian ini adalah:

1. Bagaimana algoritma Support Vector Machine dapat melakukan proses klasifikasi sentimen yang diberikan oleh pengguna layanan provider by.U dari komentar Twitter dengan baik?
2. Berapa besar tingkat akurasi algoritma Support Vector Machine dalam melakukan klasifikasi sentimen terhadap layanan *provider* by.U di Twitter?

## 1.3 Batasan Masalah

Agar penelitian ini lebih terarah, maka dibutuhkan batasan masalah, yaitu:

1. Penelitian hanya dilakukan terhadap data Twitter yang berhubungan dengan layanan telekomunikasi seluler by.U seperti *tweet* yang mengandung kata 'by.u'.
2. *Tweet* yang dianalisis menggunakan Bahasa Indonesia.
3. *Tweet* yang digunakan hanya *tweet* berupa teks, tidak mengandung gambar, video, gif, atau emoticon.



4. Klasifikasi sentimen ke dalam dua kelas yaitu positif dan negatif.
5. Data diunduh dari Twitter dari 23-02-2021 sampai 31-03-2021.
6. Jumlah data *training* adalah 1349 *tweet*.
7. Proses prediksi menggunakan model yang sudah dilatih.
8. Analisis sentimen dengan metode SVM menggunakan tools yang sudah tersedia di *library* Python.
9. Data pengujian dalam penelitian ini menggunakan data tweet dari Twitter yang diunduh secara *realtime*.
10. Pemvalidasian menggunakan *tweet* sebanyak 1750 *tweet*.
11. Bahasa pemrograman yang digunakan adalah Python.

#### **1.4 Maksud dan Tujuan Penelitian**

Maksud dan tujuan dari penelitian ini adalah untuk mengetahui kinerja algoritma Support Vector Machine dalam implementasinya melakukan klasifikasi sentimen dengan data teks berupa *tweet* berbahasa Indonesia yang berhubungan dengan layanan provider by.U ke dalam dua kelas kategori yaitu sentimen positif dan negatif, serta mencari nilai akurasi dari algoritma Support Vector Machine dalam mengklasifikasikan sentimen terhadap kepuasan layanan by.U.

## 1.5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan untuk bidang ilmu pengetahuan adalah mampu mengetahui proses implementasi *opinion mining* dan pengolahan data menggunakan algoritma Support Vector Machine untuk melakukan sentimen analisis. Sementara manfaat lain bagi perusahaan layanan telekomunikasi seluler dapat menganalisis persentase sentimen publik terhadap layanan yang diberikan sehingga dapat digunakan sebagai pengambilan keputusan langkah untuk peningkatan pelayanan dan mutu kualitas dari *provider* untuk pengguna.

## 1.6 Metode Penelitian

Dalam penelitian ini, penulis menggunakan beberapa metode, yaitu:

### 1.6.1 Metode Pengumpulan Data

#### 1.6.1.1 Metode Studi Pustaka

Penulis mempelajari berbagai macam literatur mengenai algoritma Support Vector Machine, *data mining*, *opinion mining*, dan analisis sentimen melalui berbagai sumber referensi seperti jurnal penelitian internasional dan nasional, buku teori, tesis, dan sumber informasi lain yang didapatkan melalui perpustakaan atau di internet.

#### 1.6.1.2 Metode Deskriptif

Penulis mengumpulkan data dengan cara *scraping tweet* menggunakan library Python yaitu *tweepy*. *Tweet* yang di *scrape* merupakan *tweet* berkaitan dengan provider by.U.

### 1.6.2 Metode Analisis

Dalam tahap ini, penulis melakukan:

1. Analisa tahapan yang dilakukan dari proses *text mining* hingga didapatkan hasil akhir.
2. Analisa terhadap cara kerja algoritma Support Vector Machine dalam melakukan proses klasifikasi sentimen tweet terhadap kinerja layanan telekomunikasi seluler by.U di Twitter ke dalam kelas positif atau negatif.

### **1.6.3 Metode Perancangan**

Melakukan proses persiapan data atau *preprocessing*, *term-frequency*, perancangan model dari algoritma Support Vector Machine, pembuatan *flowchart* dan antarmuka sistem untuk mempermudah proses tahap selanjutnya agar hasil penelitian dapat diperoleh sesuai dengan harapan.

### **1.6.4 Metode Implementasi**

Melakukan implementasi dari metode perancangan yang sudah dilakukan pada tahap sebelumnya. Pada tahapan ini meliputi implementasi *text preprocessing*, *term-frequency*, model dari algoritma Support Vector Machine, *flowchart*, dan antarmuka sistem.

### **1.6.5 Metode Pengujian**

Pada tahap pengujian, penulis menggunakan metode *K-fold Cross Validation* untuk memvalidasi data yang telah dilatih dan mengetahui rata-rata keberhasilan sistem pada setiap *fold* yang diujikan.

## **1.7 Sistematika Penulisan**

Penelitian ini disusun dengan menggunakan sistematika penulisan sebagai berikut:

**BAB I: PENDAHULUAN**

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

**BAB II: LANDASAN TEORI**

Pada bab ini menjelaskan tentang teori-teori yang digunakan sebagai dasar pembahasan dalam analisis penelitian.

**BAB III: ANALISIS DAN PERANCANGAN**

Pada bab ini menjelaskan tentang analisis metode penelitian yang akan dilakukan serta alur perancangan sistem yang akan dibuat.

**BAB IV: IMPLEMENTASI DAN PEMBAHASAN**

Pada bab ini menguraikan implementasi algoritma Support Vector Machine dalam mengklasifikasikan sentimen cuitan dari Twitter. Bab ini juga membahas tentang *preprocessing* data, tokenisasi, *stemming*, TF-IDF, SVM dan *Confusion Matrix*.

**BAB V: PENUTUP**

Pada bab ini menjelaskan kesimpulan yang didapat dari penelitian yang telah dilaksanakan serta saran pengembangan yang lebih baik lagi.