

BAB II

LANDASAN TEORI

2.1. Analisis Sentimen

Analisis sentimen merupakan metode yang digunakan untuk memahami sekaligus mengelompokkan kata atau kalimat ke dalam sejumlah kategori tertentu. Proses ini sering disebut juga sebagai *opinion mining*, yaitu teknik untuk menggali opini maupun emosi dari data yang dianalisis. Tujuan utama dari analisis sentimen adalah mengekstraksi atribut serta komponen penting dari berbagai komentar, misalnya yang terdapat di media sosial, sehingga dapat dipetakan ke dalam kelas positif, negatif, maupun netral [16].

Analisis sentimen yang akan dilakukan oleh penulis adalah dimulai dari tahap awal yaitu, dengan mengumpulkan dataset ulasan aplikasi Shopee yang didapatkan dari Google Play Store melalui scraping web, lalu memproses dataset tersebut melalui tahap pre-processing. Selanjutnya, ulasan kemudian diberi label sentimen secara manual. Dataset yang sudah bersih dan berlabel kemudian akan dilatih menggunakan metode *FastText* yang selanjutnya akan diuji guna mengetahui performa *FastText*.

2.2. FastText

FastText merupakan sebuah library yang dikembangkan oleh Facebook untuk kebutuhan word embedding. Teknologi ini hadir sebagai pengembangan dari Word2Vec yang sebelumnya sudah dikenal sebagai metode populer dalam representasi kata. Keistimewaan FastText terletak pada kemampuannya menangkap makna kata dengan memperhatikan bagian-bagian kecil dari kata seperti sub-kata atau n-gram. Pendekatan ini membuat FastText lebih efektif ketika berhadapan dengan kosakata baru atau kata yang mengandung kesalahan penulisan. Berbeda dengan Word2Vec yang hanya menghasilkan representasi berdasarkan konteks

global tanpa melihat struktur internal kata, FastText justru memanfaatkan elemen sub-kata untuk membangun representasi yang lebih mendalam.

Dengan adanya FastText, ulasan pengguna yang sering memuat kata tidak baku dapat tetap diubah menjadi vektor representatif, sehingga hasil analisis sentimen menjadi lebih akurat [17]. Jika kita memiliki kata w dan mengonversinya menjadi n -gram $g_1, g_2, g_3, \dots, g_k$, maka representasi vektor kata dapat dinyatakan sebagai:

$$v_w = \frac{1}{|g|} \sum_{i=1}^{|g|} v_{g_i}$$

Dimana :

$W - w$ = vektor representasi untuk kata W

$|g|$ = Jumlah n -gram dari kata W

v_{g_i} = vektor representasi untuk n -gram ke- i

2.3. Model SVM

Support Vector Machine (SVM) merupakan salah satu algoritma pembelajaran mesin yang banyak dimanfaatkan dalam proses klasifikasi, terutama pada data berukuran besar dan berdimensi tinggi. Algoritma ini diperkenalkan oleh Vladimir Vapnik dan dikenal sebagai metode yang efisien serta fleksibel sehingga dapat diaplikasikan pada berbagai bidang penelitian. SVM bekerja dengan melakukan analisis terhadap data untuk mengenali pola tertentu yang membedakan satu kelas dengan kelas lainnya.

Pada prinsipnya, SVM menerima sekumpulan data masukan sebagai data latih, kemudian menghasilkan prediksi terhadap data baru yang diberikan. Setiap data masukan akan diklasifikasikan ke dalam salah satu kelas yang tersedia. Secara dasar, SVM dirancang sebagai algoritma klasifikasi linier biner yang bersifat non-probabilistik, di mana proses pembelajaran dilakukan menggunakan data latih yang

telah diberi label dan dikelompokkan ke dalam dua kategori yang berbeda. Tujuan utama dari algoritma ini adalah untuk membentuk sebuah batas pemisah (hyperplane) yang optimal antara dua kelas tersebut. Pada klasifikasi linier menggunakan SVM, fungsi keputusan dinyatakan dalam bentuk persamaan sebagai berikut:

$$f(x) = W^T X + b$$

Dalam persamaan tersebut, fungsi $f(x)$ digunakan untuk menentukan kelas target y berdasarkan vektor fitur masukan x . Vektor W merupakan bobot yang menentukan arah hyperplane, sedangkan b adalah bias yang berfungsi sebagai konstanta penggeser. Sehingga menurut Vapnik dan Cortes (1995) diperoleh persamaan formula perumusan sebagai berikut:

$$[(W^T \cdot x_i) + b] \geq 1 \text{ untuk } y_i = +1$$

$$[(W^T \cdot x_i) + b] \leq -1 \text{ untuk } y_i = -1$$

Klasifikasi dilakukan dengan dataset latih yang dinotasikan dengan $f(x) =$ himpunan data training, ke $-i = 1, 2, \dots, n$ dan $y_i =$ label kelas dari x_i . Parameter w menotasikan vektor terhadap hyperlane dan b adalah offset [15].

2.4. Feature Selection

Salah satu pendekatan umum untuk meningkatkan akurasi klasifikasi sekaligus mengurangi dimensi fitur yang tidak relevan adalah melalui Feature Selection (FS). Tahapan ini bertujuan mengekstraksi fitur-fitur yang paling penting sehingga kinerja sistem pendeteksian dapat lebih optimal. Tidak semua fitur dalam data memiliki kontribusi yang signifikan, namun sebagian di antaranya tetap dapat memberikan dukungan dalam proses identifikasi. Oleh karena itu, diperlukan metode khusus untuk menyaring atribut yang benar-benar relevan dan mengandung informasi yang dibutuhkan guna membangun dataset yang lebih efektif.

Secara sederhana, feature selection dapat diartikan sebagai proses pemilihan subset fitur paling informatif dari keseluruhan fitur yang tersedia. Dalam analisis sentimen, peran metode ini sangat penting karena tidak hanya mampu meningkatkan performa model tetapi juga membantu mengurangi risiko overfitting [18].

2.5. Penelitian Terdahulu

Berikut adalah beberapa ringkasan hasil dari penelitian terdahulu yang peneliti gunakan sebagai referensi, diantaranya adalah:

Tabel 2 1 Hasil Penelitian Terdahulu

NO	JUDUL PENELITIAN	NAMA PENELITI	DATASET	HASIL PENELITIAN
1	Comparison of Text Representation Methods for Sentiment Analysis Using Support Vector Machine [13].	Heri Suroyo, Eric Juanda Pratama	Web Scraping	<p>Hasil penelitian pada Bag of Words menggunakan metode SVM mendapatkan accuracy sebesar 47%.</p> <p>Hasil penelitian pada TF-idf menggunakan metode SVM mendapatkan accuracy sebesar 71%.</p> <p>Hasil penelitian pada FastText menggunakan metode SVM mendapatkan accuracy sebesar 49%.</p>
2	Sentiments Analysis of Indonesian Tweet About	Dikih Arif Wibowo, Aina	Web Scraping	Hasil penelitian pada FastText menggunakan metode SVM mendapatkan accuracy sebesar 88%.

	Covid-19 Vaccine Using Support Vector Machine and Fasttext Embedding [14].	Musdholifah		Hasil penelitian pada Word2vec menggunakan metode SVM mendapatkan accuracy sebesar 76%.
3	Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM [15].	Adi Ariyo Munandar, Farikhin, Catur Edi Widodo.	Web Scraping	Hasil penelitian pada menggunakan metode SVM pada website Ruang Guru mendapatkan accuracy sebesar 99%. Hasil penelitian pada menggunakan metode SVM pada website Zenius mendapatkan accuracy sebesar 96%. Hasil penelitian pada menggunakan metode SVM pada website Quipper mendapatkan accuracy sebesar 82%.

Berdasarkan paparan ringkasan penelitian terdahulu pada tabel diatas, penulis akan menggunakan *FastText* dengan metode *SVM* menggunakan dataset yang telah didapatkan dari Google Play Store pada aplikasi Shopee. Diharapkan metode yang diusulkan dapat meningkatkan akurasi.