

1. Pendahuluan

Seiring berjalannya waktu membuat berbagai macam ilmu dan teknologi berkembang dari segi pengetahuan, transportasi, komunikasi, dll. Salah satu hal yang tidak lepas dari keseharian adalah komunikasi atau interaksi dengan orang lain karena manusia adalah makhluk sosial. Komunikasi sekarang dapat dilakukan dari jarak jauh salah satunya adalah menggunakan sosial media. Sosial media adalah media online yang penggunaannya dapat berpartisipasi, berbagi, dan membuat isi seperti forum, blog, jejaring sosial, wiki dan dunia virtual[1].

Twitter adalah salah satu media sosial dengan platform komunikasi yang banyak digunakan dalam kehidupan sehari-hari terlepas dari lokasi geografis pengguna [2]. Berbagai “tweet” panjang batas 280 karakter dengan isi “tweet” yang berbeda-beda dilontarkan setiap harinya, seperti topik terkait politik, sosial, pemerintah, pelayanan, dll. Telkomsel merupakan salah satu penyedia jasa layanan telekomunikasi yang ada di Indonesia. Banyaknya pelayanan yang disediakan Telkomsel membuat masyarakat menggunakannya, dan Telkomsel mempunyai berbagai fitur seperti SMS, telepon, internet, dll. Namun terdapat opini dari pengguna mengenai pelayanan yang Telkomsel berikan selama menggunakan Telkomsel yaitu opini dari segi positif, netral, dan negatif. Aspek yang banyak yang dibicarakan oleh pengguna Telkomsel di twitter adalah aspek sinyal dan pelayanan.

Analisis sentimen adalah studi komputasi tentang opini, emosi, dan sikap orang terhadap suatu entitas yang dapat berupa individu, peristiwa, atau topik. Isi dari topik kemungkinan besar mencakup opini atau ulasan-ulasan tertentu [3]. Salah satu metode untuk menganalisis sentimen adalah Support Vector Machine yaitu salah satu metode supervised learning dalam klasifikasi dengan mengestimasi dalam multidimensi berdasarkan saat SVM dikembangkan [4]. Penggunaan ekspansi fitur Word2Vec untuk mengasosiasikan kata-kata dengan titik-titik dalam ruang. jarak spasial antar kata menggambarkan hubungan kesamaan antara kata-kata tersebut. Word2Vec dibagi menjadi dua yaitu CBOW dan Skip-gram. Sampling pada klasifikasi digunakan untuk menyeimbangkan jumlah kelas data dengan mengurangi (*undersampling*) atau menambah (*oversampling*) biaya tahap pembelajaran, salah satu teknik *undersampling* adalah *Synthetic Minority Oversampling Technique* (SMOTE). SMOTE adalah teknik yang melakukan interpolasi di antara contoh kelas minoritas yang berdekatan. Dengan demikian, SMOTE mampu meningkatkan jumlah contoh kelas minoritas dengan memperkenalkan contoh kelas minoritas baru di lingkungan sekitar, sehingga membantu pengklasifikasi untuk meningkatkan kapasitas generalisasinya [5].

Tujuan analisis sentimen dalam penelitian ini adalah untuk mencari hasil analisis performansi terbaik menggunakan metode klasifikasi Support Vector Machine dan ekspansi fitur Word2Vec dari tweet pengguna Telkomsel di Twitter dengan empat skenario yaitu mencari *baseline* + ekstraksi fitur terbaik, *baseline* + ekstraksi fitur + seleksi fitur terbaik, *baseline* + ekstraksi fitur + seleksi fitur + ekspansi fitur word2vec + klasifikasi support vector machine, dan *baseline* + ekstraksi fitur + seleksi fitur + ekspansi fitur word2vec + klasifikasi support vector machine + *data imbalanced* dengan alasan permasalahan ketidakcocokan antar kosakata dalam kalimat atau “tweet” karena batasan karakter dalam setiap tweet adalah sebanyak 280 karakter. Penggunaan ekstraksi fitur TF-IDF digunakan sebagai pembobotan kosakata. Penelitian ini dikembangkan untuk mencari hasil dari skenario dengan hasil performansi skor *F1-score* dan ROC AUC yang tertinggi. Penggunaan SVM pada penelitian ini dikarenakan metode klasifikasi tersebut dapat menangani data yang sangat besar, terutama pada klasifikasi teks [6].

Batasan penelitian dalam penelitian ini adalah data yang dikumpulkan dan digunakan adalah data dari tweet pengguna Twitter yang membahas aspek sinyal dan pelayanan Telkomsel, data yang diambil menggunakan bahasa Indonesia, data tweet diambil dari 1 September 2021 sampai 28 Februari 2022.