

1. Pendahuluan

Media komunikasi baik media cetak, media elektronik, maupun media internet sering menyajikan hasil *review* dari sebuah film [2]. Hasil *review* yang ada pada berbagai platform tersebut berisi diskusi maupun komentar para penontonnya terhadap film-film tersebut. Seperti yang kita tahu, semakin besar dan semakin banyak jumlah informasi tersebut maka akan semakin berpengaruh informasi. Apabila masyarakat melakukan kesalahan dalam mengartikan dan memahami kumpulan informasi tersebut, kesalahan konteks dalam penyampaian akan memiliki potensi yang besar untuk terjadi. Teknik analisis sentimen digunakan untuk mencegah hal tersebut, dimana analisis sentimen akan melakukan analisa pada tiap teks untuk mengetahui aspek dan tingkatan nilai sentimen (positif atau negatif) seperti pada jurnal [3].

Penelitian mengenai analisis sentimen terhadap ulasan film berbahasa inggris sudah pernah dilakukan dengan berbagai metode seperti pada penelitian [1] dimana berfokus akan pengaruh penerapan tahap *Preprocessing* pada *dataset* menggunakan metode *Naïve Bayes* dan LSTM (*Long Short-Term Memory*) yang menghasilkan nilai akhir rata-rata *F1-Score* tertinggi sebesar 91.41% untuk *dataset* yang menggunakan proses *preprocessing*. Ada pula penelitian selanjutnya [4] dimana menggunakan *dataset* ulasan untuk film '*Quantum of Solace*' dan '*Star Trek*' dengan fokus untuk mengetahui subjektifitas dan polaritas kalimat memakai beberapa algoritma klasifikasi *supervised* dan *unsupervised*, untuk film '*Quantum of Solace*' didapatkan hasil akhir nilai akurasi yang berada pada lingkup 70% keatas dengan algoritma klasifikasi *supervised* serta akurasi yang berada di sekitar 60% kebawah untuk algoritma klasifikasi *unsupervised*. Bagi film '*Star Trek*' sendiri, didapatkan hasil akhir yang berada di sekitar nilai 60% untuk semua metode yang digunakan. Penelitian analisis sentimen pada *movie review* juga dilakukan oleh salah satu peneliti lainnya pada jurnal [27] dimana lebih berfokus pada teknik klasifikasi *Multinomial Naïve Bayes* dengan memakai *Gini Index Text Feature Selection* kemudian menggunakan *dataset* yang berasal dari IMDB, didapatkan hasil performansi terbaiknya sebesar 59.54% untuk *dataset* yang menggunakan *Multinomial Naïve Bayes* serta menerapkan *Gini Index Text Feature Selection*. Penelitian terakhir yang digunakan sebagai rujukan merupakan penelitian pada jurnal [28] dengan fokus utama melihat performansi KNN dengan bantuan *Information Gain*, didapatkan hasil akhir akurasi sebesar 60% untuk penerapan KNN tanpa bantuan IG, sedangkan untuk penerapan KNN dengan bantuan IG didapatkan hasil akurasi sebesar 96.8%.

Berdasarkan hasil-hasil penelitian diatas, penelitian ini dilakukan untuk mengetahui dan menganalisa nilai sentimen dari sebuah kalimat dalam hasil *review* film berbahasa inggris dari *website* IMDB menggunakan algoritma yang dapat memproses & memetakan informasi menjadi sebuah vektor, yaitu menggunakan metode *Word2Vec*. Penggunaan metode *Random Forest* juga diterapkan dalam penelitian ini untuk menganalisa dan melakukan klasifikasi terhadap nilai sentimen hasil *review* dalam forum *website* IMDB.

Topik yang akan dibahas pada penelitian kali ini yaitu mengetahui pengaruh beberapa faktor terhadap hasil akhir penelitian seperti penerapan *stemming* terhadap *dataset*, dan jumlah dimensi *skip-gram Word2Vec* yang digunakan. Untuk batasan pada penelitian ini diantaranya *dataset* yang digunakan diambil dari *website* IMDB, *dataset* berjumlah 50.000 *review* dengan kelas positif dan negatif, bahasa *dataset* yang digunakan adalah bahasa inggris, kemudian pada penelitian ini tidak berfokus pada pencarian nilai n-tree yang optimal pada tahap klasifikasi.

Tujuan dari penelitian tugas akhir ini berdasarkan latar belakang dan perumusan masalah yang telah dijabarkan sebelumnya yaitu melakukan analisis sistem pada *review* film berbahasa inggris menggunakan algoritma *Random Forest* dengan *Word2Vec* untuk metode ekstraksi fitur, kemudian mengukur hasil performansi dari sistem yang dibangun.

Bagian selanjutnya yang akan dibahas adalah pada bab 2 membahas studi literatur yang terkait dengan penelitian ini. Pada bab 3 membahas teori dan rancangan sistem penelitian. Hasil dan analisis dibahas pada bab 4. Kesimpulan dari penelitian yang sudah dilakukan dibahas pada bab 5.