

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam industri perfilman, menarik dan memikat penonton merupakan tujuan utama dari setiap produksi film. Seorang penulis naskah (*scriptwriter*) seringkali dihadapkan pada tugas yang rumit untuk menghasilkan naskah film yang sesuai dengan selera penonton, yang berkontribusi pada keberhasilan sebuah film di pasaran. Sebuah film harus disukai dan dapat dipahami maksudnya oleh penonton serta membuat penonton penasaran untuk terus mengikuti adegan demi adegan dalam cerita film (Saleh, 2016). Namun penonton memiliki beragam preferensi genre film dan ekspektasi ketika memilih film yang akan mereka tonton, yang digambarkan dengan rating penonton dari skala 1-10. Rating sebuah film sebenarnya tidak hanya bergantung pada sinopsis, tetapi juga dipengaruhi beberapa faktor misalnya aktor, sutradara, *scriptwriter*, cameraman, bahkan produser film tersebut. Namun, hampir di semua kasus, sinopsis dari film tersebut memberikan dampak yang sangat besar pada rating penonton (Battu, 2018).

Selama ini, penentuan genre film masih dilakukan secara manual oleh produser film (Hoang, 2018). Menurut paper tersebut juga disebutkan bahwa masih terdapat kesalahan terutama dalam film yang multi-label atau lebih dari satu genre. Genre film memiliki peran yang krusial dalam memprediksi rating penonton karena setiap genre cenderung memiliki karakteristik dan ekspektasi tertentu yang secara langsung mempengaruhi bagaimana penonton akan menilai sebuah karya. Sebagai contoh, film bergenre drama seringkali menarik perhatian penonton dengan narasi yang mendalam dan karakter yang kompleks, sementara film aksi mungkin lebih dihargai untuk adegan-adegan ketegangan dan kecepatan. Dengan mengetahui genre film, kita dapat mengidentifikasi elemen-elemen khas yang diharapkan oleh penonton dalam konteks tertentu. Oleh karena itu, untuk memprediksi genre dan rating berdasarkan sinopsis sangat penting agar sesuai dengan preferensi dan harapan audiens potensial.

Teknologi kecerdasan buatan berkembang sangat pesat, sehingga pada tahun 2017 ditemukan metode “Transformer” yang lebih cepat dan efisien. Metode “Transformer” bergantung pada *parallel multi-head attention mechanism* sehingga membutuhkan lebih sedikit

training time dibandingkan metode lain seperti LSTM (Vaswani, 2017). Salah satu turunan dari model Transformer yang terbaru adalah Bidirectional Encoder Representations from Transformers atau disingkat BERT. BERT diciptakan untuk membuat pemahaman teks yang lebih dalam dari teks yang tidak memiliki label dengan mempertimbangkan konteks di sebelah kiri dan kanan sehingga kita bisa menggunakan model BERT yang sudah dilatih dengan hanya menambahkan satu lapisan tambahan untuk menjawab pertanyaan dan tugas-tugas lainnya tanpa perlu merombak strukturnya (Devlin, 2018).

Seiring evolusi teknologi, varian-varian seperti RoBERTa telah berhasil menunjukkan performa yang lebih unggul dibandingkan dengan model dasarnya. RoBERTa merupakan varian BERT yang diperkenalkan oleh Liu et al. (2019), yang menggunakan lebih banyak data latihan sehingga memungkinkan model memahami lebih banyak jenis teks dengan lebih baik. Selain itu RoBERTa juga memiliki peningkatan langkah-langkah *preprocessing* sehingga dapat menyaring noise atau informasi yang kurang relevan dengan lebih baik. Indonesian RoBERTa adalah versi RoBERTa yang telah dilatih khusus dengan data Wikipedia dalam bahasa Indonesia. Model ini menggunakan pendekatan Masked Language Modeling (MLM), sama seperti RoBERTa, tetapi khusus untuk bahasa Indonesia (Suwarningsih et al., 2022). Paper tersebut juga menyebutkan bahwa model ini tidak membedakan antara “indonesia” dan “Indonesia”, yang berarti model ini tidak memperhatikan kapitalisasi atau “*uncased*”.

Dengan adanya metode AI yang menggunakan analisis naskah untuk memberikan prediksi kinerja dan evaluasi konten, dapat membantu produser dan investor dalam mengambil keputusan berdasarkan data yang akurat (Hanifa, H., 2023). Dengan memanfaatkan Transformer, penulis naskah dapat melakukan prediksi genre film yang paling cocok dan prediksi rating dari penonton berdasarkan analisis mendalam terhadap elemen-elemen dalam sinopsis film sehingga dapat membuat keputusan yang lebih tepat sesuai preferensi dan ekspektasi penonton. Dengan adanya alat bantu AI ini, produser film, investor, dan penulis naskah dapat mengoptimalkan kreativitas mereka sambil mempertimbangkan preferensi audiens, sehingga dapat membantu menghasilkan film-film yang lebih sukses dan disukai oleh penonton.

1.2 Perumusan masalah

Berdasarkan latar belakang yang ada, dapat dirumuskan permasalahan

sebagai berikut:

1. Seberapa akurat metode IndoBERT atau Indonesian RoBERTa dalam memprediksi genre film?
2. Seberapa akurat metode IndoBERT atau Indonesian RoBERTa dalam memprediksi rating penonton?

1.3 Tujuan Penelitian

Tujuan skripsi ini adalah untuk memanfaatkan teknologi kecerdasan buatan (AI) terutama IndoBERT atau Indonesian RoBERTa untuk membantu mengevaluasi sinopsis cerita yang akan dibuat apakah sesuai dengan minat pasar.

1.4 Ruang Lingkup

Ruang lingkup dibatasi pada:

- a. Aplikasi dikembangkan dengan bahasa pemrograman Python
- b. Aplikasi diimplementasikan dalam bentuk website sederhana dengan framework Flask
- c. Pengembangan model prediksi menggunakan model IndoBERT dan Indonesian RoBERTa.
- d. Menggunakan library dari "huggingface"
- e. Dataset menggunakan API dari TMDb (The Movie Database) yang didapatkan dari <https://developer.themoviedb.org/> dan kaggle (1000 data, untuk tambahan dataset prediksi genre).
- f. Data yang digunakan hanya yang sesuai dengan kriteria berikut:
 - i. Bahasa Indonesia
 - ii. Berbentuk Film (bukan *tv series / mini series*)
- g. Jumlah data sekitar 3,000 row (film) dengan atribut antara lain:
 - i. Title
 - ii. Vote Average / Rating
 - iii. Genre
 - iv. Synopsis / Overview
- h. Data di tokenize dengan *BERT Subword Tokenizer*.
- i. Data di vectorize menggunakan *Contextual Embeddings*.
- j. Data null akan di drop / tidak digunakan.
- k. Pengukuran akurasi prediksi genre menggunakan *Confusion Matrix*.

- l. Pengukuran akurasi prediksi rating penonton menggunakan *Mean Absolute Error* (MAE).
- m. Untuk memverifikasi performa aplikasi, akan diujicobakan pada produser/sutradara/scriptwriter film indie yang pernah membuat film pendek dan layar lebar.
- n. Cara kerja aplikasi:
 - i. Membuka halaman web
 - ii. User memasukkan sinopsis yang ingin diperiksa
 - iii. Aplikasi menampilkan 1-3 prediksi genre film yang paling sesuai
 - iv. Aplikasi menampilkan prediksi rating penonton (dengan skala 1-10)

1.5 Metodologi Penelitian

Langkah-langkah dalam pengerjaan Skripsi:

1. Studi Literatur
 - 1.1. Teori mengenai peringkat atau *rating* TMDb
 - 1.2. Teori mengenai *Bidirectional Encoder Representations from Transformers* (BERT)
 - 1.3. Teori mengenai IndoBERT
 - 1.4. Teori mengenai Indonesian RoBERTa
2. *Preprocessing Data*
 - 2.1. Melakukan cleaning parameter yang tidak diperlukan
3. Pembuatan Model
 - 3.1. Membuat model dengan IndoBERT dan RoBERTa
 - 3.2. Melakukan *training data* pada model
4. Pengujian dan Analisis Program
 - 4.1. Menganalisa hasil *output* program
 - 4.2. Melakukan uji coba program kepada user
 - 4.3. Menganalisa hasil respon user
5. Pengambilan Kesimpulan
 - 5.1. Mengukur akurasi *output* program
 - 5.2. Membuat kesimpulan berdasarkan hasil penelitian dan analisa yang sudah dilakukan

6. Pembuatan Website

- 6.1. Pembuatan *User Interface* dalam bentuk website untuk implementasi hasil

7. Pembuatan Laporan

- 7.1. Pembuatan laporan dari hasil yang diperoleh

1.6 Sistematika Penulisan

Penulisan laporan skripsi ini dibagi menjadi beberapa bab, yaitu :

- Bab I : Pendahuluan
Bab ini berisikan judul, latar belakang, perumusan masalah, ruang lingkup, tujuan skripsi, dan metodologi penelitian yang akan digunakan dalam skripsi ini.
- Bab II : Landasan Teori
Bab ini berisikan teori-teori yang digunakan dan diterapkan dalam skripsi ini.
- Bab III : Analisis dan Desain Sistem
Bab ini menjelaskan analisis masalah yang dihadapi dan perencanaan pembuatan keseluruhan system dalam aplikasi yang akan dibuat.
- Bab IV : Implementasi Sistem
Bab ini berisikan tentang implementasi system berdasarkan desain
- Bab V : Pengujian Sistem
Bab ini berisi tentang hasil pengujian yang dilakukan terhadap aplikasi yang telah dibuat berdasarkan implementasi pada system yang telah dirancang dan dibuat pada Bab IV.
- Bab VI : Kesimpulan dan Saran
Bab ini berisikan kesimpulan yang dapat diambil terhadap hasil yang dicapai, dan saran-saran yang berguna bagi pengembangan selanjutnya.