

# **BAB I**

## **PENDAHULUAN**

### **I.1 Latar Belakang**

Analisis sentimen adalah bagian dari pemrosesan bahasa alami dan *machine learning*. Ini berfungsi dengan mengklasifikasikan semua data *tweet* ke dalam kategori yang dibuat. Analisis sentimen dalam menganalisis data *tweet* yang meliputi rating, penilaian, sikap dan emosi terhadap entitas seperti produk, layanan, organisasi, orang, topik, dan peristiwa (Cahyono, 2017). Dari beberapa teknik klasifikasi yang digunakan untuk klasifikasi data adalah metode *Support Vector Machine*. Algoritma *Support Vector Machine* dipilih dikarenakan algoritma ini cocok untuk memisahkan dua buah kelas data atau lebih pada ruang *input*. Kelebihan *Support Vector Machine* adalah metode yang sederhana tetapi memiliki akurasi dalam pengklasifikasian teks dalam jumlah dimensi lebih besar dari jumlah *sampel* (Chairunnisa dkk, 2022).

Perkembangan media *online* media sosial *Twitter* mendukung munculnya informasi tekstual dalam jumlah yang tidak terbatas, sehingga nilai dari informasi ini perlu dikaji. Istilah di *Twitter* yang disebut *Tweet* yang merupakan pesan atau status yang dibuat oleh penggunanya. *Tweet* dapat mengungkapkan perasaan atau situasi pengguna *Twitter* maupun informasi (Cahyono, 2017).

Analisis sentimen adalah teknik yang digunakan untuk mengekstraksi dan memahami nuansa emosional dari data teks. Ini digunakan untuk menentukan apakah suatu komentar memiliki konotasi positif, negatif, atau netral (T. D. Putra dkk, 2022)

Seleksi fitur menjadi salah satu faktor yang paling penting dalam mempengaruhi tingkat akurasi klasifikasi karena jika *dataset* berisi sejumlah fitur, dimensi ruang akan menjadi besar, merendahkan tingkat akurasi klasifikasi. Seleksi fitur merupakan proses optimasi untuk mengurangi *dataset* yang besar memiliki fitur besar dari sumber asli agar subset fitur yang relatif kecil yang signifikan untuk meningkatkan akurasi klasifikasi yang cepat dan efektif (Rusardi dkk, 2022).

Metode Seleksi Fitur yang umum digunakan di dalam klasifikasi teks adalah *Term Frequency-Inverse Document Frequency* (TF-IDF), *Information Gain* (IG), *Mutual Information* (MI), *Chi-square* ( $\chi^2$ ), *Ambiguity Measure* (AM), *Term Strength* (TS), *Term Frequency-Relevance Frequency* (TF-RF) dan *Symbolic Feature Selection* (SFS) (Harish & Revanasiddappa, 2017). Memilih satu metode Seleksi Fitur yang ideal untuk diterapkan dan dikombinasikan dengan semua algoritma *machine learning* adalah sulit dilakukan. (Bahassine dkk, 2020) telah melakukan penelitian terkait metode Seleksi Fitur yang ideal pada klasifikasi teks dan memperoleh hasil bahwa jika jumlah atribut atau fitur di bawah 20 maka *Chi-Square* telah memberikan akurasi yang sangat baik.

Metode *Chi-Square* merupakan metode yang banyak digunakan di dalam klasifikasi teks oleh sejumlah peneliti karena mudah diterapkan dan melakukan peninjauan kembali terhadap *data sampling* berdasarkan pada *Chi-Square Metric* yang diukur berdasarkan nilai *True Positives* (TP), *False Positives* (FP), *True Negatives* (TN), *False Negatives* (FN), *Probability of Number of Positive Cases* ( $P_{Pos}$ ) dan *Probability of Number of Negative Cases* ( $P_{Neg}$ ) (Sumaiya Thaseen & Aswani Kumar, 2017).

Sejumlah peneliti telah menyarankan penggunaan *Feature Selection* pada SVM untuk membatasi jumlah fitur *input* di dalam *classifier* untuk memperoleh akurasi dan waktu komputasi yang baik (Niu dkk, 2019). Seleksi fitur sekaligus memberikan setelan *parameter* pada *Support Vector Machine* secara signifikan yang berpengaruh pada hasil akurasi klasifikasi (Rusardi dkk, 2022).

Maka dari itu dalam penelitian ini akan berfokus pada analisis sentimen *tweet* menggunakan *Support Vector Machine* (SVM) dan seleksi fitur dengan *Chi Square*. Dari analisis tersebut dapat dilihat seberapa pengaruhnya nilai akurasi SVM dengan menggunakan seleksi fitur *Chi Square*.

## **I.2 Rumusan Masalah**

Rumusan masalah yang terdapat pada penelitian tesis ini ialah :

1. Bagaimana pengaruh nilai akurasi *Support Vector Machine* (SVM) terhadap seleksi fitur menggunakan *Chi Square* pada analisis sentimen *tweet*.
2. Apakah hasil uji algoritma *Chi Square* dapat meningkatkan akurasi dalam *feature selection* pada SVM di dalam analisis sentimen *tweet*.

## **I.3 Tujuan**

Tujuan dari penelitian tesis ini adalah :

1. Untuk mengetahui nilai akurasi proses algoritma *Support Vector Machine* (SVM) dengan seleksi fitur menggunakan *Chi Square* yang diterapkan pada analisis sentimen *tweet*.
2. Untuk memberikan hasil yang optimal pada klasifikasi dengan *Support Vector Machine* (SVM) dalam *feature selection* pada *Chi Square* dalam analisis sentimen *tweet*.

#### **I.4 Manfaat**

Adapun manfaat dari penelitian tesis ini adalah :

1. Menambah wawasan dari analisis sentimen pada *Twitter* dengan algoritma *Support Vector Machine* dengan seleksi fitur *Chi-square*.
2. Memberikan sumbangsih bagi pengembangan teori terutama dalam bidang analisis sentimen *tweet* maupun pengembangan dalam bidang seleksi fitur pada algoritma *Support Vector Machine* (SVM).

#### **I.5 Batasan Masalah**

Batasan masalah yang terdapat pada penelitian tesis ini ialah :

1. Algoritma yang digunakan *machine learning* dalam klasifikasi dengan algoritma *Support Vector Machine*.
2. Algoritma dalam *feature selection* menggunakan algoritma *Chi-square*.
3. *Dataset* dengan nama *Twitter Sentiment Analysis* diperoleh dari *kaggle.com* yang memiliki kapasitas 3.49 mb yang berformat csv.

#### **I.6 Sistematika Pembahasan**

Sistematika dari penulisan tesis ini terdiri dari enam (6) bab, dimana setiap bab terdiri dari sub dan sub-sub yang dapat diuraikan sebagai berikut :

#### **BAB I : PENDAHULUAN**

Bab ini membahas latar belakang masalah, rumusan, tujuan penelitian, manfaat, batasan masalah dan sistematikan pembahasan.

**BAB II : TINJAUAN PUSTAKA**

Bab ini berisi tentang kajian pustaka sumber acuan dari teori yang relevan, literatur ilmiah yang berkaitan dengan penelitian.

**BAB III : METODOLOGI PENELITIAN**

Bab ini menjelaskan langkah-langkah penelitian beserta metode yang digunakan dan teknik pengujian metode.

**BAB IV : HASIL**

Bab ini merupakan hasil pengujian implementasi program dari hasil yang diteliti.

**BAB V : PEMBAHASAN**

Bab ini merupakan hasil pembahasan dari hasil analisa metode tahapan-tahapan yang digunakan dalam penelitian.

**BAB VI : KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan juga saran bagi penelitian mendatang yang berasal dari kekurangan ataupun temuan penelitian lain.