

1. Pendahuluan

1.1 Latar Belakang

Pada masa kini media sosial memberikan peran dan pengaruh yang sangat besar dalam perkembangan teknologi komunikasi. Twitter merupakan salah satu media sosial microblogging yang memungkinkan penggunanya dapat membuat, melihat dan membalas postingan yang lebih dikenal dengan sebutan tweet. Pengguna Twitter sering kali menggunakan Twitter sebagai media untuk mengekspresikan diri dalam menanggapi suatu kejadian ataupun hal – hal yang terjadi di lingkungannya[1]. Hal ini dapat dijadikan sebagai acuan untuk mengetahui sentiment dan menentukan kecenderungan opini masyarakat terhadap suatu kejadian yang terjadi disekitar masyarakat. Salah satu kejadian yang ditanggapi oleh pengguna Twitter adalah proyek kereta cepat Jakarta – Bandung yang sudah dibangun sejak awal tahun 2016.

Proyek kereta cepat Jakarta – Bandung telah dimulai sejak tanggal 21 januari 2016 dengan dilakukannya groundbreaking oleh Jokowi di Perkebunan Mandalawangi Maswati, Cikalong Wetan, Bandung Barat, Jawa Barat[2]. Menurut Rini mantan menteri BUMN, keuntungan dibangunnya kereta cepat Jakarta – Bandung diantaranya akan meningkatkan perekonomian, mengangkat sektor pariwisata, dan membuka lapangan pekerjaan yang baru[3]. Namun, dalam pembangunannya terdapat beberapa masalah, seperti banjir yang terjadi di Bekasi dan menyebabkan kemacetan dan mengganggu kelancaran logistik. Permasalahan yang lain dalam proyek kereta cepat ini kurang memperhatikan kelancaran akses keluar – masuk jalan tol, pembiaran penumpukan material yang mengganggu fungsi drainase, pembangunan pilar LRT tanpa izin, sampai persoalan keselamatan dan Kesehatan kerja (K3)[4]. Oleh karena itu, diperlukan analisis sentimen untuk mengetahui bagaimana sentimen yang ada pada media sosial mengenai pembangunan proyek kereta cepat Jakarta – Bandung.

Analisis sentiment adalah suatu cara untuk mendapatkan sebuah informasi sentiment pada suatu media sosial untuk memahami suatu opini atau preferensi dari pengguna media sosial. Informasi yang didapat bisa berupa sentimen yang bernilai positif atau negatif. Salah satu metode yang dapat digunakan untuk analisis sentiment adalah metode yang terdapat pada Machine Learning seperti, KNN, Naïve Bayes dan SVM. Tahapan – tahapan dasar yang dilakukan dalam menjalankan metode ini berupa crawling data, preprocessing, menambahkan label, menambahkan fitur, melakukan klasifikasi dan melakukan perhitungan sentimen. Seperti pada penelitian yang dilakukan oleh Pulung Hendro Prastyo menghasilkan tingkat rata – rata yang tinggi pada metode SVM untuk confusion matrix[6]. Namun, pada penelitian yang dilakukan oleh [5] Mohammad Rezwanul Huq menunjukkan metode KNN memiliki hasil yang lebih tinggi dibandingkan dengan metode SVM pada penelitiannya. Oleh karena itu, jika hanya melihat dari tingkat akurasi, maka akan terjadi kekeliruan dalam penelitian, maka dalam melakukan penelitian perlu menambahkan precision atau recall sebagai evaluasi[5].

Dalam analisis sentimen dapat menggunakan word embedding sebagai fitur untuk merepresentasi kata berbentuk vektor bernilai riil. Pada penelitian yang dilakukan oleh [19] Liang-Chih Yu membahas tentang word embedding seperti, GloVe dan Word2vec pada analisis sentiment untuk menangkap sintaks dan semantic kata. Penelitian ini menggunakan word embedding pada klasifikasi Stanford Sentiment Treebank (SST) dan menghasilkan performansi yang lebih baik. Penelitian lainnya yang dilakukan oleh [20] Ru Ni dan Huan Cao memiliki hasil yang baik pada LSTM-GRU menggunakan GloVe sebesar 87.10% untuk akurasi dan 86.76% untuk F1-score.

Dengan adanya sentimen yang diungkapkan di media sosial *twitter* tentang proyek pembangunan kereta cepat Jakarta – Bandung akan dilakukan analisis sentimen menggunakan klasifikasi SVM dan GloVe *word embedding*. Oleh karena itu, Tugas Akhir ini akan mengangkat permasalahan mengenai pembangunan kereta Cepat Jakarta – Bandung dengan menggunakan metode klasifikasi Support Vector Machine (SVM). SVM memiliki kelebihan dalam menentukan hyperplane dengan memilih bidang yang optimal pada ruang input antar kelas dengan mencari margin yang merupakan titik data yang dekat dengan hyperplane[18]. Selain menggunakan SVM, penulis menggunakan ekspansi fitur Global Vector (GloVe) pada penelitian ini karena pada penelitian yang dilakukan oleh Ru Ni dan Liang-Chih Yu menghasilkan nilai performansi yang baik [19][20], serta menggunakan confusion matrix untuk melakukan pengukuran performansi.

1.2 Topik dan Batasan

Topik yang akan dibahas pada penelitian ini adalah analisis sentiment menggunakan metode klasifikasi *support vector machine* dan *Global Vector* untuk *word embedding* pada *tweet* masyarakat di media sosial *twitter* yang membahas seputar pembangunan kereta cepat Jakarta – Bandung. Batasan pada penelitian ini adalah menggunakan data set yang bersumber dari media sosial *twitter* dan hanya menggunakan kategori “proyek kereta cepat Jakarta – Bandung” dan data set yang diambil hanya menggunakan Bahasa Indonesia dengan format data .csv.

1.3 Tujuan

Tujuan dari penelitian ini adalah melakukan analisis sentiment pada media sosial *twitter* yang membahas tentang kereta cepat Jakarta – Bandung menggunakan metode klasifikasi *support vector machine* dan metode

word embedding global vector untuk mengetahui performa model yang didapatkan dari metode SVM dan GloVe dengan menentukan kombinasi nilai parameter GloVe yang terbaik dan untuk mengetahui bagaimana sentiment publik mengenai proyek kereta cepat Jakarta – Bandung.

1.4 Organisasi Tulisan

Bagian dari laporan TA adalah sebagai berikut. Bab 2 yang membahas studi literatur yang terkait dengan penelitian ini. Kemudian pada bab 3 membahas sistem yang dibangun pada penelitian ini. Bab 4 membahas tentang hasil dan analisis dari penelitian dan bab 5 akan membahas kesimpulan dari penelitian yang sudah dilakukan.