

## 1. Pendahuluan

Kegiatan jual beli barang adalah kegiatan yang umum dilakukan oleh masyarakat sehari-hari, kini dengan banyaknya aplikasi online marketplace kegiatan tersebut dapat dilakukan dengan mudah tanpa harus bertatap muka dan dapat dilakukan dimana saja. Di era informasi, Internet, dan media sosial, kebutuhan untuk mengumpulkan dan menganalisis sentimen tidak pernah sebesar ini [1].

Kini dengan pengguna ponsel pintar yang semakin ekspresif dalam memberikan ulasan dan komplain terhadap suatu aplikasi baik dalam bentuk teks atau video. Calon pengguna dapat dengan mudah terpengaruhi oleh ulasan yang ditulis dan didengar sehingga mampu mengurangi jumlah aktivitas di suatu aplikasi. Oleh karena itu para pengembang aplikasi seperti online marketplace diharuskan mampu mengolah ulasan para pengguna yang umumnya dalam bentuk teks. Tetapi dengan banyaknya ulasan dan komplain yang dibuat dalam satu hari, tentu mustahil apabila harus membaca tiap ulasan atau komplain tersebut dalam satu hari.

Sehingga kebutuhan akan teknik serta tools text mining yang efektif dan efisien semakin meningkat dengan jumlah data teks yang dihasilkan semakin banyak. Sentimen Analisis sendiri merupakan salah satu teknik dalam machine learning yang dapat digunakan untuk menganalisa data yang tersimpan dalam bentuk text seperti opini, ulasan, evaluasi, penelitian, penilaian, sikap, dan emosi orang terhadap suatu entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atribut lainnya [2].

Dengan penerapan analisis sentimen, *online marketplace* dapat mengambil langkah-langkah efektif untuk menilai produk atau layanan mana yang memerlukan perbaikan, dan fitur baru apa yang diinginkan pengguna, dan manfaat lainnya. Sentimen Analisis menggunakan NLP dengan kecerdasan buatan untuk mengelompokan data ulasan yang ada berdasarkan kelas sentimen positif dan negatif [3].

Terdapat beberapa metode pendekatan yang dapat digunakan dalam Sentimen analisis. Seperti Ekstraksi fitur dengan Word2Vec yang digunakan untuk memetakan kata-kata yang ada dalam suatu kalimat ke dalam bentuk vektor. Pada penelitian [4] menggunakan pendekatan Word2Vec dinilai memiliki kelebihan dalam memperoleh akurasi yang lebih baik dengan metode klasifikasi Support Vector Machine yang menghasilkan akurasi terbaik.

Metode klasifikasi utama yang dipakai yaitu Long Short-Term Memory (LSTM), LSTM adalah variasi pengembangan dari algoritma Recurrent Neural Network (RNN), dengan menggantikan simpul hidden layer dengan sel LSTM yang berfungsi untuk menyimpan informasi sebelumnya. Dimana pada penelitian sebelumnya [6] yaitu pada *aspect-based sentiment analysis* metode klasifikasi LSTM mendapatkan hasil akurasi 89,32%.

Metode klasifikasi lainnya yang digunakan yaitu Support Vector Machine dengan menemukan hyperplane yang memisahkan data ke dalam kelas-kelas yang berbeda. Hyperplane yang dipilih memiliki margin terbesar, yang merupakan jarak antara hyperplane dan titik data terdekat dari setiap kelas.

Metode ekstraksi fitur yang digunakan pada model SVM yaitu TF-IDF, adalah metode statistik yang mengukur tingkat kepentingan sebuah kata dalam dokumen. Metode ini memperhitungkan frekuensi sebuah kata dalam dokumen dan jumlah dokumen tempat kata tersebut muncul. Metode ini telah banyak digunakan untuk ekstraksi fitur dalam model pembelajaran mesin tradisional, termasuk SVM, untuk tugas-tugas NLP, termasuk analisis sentimen. Yang mana SVM merupakan model linear yang dapat dilatih dengan cepat dan mampu memberikan hasil akurasi yang tinggi secara konsisten.

### 1.1 Latar Belakang

Berdasarkan beberapa penelitian sebelumnya [5][6], penulis termotivasi untuk melakukan eksperimen untuk mengetahui pengaruh performansi penerapan teknik word embedding dengan pendekatan Word2Vec pada metode klasifikasi LSTM (Long-Short Term Memory).

Algoritma LSTM dapat menangkap informasi secara *sequential* atau berurutan dalam teks, dan dengan bantuan word2vec sebagai metode ekstraksi fitur mampu memperkuat kemampuan LSTM dalam menangkap makna semantik dari teks, sehingga dapat menghasilkan performansi yang tinggi terhadap task NLP seperti analisis sentimen.

Pada penelitian ini masalah yang dibahas adalah bagaimana pengaruh dan tingkat performansi model LSTM setelah diterapkan teknik word embedding dengan pendekatan Word2Vec serta bagaimana performa dari model LSTM tersebut dibandingkan terhadap model SVM dengan ekstraksi fitur TF-IDF.

Model SVM dan pembobotan TF-IDF digunakan sebagai pembanding karena, algoritma SVM merupakan model linear yang dapat dilatih dengan cepat dan pada penelitian sebelumnya mampu mendapat hasil akurasi sebesar 79%, sehingga sangat cocok untuk task analisis sentimen jika digunakan dengan metode ekstraksi fitur yang sesuai seperti TF-IDF.

## **1.2 Topik dan Batasannya**

Pada penelitian ini masalah yang akan dibahas adalah bagaimana pengaruh dan tingkat performansi sistem setelah diterapkan teknik word embedding pada model klasifikasi LSTM, terhadap performansi model klasifikasi SVM dengan pembobotan TF-IDF.

Kemudian batasan penelitian penelitian ini, yaitu dataset sentimen bahasa Indonesia sebanyak 8000 ulasan pengguna dari empat online marketplace di Indonesia yang diambil dari hasil crawling dari Google Play Review, dengan metode klasifikasi yang digunakan SVM dan LSTM.

## **1.3 Tujuan**

Pengujian pada kedua model dilakukan pada environment python memakai library yang sudah ada, dengan data yang dipakai dikumpul dari database ulasan google play untuk online marketplace di Indonesia, dengan tujuan mengetahui performa dari model LSTM dengan ekstraksi fitur menggunakan pendekatan Word2Vec.

Model lain memakai metode klasifikasi SVM dengan ekstraksi fitur memakai pendekatan TF-IDF yang mampu memberikan hasil yang cukup tinggi pada penelitian sebelumnya, sehingga model SVM tersebut digunakan dengan tujuan untuk menganalisa perbandingan performa dari model LSTM dengan harapan kedua model mampu memberikan hasil pengujian yang tinggi.

## **1.4 Organisasi Tulisan**

Pada bab 2 berisikan studi terkait yaitu hasil literatur yang memiliki kontribusi terbesar yang didapatkan selama penyusunan laporan. Pada bab 3 akan dijelaskan sistem yang dibangun pada penelitian ini, dari gambaran alur proses dari rancangan sistem, hingga penjelasan mengenai tiap metode. Dan pada bab 4 akan dipaparkan hasil pengujian yang sudah dilakukan serta analisa dari hasil pengujian. Kemudian pada bab 5 yaitu kesimpulan yang didapat dari penelitian.