

1. Pendahuluan

Telkomsel merupakan salah satu perusahaan telekomunikasi penyedia layanan internet terkemuka di Indonesia yang menciptakan aplikasi berbasis mobile bernama MyTelkomsel. Aplikasi Mytelkomsel diluncurkan oleh Telkomsel secara komersial sebagai saluran digital dengan tujuan memberikan kemudahan bagi pelanggan dalam melakukan layanan mandiri secara online dan diharapkan pengguna leluasa berinteraksi dengan aplikasi [1], [2]. Seiring dengan berkembangnya aplikasi MyTelkomsel yang banyak digunakan oleh masyarakat mengakibatkan munculnya ulasan dari penggunanya. Maka dari itu, diperlukan pendekatan dengan menggunakan analisis sentimen. Dimana tujuan utama dari analisis sentimen adalah untuk menentukan sudut pandang dan mengkategorikan emosi atau pendapat pengguna mengenai sesuatu [3], [4], [5]. Hal ini dapat membantu untuk mengukur sentimen dari pengguna terhadap aplikasi tersebut.

Riset mengenai sentimen pengguna aplikasi MyTelkomsel sudah pernah dilakukan oleh R. Indra Kurnia dan A. Suganda Girsang pada penelitian [6]. Penelitian tersebut menggunakan word embedding Word2vec dan LSTM. Jumlah data yang digunakan sebanyak 16,359 data yang dibagi menjadi 80:20 untuk data latih dan data uji. Didapatkan hasil f1-score dan presisi terbaik pada 5 kelas menggunakan LSTM-SMOTE, yaitu 0.62 dan 0.70. Sedangkan hasil f1-score dan presisi terbaik pada 3 kelas menggunakan LSTM-SMOTE, yaitu 0.86 dan 0,87. Penambahan algoritma SMOTE dapat mempengaruhi nilai akurasi dikarenakan meningkatkan nilai akurasi pada setiap percobaan. Word2vec merupakan salah satu metode word embedding yang dibuat oleh Tomas Mikolov dari tim Google [7]. Teknik Word2Vec dibuat untuk penyematan kata dengan dua model: Skip-Gram and Continuous Bag of Words (CBOW). Pada penelitian [8] menunjukkan bahwa penggunaan algoritma CBOW memberikan hasil yang cukup baik pada artikel berita dan algoritma Skip-Gram memberikan hasil yang lebih baik pada tweet.

Terdapat teknik word embedding lainnya yang populer digunakan pada penelitian yaitu word embedding GloVe. Pennington, Socher dan Manning memperkenalkan GloVe pada tahun 2014, dimana teknik GloVe merupakan sebuah metode baru untuk mencocokkan kata yang menggunakan rasio probabilitas kemunculan bersama antar katanya [9]. Dua kelas representasi kata digunakan pada model ini yaitu faktorisasi matriks global dan Skip-Gram yang di gunakan untuk mengekstraksi fitur yang lebih baik dengan memeriksa hubungan antar katanya [10]. Pada penelitian [11] melakukan perbandingan metode word embedding untuk klasifikasi sentimen pada data ulasan produk. Data yang digunakan berisi 19.977 data dan dibagi menjadi 11,986 data latih dan 7,991 data uji. Penelitian ini menggunakan CNN untuk klasifikasi sentimen dengan membandingkan tiga word embedding, yaitu Word2vec, GloVe, FastText untuk di cari yang mana merupakan word embedding terbaik pada pengklasifikasian ini. Hasil akurasi yang diperoleh dari penelitian ini adalah 92.5% pada Word2vec, 95.8% pada GloVe, 97.2% pada FastText. Didapatkan bahwa word embedding FastText lebih unggul dari pada word embedding lainnya pada pengklasifikasian menggunakan CNN.

Bojanowski dkk. mengembangkan teknik FastText pada tahun 2017, sebuah penyempurnaan model skip-gram yang digunakan dalam metode Word2vec, dimana pendekatan FastText ini menganalisis representasi kata dengan mempertimbangkan informasi subkata dari kata tersebut [12]. Pada penelitian [13] dilakukan perbandingan metode deep learning dalam pengklasifikasian teks menggunakan CNN, RNN, dan LSTM dengan word embedding Word2vec dan Fasttext. Didapatkan bahwa kombinasi FastText dan LSTM memiliki nilai akurasi terbaik dengan nilai sebesar 97.86%. penemuan ini mengindikasikan bahwa kombinasi tersebut unggul dalam menangani tugas klasifikasi teks dibandingkan dengan kombinasi lainnya. Pada penelitian [14], [15] dibuktikan juga bahwa performa dari model LSTM sering kali lebih unggul dibandingkan dengan CNN dan RNN dalam tugas pengklasifikasian teks.

Pada penelitian ini penulis membandingkan beberapa model word embedding seperti Word2vec, GloVe, dan FastText sebagai optimasi dari word embedding pada pengklasifikasian menggunakan LSTM dan LSTM yang ditambah dengan algoritma SMOTE untuk mengatasi ketidakseimbangan. Penelitian mengenai hal ini belum pernah dilakukan pada riset-riset terdahulu. Dataset pada penelitian ini diambil dari ulasan pengguna aplikasi MyTelkomsel berbahasa Indonesia pada Google Play Store. Dataset yang digunakan memiliki memiliki kecenderungan terhadap salah satu kategori dan juga kategori lainnya yang kurang terwakilkan dengan jumlah yang tidak seimbang antara sentimen positif dan negatif.