BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Masyakarat Indonesia saat ini gemar melakukan transaksi jual beli secara *online*, karena mudah dijangkau seperti wilayah dan waktu yang fleksibel saat bertransaksi online melalui *e-commerce* [1]. *E-commerce* di Indonesia semakin berkembang akibat meledaknya jumlah pengguna, itu sebabnya jasa pengiriman barang sangat dibutuhkan [2], salah satu jasa pengiriman barang di Indonesia yaitu Sicepat.

PT Sicepat adalah jasa ekspedisi yang berdiri pada tahun 2014, perusahaan jasa pengiriman yang berpusat di Jakarta, Indonesia [3]. Sicepat mempunyai cabang dan gerai yang tersebar di seluruh kota di Indonesia. Pelanggan Sicepat dapat memberikan komentar baik positif, negatif, dan netral terhadap pelayanannya melalui Twitter @sicepat_ekspres [3]. Twitter memiliki Fitur trending topik dan hashtag untuk penyebaran informasi secara cepat. Dengan adanya fitur ini, algoritma Twitter dapat diketahui apa yang sedang hangat dibicarakan diseluruh dunia [4]. Selain Twitter ada media sosial lainnya seperti Instagram. Instragram merupakan aplikasi berbagi foto dan video, tidak bisa hanya posting teks seperti Twitter [5].

Twitter hanya dapat memuat 140 karakter kata untuk mengirim satu pesan[6]. Pesan akan terkirim dan ditampilkan pada profil pengguna, sehingga siapa saja dapat melihat jumlah postingannya [1]. Pendapat pelanggan akan mengambarkan pandangan mereka terhadap Sicepat dan mempengaruhi pandangan masyarakat umum. Kemudahan pelanggan berkomentar di Twitter menghasilkan data komentar yang sangat banyak. Data komentar adalah data tidak terstruktur, sehingga memerlukan metode tertentu untuk menggali informasi di dalamnya. Salah satunya mengenai informasi sentimen pelanggan. Dengan demikian dibutuhkan sebuah metode yang mampu mengklasifikasi kalimat komentar pelanggan Sicepat tersebut kedalam kelas sentimen positif atau negatif. Klasifikasi tersebut dapat memberikan infromasi seberapa baik

atau buruk kualitas layanan Sicepat menurut konsumen.

Pada tahun 2020 penelitian terkait data komentar Sicepat oleh Aninda Cahya Putri M. Penelitian tersebut membangun aplikasi analisis sentimen, menggunakan 300 data komentar yang diambil dari akun @sicepat_ekspres. Metode klasifikasi yang digunakan adalah Support Vector Machine, tingkat akurasinya adalah sebesar 92%. Penelitian tersebut tidak menjelaskan penggunaan teknik vektorisasi yang digunakan [7].

Vektorisasi fitur kata pada analisis sentimen umumnya menggunakan TF-IDF[6] [8]. TF-IDF memilki kelemahan yaitu tidak dapat memahami konteks kalimat atau makna semantik[9]. Menurut penelitian Kurniawan dan Maharani, word2vec dapat mengatasi kelemahan TF-IDF[6]. Word2vec adalah metode *word embedding.* Algoritma Word2vec ditemukan oleh Mikolov dkk. pada tahun 2013. Keunggulan dari Word2vec adalah dapat melihat hubungan semantik antar kata. Beberapa penelitian mengenai word2vec untuk sentimen analisis sebagai berikut.

Kurniawan dan Maharani pada 2020 menerapkan word2vec untuk mengekstraksi fitur dan support vector machine (SVM) untuk klasifikasi sentimen komentar ditwitter. Penelitian tersebut bertujuan menganalisis faktor-faktor yang mempengaruhi akurasi model klasifikasi menggunakan gabungan skip-gram dan SVM. Hasil penelitian menyatakan bahwa model skip-gram dengan parameter 100 dimensi memberikan hasil klasifikasi paling baik, ditunjukan dengan nilai tingkat presisi sebesar 64,4%, recall sebesar 58%, dan f-score sebesar 61,1%[6].

Rusli dkk pada 2020 menerapkan Word2vec pada sentimen analisis komentar mahasiswa, menghasilkan akurasi sebesar 84,8%. Metode perhitungan vektor output yaitu Centroid Based sedangkan klasifikasinya menggunakan SVM [10]. Metode lain perhitungan vektor output bisa menggunakan *average based* [11].

Berdasarkan latar belakang tersebut, penelitian ini menggabungkan SVM dan Word2vec untuk klasifikasi sentimen pelanggan sicepat. Data yang

digunakan 945 data komentar. Tahapan teks pre-processing antara lain *cleansing, filtering,* dan tokenisasi. Word2vec digunakan untuk vektorisasi sedangkan svm untuk klasifikasi.

1.2 Rumusan Masalah

Word2vec dikatakan dapat menghasilkan *F1-Score classifier* sentimen yang lebih baik, namun bagaimana akurasi *classifier* gabungan metode word2vec dan SVM untuk model klasifikasi sentimen pada komentar pelanggan sicepat.

1.3 Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan, didapatkan suatu pertanyaan penelitian, yaitu:

1. Bagaimana sentiment pelanggan sicepat jika digali berdasarkan data cuitan konsumen di Twitter?
2. Bagaimana akurasi model klasifikasi sentimen pada Sicepat ekspres menggunakan *Support Vector Machine* dan Word2vec?

1.4 Batasan Masalah

Berdasarkan rumusan masalah, tujuan penelitian dan manfaat penelitian, maka penelitian yang sesuai dengan masalah yang ada diperoleh batasan-batasan masalah penelitian sebagai berikut :

1. Data yang diklasifikasi hanya data ulasan pengguna Sicepat pada Twitter.
2. Ulasan yang diambil menggunakan Bahasa Indonesia.
3. Data yang digunakan hanya tweet berupa text, tidak mengandung gambar, emoji, hastag.
4. Jumlah data training yang digunakan 945 tweet.
5. Hanya menggunakan dua kelas sentimen, yaitu positif, negatif, dan netral

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui gambaran umum persepsi pengguna Sicepat berdasarkan ulasan pengguna di Twitter yang terbagi menjadi kelas positif, negatif, dan netral.

2. Menerapkan ekstraksi fitur pada ulasan pengguna di Twitter dalam teks Bahasa Indonesia menggunakan Word2vec.
3. Mengetahui akurasi model klasifikasi sentimen menggunakan *Support Vector Machine* dan Word2Vec Skip Gram, khususnya untuk analisis sentimen pada Sicepat ekspres.

1.6 Manfaat Penelitian

Manfaat penelitian dilakukan adalah sebagai berikut:

1. Membantu Sicepat mengetahui tingkat kepuasan pelanggan melalui platfrom Twitter.
2. Dapat memudahkan pihak melihat persepsi pengguna dalam bentuk sentimen positif, negatif, dan netral sehingga dapat fokus melakukan evaluasi serta perbaikan ke arah yang lebih baik.
3. Mengetahui berapa keberhasilan (akurasi) algoritma Word2vec pada pengklasifikasian teks berbahasa Indonesia.