

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sepak bola adalah satu-satunya budaya populer dalam sejarah olahraga dunia yang dapat menimbulkan gairah kebersamaan. Sepak bola menghubungkan banyak budaya dari Eropa dan Amerika Selatan ke Australia, Afrika, Asia, dan bahkan Amerika Serikat. Sepak bola telah menyebar ke seluruh penjuru dunia, memungkinkan budaya yang berbeda bersatu untuk membangun identitas tertentu melalui praktik dan interpretasi permainan. Federation of Football Association (FIFA), dengan kantor pusat di Swiss, bertanggung jawab atas semua aspek yang berkaitan dengan sepak bola di seluruh dunia. Dalam Sejarahnya sepak bola butuh waktu yang lama untuk berkembang menjadi konsep dan permainan *modern* seperti saat ini (Lukyani & Agustina, 2020).

Tahun 2023, Indonesia memiliki peluang emas untuk mengikuti kompetisi sepak bola kelas dunia pada ajang Piala Dunia Under 20 (U-20) yang diadakan oleh FIFA. Negara Indonesia berhasil mengatasi persaingan Brasil dan Peru sebagai pesaing tuan rumah Piala Dunia U-20, sehingga Indonesia terpilih sebagai tuan rumah turnamen tersebut. Pengumuman resmi Indonesia sebagai tuan rumah dilakukan oleh presiden FIFA Gianni Infantino dalam FIFA *council meeting* yang diadakan pada tanggal 24 November 2019 di Shanghai, China. Namun, mendekati pelaksanaan Piala Dunia U-20 beberapa pihak menolak keikutsertaan tim nasional Israel untuk bermain di negara Indonesia. Suara penolakan ini mendorong FIFA untuk membatalkan *drawing* piala dunia yang akan dilaksanakan di Bali pada tanggal 31 Maret 2023, munculnya permasalahan tersebut membuat FIFA menganggap Indonesia tidak mampu untuk menjadi tuan rumah Piala Dunia U-20. Hal ini terjadi karena FIFA memiliki prinsip kesetaraan, *fairplay*, dan tidak ada diskriminasi. FIFA akan mengirimkan pesan yang jelas bahwa diskriminasi tidak dapat ditoleransi dalam bentuk apapun (Kassimeris, 2021). Akibat munculnya permasalahan tersebut FIFA secara resmi memberi keputusan pencabutan status Indonesia sebagai tuan rumah Piala Dunia U-20 melalui *website* resmi *fifa.com* pada hari Rabu tanggal 29 Maret 2023, keputusan yang dibuat menjelang dua

bulan sebelum turnamen dimulai membuat netizen Indonesia merasa kecewa dan menjadi *trending* Twitter (Khairil Rahim, 2023).

Twitter merupakan salah satu media yang digunakan untuk memberikan opini melalui *tweet* (Pradana dkk., 2020). Inti dari Twitter adalah *tweet*, sejak tahun 2017 *tweet* telah diperluas menjadi 280 karakter. *Tweet* dapat berisikan komentar, opini, ataupun percakapan secara online untuk menyebarkan sebuah informasi. Dengan adanya sosialisasi secara maya ini maka Twitter juga dapat dikatakan sebagai media untuk bertukar informasi antar penggunanya dan di dalamnya banyak pengguna yang menyebarkan informasi mengenai permasalahan yang sedang *trending* seperti pembatalan tuan rumah Indonesia di Piala Dunia U-20 yang tentunya mengandung unsur opini saat merangkainya. Seringkali pengguna Twitter sulit untuk dapat memahami informasi yang disampaikan oleh pengguna lain karena informasi yang dibagikan hanya dalam bentuk teks. Meskipun teks dapat dijadikan sebagai representasi dari opini yang ingin disampaikan oleh penulis, sering terjadi ketimpangan antara informasi yang sedang dibahas dan informasi yang diinterpretasikan oleh pembaca. Fenomena ini dapat menyebabkan banyaknya penyebaran informasi yang tidak benar (hoaks) karena kita tidak benar-benar bisa memahami perasaan dan nuansa yang ingin disampaikan, walaupun pengguna Twitter telah menggunakan tipografi dan simbol-simbol tertentu untuk merepresentasikan ekspresi wajah di dalam tulisannya. Oleh sebab itu, banyak penelitian pembelajaran mesin yang dilakukan untuk menghasilkan sebuah program yang dapat membantu menganalisis hal tersebut.

Analisis sentimen adalah suatu proses yang melibatkan ekstraksi, pengolahan, dan pemahaman data teks yang bersifat tidak terstruktur secara otomatis, dengan tujuan untuk mencari informasi mengenai sentimen yang terdapat dalam suatu kalimat, pendapat, atau opini (Arsi & Waluyo, 2021). Informasi sentimen yang dihasilkan dapat berupa opini positif, opini negatif, ataupun netral. Analisis ini dapat dilakukan dengan menerapkan beberapa algoritma seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Deep Neural Network (DNN)*, *Recurrent Neural Network (RNN)* dan masih banyak lagi algoritma pembelajaran mesin lainnya. Bahkan banyak peneliti yang memadukan

ataupun membandingkan beberapa algoritma guna memperoleh hasil analisis yang optimal. Terdapat banyak pilihan yang bisa diterapkan dalam melakukan analisis, salah satunya dengan menerapkan *word embeddings*. *Word embeddings* adalah metode yang digunakan untuk menggambarkan sebuah kata yang serupa dengan mengubahnya menjadi kata yang memiliki arti sama, dengan cara memetakan kata-kata tersebut ke dalam sebuah dokumen vektor. Dokumen tersebut merupakan representasi dari kata-kata di dalam ruang vektor (Dewi dkk., 2022). Beberapa metode *word embeddings* yang dapat diterapkan, antara lain *Term Frequency-Inverse Document Frequency* (TF-IDF), *Global Vectors for Word Representation* (GloVe), *Word to Vector* (Word2Vec), FastText, dan sebagainya (Nurdin dkk., 2020).

Terdapat beberapa penelitian terdahulu yang juga mengangkat analisis sentimen sebagai topik penelitian. Pada tahun 2020, Faisal Faturohman, Budi Irawan dan Cesi Setianingsih melakukan penelitian untuk menganalisis sentimen pada BPJS Kesehatan menggunakan RNN. *Dataset* yang digunakan diambil dari media sosial Twitter sejumlah 1400 data, selanjutnya data masuk ke tahap *preprocessing* agar data siap untuk proses klasifikasi. Pemberian bobot *dataset* dilakukan menggunakan metode Word2Vec dimana setiap kata mempunyai nilai *vector* masing-masing. Setelah setiap kata memiliki bobot *dataset* akan dilakukan proses klasifikasi dengan menggunakan Algoritma RNN menggunakan tipe *Long Short-Term Memory* (LSTM). Hasil dari klasifikasi berupa 700 data berlabel positif dan 700 data berlabel negatif. Selanjutnya data diuji menggunakan sistem partisi sebanyak 9 kali dengan membagi data uji dan data latih dengan porsi yang berbeda, dari hasil pengujian didapatkan partisi 90% data latih dan 10% data uji dengan tingkat akurasi paling tinggi mencapai 86.67% (Faturohman dkk., 2020).

Penelitian lain yang membahas tentang analisis sentimen dilakukan oleh Meytry Petronella Purba dan Yuliagnis Transver Wijaya pada tahun 2022. Purba dan rekannya menguji seberapa akurat Metode LSTM dan FastText untuk menganalisis *basic emotion* masyarakat pada masa Pandemi Covid-19. Pada penelitian tersebut, data yang digunakan sebanyak 4.527 diambil dari Twitter dengan cara di-*scraping* menggunakan kata kunci “kesehatan mental” dan “kesehatan jiwa”, selanjutnya dilakukan *preprocessing*. Setelah data diuji dengan

menggunakan metode LSTM dan menerapkan FastText didapatkan tingkat akurasi sebesar 74,48% hanya dengan 5 *epoch* saja (Purba & Wijaya, 2022). Sedangkan untuk penelitian tugas akhir ini, peneliti akan melakukan analisis dengan menggunakan 9.645 data *tweet* bersumber dari Twitter yang terbatas hanya pada opini berbahasa Indonesia. Sentimen pada opini tersebut nantinya akan diklasifikasi ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif menggunakan kombinasi metode FastText sebagai metode *word embedding* dan algoritma RNN untuk klasifikasinya, sehingga didapatkan kesimpulan opini publik menyukai atau tidak suka dengan permasalahan tersebut, mendapatkan informasi penyebab utama permasalahan, dan dampak setelahnya berdasarkan hasil opini yang telah dianalisis. FastText merupakan bagian metode *word embedding* Word2Vec yang dikembangkan, dengan menggunakan n-gram yang tidak hanya menggunakan satu kata secara utuh untuk diproses (Bojanowski dkk., 2017). Sebagai contoh pada metode FastText dengan kata “pembatalan” menggunakan trigram (n=3), maka kata tersebut akan dipecah menjadi sub-kata "pem", "emb", "mba", "bat", "ata", "tal", "ala", "lan". Setiap sub-kata tersebut akan diberi vektor representasi tersendiri. Sehingga dengan memperhitungkan informasi sub-kata, FastText dapat menghasilkan representasi yang lebih baik, terutama untuk kata-kata yang jarang atau kata-kata yang tidak terlihat dalam data pelatihan dan menurut penelitian terdahulu di atas klasifikasi sentimen menggunakan algoritma RNN terbukti memiliki akurasi yang cukup tinggi.

Berdasarkan paparan di atas, maka peneliti akan mengangkat sebuah judul penelitian “Analisis Sentimen Publik Pada Pembatalan Tuan Rumah Indonesia Di Piala Dunia U-20 Menggunakan Metode FastText Dan *Algoritma Recurrent Neural Network* (RNN)”.

1.2. Rumusan Masalah

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

1. Bagaimana cara implementasi FastText *embeddings* pada algoritma *Recurrent Neural Network* untuk melakukan analisis sentimen *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20?

2. Bagaimana tingkat akurasi penggunaan FastText *embeddings* pada algoritma *Recurrent Neural Network* dalam melakukan analisis sentimen *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20?
3. Bagaimana hasil analisis sentimen yang dilakukan terhadap *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20?

1.3. Tujuan Penelitian

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

1. Mengetahui cara implementasi FastText *embeddings* pada algoritma *Recurrent Neural Network* untuk melakukan analisis sentimen *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20.
2. Mengetahui tingkat akurasi penggunaan FastText *embeddings* pada algoritma *Recurrent Neural Network* dalam melakukan analisis sentimen *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20.
3. Mengetahui hasil analisis sentimen yang dilakukan terhadap *tweet* tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20.

1.4. Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat, diantaranya:

1. Penelitian ini diharapkan dapat memberikan pengetahuan tentang kombinasi FastText *embeddings* pada model algoritma *Recurrent Neural Network*.
2. Menghasilkan model algoritma untuk klasifikasi teks yang cukup akurat dalam menganalisis sentimen publik tentang permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20.
3. Mengetahui opini publik terhadap permasalahan pembatalan tuan rumah Indonesia di Piala Dunia U-20 sehingga didapatkan kesimpulan apakah opini publik di media sosial Twitter menyukai (positif) atau membenci

(negatif). Dengan melihat seberapa banyak opini publik yang menyukai atau membencinya serta mengetahui penyebab utama permasalahan ini terjadi.

1.5. Batasan Masalah

Analisis sentimen dapat dilakukan dengan banyak algoritma. Sehingga peneliti membuat batasan masalah penelitian untuk mengurangi ketidakpastian yang timbul, yaitu sebagai berikut:

1. Data yang diambil berasal dari media sosial Twitter sebanyak 9.645 data *tweet* dan berfokus pada satu topik yaitu pembatalan Piala Dunia U-20 dengan menggunakan tiga *keyword* yaitu “pembatalan piala dunia”, “piala dunia u 20”, “fifa batal u-20”.
2. Data opini publik yang digunakan dalam penelitian hanya berfokus pada ulasan berbahasa Indonesia.
3. Penelitian hanya berfokus pada implementasi FastText *embeddings* dan algoritma *Recurrent Neural Network*.
4. Jenis algoritma *Recurrent Neural Network* yang dipakai adalah LSTM.
5. Implementasi dari algoritma ini dibangun menggunakan bahasa pemrograman Python.
6. Sentimen data opini publik akan diklasifikasikan ke dalam dua kategori, yaitu sentimen positif dan sentimen negatif.