

1. Pendahuluan

Di era digital saat ini, semua orang dapat mengakses informasi dengan mudah menggunakan internet. Internet unik karena menggabungkan dua metode komunikasi yang beragam dan beberapa jenis konten menjadi satu media [1]. Sebelum menonton film, seseorang akan mencari ulasan tentang film untuk ditonton. Internet digunakan untuk membagikan ulasan ini, salah satu platform yang menyediakan tempat untuk berbagi ulasan film adalah Rotten Tomatoes. Informasi dapat diproses menggunakan analisis sentimen berdasarkan penilaian ini.

Analisis Sentimen adalah analisis pendapat, sikap, dan emosi seseorang menggunakan komputer. Tujuan dari analisis sentimen adalah untuk menemukan pendapat, mengidentifikasi emosi yang diungkapkan oleh seseorang, dan mengkategorikan kelas pendapat [2]. Ulasan di situs ulasan film dapat digunakan sebagai sumber daya bagi penggemar film untuk menemukan rekomendasi film sebagai alat bagi produser film untuk menentukan reaksi penonton terhadap film yang baru dirilis [3].

Dalam beberapa tahun terakhir, sejumlah studi telah mengusulkan analisis perasaan berbasis *deep learning* dengan karakteristik dan efektivitas yang bervariasi [4]. Dalam pengenalan pola dan penglihatan komputer, jaringan saraf pembelajaran mendalam telah membuat kemajuan yang luar biasa. Sejumlah algoritma *deep learning* yang canggih, termasuk analisis sentimen, diperkenalkan dari waktu ke waktu untuk menyelesaikan tugas-tugas NLP yang kompleks [5].

Berdasarkan penelitian [6] atas dasar pengembangan penelitian dan aplikasi *deep learning*, diharapkan bahwa penelitian *deep learning* tentang analisis sentimen akan segera muncul. Sebelum proses klasifikasi dilakukan, data yang telah diperoleh akan diproses terlebih dahulu pada tahap *preprocessing* dan dilanjutkan dengan penggunaan ekstraksi fitur TF-IDF yang digunakan. Selain itu, Word2Vec mendukung ekspansi dan *balancing* data melalui SMOTEN. Tujuan akhir dari prosedur klasifikasi ini adalah untuk menghasilkan nilai akurasi. Dalam penelitian ini, akurasi dan *F1-score* digunakan untuk mengevaluasi kinerja dua model klasifikasi.

Tujuan dari penelitian ini adalah untuk mencari tahu metode klasifikasi terbaik, antara CNN dan LSTM, dalam penilaian perasaan pada *dataset* bahasa Inggris yang diperoleh dari situs *website* Rotten Tomatoes. Klasifikasi sentimen dibagi menjadi dua kelas: kelas positif dan kelas negatif. Semakin besar *dataset* yang digunakan, dikombinasikan dengan *deep CNN* yang diusulkan dan strategi pelatihan, akan menghasilkan kemampuan model yang lebih baik untuk menyederhanakan, dan meningkatkan kepercayaan dalam generalisasi [7]. LSTM telah terbukti menjadi alat yang tidak ternilai dalam mempelajari tugas pemodelan urutan yang melibatkan panjang yang tidak diketahui. Keuntungan utamanya adalah kemampuan untuk menyimpan memori jangka panjang [8]. TF-IDF adalah istilah untuk *frekuensi-inverse frequency*. Teknik ini adalah salah satu yang paling banyak digunakan di bidang pencarian informasi dan analisis teks. TF-IDF adalah metrik untuk menentukan makna kata-kata dokumen [9]. Word2Vec dapat memberikan perkiraan yang akurat dari makna sebuah kata dengan data, penggunaan, dan konten yang cukup. Salah satu keuntungan penting dari Word2Vec adalah kecepatannya yang tinggi bahkan ketika digunakan dengan *dataset* besar. Dalam konteks *deep learning*, makna kata menjadi sinyal sederhana yang berguna dalam mengklasifikasikan entitas yang lebih kompleks [7]. SMOTEN adalah varian dari algoritma SMOTE yang menggabungkan *oversampling* dengan penghapusan data. Selain *over sample* kelas minoritas, metode ini juga menghapus instansi atau data yang berdekatan dengan kelompok minoritas. Tujuan utama dari metode ini adalah untuk mengidentifikasi dan menghilangkan tetangga kelas mayoritas yang berdekatan sebelum mengambil sampel yang berlebihan dari kelas minoritas, dengan demikian menghapus data yang paling sering sebelum menyeimbangkan kelas yang kurang diwakili [10].