

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, klasifikasi genre film berdasarkan sinopsis menggunakan algoritma *Complement Naive Bayes*, dapat disimpulkan hal-hal sebagai berikut:

1. Model klasifikasi genre film berhasil dibangun dengan menggunakan teks sinopsis dan algoritma *Complement Naive Bayes*. Setelah melalui preprocessing teks, ekstraksi fitur TF-IDF, dan penanganan data tidak seimbang dengan *SMOTE-Tomek*, model yang dihasilkan menunjukkan kinerja yang baik dan konsisten.
2. Kinerja model yang dihasilkan mencapai akurasi sebesar 76,87% dengan nilai *Macro F1-Score* 76,27%. Hasil *5-fold cross-validation* yang stabil (akurasi rata-rata $76,21\% \pm 0,0123$) membuktikan bahwa tidak bergantung pada pembagian data tertentu dan memiliki konsistensi yang tinggi.
3. Hasil perbandingan dengan dua metode lain yaitu :
Multinomial Naive Bayes (MNB): akurasi 78,23%, dan *Support Vector Machine* (SVM): akurasi 84,01% dan F1-score 83,34%, menunjukkan bahwa SVM memberikan performa terbaik, sedangkan CNB tetap unggul dalam efisiensi dan interpretabilitas model.
4. Genre Horror dan Romance memiliki performa tertinggi ($F1 > 0,80$), sedangkan Action menjadi yang terendah ($F1 = 0,64$). Perbedaan ini menunjukkan variasi kompleksitas leksikal antar genre.
5. Kinerja model dipengaruhi oleh beberapa faktor kunci, yaitu: kualitas preprocessing teks yang menghilangkan *noise*, penanganan data tidak seimbang (*SMOTE-Tomek*) yang meningkatkan representasi sampel kelas minoritas, dan konfigurasi parameter TF-IDF.

Dengan demikian, penelitian ini menunjukkan bahwa teks sinopsis film mengandung pola linguistik yang cukup kuat untuk mengklasifikasikan genre film, serta CNB masih layak digunakan sebagai baseline yang efisien dibandingkan metode yang lebih kompleks seperti SVM.

5.2 Saran

Berdasarkan keterbatasan dan temuan selama proses penelitian, berikut adalah beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian selanjutnya:

1. Perbaikan dan Perluasan Dataset:

Mengumpulkan dataset yang lebih besar dan lebih seimbang untuk semua genre, sehingga tidak diperlukan penyaringan genre yang ketat. Ini akan membuat model lebih komprehensif. Dan mempertimbangkan pendekatan *multi-label classification* karena sebuah film seringkali memiliki lebih dari satu genre. Penelitian ini hanya menggunakan genre utama, yang berpotensi mengabaikan informasi penting.

2. Eksplorasi Teknik dan Model yang Lebih Maju:

Mengeksplorasi model berbasis deep learning, seperti LSTM (*Long Short-Term Memory*) atau Transformer, yang mampu memahami konteks dan urutan kata dalam sinopsis, sesuatu yang tidak dapat ditangani oleh model Naive Bayes. Melakukan optimasi hyperparameter yang lebih menyeluruh tidak hanya pada TF-IDF tetapi juga pada model klasifikasi itu sendiri untuk mendapatkan kombinasi parameter terbaik. Membandingkan teknik penanganan data tidak seimbang lainnya, seperti SMOTE-ENN atau ADASYN, untuk melihat efektivitasnya pada dataset.

3. Penyempurnaan Fitur dan Preprocessing:

Menggunakan kamus stemming yang lebih khusus untuk domain film dan sinopsis untuk meningkatkan akurasi pengembalian kata dasar. Mengeksplorasi metode ekstraksi fitur lain selain TF-IDF, seperti Word2Vec atau GloVe, yang dapat menangkap hubungan semantik antar kata. Menambahkan fitur tambahan di luar sinopsis, seperti nama judul, sutradara, atau tahun rilis, yang mungkin memiliki hubungan dengan genre tertentu.

4. Implementasi dan Deployment:

Membangun antarmuka (*interface*) seperti aplikasi web atau mobile, sehingga model yang telah dibuat dapat digunakan secara praktis oleh orang lain untuk memprediksi genre film berdasarkan sinopsis.

