

BAB I PENDAHULUAN

1.1 Latar Belakang

Memprediksi kesuksesan sebuah film merupakan aspek penting karena risiko tinggi dan investasi signifikan yang terlibat dalam produksi film [1]. Pembuat film dan investor perlu membuat keputusan yang tepat sejak awal proses produksi untuk meningkatkan peluang kesuksesan sebuah film. Namun, ada dua tantangan utama dalam hal ini. Tantangan pertama adalah ketidakpastian dalam industri film, dengan banyak parameter yang terlibat, seperti genre film, negara produksi, pemeran, sutradara, komentar *trailer*, dampak media sosial dan preferensi penonton [2]. Tantangan kedua adalah memilih model *machine learning* yang tepat [3]. Kita dapat menggabungkan fitur film dan analisis sentimen dari media sosial menggunakan teknik *machine learning* untuk memprediksi kesuksesan sebuah film [4].

Terdapat dua penelitian terdahulu yang dijadikan sebagai rujukan. Penelitian pertama adalah penelitian yang dilakukan oleh Tripathi *et al* [4]. Mereka menggunakan *dataset* IMDb dan *dataset* ulasan film. Hasilnya menunjukkan bahwa *Linear Regression* memberikan kinerja terbaik untuk prediksi menggunakan fitur film, sedangkan *Linear SVC* adalah algoritma terbaik untuk analisis sentimen dengan akurasi sebesar 88,47%. Penelitian kedua adalah penelitian yang dilakukan oleh Gandasari *et al* [5]. Mereka menggunakan *dataset* Netflix yang diperoleh melalui *data scraping* dengan Python. Hasilnya menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi sebesar 81,59%.

Algoritma *ensemble*, seperti *Random Forest* dan *Gradient Boosting*, menggabungkan kekuatan beberapa model untuk meningkatkan akurasi prediksi. Metode *ensemble* membangun model prediktif dengan menggunakan beberapa prediktor berbeda yang dikenal sebagai *base learners*. *Base learners* adalah model sederhana yang biasanya tidak berkinerja lebih baik daripada tebakan acak, sehingga disebut sebagai *weak learners*. Dengan menggabungkan hasil dari *base learners*, metode *ensemble* menciptakan *strong learner* yang mampu menghasilkan prediksi yang lebih akurat [6]. Penelitian di berbagai sektor, termasuk industri film, telah menunjukkan efektivitas pendekatan *machine learning ensemble* dalam meningkatkan akurasi prediksi [7]. Pada tahun 2022, Zhang *et al.* [8] mengevaluasi model *Gradient Boosting Random Forest*, yang mengoptimalkan *decision trees* dalam *Random Forest* menggunakan *Gradient Boosting*. Hasil evaluasi menunjukkan bahwa *Gradient Boosting Random Forest* mencapai akurasi 5% lebih tinggi dibandingkan dengan *Random Forest* saja.

Algoritma LSTM telah banyak digunakan dalam analisis sentimen ulasan film dan telah menunjukkan kinerja yang baik. Bilen dan Horasan [9] membandingkan algoritma LSTM dengan teknik *machine learning* lainnya untuk analisis sentimen pada *dataset* IMDb dan menemukan bahwa LSTM memiliki performa yang lebih baik.

Penelitian ini bertujuan untuk memprediksi kesuksesan film menggunakan fitur-fitur film dan komentar *trailer* yang diunggah di YouTube. Penelitian ini merujuk pada penelitian Tripathi *et al.* [4], yang menyarankan penggunaan algoritma *machine learning* lainnya dan analisis sentimen dari objek yang berbeda. Selain itu, penelitian oleh Gandasari *et al.* [5] juga dijadikan referensi, menggunakan objek yang sama, yaitu Netflix, tetapi dengan model *ensemble*. Model *ensemble* lebih unggul daripada model tunggal dalam memprediksi kesuksesan film karena dapat menggabungkan kekuatan beberapa model.

Dari permasalahan dan kajian literatur yang telah dijabarkan, maka dilakukanlah penelitian ini dengan judul “**Prediksi Kesuksesan Film Berdasarkan Fitur dan Komentar Trailer Menggunakan Model Ensemble+LSTM**”

1.2 Rumusan Masalah

Berdasarkan latar belakang, prediksi kesuksesan film perlu dilakukan untuk membantu pembuat film dalam meningkatkan peluang kesuksesan film yang diproduksi. Maka permasalahan yang dihadapi pada penelitian ini adalah apakah penggabungan fitur film dan analisis sentimen dari komentar *trailer* dapat meningkatkan hasil prediksi kesuksesan film? Kemudian apakah model *ensemble (Random Forest+Gradient Boosting)* dan LSTM memiliki tingkat akurasi yang lebih tinggi jika dibandingkan dengan algoritma *machine learning* lainnya?

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah melakukan prediksi kesuksesan film berdasarkan fitur dan komentar *trailer* dengan menggunakan model *ensemble Random Forest Gradient Boosting* dan LSTM.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat:

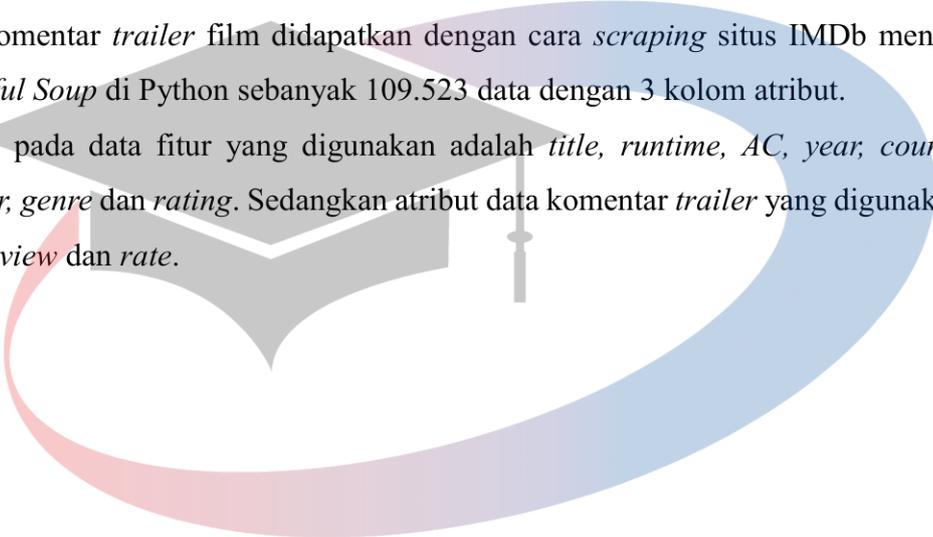
1. Memberikan kesempatan kepada pembuat film untuk mengubah strategi pemasaran berdasarkan prediksi kesuksesan film yang didapatkan setelah *trailer* diunggah.
2. Memberikan panduan kepada pembuat film dan investor dalam membuat keputusan terkait produksi film selanjutnya.

3. Sebagai literatur untuk penelitian yang akan datang.

1.5 Ruang Lingkup

Batasan dan ruang lingkup yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. Data film yang digunakan adalah film yang tayang di Netflix dan rilis pada rentang waktu Januari 2000 sampai dengan Juni 2024.
2. Data fitur film dan pemeran film didapatkan dengan cara *scraping* situs IMDb menggunakan Apify (diakses melalui <https://console.apify.com/>) sebanyak 4.021 data dengan 15 kolom atribut.
3. Data komentar *trailer* film didapatkan dengan cara *scraping* situs IMDb menggunakan *Beautiful Soup* di Python sebanyak 109.523 data dengan 3 kolom atribut.
4. Atribut pada data fitur yang digunakan adalah *title, runtime, AC, year, country, cast, director, genre* dan *rating*. Sedangkan atribut data komentar *trailer* yang digunakan adalah *title, review* dan *rate*.



UNIVERSITAS
MIKROSKIL