BAB 5

KESIMPULAN DAN SARAN

Bab ini akan membahas kesimpulan yang didapat dari hasil penelitian ini dan saran yang dapat diberikan untuk pengembangan penelitian ini lebih lanjut.

5.1 Kesimpulan

Kesimpulan yang dihasilkan dari penelitian menggunakan *dataset* yang digunakan adalah :

- Faktor yang paling berpengaruh dalam menentukan kesuksesan film berdasarkan *dataset* yang digunakan antara lain adalah *votes*, *budget* dan jumlah *view trailer* Youtube. Evaluasi akurasi prediksi *revenue* menggunakan R2 dapat mencapai 0.63 berdasarkan penggunaan fitur tersebut. *Votes*, *Budget* dan Youtube adalah 3 fitur yang memiliki korelasi tertinggi dengan *revenue* dibanding fitur lain berdasarkan pengujian *pearson*.
- Berdasarkan *dataset* yang dianalisis, selera penonton berbeda dengan selera kritikus *review* dalam menilai bagus tidaknya sebuah film. Selera penonton (*votes*) memiliki hubungan korelasi positif dengan *revenue* yang lebih tinggi yaitu 0.6 . Nilai korelasi *pearson votes* memiliki nilai yang lebih tinggi dibanding selera kritikus (*review*) yaitu 0.2.
- Grafik tren nilai akumulasi *revenue* , *profit* dan *budget* dari tahun ke tahun meningkat berdasarkan *dataset* yang dianalisis. Pada tahun 2011, terjadi penurunan yaitu nilai akumulasi *revenue* yang menurun dibanding tahun sebelumnya. Hal ini disebabkan oleh film-film tahun 2011 yang lebih banyak menghasilkan *revenue* yang lebih kecil dari tahun 2010. Terjadi peningkatan jumlah film yang dibuat tiap tahunnya.
- Berdasarkan *dataset* yang dianalisis, tiap kombinasi *genre* film memiliki rentang pendapatan yang berbeda. Visualisasi distribusi *revenue boxplot* tiap kombinasi *genre* seperti contoh kombinasi *Action,Adventure,Mystery* memiliki nilai Q2 yang lebih besar dari *Action,Drama,Fantasy*. Pemilihan *genre* pada pembuatan film mempengaruhi rentang *revenue* yang dapat diperoleh.
- Berdasarkan *dataset* yang dianalisis, *budget* yang besar tidak menjamin *profit* yang diperoleh akan besar. Visualisasi *barchart* perbandingan 10 *profit* tertinggi pada tiap kombinasi *genre* menunjukkan film dengan kombinasi *genre Horror,Mystery,Thriller* memiliki *budget* yang sangat kecil tetapi mendapatkan keuntungan yang besar.
- Tiap aktor memiliki *genre* favorit. *Genre* favorit aktor adalah *genre* yang paling sering dimainkan seorang aktor dan memliki kontribusi jumlah film paling banyak. *Genre* favorit aktor cenderung berkontribusi menghasilkan *revenue* yang tinggi dibanding *genre* lain berdasarkan pengujian *clustering* aktor.
- Algoritma *Agglomerative* lebih cepat dibandingkan dengan algoritma *K-Means* dalam melakukan *clustering*. Berdasarkan pengujian *clustering* pada *dataset*, waktu yang dibutuhkan *Agglomerative* adalah 9 detik sedangkan *K-Means* adalah 752 detik. *Agglomerative* lebih cepat dari *K-Means* karena *Agglomerative* tiap iterasinya akan menggabungkan 2 data objek dan

mengurangi jumlah *cluster* terpisah sedangkan *K-Means* tiap iterasi akan menghitung jarak data objek dengan *centroid*.

- *Hashtag* Instagram pada *dataset* yang dianalisis tidak memiliki korelasi positif yang kuat dengan *revenue*. *Hashtag* Instagram mengandung kata-kata yang orang sering gunakan seperti 'Red' dan 'Vacation'. Hal ini menyebabkan data *Hashtag* mengandung noise
- Hubungan korelasi positif Youtube dengan *Revenue* lebih tinggi dari Instagram berdasarkan pengujian korelasi dengan *pearson*. Hal ini disebabkan oleh pengaruh kesalahan data *hashtag* judul film di Instagram pada kesimpulan sebelumnya.
- Pengujian prediksi *revenue* berdasarkan *cluster* menghasilkan nilai evaluasi akurasi R2 yang sangat kecil yaitu 0.23 . Hal ini disebabkan oleh jumlah data *train* yang sangat sedikit setelah di*cluster* sehingga model prediksi tidak valid untuk diuji.

5.2 Saran

Saran yang dapat dilakukan untuk memperbaiki dan mengembangkan penelitian ini lebih lanjut :

- Merubah metode prediksi dengan mengubah model regresi menjadi model klasifikasi. Film-film pada *dataset* dapat dikelompokkan berdasarkan *revenue* / *profit*.
- Menambah ukuran *dataset* yang dianalisis. Penambahan jumlah film pada *dataset* akan membantu mengatasi kendala ketika *dataset* sudah di*cluster* tidak mengalami kekurangan data *train*.

DAFTAR REFERENSI

[1] Han, J., Kamber, M., dan Pei, J. (2012) Data mining concepts and techniques, third edition.

[2] Tan, P.-N., Steinbach, M., Karpatne, A., dan Kumar, V. (2020) *Introduction to data mining.* Pearson.

[3] Draper, N. R. dan Smith, H. (1998) *Applied regression analysis.* John Wiley & Sons.

[4] Manning, C. D., Raghavan, P., dan Schutze, H. (2018) *Introduction to information retrieval.* Cambridge University Press.