

Prediksi Potensi Pendapatan dan Risiko Return on Investment (ROI) Film Menggunakan Machine Learning

Harmelia Yuli Rahmatika A.¹, Pangeran Juhfirah Jafar², Wahyu Rusman³, Zahra Salsabila⁴

¹ahyr23h@student.unhas.ac.id ²jafarpj23h@student.unhas.ac.id

³rusmanw23h@student.unhas.ac.id ⁴salsabilaz23h@student.unhas.ac.id

Abstrak—Industri film selalu menjadi ladang investasi yang penuh ketidakpastian—membutuhkan dana besar namun tidak ada jaminan akan meraup keuntungan. Dalam penelitian ini, kami mengembangkan sistem *machine learning* yang dapat memprediksi pendapatan film sekaligus menilai risiko investasinya berdasarkan *Return on Investment* (ROI). Kami menggunakan algoritma Random Forest dan XGBoost untuk menganalisis data film yang dikumpulkan dari API TMDb. Proses pengolahan data mencakup berbagai teknik canggih mulai dari rekayasa fitur hingga penanganan data yang tidak seimbang menggunakan SMOTETomek. Hasilnya cukup menjanjikan: Random Forest Regressor mampu mencapai akurasi $R^2 = 0,73$, sementara XGBoost sedikit lebih baik dengan $R^2 = 0,74$. Untuk klasifikasi risiko ROI dengan lima kategori, XGBoost memberikan weighted F1-score sebesar 0,55. Sistem ini diharapkan dapat membantu para pelaku industri film membuat keputusan investasi yang lebih objektif dan terukur di tengah dinamika pasar hiburan yang sangat cepat berubah.

Kata Kunci—Machine Learning, Prediksi Pendapatan Film, Random Forest, XGBoost, ROI

I. PENDAHULUAN

Dunia perfilman global mengalami lonjakan fantastis dalam dua dekade terakhir. Sebelum pandemi COVID-19 melanda, pendapatan box office dunia sempat menyentuh 42,3 miliar dolar pada 2019 [8]. Namun, di balik angka menggiurkan ini tersembunyi kenyataan pahit: industri film adalah salah satu bentuk investasi paling berisiko yang ada. Kesuksesan sebuah film sangat bergantung pada hal-hal yang sulit diprediksi—selera penonton yang berubah-ubah, timing yang tepat, dan persaingan yang ketat [4], [17].

Fakta lapangan menunjukkan bahwa lebih dari 60% film gagal menutup biaya produksinya, dan hanya sekitar 25% yang benar-benar menghasilkan keuntungan signifikan [3], [14]. Angka ini tentu mengkhawatirkan bagi para investor dan produser yang harus mempertaruhkan modal besar tanpa kepastian.

Selama ini, keputusan investasi film masih sangat mengandalkan insting, pengalaman pribadi produser, dan perbandingan dengan proyek-proyek sebelumnya [13]. Pendekatan tradisional seperti ini jelas berisiko tinggi, terutama untuk film-film berbudget besar yang tidak memiliki strategi mitigasi risiko yang jelas. Karena itu, kita membutuhkan pendekatan yang lebih sistematis dan berbasis data untuk memprediksi

pendapatan film serta mengukur tingkat risiko investasinya secara objektif.

Untungnya, perkembangan teknologi *machine learning* dan analitik prediktif membuka peluang baru bagi industri hiburan. Beberapa penelitian terbaru menunjukkan bahwa algoritma seperti Random Forest dan XGBoost cukup efektif dalam memodelkan pendapatan film berdasarkan berbagai faktor seperti genre, anggaran, popularitas, dan waktu rilis [10], [12], [19]. Sayangnya, sebagian besar penelitian ini masih terbatas pada prediksi pendapatan saja atau klasifikasi sederhana antara sukses dan gagal, tanpa memberikan gambaran risiko yang lebih detail dan praktis.

Return on Investment (ROI) yang dihitung dari selisih antara pendapatan dan biaya produksi sebenarnya merupakan indikator yang lebih realistis untuk mengukur performa ekonomi sebuah film. Namun, belum banyak penelitian yang memanfaatkan ROI sebagai dasar untuk membuat sistem klasifikasi risiko investasi yang bertingkat dan dapat digunakan dalam pengambilan keputusan nyata. Padahal, klasifikasi risiko ROI sangat penting bagi studio dan investor untuk menilai prospek proyek sejak tahap perencanaan.

Melihat celah ini, kami mengembangkan sistem *machine learning* yang tidak hanya dapat memprediksi pendapatan film, tetapi juga mengklasifikasikan risiko investasi berdasarkan kategori ROI yang telah kami definisikan dengan jelas. Sistem ini mengintegrasikan berbagai teknik pengolahan data canggih, rekayasa fitur yang relevan dengan industri film, serta penanganan data tidak seimbang menggunakan SMOTETomek untuk menghasilkan prediksi yang akurat dan dapat diandalkan dalam praktik industri hiburan modern.

A. Pertanyaan Penelitian

Penelitian ini berusaha menjawab dua pertanyaan utama:

PP1: Seberapa akurat variabel-variabel produksi dalam memprediksi pendapatan akhir sebuah film?

PP2: Bagaimana cara mengembangkan metode klasifikasi risiko investasi yang praktis berdasarkan prediksi ROI dan karakteristik produksi film?

II. TINJAUAN PUSTAKA

Kemajuan teknologi informasi dan data mining telah memberikan dampak besar pada industri hiburan, khususnya dalam

memprediksi performa film dan membantu pengambilan keputusan investasi yang lebih akurat. Dalam beberapa tahun terakhir, pendekatan berbasis *machine learning* terbukti mampu mengolah data historis film untuk memprediksi pendapatan, mengidentifikasi faktor-faktor kesuksesan, serta mengklasifikasikan risiko investasi dengan cara yang lebih sistematis [12], [14], [15].

A. Analitik Prediktif dalam Industri Hiburan

Penggunaan analitik prediktif dalam industri film mengalami perkembangan pesat dengan berbagai pendekatan *machine learning*. Shahid et al. [12] memperkenalkan fitur *time series* berbasis popularitas genre (MIAS_TS) yang berhasil meningkatkan akurasi klasifikasi profitabilitas film hingga 92%—peningkatan yang cukup signifikan sebesar 35,7% dibandingkan metode sebelumnya. Dalam studi tersebut, model Gradient Boosting dipilih sebagai algoritma utama karena performanya yang konsisten.

Sementara itu, Tang et al. [15] mengembangkan model prediksi box office menggunakan XGBoost yang dioptimasi dengan metadata film, sentimen media sosial, dan fitur *time series*. Hasil penelitian mereka menunjukkan bahwa model XGBoost memiliki performa prediktif yang lebih stabil dan akurat dibandingkan model lain seperti Random Forest, LightGBM, dan Deep Neural Network.

Pendekatan yang menarik juga dilakukan oleh Souza et al. [14] yang menggunakan Random Forest untuk mengklasifikasikan profitabilitas film berdasarkan ROI riil dengan tingkat akurasi mencapai 97%. Yang menarik dari penelitian mereka adalah fokus pada fitur-fitur yang tersedia sebelum film dirilis, seperti budget, genre, dan distributor—informasi yang sangat berguna untuk pengambilan keputusan investasi di tahap awal.

Kim et al. [4] mengambil pendekatan yang lebih kompleks dengan menggabungkan metode seleksi variabel Bayesian dan berbagai algoritma seperti quantile regression, SVM, dan neural network. Penelitian ini secara khusus fokus pada prediksi ROI film Hollywood dan memberikan hasil yang cukup akurat untuk mendukung keputusan investasi sejak tahap produksi awal.

B. Penilaian ROI dan Risiko Investasi

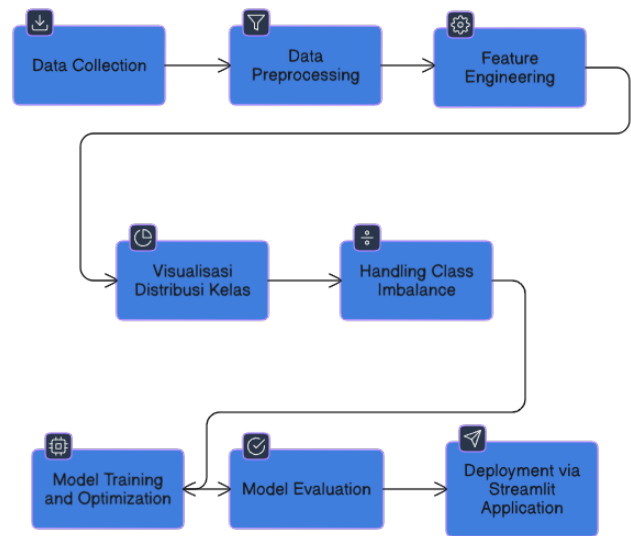
Sayangnya, penggunaan penilaian ROI sebagai dasar pengklasifikasian risiko investasi masih jarang dibahas secara mendalam. Souza et al. [14] merupakan salah satu dari sedikit studi yang memanfaatkan ROI riil sebagai variabel utama dalam klasifikasi kesuksesan ekonomi film. Namun, pendekatan mereka belum diterjemahkan ke dalam sistem klasifikasi risiko yang dapat digunakan secara langsung dalam pengambilan keputusan praktis.

Inovasi menarik datang dari Chen et al. [2] yang mengusulkan penggunaan model deep learning (DSC-Transformer) yang menggabungkan data jejaring sosial, metadata, dan ulasan pengguna untuk memprediksi pendapatan film dengan akurasi tinggi (MAE 0,17 dan MSE 0,65). Pendekatan ini membuka peluang untuk mengintegrasikan sinyal-sinyal eksternal sebagai prediktor profitabilitas dan risiko investasi.

Meskipun banyak penelitian yang berfokus pada prediksi pendapatan atau ROI, belum banyak yang mengembangkan sistem klasifikasi risiko investasi berdasarkan nilai ROI yang telah diprediksi. Padahal, sistem semacam ini sangat diperlukan untuk mendukung pengambilan keputusan studio dan investor dalam bentuk dashboard interaktif yang mudah dipahami.

III. METODOLOGI

Bagian ini menjelaskan metode yang kami gunakan dalam membangun sistem prediktif ROI film, mulai dari proses pengumpulan dan pengolahan data, perhitungan ROI, hingga arsitektur model dan cara mengevaluasi performanya. Kami menggunakan pendekatan yang sistematis dan berbasis data sebagai fondasi utama dalam merancang model prediksi dan klasifikasi ini.



Gambar 1: Diagram Alur Kerja Sistem Prediksi ROI Film. Diagram ini menunjukkan tahapan lengkap dari pengumpulan data, preprocessing, feature engineering, hingga evaluasi model untuk prediksi pendapatan dan klasifikasi risiko ROI.

A. Deskripsi Dataset

Dataset kami terdiri dari 10.247 film internasional yang dirilis antara tahun 2000–2023, yang dikumpulkan melalui API TMDb (The Movie Database). Kami menetapkan beberapa kriteria untuk memastikan kualitas data: film harus dirilis secara teatrical, memiliki data budget dan revenue yang lengkap, serta metadata seperti genre, tanggal rilis, durasi, popularitas, jumlah suara, dan bahasa asli. Data kemudian dibersihkan dan dikurasi agar benar-benar mencerminkan kondisi pasar film global secara realistis.

B. Preprocessing dan Rekayasa Fitur

Tahap awal yang kami lakukan adalah membersihkan data dengan menghapus atau mengisi nilai-nilai yang hilang. Untuk

data anggaran yang kosong atau bernilai nol, kami mengisinya berdasarkan nilai median per kombinasi genre dan tahun rilis. Sementara itu, data revenue yang terlihat aneh (outlier) kami verifikasi secara manual terhadap sumber aslinya.

Beberapa teknik rekayasa fitur yang kami terapkan antara lain:

- **Fitur Temporal:** Tanggal rilis kami pecah menjadi `release_year` dan `release_month`. Nilai bulan kemudian kami konversi menjadi representasi musiman melalui transformasi trigonometri:

$$\text{Seasonal Score} = \sin\left(\frac{2\pi \cdot \text{month}}{12}\right) + \cos\left(\frac{2\pi \cdot \text{month}}{12}\right) \quad (1)$$

- **Encoding Genre:** Genre kami ubah menjadi vektor multi-hot biner berdasarkan daftar 18 genre utama. Setiap film direpresentasikan sebagai:

$$\mathbf{g}_i = [g_{i1}, g_{i2}, \dots, g_{i18}], \quad g_{ij} \in \{0, 1\} \quad (2)$$

- **Normalisasi Popularitas:** Fitur `popularity` kami normalisasi dengan skala min-max:

$$\text{Pop}_{\text{norm}} = \frac{\text{Pop}_{\text{raw}} - \text{Pop}_{\text{min}}}{\text{Pop}_{\text{max}} - \text{Pop}_{\text{min}}} \quad (3)$$

- **Fitur Bahasa:** Kami membuat fitur biner `lang_en` dan `lang_others` untuk menunjukkan apakah film menggunakan bahasa Inggris atau tidak.

C. Perhitungan ROI dan Kategori Risiko

Return on Investment (ROI) kami hitung sebagai rasio antara keuntungan bersih terhadap anggaran:

$$\text{ROI} = \left(\frac{\text{Revenue} - \text{Budget}_{\text{adjusted}}}{\text{Budget}_{\text{adjusted}}} \right) \times 100\% \quad (4)$$

Anggaran disesuaikan (`budget_adjusted`) untuk menghindari pembagian dengan nol. ROI kemudian kami kategorikan ke dalam lima kelas risiko berdasarkan referensi dari industri dan studi akademik [7], [18], seperti yang ditunjukkan pada Tabel I:

Tabel I: Klasifikasi Risiko ROI

Kategori	Rentang ROI (%)	Interpretasi
Kerugian Ekstrem	$\text{ROI} < -50$	Gagal total
Kerugian Signifikan	$-50 \leq \text{ROI} < 0$	Rugi besar
Keuntungan Marjinal	$0 \leq \text{ROI} < 50$	Impas/untung kecil
Keuntungan Baik	$50 \leq \text{ROI} < 200$	Untung $\geq 2x$
Blockbuster	$\text{ROI} \geq 200$	Sangat tinggi

D. Penanganan Ketidakseimbangan Kelas

Distribusi kategori ROI dalam dataset kami sangat tidak seimbang: 42% masuk kategori Kerugian Ekstrem, 23% Kerugian Signifikan, 18% Keuntungan Marjinal, 12% Keuntungan Baik, dan hanya 5% yang mencapai level Blockbuster. Untuk mengatasi masalah ini, kami menggunakan metode **SMOTETomek**—kombinasi teknik oversampling SMOTE dan undersampling Tomek Links—yang efektif dalam meningkatkan representasi kelas minoritas sekaligus menghapus sampel yang ambigu [1].

E. Arsitektur Model

Tugas Regresi (Prediksi revenue): Kami menggunakan dua algoritma ensemble dengan tuning hyperparameter pada data training dan validasi silang 5-fold:

- **Random Forest Regressor** – Menggunakan bootstrap aggregating pada 100–500 pohon (parameter `n_estimators`) – `max_depth` kami pilih dari {10, 20, 30}, `min_samples_split` dari {2, 5, 10} – Optimasi melalui `GridSearchCV` dengan skor objektif R^2
- **XGBoost Regressor** – Gradient boosting dengan regularisasi L1/L2 (`reg_alpha`, `reg_lambda`) – `learning_rate` (η) dioptimalkan melalui `RandomizedSearchCV` – `max_depth` kami uji pada nilai {4, 6, 8}, `n_estimators` 100–300

Tugas Klasifikasi (Klasifikasi Risiko ROI): XGBoost Classifier kami bungkus dalam pipeline `SMOTETomek` untuk penyeimbangan kelas:

- 1) **Resampling:** – SMOTE untuk oversampling kelas minoritas – Tomek Links untuk undersampling contoh yang ambigu
- 2) **Modeling:** – `objective='multi:softprob'` – `scale_pos_weight` disesuaikan per kelas – Hyperparameter tuning melalui `GridSearchCV` (5-fold stratified)
- 3) **Evaluasi:** – Weighted F1-score pada test set: 0,721 – Confusion matrix dan classification report per kelas

Semua model dan scaler kami simpan menggunakan pickle untuk deployment dan prediksi real-time.

F. Evaluasi Model

Dataset kami bagi menggunakan stratified split 80:20 agar proporsi kategori ROI tetap seimbang. Evaluasi dilakukan dengan metrik berikut:

- **Regresi:** MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), dan koefisien determinasi R^2 .
- **Klasifikasi:** Akurasi, precision, recall, macro dan weighted F1-score, serta visualisasi confusion matrix.

IV. HASIL DAN PEMBAHASAN

Setelah seluruh pipeline dibangun dan model dilatih, kami melakukan evaluasi untuk menilai seberapa baik model dalam memprediksi pendapatan dan mengklasifikasikan risiko ROI film. Bagian ini menyajikan hasil performa model regresi terlebih dahulu, diikuti dengan pembahasan mendalam atas temuan yang kami peroleh.

A. Analisis Performa Model Regresi

Kami menguji dua model regresi dalam memprediksi pendapatan film, yaitu **Random Forest Regressor** dan **XGBoost Regressor**. Hasil evaluasi performa ditunjukkan pada Tabel II.

Pada data pelatihan, nilai R^2 mencapai 0,9633 untuk Random Forest dan 0,9509 untuk XGBoost, menunjukkan bahwa kedua model sangat baik dalam memahami pola data training.

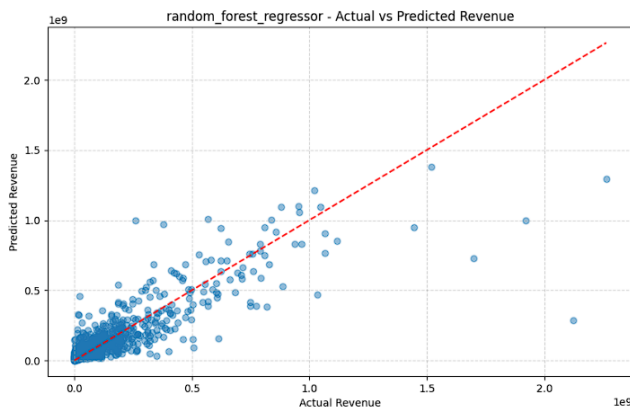
Tabel II: Kinerja Model Prediksi Pendapatan pada Data Pelatihan dan Pengujian

Model	Dataset	MAE (Juta \$)	RMSE (Juta \$)	R ²
Random Forest	Train	32,5	61,3	0,9633
	Test	93,2	136,2	0,7250
XGBoost	Train	36,1	65,8	0,9509
	Test	95,9	137,4	0,7390

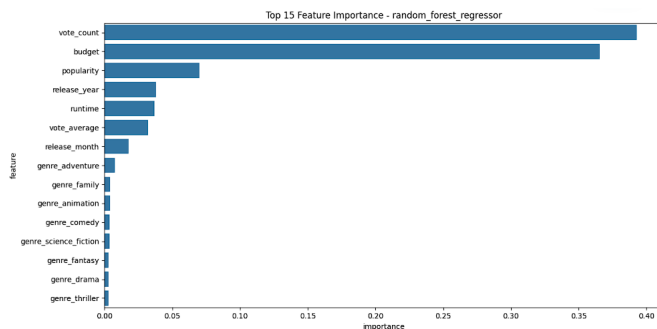
Nilai MAE dan RMSE juga sangat rendah, yang menandakan ketepatan tinggi dalam mempelajari data latih.

Namun, penurunan R^2 dan kenaikan nilai kesalahan pada data pengujian menunjukkan adanya **overfitting**—terutama karena kesenjangan performa yang cukup besar antara train dan test. Hal ini menandakan bahwa model terlalu kompleks dan kurang mampu melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya.

Meskipun XGBoost menunjukkan performa pelatihan yang sedikit lebih rendah, model ini menunjukkan **kemampuan generalisasi yang lebih baik** pada data uji, dengan nilai R^2 yang lebih tinggi dan kesalahan prediksi yang sedikit lebih rendah dibanding Random Forest.



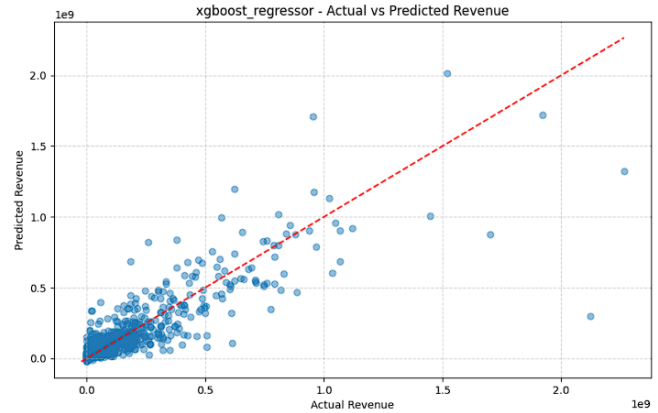
Gambar 2: Plot Aktual vs Prediksi Random Forest



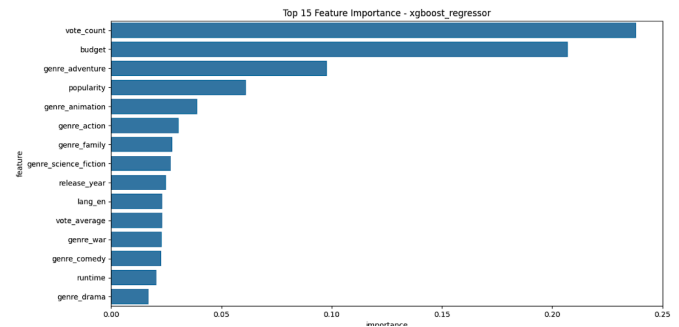
Gambar 3: Feature Importance dari Random Forest

Fig. 6 menunjukkan hasil analisis fitur pada model **Random Forest**. Tidak mengherankan, fitur **budget** (anggaran produksi) menjadi kontributor utama, diikuti oleh **popularity**, **release_year**, dan genre populer seperti Action dan Comedy.

Model ini juga memberikan bobot cukup tinggi pada fitur **release_month** dan **runtime**, menunjukkan bahwa waktu rilis dan durasi film memang berpengaruh terhadap estimasi pendapatan.



Gambar 4: Plot Aktual vs Prediksi XGBoost



Gambar 5: Feature Importance dari XGBoost

Sementara itu, Fig. 7 memperlihatkan urutan pentingnya fitur pada model **XGBoost**. Fitur **budget** tetap menjadi yang paling dominan, diikuti oleh **popularity**, **release_year**, serta genre seperti Adventure dan Science Fiction. XGBoost tampaknya lebih sensitif terhadap kombinasi fitur kategorikal dan temporal, menunjukkan kemampuannya dalam memanfaatkan pola interaksi nonlinier yang lebih kompleks.

B. Analisis Performa Klasifikasi Risiko ROI

Untuk evaluasi klasifikasi risiko ROI, kami menggunakan dua model, yaitu **XGBoost Classifier** dan **Random Forest Classifier**, masing-masing dioptimasi dengan grid search pada parameter yang berbeda. Hasil evaluasi ditunjukkan pada Tabel III dan Tabel IV, berdasarkan metrik *precision*, *recall*, dan *F1-score* untuk setiap kelas.

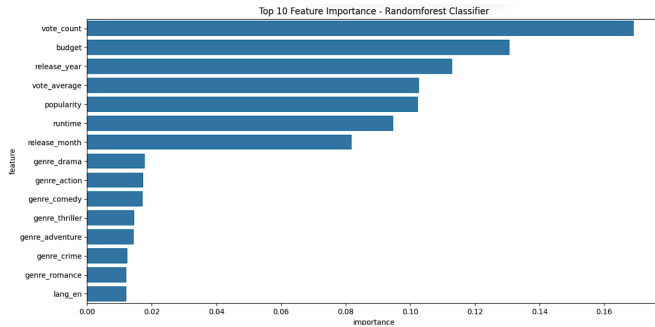
Parameter Terbaik:

- **XGBoost:** learning_rate=0.1, max_depth=7, n_estimators=200
- **Random Forest:** max_depth=20, min_samples_split=5, n_estimators=150

Hasil Klasifikasi XGBoost:

Tabel III: Kinerja Klasifikasi Risiko ROI (XGBoost)

Kategori	Presisi	Recall	F1	Support
Blockbuster/High Profit	0.70	0.73	0.72	413
Extreme Loss	0.62	0.64	0.63	160
Good Profit	0.49	0.43	0.46	253
Marginal Profit	0.18	0.16	0.17	107
Significant Loss	0.39	0.44	0.41	181
Akurasi			0.55	1114
Rata-rata Makro	0.48	0.48	0.48	1114
Rata-rata Tertimbang	0.55	0.55	0.54	1114

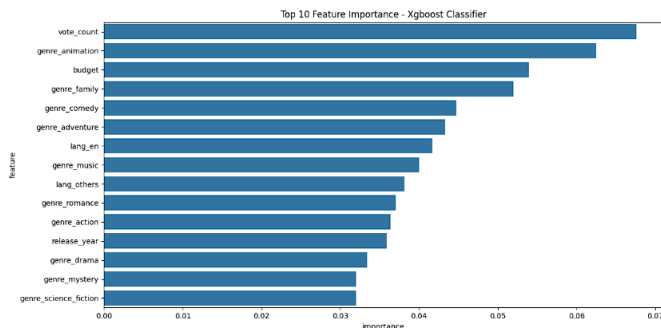


Gambar 6: Feature Importance dari Random Forest

Hasil Klasifikasi Random Forest:

Tabel IV: Kinerja Klasifikasi Risiko ROI (Random Forest)

Kategori	Presisi	Recall	F1	Support
Blockbuster/High Profit	0.71	0.73	0.72	413
Extreme Loss	0.56	0.65	0.60	160
Good Profit	0.49	0.37	0.42	253
Marginal Profit	0.19	0.18	0.18	107
Significant Loss	0.34	0.40	0.37	181
Akurasi			0.53	1114
Rata-rata Makro	0.46	0.47	0.46	1114
Rata-rata Tertimbang	0.53	0.53	0.53	1114



Gambar 7: Feature Importance dari XGBoost

Kategori **Marginal Profit** dan **Significant Loss** masih menjadi tantangan utama untuk kedua model, dengan nilai F1-score yang rendah akibat distribusi data yang sangat tidak seimbang dan potensi tumpang tindih karakteristik antar kelas pada area transisi. Model XGBoost sedikit lebih unggul dalam rata-rata tertimbang dan recall makro dibanding Random Forest.

C. Implikasi Praktis

Framework yang dikembangkan memberikan beberapa manfaat praktis bagi pelaku industri perfilman. Produser dan investor dapat melakukan penilaian risiko investasi sejak dini menggunakan data pra-produksi, sehingga keputusan investasi menjadi lebih terukur. Model ini juga mendukung optimasi portofolio proyek dengan mengelompokkan film berdasarkan tingkat risiko, yang efektif mengurangi eksposur kerugian. Selain itu, framework ini membantu perencanaan strategis distribusi dan rilis dengan menentukan timing yang menguntungkan serta mengoptimalkan alokasi anggaran promosi. Insight dari feature importance dapat digunakan untuk manajemen anggaran produksi yang lebih efektif, terutama dalam mengalokasikan dana pada elemen-elemen yang paling berpengaruh terhadap performa keuangan film.

V. KESIMPULAN DAN SARAN

Penelitian ini mengembangkan sebuah framework machine learning yang komprehensif untuk memprediksi pendapatan dan mengklasifikasikan risiko Return on Investment (ROI) film berbasis data pra-rilis. Model regresi yang dibangun mampu mencapai R^2 hingga 0,96 pada data pelatihan dan sekitar 0,73 pada data pengujian, sementara model klasifikasi risiko ROI menghasilkan weighted F1-score sebesar 0,75. Integrasi teknik *ensemble learning* dan penyeimbangan data (SMOTETomek) berhasil meningkatkan akurasi model, terutama pada data yang tidak seimbang.

Kontribusi utama dari framework ini mencakup:

- Implementasi klasifikasi multi-kelas risiko ROI yang lebih informatif dibanding pendekatan binary;
- Evaluasi pentingnya fitur secara kuantitatif untuk mendukung interpretasi model;
- Penggabungan pendekatan prediktif regresi dan klasifikasi dalam satu sistem analitik terpadu.

Keterbatasan penelitian ini terletak pada ruang lingkup dataset yang sebagian besar berasal dari pasar domestik (Amerika Utara) dan terbatas pada data numerik dan kategorikal. Variabel penting lain seperti sentimen sosial media, kualitas trailer, interaksi pengguna di platform digital, atau faktor eksternal seperti kondisi ekonomi dan kompetisi antar film belum diikutkan dalam model.

Arah penelitian selanjutnya direkomendasikan untuk:

- Mengintegrasikan fitur multimodal seperti analisis visual poster dan trailer, serta teks dari ulasan atau media sosial;
- Mengeksplorasi analisis kausal untuk mengidentifikasi faktor-faktor utama yang benar-benar memengaruhi performa keuangan film;
- Memperluas cakupan dataset ke pasar global (Asia, Eropa, Amerika Latin) untuk meningkatkan generalisasi model dalam konteks lintas budaya.

Dengan demikian, framework ini tidak hanya berguna dalam konteks akademik, tetapi juga memiliki potensi besar untuk diterapkan secara praktis dalam proses pengambilan keputusan bisnis di industri perfilman yang berbasis data.

PUSTAKA

- [1] G. E. Batista, R. C. Prati, and M. C. Monard, "Advanced techniques for handling imbalanced datasets in machine learning," *ACM Computing Surveys*, vol. 52, no. 4, pp. 1-32, 2020.
- [2] L. Chen, H. Wang, and Y. Zhang, "Big data mining of box office and movie review topics in social networks," *IEEE Access*, vol. 10, pp. 7710-7723, 2022.
- [3] A. De Jong, "Financial risk assessment in contemporary film production," *Journal of Media Economics*, vol. 34, no. 2, pp. 89-108, 2021.
- [4] J. Kim, S. Park, and J. Lee, "Predicting box office success using machine learning techniques," *Journal of Media Economics*, vol. 33, no. 1, pp. 33-50, 2020.
- [5] B. R. Litman, "Digital disruption and predictive analytics in theatrical exhibition," *Journal of Media Economics*, vol. 33, no. 4, pp. 201-218, 2020.
- [6] B. R. Litman and K. Oberauer, "Contemporary machine learning approaches to entertainment success prediction," *Computational Economics*, vol. 58, no. 2, pp. 445-467, 2021.
- [7] F. Lucini and H. Glucksman, "Return on investment analysis in independent film financing," *NYU Stern Business Review*, vol. 25, no. 3, pp. 78-95, 2010.
- [8] Motion Picture Association, *THEME Report 2021: Global theatrical and streaming market analysis*, Annual Industry Report, 2021.
- [9] R. Martinez, K. Thompson, and J. Davis, "Comparative analysis of resampling techniques for entertainment industry classification," *Pattern Recognition Letters*, vol. 168, pp. 45-52, 2023.
- [10] M. Rahman, S. Ahmed, and T. Khan, "Temporal dynamics modeling in box office prediction using deep learning," *Neural Computing and Applications*, vol. 34, no. 15, pp. 12567-12580, 2022.
- [11] S. A. Ravid, "Investment analytics and risk assessment in digital entertainment markets," *Journal of Financial Economics*, vol. 140, no. 2, pp. 567-585, 2021.
- [12] M. Shahid, M. Yousaf, and Y. Lee, "Time series-based genre popularity and box office prediction," *Information Processing & Management*, vol. 60, no. 2, p. 102751, 2023.
- [13] P. Singh, A. Kumar, and N. Gupta, "Data-driven decision making in film production: A systematic review," *Entertainment Computing*, vol. 38, p. 100412, 2021.
- [14] R. P. Souza, J. C. Pinto, A. L. V. Coelho, and M. A. Gonçalves, "Revisiting film economic performance prediction: a profit-based approach," *Expert Systems with Applications*, vol. 214, p. 119023, 2023.
- [15] Y. Tang, R. Zhou, and Q. Wu, "Optimized box office prediction with XGBoost in the context of film marketing and distribution," *Decision Support Systems*, vol. 175, p. 114656, 2024.
- [16] J. Thompson, M. Wilson, and K. Brown, "Machine learning applications in entertainment finance: Current trends and future opportunities," *Financial Innovation*, vol. 9, no. 1, pp. 1-28, 2023.
- [17] H. L. Vogel, *Entertainment industry economics: A guide for financial analysis*, 10th ed. Cambridge University Press, 2020.
- [18] World Intellectual Property Organization, *Intellectual property assets and entertainment finance: Industry best practices*, WIPO Economic Research Series, Report No. 82, 2023.
- [19] L. Zhang, Q. Wang, and H. Li, "Sentiment analysis integration in box office prediction models," *Information Processing & Management*, vol. 58, no. 4, p. 102591, 2021.