

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dunia sepak bola Indonesia berduka dimana pada tanggal 1 Oktober 2022 telah terjadi tragedi kelam ketika pertandingan Arema FC vs Persebaya di Stadion Kanjuruhan, Kabupaten Malang, Jawa Timur, Tragedi yang memakan korban setidaknya 714 jiwa, terdiri dari korban meninggal dunia 131 orang, korban luka 583 orang (Irjen Pol Dedi Prasetyo) per 9 Oktober 2022 [1]. Komnas HAM menyatakan penyebab banyaknya korban diduga karena adanya penembakan gas air mata ke tribun penonton yang menyebabkan sesak dan juga ditutupnya pintu keluar stadion sehingga penonton tidak bisa keluar dan saling berdesakan [2].

Tragedi ini memicu berbagai reaksi dan opini yang menyebar luas di kalangan masyarakat, termasuk melalui media sosial seperti *Twitter* dan *Youtube*, dimana tragedi Kanjuruhan sempat menjadi *trending topic* di kedua platform tersebut. Pendapat-pendapat ini juga dapat disebut sebagai sentimen [3]. Banyaknya masyarakat yang mengemukakan sentimennya masing-masing mengakibatkan kita sulit untuk mengetahui sentimen seperti apa yang masyarakat sampaikan. Analisis sentimen merupakan suatu kegiatan untuk mengetahui dan mengklasifikasikan emosi seseorang seperti emosi baik, buruk, ataupun emosi netral yang dimuat dalam suatu tulisan atau komentar seseorang [4]. Dengan memahami sentimen yang diungkapkan oleh masyarakat terkait peristiwa tragis seperti yang terjadi di Kanjuruhan, akan memungkinkan peningkatan upaya penanganan dan pengambilan tindakan yang lebih efektif dalam situasi serupa di masa depan.

Dalam mengidentifikasi sentimen dari sebuah teks atau dokumen, dapat dilakukan dengan menggunakan beberapa pendekatan. Pendekatan-pendekatan tersebut antara lain adalah *unsupervised learning* atau *lexicon based*, *supervised learning* atau dengan *machine learning*, dan ada juga yang menggunakan

gabungan dari keduanya [5]. Dikenal sebagai *dictionary based approach* pendekatan *lexicon based* adalah suatu metode atau pendekatan yang berbasis pada kamus [6]. Teknik ini menghasilkan daftar kata yang sering digunakan untuk mengekspresikan opini oleh orang-orang. Kalimat-kalimat ini umumnya mengandung kata-kata sifat yang dapat dijadikan sebagai ukuran atau tolak ukur dari sentimen yang terdapat dalam sebuah kalimat opini. Contohnya bagus, jelek, indah, dll. Berbeda dengan *lexicon based* pendekatan *supervised learning* mengandalkan *machine learning* untuk menganalisa sentiment dalam pendekatan ini banyak metode yang digunakan seperti *naive Bayes*, *support vector machine (SVM)*, *k-nearest neighbors (KNN)*, *logistic regression (LR)* dan *decision tree (DT)*[7].

Beberapa penelitian terakhir telah banyak mengimplementasikan analisis sentimen dengan menggunakan metode *machine learning* seperti penerapan algoritma *naive Bayes* pada *twitter sentiment analysis* Covid-19 di Filipina [8]. Menggunakan metode yang sama namun dikombinasikan dengan *K-means* dilakukan juga penelitian mengenai analisis sentimen pada *review* produk [9]. Penerapan algoritma *random forest* dan *support vector machine* berbasis pendekatan *hybrid* pada analisis sentimen [10]. Penelitian yang memprediksi arah pergerakan pasar menggunakan analisis sentimen dengan metode *SVM* [11]. Metode *Naive bayes* merupakan metode yang cukup baik dalam untuk melakukan klasifikasi data dari pada metode lainnya dalam hal akurasi dan komputasi, Namun, pada beberapa kasus tertentu, metode *naive Bayes* memiliki kelemahan dalam menghadapi data yang kompleks atau dengan fitur yang sangat banyak dan berkorelasi tinggi. Kekurangan ini dapat menyebabkan performa *naive Bayes* menurun secara signifikan ketika dihadapkan pada dataset yang kompleks[12]. Sementara itu pada penelitian yang berbeda dikatakan bahwa *SVM* memiliki teori dan juga konsep yang terstruktur sehingga metode ini adalah salah satu metode dengan akurasi terbaik dalam bidang pengklasifikasian teks yang digunakan pada banyak jenis kasus untuk klasifikasi sentimen [13],[14]. Penelitian pada tahun 2019 yang dilakukan untuk membandingkan analisis sentimen Pilpres 2019 dengan *SVM* dan *KNN* didapat bahwa algoritma *SVM* lebih cepat dalam pengujian

dan memiliki rata-rata akurasi yang lebih tinggi daripada *KNN* yang cenderung lebih lambat dan memerlukan komputasi yang lebih intensif, terutama ketika jumlah data dan fitur semakin besar [15].

Berdasarkan hasil dari penelitian-penelitian sebelumnya metode *support vector machine* dapat dengan baik dalam melakukan klasifikasi teks, memiliki waktu pengujian yang cepat, dan memiliki akurasi yang baik. Namun belum ada penelitian terakhir yang menggunakan metode ini dalam melakukan analisis sentimen pada tragedi kanjuruhan. Oleh karena itu, penelitian ini dituangkan ke dalam Tugas Akhir yang berjudul “**ANALISIS SENTIMEN MASYARAKAT TERHADAP TRAGEDI KANJURUHAN DENGAN METODE *SUPPORT VECTOR MACHINE***”.

## **1.2 Pertanyaan Penelitian**

Peneliti merumuskan beberapa pertanyaan terkait penelitian yang akan dilakukan diantaranya:

1. Bagaimana polaritas yang lebih dominan dari sentimen masyarakat mengenai tragedi Kanjuruhan?
2. Bagaimana tingkat akurasi yang dihasilkan oleh algoritma *support vector machine (SVM)* untuk klasifikasi sentimen mengenai tragedi kanjuruhan?

## **1.3 Perumusan Masalah**

Berdasarkan latar belakang yang telah disampaikan maka permasalahan dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Tragedi Kanjuruhan yang terjadi pada 1 Oktober 2022 mendapat berbagai tanggapan positif dan juga negatif dari masyarakat, dengan demikian dapat dilakukan analisis sentimen untuk mengetahui polaritas masyarakat terhadap topik ini.
2. Algoritma *support vector machine (SVM)* memperoleh akurasi yang tinggi dalam melakukan klasifikasi teks pada berbagai analisis sentimen, sehingga perlu dibuktikan akurasi yang didapatkan dengan algoritma *SVM* pada analisis sentimen tragedi kanjuruhan.

#### 1.4 Batasan Masalah

Berdasarkan permasalahan yang ada, Agar permasalahan yang diangkat dalam penelitian ini lebih terukur maka dibuat batasan-batasan masalah sebagai berikut:

1. Data penelitian diambil dari data *tweet* pengguna *twitter* mengenai tragedi kanjuruhan dan data komentar *youtube* “Tragedi Kanjuruhan #UsutsampaiTuntas | Mata Najwa”.
2. *Tweet* yang diambil adalah *tweet* dari 1 oktober - 31 oktober 2022 dan komentar *youtube* “Tragedi Kanjuruhan #UsutsampaiTuntas | Mata Najwa” sampai pada tanggal 20 November 2022.
3. Fokus penelitian ini adalah bagaimana algoritma *Support Vector Machine* (SVM) melakukan klasifikasi pada dataset *tweet* *twitter* dan komentar *youtube* tentang tragedi kanjuruhan.

#### 1.5 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dari penelitian ini dapat dijabarkan sebagai berikut:

1. Menganalisis polaritas yang lebih dominan dari sentimen masyarakat mengenai tragedi kanjuruhan.
2. Menganalisis akurasi yang dihasilkan algoritma *Support Vector Machine* dalam klasifikasi sentimen tragedi kanjuruhan.

#### 1.6 Manfaat Penelitian

Berdasarkan rumusan masalah, batasan masalah, dan tujuan penelitian yang telah diuraikan, maka dapat diketahui manfaat dari penelitian ini adalah:

1. Manfaat Teoritis
  - a. Bagi Penulis, dapat memberikan pengetahuan baru dalam ilmu dibidang *machine learning*, *natural language processing*, dan metode *SVM*.
  - b. Bagi Masyarakat, dapat memberikan pemahaman kepada masyarakat tentang polaritas sentimen tragedi kanjuruhan.

- c. Bagi IT Telkom Purwokerto, dapat dipergunakan sebagai bahan referensi maupun pembandingan untuk penelitian selanjutnya dalam klasifikasi sentimen.

## 2. Manfaat Praktis

- a. Bagi Penulis, dapat menjadi alat dalam implementasi algoritma *SVM* untuk klasifikasi sentimen.
- b. Bagi Masyarakat, dapat digunakan sebagai acuan bagi masyarakat untuk memahami bagaimana polaritas sentimen tragedi kanjuruhan.
- c. Bagi IT Telkom Purwokerto, dapat dipergunakan sebagai model klasifikasi sentimen lain pada penelitian selanjutnya.