

1. Pendahuluan

1.1 Latar Belakang

Seiring bertambahnya tahun, perkembangan industri perfilman menjadi semakin pesat dan banyak diminati. Akses yang mudah untuk menikmati suatu film merupakan salah satu faktor yang menjadikan penggemar film meningkat. Ulasan film merupakan wadah untuk para penikmat film dapat menyampaikan suatu komentar atau kritik terhadap berbagai film. Ulasan film tersebut dapat berisikan suatu komentar positif atau negatif yang dapat memberikan informasi mengenai suatu film. Namun, terdapat beberapa ulasan film yang terlihat kurang jelas apakah ulasan tersebut termasuk ke dalam kategori ulasan positif atau negatif.

Salah satu teknik yang digunakan untuk mengkategorikan apakah ulasan film tersebut tergolong kelas positif atau negatif adalah analisis sentimen atau *opinion mining*. Analisis sentimen atau *opinion mining* adalah suatu bidang studi yang dilakukan untuk memahami, mengekstrak, serta mengolah suatu opini, sentimen, sikap, dan emosi terhadap entitas seperti produk, layanan, maupun topik dari suatu kasus. Pada saat ini, analisis sentimen sering dilakukan untuk memperoleh informasi serta polaritas mengenai suatu entitas atau objek dalam bentuk tekstual untuk mengetahui apakah kalimat tersebut cenderung positif atau negatif [1]. Dalam analisis sentimen terdapat pendekatan *lexicon* dan *machine learning* untuk melakukan pengklasifikasian. Pendekatan *lexicon* didasarkan pada penggunaan kamus yang terdapat pada kata-kata opini atau dengan nama lain kamus sentimen, yang digunakan untuk menetapkan sentimen kelas positif atau negatif. Sedangkan untuk pendekatan *machine learning*, merupakan pendekatan yang dalam penggunaannya memerlukan data manual dari *dataset* dan membuat model untuk melatih *classifier* yang dikenal dengan *data training* yang nantinya akan diujikan dengan *data testing* pada proses klasifikasi sentimen.

Sudah banyak penelitian mengenai analisis sentimen menggunakan pendekatan *machine learning*, seperti analisis sentimen terhadap ulasan *e-commerce* menggunakan *K-Nearest Neighbor*, analisis mengenai ulasan produk *online* menggunakan *K-nearest Neighbor* dan *Naïve Bayes*, dan lain sebagainya. *Support Vector Machine* merupakan algoritma klasifikasi yang mempunyai nilai akurasi yang tinggi dan akurat yaitu 89% lebih baik dari pada *Naïve Bayes*, *Random Forest*, dan *Decision Tree* [2][3]. Sehingga, *Support Vector Machine* digunakan untuk mencapai kinerja performansi yang baik untuk sistem klasifikasi yang dibangun serta *Support Vector Machine* mampu bekerja untuk data yang mempunyai dimensi yang tinggi. Namun, dikarenakan *Support Vector Machine* dapat bekerja untuk data yang mempunyai dimensi yang tinggi, akan menjadi salah satu faktor yang mempengaruhi hasil dari proses klasifikasi. Hal itu disebabkan karena banyaknya fitur yang terkandung dalam *dataset* serta terdapat fitur yang kurang relevan terhadap kelas keluaran tertentu akan mengakibatkan menurunnya performansi dalam proses klasifikasi yang dibangun. Untuk mengatasi hal tersebut, perlu adanya penggunaan seleksi fitur [4].

Seleksi fitur merupakan salah satu teknik yang digunakan untuk mengurangi dimensi fitur yang tinggi dalam data yang ada dengan menghapus fitur-fitur yang kurang relevan. Salah satu metode seleksi fitur yaitu *Gini Index*. Pada penelitian [5] yang membandingkan penggunaan metode seleksi fitur, didapatkan bahwa metode *Gini Index* dapat menghasilkan performansi yang lebih baik dibandingkan dengan metode seleksi fitur lainnya. *Gini Index* digunakan untuk memisahkan fitur-fitur yang ada dari yang relevan ke fitur yang *irrelevant* terhadap kelas keluaran tertentu. Sehingga, penggunaan metode seleksi fitur *Gini Index* diharapkan mampu untuk meningkatkan kinerja dari algoritma klasifikasi *Support Vector Machine* yang bekerja pada dimensi fitur yang tinggi untuk mereduksi dimensi fitur tersebut dan dapat meningkatkan kinerja dari klasifikasi serta dapat meningkatkan perolehan akurasi yang dibangun.

Berdasarkan kondisi tersebut, penelitian ini menggunakan pendekatan *machine learning* yang berfokus pada algoritma *Support Vector Machine* dengan *Gini Index Feature Selection* dalam melakukan analisis sentimen terhadap ulasan film. Adapun penelitian ini dilakukan untuk dapat mengelompokkan suatu pendapat atau opini terhadap suatu film, dengan melihat kecenderungan opini tersebut apakah tergolong suatu komentar positif atau negatif, sehingga dapat memudahkan pengguna atau penikmat film dapat mengetahui kualitas dari film berdasarkan pengalaman dari penonton yang diluapkan ke dalam ulasan film. Serta dilakukan pengujian pada penelitian ini yang berdasarkan pada skenario pengujian untuk mengetahui performansi dari sistem yang dibangun dengan menggunakan evaluasi *F1-Score*, dimana semakin tinggi hasil yang diperoleh maka semakin baik sistem yang dibangun.

1.2 Topik dan Batasannya

Pada penelitian ini penulis mengangkat permasalahan terkait bagaimana membangun sistem klasifikasi sentimen analisis terhadap ulasan film dengan tepat. Ketepatan dalam pembangunan sistem klasifikasi analisis sentimen ini dapat dipengaruhi oleh sejumlah tahapan yang ada dalam sistem, seperti *preprocessing*, proses pemilihan fitur, dan penggunaan metode klasifikasi yang digunakan. Tahapan-tahapan tersebut menjadi titik fokus dalam penelitian ini. Dalam tahap *preprocessing*, penulis menggunakan *lemmatization* sebagai pengujian dalam tahapan *preprocessing*. Pada proses pemilihan fitur, akan digunakan perbandingan terhadap penggunaan *Gini Index* sebagai metode seleksi fitur dan tanpa penggunaan *Gini Index* sebagai metode seleksi fitur. Pada saat melakukan perbandingan mengenai penggunaan seleksi fitur, diperhatikan jumlah fitur yang digunakan dalam tahapan klasifikasi untuk mencari perolehan *F1-Score* yang optimal. Kemudian, dalam tahapan metode

klasifikasi, menggunakan metode *Support Vector Machine kernel linear*, sehingga didapatkan performansi dari sistem yang telah dibangun. Adapun batasan-batasan yang terdapat dalam penelitian Tugas Akhir ini. Pertama, menggunakan *dataset* ulasan film berbahasa Inggris yang didapat dari situs *website IMDb*. Kedua, menggunakan *level* dokumen dalam melakukan klasifikasi sentimen terhadap ulasan film. Ketiga, dikarenakan keterbatasan spesifikasi perangkat yang digunakan maka, *dataset* yang digunakan dalam penelitian ini berjumlah 3000 ulasan film dengan berlabel positif atau negatif.

1.3 Tujuan

Tujuan dalam penelitian Tugas Akhir ini untuk menganalisis pengaruh yang ditimbulkan dalam menggunakan *lemmatization* pada tahap *preprocessing*, menganalisis perbandingan antara penggunaan *Gini Index* sebagai metode seleksi fitur dan tanpa penggunaan *Gini Index* sebagai metode seleksi fitur, dan menganalisis perbandingan penggunaan fitur yang telah ditentukan.

1.4 Organisasi Tulisan

Terkait dengan Analisis Sentimen Terhadap Ulasan Film Menggunakan *Support Vector Machine* dengan *Gini Index Feature Selection*, maka dalam penelitian ini terdapat pemaparan isi jurnal tugas akhir seperti halnya, bagian I Pendahuluan, menjelaskan mengenai latar belakang, topik dan batasan masalah, serta solusi yang dijelaskan secara rinci. Dalam bagian II Studi Terkait, memaparkan berbagai penelitian yang telah dilakukan sebelumnya serta hal-hal yang terdapat kaitannya dengan penelitian tugas akhir ini. Selanjutnya bagian III Sistem yang dibangun, yang menjelaskan metode penelitian yang digunakan dalam suatu sistem yang dihasilkan serta menunjukkan implementasinya secara rinci. Kemudian bagian IV Evaluasi, memaparkan hasil pengujian dan analisis hasil pengujian dalam penelitian Tugas Akhir. Terakhir bagian V Kesimpulan, memberikan kesimpulan singkat terkait hasil pengujian dan analisis pengujian serta memberikan saran untuk dilakukan pada penelitian selanjutnya.