

BAB I

PENDAHULUAN

1.1. Latar Belakang

Media sosial merupakan media yang digunakan pengguna dalam rangka mempresentasikan dirinya, berinteraksi, bekerja sama, berbagi informasi, maupun berinteraksi dengan pengguna lainnya (Krisdiyanto dan Nurhayanto, 2021). Salah satu media sosial yang ada di internet saat ini adalah Twitter. Menurut survei yang dilakukan oleh lembaga We Are Social pada tahun 2020, Twitter merupakan salah satu sosial media yang paling sering digunakan di Indonesia dengan menduduki peringkat ke-5 dengan persentase pengguna 56% (Katadata, 2020). Berdasarkan data yang disebutkan sebelumnya, Twitter menjadi salah satu platform media sosial yang cukup berpengaruh bagi penggunanya di Indonesia. Sehingga, banyak pengguna Twitter di Indonesia memposting kicauan atau *tweet* mengenai perasaan atau pendapat mereka terhadap suatu hal yang terjadi di Indonesia seperti peristiwa Tragedi Kanjuruhan.

Tragedi Kanjuruhan merupakan salah satu tragedi di dunia sepakbola yang terjadi pada tanggal 1 Oktober 2022. Tragedi ini terjadi setelah pertandingan sepak bola Liga 1 antara Arema FC dan Persebaya di Stadion Kanjuruhan Malang yang menyebabkan adanya kerusuhan. Menurut laporan dari Tim Gabungan Independen Pencari Fakta Tragedi Stadion Kanjuruhan Malang, tragedi tersebut menyebabkan 712 orang menjadi korban, termasuk 132 orang yang meninggal dunia, 96 orang luka berat, dan 484 orang luka ringan atau sedang. Tragedi Kanjuruhan menempati urutan kedua sebagai peristiwa tragis di dunia sepak bola dunia dan menjadi peristiwa dengan jumlah korban terbanyak dalam sejarah sepak bola Indonesia. (Humas Kemenko Polhukam RI, 2022). Tragedi tersebut menimbulkan berbagai reaksi dari kalangan masyarakat Indonesia terutama di media sosial Twitter. Banyak pengguna Twitter di Indonesia yang mengutarakan berbagai perasaan atau pendapat melalui *tweet* yang mereka posting di Twitter terkait tragedi ini, ada yang mengutarakan perasaan duka dan belasungkawa atas tragedi ini, namun tidak menutup kemungkinan juga ada masyarakat Indonesia yang mengutarakan perasaan lain pada postingan *tweet* mereka, entah *tweet* tersebut memiliki sentimen

positif, negatif, atau netral. Oleh karena itu, penelitian ini mencoba untuk melakukan penelitian untuk menganalisa sentimen pada *tweet* tentang tragedi Kanjuruhan dengan berbagai metode.

Analisis sentimen merujuk pada proses otomatis untuk memahami, mengekstrak, dan memproses data tekstual guna mendapatkan informasi mengenai sentimen yang terdapat dalam suatu kalimat opini. Tujuan dari analisis sentimen adalah untuk mengetahui pendapat atau kecenderungan opini seseorang terhadap suatu masalah atau objek, apakah bersifat positif, negatif, atau netral (Rozi dkk, 2012). Analisis sentimen dapat dilakukan dengan menerapkan beberapa algoritma seperti *Naïve Bayes*, *Support Vector Machine*, *Deep Neural Network*, dan masih banyak lagi algoritma pembelajaran mesin lainnya. Penelitian ini mengambil sentimen dari *tweet* pengguna Twitter dengan topik “Kanjuruhan”. Kemudian, topik tersebut dianalisis sentimennya dan data dikelompokkan atau diklasifikasikan menjadi tiga bagian yang terdiri dari sentimen positif, negatif, dan netral.

Terdapat beberapa penelitian terdahulu yang mengangkat analisis sentimen sebagai topik penelitiannya. Pada penelitian yang dilakukan oleh Chetanpal Singh, Tassaduq Imam, Santoso Wibowo, dan Sriman Grandhi tentang pendekatan *Deep Learning* untuk analisis sentimen *review* COVID-19 menggunakan data Twitter menunjukkan bahwa pendekatan *Deep Learning* dapat diimplementasikan untuk analisis sentimen *review* COVID-19 dimana pada penelitian tersebut memiliki hasil akurasi dari berbagai algoritma pembelajaran mesin yang telah dicoba. *Naïve Bayes* menghasilkan akurasi 67%, *Random Forest* 60%, *SVM* 62%, *Logistic Regression* 70%, *Long Short-Term Memory-Recurrent Neural Network (LSTM-RNN)* 76%, dan yang terakhir dengan *LSTM-RNN* dengan pendekatan yang diusulkan oleh penelitian tersebut memiliki hasil akurasi 84% (Singh dkk, 2022).

Penelitian lain yang dilakukan oleh Nhan Cach Dang, Maria N. Moreno-Garcia, dan Fernando De la Prieta pada tahun 2020 dengan judul *Sentiment Analysis Based on Deep Learning: A comparative study*, mereka melakukan analisis sentimen pada berbagai *dataset* dengan metode *word embedding* dan *term frequency-inverse document frequency (TF-IDF)* yang nantinya digabungkan dengan berbagai algoritma *Deep Learning*. Hasil penelitian tersebut menunjukkan bahwa pada *dataset* “Tweets SemEval” menghasilkan nilai akurasi yang berbeda

dari setiap metode yang diterapkan. Pada metode TF-IDF yang digabungkan dengan berbagai algoritma *Deep Learning* yaitu *Deep Neural Network* (DNN), *Convolutional Neural Network* (CNN), dan *Recurrent Neural Network* (RNN) menghasilkan nilai akurasi masing – masing adalah DNN 83%, CNN 81%, dan RNN 54%. Selanjutnya pada metode lain yaitu *word embedding* yang digabungkan dengan berbagai algoritma *Deep Learning* yang sama menghasilkan nilai akurasi masing – masing adalah DNN 83%, CNN 84%, dan RNN 85% (Dang dkk, 2020). Namun pada penelitian tersebut hanya melakukan analisis sentimen menggunakan metode TF-IDF dan *word embedding* sedangkan metode *lexicon-based* masih belum digunakan dalam melakukan analisis sentimen.

Pada penelitian yang dilakukan oleh Mubarik Ahmad, M. Ferdy Octaviansyah, Aan Kardiana, dan Kukuh Fadli Prasetyo tentang sentimen analisis pada *tweet* berbahasa Indonesia dengan *dataset* sebanyak 500 *tweet*, Mereka menggunakan metode yang berbeda dimana metode yang digunakan adalah *Lexicon* dan *Naïve Bayes* dimana *Lexicon* menghasilkan nilai akurasi 72% dan *Naïve Bayes* mendapat hasil akurasi 84% (Ahmad dkk, 2019). Penelitian lain yang dilakukan oleh Shinta Yuan Ayu Pratiwi dan Salamun Rohman Nudin tentang analisis sentimen pada Facebook *Marketplace* menggunakan *Lexicon-Based* dan *Support Vector Machine* (SVM) dimana *dataset* sebanyak 478 *tweet* akan dihitung skor sentimen kata yang tergolong positif atau negatif menggunakan *Lexicon* lalu dilakukan klasifikasi untuk mendapat nilai akurasi dengan SVM. Hasil dari *Lexicon-Based* menunjukkan hasil sentimen positif sebesar 82,6% dan sentimen negatif sebesar 17,4%, lalu hasil nilai akurasi klasifikasi menggunakan SVM pada *kernel* linear adalah 51% dan pada *kernel polynomial* adalah 52% (Pratiwi dan Nudin, 2021). Kedua penelitian tersebut memiliki kesamaan yaitu penggunaan *dataset* yang sedikit yaitu tidak lebih dari 500 data. Kedua penelitian tersebut juga menggunakan *lexicon-based* dalam melakukan analisis sentimen, namun penggunaannya secara terpisah dengan algoritma pembelajaran mesin lain dan cenderung memiliki hasil yang rendah terutama dalam penggunaan pembelajaran mesin SVM.

Penelitian lain tentang *Analysis of The Multilayer Perceptron Algorithm on Twitter User's Sentiment Towards The COVID-19 Vaccine* yang dilakukan oleh

Fordinand Halomoan Pasaribu, Nurul Khairina, Dian Noviandri, Susilawati, dan Rahmad Syah pada tahun 2023 bertujuan untuk mengetahui pendapat masyarakat tentang vaksin COVID-19 di Twitter. Penelitian tersebut menggunakan algoritma *Multilayer Perceptron* (MLP) dengan fungsi aktivasi yang digunakan adalah *sigmoid* dan pengujian dengan *confusion matrix*. Hasil terbaik yang didapat dari berbagai model yang dibuat adalah 81.6% *accuracy*, 84.2% *precision*, dan 71.6% *recall* saat training model, dan 81.2% *accuracy*, 83.8% *precision*, dan 71.2% *recall* untuk hasil tes model (Pasaribu dkk, 2023). Penelitian tersebut menggunakan *dataset* sebanyak 228.208, namun seluruh hasil *recall* memiliki jarak yang jauh lebih rendah disbanding akurasi dan presisi-nya.

Berdasarkan paparan yang telah disebutkan, maka akan dilakukan penelitian untuk menerapkan *deep learning* untuk *lexicon-based sentiment analysis* tentang tragedi Kanjuruhan pada Twitter, dimana penelitian akan melakukan analisis sentimen dengan topik Tragedi Kanjuruhan di media sosial Twitter dengan *dataset* yang lebih banyak, memiliki hasil akhir yang seimbang, menggunakan berbagai parameter yang berbeda, dan akan menggunakan berbagai metode *deep learning*. Metode yang akan digunakan dalam melakukan *Lexicon-based Sentiment Analysis* nanti adalah *Multilayer Perceptron* (MLP), dan *Convolutional Neural Network* (CNN). Nantinya dari kedua metode tersebut, akan menghasilkan nilai akurasi dari masing – masing metode lalu akan diketahui metode mana yang memiliki kinerja paling baik dalam melakukan analisis sentimen Tragedi Kanjuruhan pada media sosial Twitter.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat dibentuk beberapa poin rumusan masalah yaitu :

1. Bagaimana tingkat akurasi penggunaan metode MLP dan CNN serta metode mana yang memiliki kinerja paling baik dalam melakukan *lexicon-based* analisis sentimen untuk *tweet* tentang tragedi Kanjuruhan pada media sosial Twitter?
2. Bagaimana hasil analisis sentimen yang dilakukan terhadap *tweet* tentang tragedi Kanjuruhan pada media sosial Twitter?

1.3. Tujuan

Berdasarkan rumusan masalah yang ada, maka tujuan dari penelitian ini adalah:

1. Mengetahui tingkat akurasi metode MLP dan CNN serta mengetahui metode mana yang memiliki kinerja paling baik dalam melakukan *lexicon-based* analisis sentimen *tweet* tentang tragedi Kanjuruhan pada media sosial Twitter.
2. Mengetahui hasil analisis sentimen *tweet* tentang tragedi Kanjuruhan pada media sosial Twitter.

1.4. Manfaat

Dengan dilakukannya penelitian ini diharapkan dapat metode mana yang cocok untuk melakukan analisis sentimen *tweet* tentang tragedi Kanjuruhan pada media sosial Twitter. Manfaat yang lain dari penelitian ini terhadap lingkungan atau publik adalah untuk mengetahui sentimen atau opini publik terhadap tragedi Kanjuruhan lalu didapatkan kesimpulan apakah asumsi publik di media sosial Twitter menyukai, membenci, atau biasa saja terhadap tragedi tersebut, dengan melihat seberapa besar banyaknya opini publik yang baik (positif), buruk (negatif) atau biasa saja (netral).

1.5. Batasan Masalah

Batasan masalah yang ditentukan pada penelitian ini dalam melakukan penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data *tweet* yang terdeteksi sebagai bahasa Indonesia yang didapat di media sosial Twitter dengan satu topik yaitu “Kanjuruhan”.
2. Penelitian ini menggunakan metode MLP dan CNN untuk melakukan *lexicon-based* analisis sentimen.