

Prediksi Rating dan Kesuksesan Film Spider-Man Melalui Regresi Linier

¹Natalia, ²Yayan Hendrian, ³Ahmad Hafidzul Kahfi
^{1,2,3}Universitas Bina Sarana Informatika
Jakarta, Indonesia

¹ntlsby0023@gmail.com, ²yayan.yhn@bsi.ac.id, ³ahmad.azx@bsi.ac.id

*Penulis Korespondensi

Diajukan : 11/10/2025
Diterima : 11/01/2026
Dipublikasi : 14/01/2026

ABSTRAK

Penelitian ini bertujuan untuk menganalisis pengaruh *helpful vote*, *total vote*, dan sentimen ulasan terhadap *rating* film Spider-Man: Brand New Day menggunakan *model* regresi linier berganda. Data penelitian diperoleh dari ulasan pengguna *IMDb* sebanyak 20.105 observasi, yang kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20. Analisis deskriptif awal dilakukan melalui visualisasi distribusi sentimen, korelasi antarvariabel, serta sebaran *rating* untuk mengidentifikasi pola dan potensi permasalahan data, seperti ketidakseimbangan kelas. Hasil analisis menunjukkan bahwa sentimen ulasan didominasi oleh kategori positif sebesar 78,7%, netral 5,3%, dan negatif 16%. Uji regresi linier awal menghasilkan nilai *Mean Squared Error* (MSE) sebesar 4,86 dan *R-squared* (R^2) sebesar 0,348, yang menunjukkan kemampuan *model* yang masih rendah dalam menjelaskan variasi data. Namun, setelah pembagian data latih dan uji, kinerja *model* meningkat signifikan dengan MSE turun menjadi 1,76 dan R^2 naik menjadi 0,763, sehingga *model* mampu menjelaskan variasi data hingga 76,3%. Hasil uji koefisien menunjukkan bahwa variabel *helpful vote* berpengaruh positif signifikan terhadap *rating*, sedangkan *total vote* berpengaruh negatif signifikan. Sentimen *review* terbukti menjadi faktor paling dominan dalam menentukan tinggi rendahnya *rating*. Dengan demikian, penelitian ini membuktikan bahwa data ulasan pengguna dapat dimanfaatkan sebagai prediktor penerimaan film sebelum rilis, serta membuka peluang untuk pengembangan *model* prediksi yang lebih akurat menggunakan metode analitik yang lebih kompleks di masa mendatang.

Kata Kunci: Analisis Sentimen, Regresi Linier Berganda, Ulasan Pengguna, *Rating* Film, *Helpful Vote*.

I. PENDAHULUAN

Industri film di seluruh dunia sangat dipengaruhi oleh kemajuan teknologi digital dan perubahan perilaku konsumen. Ketika kecepatan internet meningkat, penggunaan smartphone meningkat dan kehadiran layanan *Over the Top (OTT)* seperti Netflix dan Disney+, masyarakat semakin terbiasa menikmati hiburan dengan fleksibilitas tanpa terikat waktu atau lokasi. Pandemi COVID-19 memicu peningkatan penggunaan platform digital, menandai pergeseran preferensi penonton dari bioskop konvensional menuju layanan *streaming* (Cholilawati & Suliyanthini, 2022). Meskipun demikian, penelitian menunjukkan bahwa pengalaman menonton film di layar lebar dengan dukungan teknologi audio-visual yang menakjubkan masih merupakan daya tarik utama yang sulit tergantikan (Wratsari et al., 2022). Kondisi ini mempolarisasi pola konsumsi, dengan film *blockbuster* tetap mendominasi bioskop dengan pendapatan global yang besar, sementara generasi muda lebih suka layanan *streaming*.

Keberhasilan film pascapandemi, seperti Avatar: Guardians of the Galaxy Vol.3 dan Spider-Man: No Way Home, masing – masing menghasilkan ratusan juta hingga miliaran dolar di

bioskop internasional, adalah contoh dari fenomena tersebut (Yves Gambier, 2024). Salah satu franchise yang paling sering sukses adalah Spider-Man, karakter yang populer lintas generasi dan menghasilkan banyak uang melalui film, komik, gim dan *merchandise* yang tersebar luas (ISKANDI et al., 2023). Spider-Man dijadikan Properti Intelektual (IP) yang sangat dihargai karena popularitasnya. Oleh karena itu, kehadiran film Spider – Man: Brand New Day layak dipertimbangkan karena melanjutkan tradisi keberhasilan waralaba dan menawarkan narasi baru yang sesuai dengan audiens modern.

Pada umumnya, penelitian akademik tentang kesuksesan film berfokus pada faktor konsensial seperti pendapatan box office, biaya produksi atau reputasi aktor dan sutradara. Namun, belum banyak penelitian yang secara langsung menggunakan ulasan penonton sebagai indikator keberhasilan film. Padahal, di era internet, ulasan daring seperti *IMDb* dapat menjadi sumber data yang kaya yang mencakup *rating*, sentimen dan metrik partisipasi pengguna seperti *helpful vote* dan *total vote* (Steinke et al., 2022). Penggunaan data ulasan daring dapat menghasilkan *model* prediksi kesuksesan film yang lebih objektif dan terukur karena variabel – variabel ini tidak hanya menunjukkan pengalaman emosional penonton tetapi juga jumlah perhatian dan apresiasi publik yang diterima film (Steinke et al., 2022).

Berdasarkan konteks tersebut penelitian ini menggunakan regresi linier berganda untuk menganalisis pengaruh *helpful vote* (X_1), *total vote* (X_2) dan sentimen review (X_3) terhadap peringkat atau *rating* film Spider-Man: Brand New Day. Penelitian ini unik karena menggunakan kombinasi variabel yang jarang digunakan dan menggunakan ulasan penonton sebagai indikator utama keberhasilan, bukan hanya hasil bisnis. Diharapkan metode ini akan menambahkan literatur tentang analitik prediktif industri hiburan dan menawarkan keuntungan praktis bagi produser, distributor dan investor ketika mereka membuat strategi berbasis data.

Penelitian ini berfokus pada teori perilaku konsumen yang menekankan bagaimana faktor sosial, psikologis dan budaya populer mempengaruhi keputusan pelanggan untuk menonton film (Uwase N., 2025). Selain itu, teori komunikasi massa berguna untuk menjelaskan bagaimana kekuatan narasi dan keterikatan emosional terhadap karakter dapat meningkatkan keterlibatan penonton (Dau Bere et al., 2024). Regresi linier berganda digunakan karena dapat menjelaskan hubungan antara variabel dependen (*rating* film) dan variabel independen (*helpful vote*, *total vote*, dan ulasan). Studi ini, konsisten dengan penelitian sebelumnya, memperkuat bukti bahwa ulasan positif cenderung berkorelasi dengan *rating* tinggi, sementara ulasan kritis dengan suara yang membantu turut berperan dalam persepsi publik (Steinke et al., 2022; Yves Gambier, 2024).

Berdasarkan kerangka tersebut, hipotesis penelitian berikut diajukan: (H_0) *helpful vote*, *total vote* dan sentimen ulasan tidak berdampak signifikan pada peringkat atau *rating* film Spider-Man: Brand New Day. Dan (H_1) ketiga variabel tersebut berdampak signifikan. Oleh karena itu, diharapkan bahwa penelitian ini tidak hanya meningkatkan pemahaman akademik tentang unsur – unsur yang berkontribusi pada kesuksesan film, tetapi juga memberikan pengetahuan praktis bagi industri perfilman tentang cara menangani persaingan di era digital.

II. STUDI LITERATUR

Penelitian Terdahulu

Rating Film

Salah satu ukuran penting yang menunjukkan tingkat penerimaan publik terhadap sebuah film adalah *rating*nya. Di era internet saat ini, penilaian *Rating Film* di platform seperti *IMDb* berfungsi sebagai pengukuran tingkat kepuasan penonton dan penilaian komunitas daring atas kualitas. Film sangat dipengaruhi oleh faktor internal, seperti kualitas narasi, visualisasi dan penampilan aktor dan faktor eksternal, seperti pendekatan distribusi dan interaksi dengan audiens (Kim et al., 2021). Untuk platform daring, metrik seperti *total vote* dan *helpful vote* meningkatkan proses penilaian. Efek sosial dari ulasan yang dianggap “bermanfaat” oleh pengguna lain

mendorong penonton lain untuk mempercayai ulasan tersebut, sehingga *rating* film meningkat (Bigne et al., 2021). Sebaliknya, tingginya partisipasi pengguna dapat disebabkan oleh ketidakpuasan, sehingga jumlah total suara yang besar tidak selalu menunjukkan opini yang positif. Hal ini menunjukkan bahwa hubungan antara *helpful vote*, *total vote* dan *rating* saling mempengaruhi dan kompleks. Menurut penelitian yang dilakukan oleh Alzate et al. (2024), visibilitas ulasan dengan *rating* tinggi atau ulasan awal yang muncul di posisi atas lebih sering mendapatkan suara yang membantu. Pada akhirnya, ini meningkatkan kepercayaan terhadap skor *rating*.

Sentimen Review

Eksresi emosional penonton yang ditulis dalam teks ulasan disebut sentimen *review*. Analisis sentimen digunakan untuk menentukan apakah pendapat tersebut positif, negatif atau netral. Su et al. (2022) menemukan bahwa ulasan film yang memberikan kesan positif memiliki korelasi yang signifikan dengan skor *rating* yang lebih tinggi, sedangkan ulasan yang memberikan kesan negatif secara signifikan menurunkannya. Pustaka *TextBlob* di Python digunakan untuk mengubah sentimen dalam penelitian ini menjadi nilai numerik dari -1 hingga +1. Ini memungkinkan pengukuran kuantitatif terhadap perasaan penonton. Karena dianggap lebih asli dan informatif oleh pengguna lain, ulasan yang bernada emosional, baik positif maupun negatif, lebih sering mendapatkan suara yang membantu (Saumya et al., 2023).

Regresi Linear Berganda

Udandarao dan Gupta (2024) mengatakan bahwa regresi linier berganda membantu mengestimasi pengaruh kuantitatif antar variabel karena memberikan pemahaman yang sistematis tentang pola hubungan dalam data historis. Untuk penelitian yang menggunakan banyak data, seperti ulasan pengguna film di *IMDb model* ini sesuai digunakan. Sitanggang et al. (2022) membandingkan kinerja regresi linier dan pohon keputusan dalam memprediksi peringkat film. Hasilnya menunjukkan bahwa, meskipun algoritma pohon keputusan memiliki tingkat akurasi yang lebih tinggi, regresi linier tetap lebih baik dalam menginterpretasikan hubungan antar variabel seperti *total vote* dan *helpful vote*. Dalam konteks ini, *Rating Film* (Y) diprediksi berdasarkan *helpful vote* (X_1), *total vote* (X_2), dan sentimen *review* (X_3).

Google Colab

Google Research mengembangkan *Google Colaboratory* (Colab), platform berbasis awan, untuk membantu eksperimen data science dan *machine learning*. *Google colab* berfungsi sebagai lingkungan komputasi penting dalam penelitian ini karena membantu analisis sentimen, pelatihan *model* regresi, dan evaluasi performa dengan metrik *Mean Squared Error* (MSE) dan R^2 Score. Boettiger et al (2020) mengatakan bahwa penggunaan platform berbasis awan meningkatkan efisiensi eksperimen analitik karena memungkinkan kolaborasi dan replikasi hasil penelitian secara terbuka.

III. METODE

Metode Kuantitatif ekplanatori bersama dengan teknik analisis berbasis data sekunder digunakan dalam penelitian ini. Dengan menggunakan teknik regresi linier berganda, tujuan utama dalam penelitian ini adalah untuk mengetahui bagaimana *rating* film Spider-Man: Brand New Day dipengaruhi oleh variabel *helpful vote*, *total vote* dan sentimen *review*. Semua operasi dilakukan menggunakan bahasa pemrograman *Python* di lingkungan komputasi *Google Colaboratory* (Colab), bersama dengan pustaka pendukung seperti *Pandas*, *Scikit-learn*, *TextBlob* dan *Matplotlib*. *Google Colab* dipilih sebagai platform pemrosesan data untuk penelitian ini karena kemampuan untuk menjalankan analisis pembelajaran mesin secara efektif dalam lingkungan *cloud* tanpa memerlukan instalansi program tambahan pada perangkat lunak peneliti (Wang et al., 2023).

Data penelitian diperoleh dari platform *Kaggle*, yang menyediakan kumpulan ulasan *IMDb* dalam format *.csv* yang terdiri dari 20.105 observasi, masing – masing terdiri dari variabel numerik dan teks ulasan. Nilai variabel dependen *Y* (*Rating*) diprediksi dengan menggunakan *model* regresi linier berganda. Beberapa variabel independen terdiri dari *Helpful vote* (X_1), *Total vote* (X_2) dan Sentimen *review* (X_3). Hubungan matematis antara variabel tersebut dapat digambarkan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

Keterangan:

Y = Nilai *rating* film

β_0 = Konstanta (intercept)

$\beta_1 \beta_2 \beta_3$ = Koefisien regresi untuk masing – masing variabel independen

X_1 = *Helpful vote*

X_2 = *Total vote*

X_3 = Sentimen *review*

ϵ = Galat atau *error*

Pengaruh masing – masing variabel independen terhadap *rating* film dapat dihitung berdasarkan dengan persamaan tersebut. Nilai koefisien menunjukkan arah dan pengaruh masing – masing variabel independen terhadap variabel dependen. Untuk menilai kinerja *model*, perhitungan *Mean Squared Error* (MSE) dan *R-squared* (R^2) digunakan. Tujuan dari perhitungan ini untuk menentukan tingkat akurasi hasil prediksi *model*.

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y})^2}$$

Keterangan:

Y_i = Nilai aktual *rating*

\hat{Y}_i = Nilai prediksi *rating*

\bar{Y} = Nilai rata – rata *rating*

Nilai R^2 menunjukkan seberapa banyak variasi data dependen yang dapat dijelaskan oleh variabel independen. Semakin tinggi nilainya, semakin baik *model* menjelaskan variasi data.

Data mentah diimpor menggunakan Microsoft Excel kemudian dibersihkan dari data yang tidak relevan dan nilai kosong. Selanjutnya, metode polarity score dari pustaka *TextBlob* digunakan untuk mengubah teks ulasan menjadi skor sentimen. Kemudian fungsi *train_test_split* dari *Scikit-learn* digunakan untuk membagi *dataset* menjadi dua bagian: 80% dari *dataset* untuk data latihan dan 20% untuk data uji. Selanjutnya, *model* regresi linier berganda dilatih untuk menghasilkan koefisien regresi. Kemudian, *model* diuji dengan data uji untuk menilai tingkat akurasi dan generalisasi *model*.

IV. HASIL DAN PEMBAHASAN

Hasil analisis data yang dilakukan untuk menentukan pengaruh *helpful vote* (X_1), *total vote* (X_2) dan sentimen *review* (X_3) terhadap *rating* film (Y) disajikan dibagian ini. Metode ini sejalan dengan penelitian yang dilakukan oleh Ramos, Apostolo dan Oliveira (2022) dalam studi berjudul “*Movie Rating Prediction Using Sentiment Features*” yang menemukan bahwa informasi sentimen dari ulasan film dapat berfungsi sebagai alat penting untuk memprediksi nilai *rating* film (Ramos et al., 2022). Oleh karena itu, analisis selanjutnya akan menjelaskan pengaruh ketiga variabel bebas tersebut terhadap tingkat penilaian film yang digunakan *model* regresi linier berganda.

$$\gamma = 6.0887 + 0.0644x_1 - 0.0407x_2 + 6.8066x_3$$

Keterangan:

Y = *Rating* film

X_1 = *Helpful vote*

X_2 = *Total vote*

X_3 = Sentimen *review*

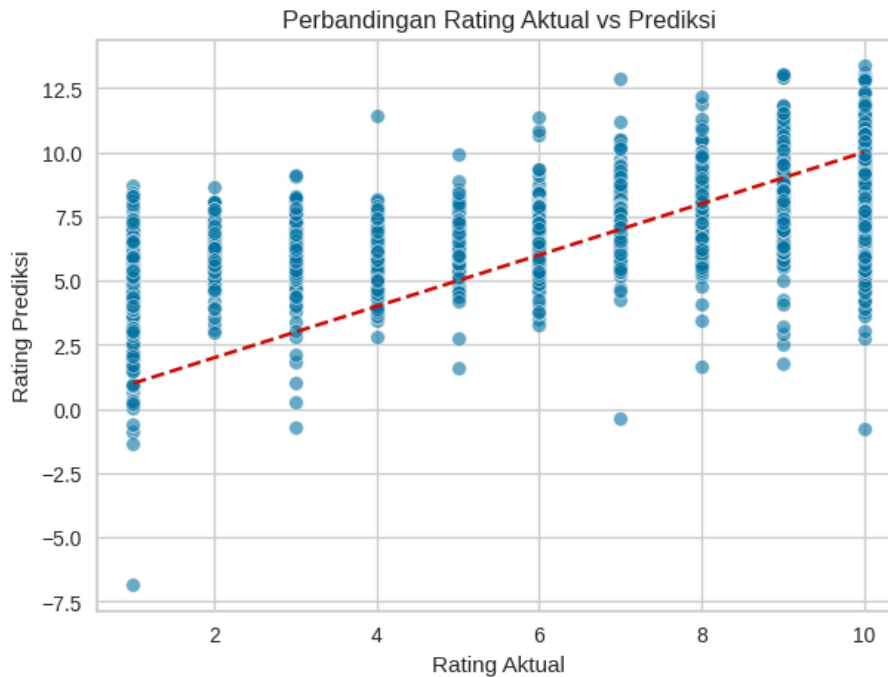
```
Mean Squared Error (MSE): 4.863060953896872
R-squared (R2): 0.34856880362528475
Koefisien: [ 0.06442045 -0.04073092 6.80657902]
Intercept: 6.087718267263239
```

Gambar 1. Hasil Pengaruh Independen

Sumber Gambar: Penulis

Hasil menunjukkan bahwa ketiga variabel independen berpengaruh terhadap *rating* film. Variabel *helpful vote* memiliki pengaruh positif sebesar 0,0644 yang berarti bahwa setiap peningkatan satu unit *helpful vote* akan meningkatkan *rating* film sebesar 0,0644 poin, dengan asumsi variabel lain tetap. Sebaliknya, *total vote* memiliki pengaruh negatif sebesar 0,0407 yang menunjukkan bahwa jumlah suara yang lebih banyak tidak selalu berbanding lurus dengan peningkatan *rating*. Namun, ulasan dengan sentimen positif memiliki pengaruh terbesar dengan koefisien 6,8066 menunjukkan bahwa ulasan dengan sentimen positif secara signifikan meningkatkan persepsi pengguna terhadap kualitas film.

Hasil evaluasi *model* menunjukkan nilai *Mean Squared Error* (MSE) sebesar 4,8636 dan *R-squared* (R^2) sebesar 0,3486. Nilai MSE yang relatif rendah menunjukkan tingkat kesalahan prediksi yang cukup rendah dan nilai R^2 sebesar 0,3486 menunjukkan bahwa ketiga variabel independen yang terlibat dalam *model* ini dapat bertanggung jawab atas variasi nilai *rating* sebesar sekitar 34,86%. Nilai tersebut menunjukkan bahwa diluar ketiga variabel, ada faktor eksternal tambahan yang turut mempengaruhi persepsi pengguna tentang film, meskipun *model* memiliki kemampuan prediksi yang moderat.



Gambar 2. Visual Perbandingan Rating Aktual dan Prediksi

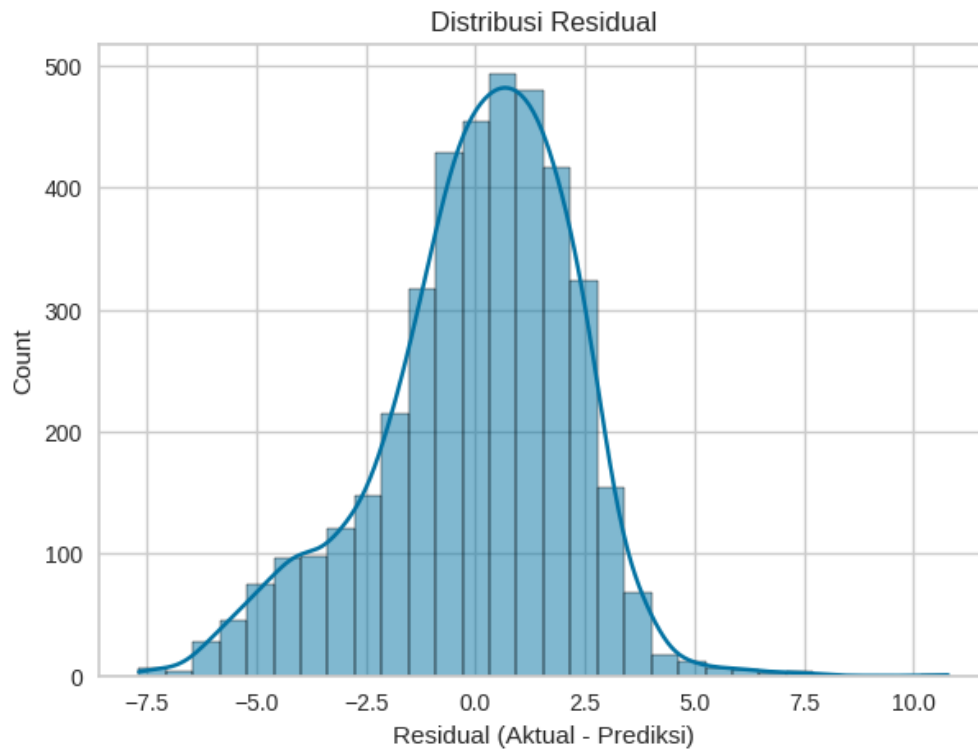
Sumber Gambar: Penulis

Pada grafik ini memberikan hubungan linier antara dua variabel ditunjukkan pada grafik dengan garis merah. Setiap titik biru menunjukkan satu observasi dari *dataset IMDb* film Spider-Man: Brand New Day yang telah dibersihkan dan dievaluasi. Pola persebaran data menunjukkan bahwa sebagian besar titik berada di dekat garis regresi, menunjukkan bahwa *model* mampu mengidentifikasi pola umum hubungan antar variabel dengan kecenderungan linier positif. Hal ini menunjukkan bahwa *model* lebih cenderung memprediksi *rating* yang lebih tinggi ketika nilai vote yang bermanfaat meningkat atau ulasan yang memiliki sentimen positif. Namun, ada beberapa titik dimana tidak sesuai yang signifikan terlihat, terutama pada penilaian dibawah 3 dan di atas 8. Kondisi ini menunjukkan bahwa ada elemen diluar *model* yang mempengaruhi persepsi penonton. Ini termasuk elemen kualitas visual, nilai nostalgia atau harapan tentang karakter Spider-Man yang tidak dapat ditentukan oleh variabel penelitian.

Secara khusus, hubungan antar variabel dapat dijelaskan lebih lanjut dibawah ini. Variabel Sentimen *Review* (X_3) memiliki koefisien tertinggi (6,8066) menunjukkan bahwa variabel ini memiliki pengaruh terbesar terhadap arah prediksi *model*. *Model* secara otomatis meningkatkan nilai prediksi *rating* ketika ulasan pengguna menunjukkan kecenderungan positif.

Namun, variabel *helpful vote* (X_1) dengan koefisien 0,0644 menunjukkan bahwa semakin banyak ulasan yang dianggap membantu oleh pengguna lain, semakin besar film mendapatkan *rating* yang lebih baik. Pola ini dapat dilihat pada sebaran titik di area tengah grafik, dimana ulasan *helpful* yang lebih tinggi biasanya lebih dekat dengan *rating* aktual.

Sebaliknya, variabel *total vote* (X_2) memiliki koefisien negatif sebesar -0,0407 menunjukkan bahwa banyaknya suara secara keseluruhan tidak selalu berkorelasi dengan *rating* yang lebih tinggi. Hal ini dapat terjadi karena ulasan dengan banyak suara tidak selalu menunjukkan ulasan yang baik; pengguna sering memberikan suara tanpa membaca isi ulasan, atau film yang populer justru menarik lebih banyak ulasan negatif.



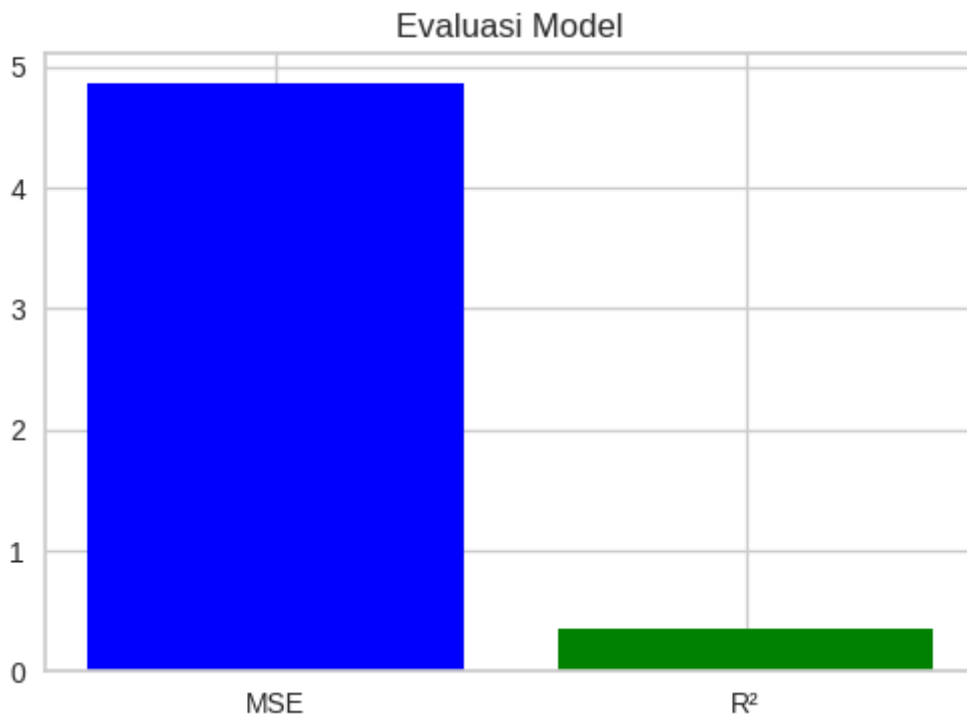
Gambar 3. Distribusi Residual

Sumber Gambar: Penulis

Pola yang menyerupai kurva lonceng atau kurva berbentuk *bell*, terlihat pada grafik distribusi residual, yang menunjukkan bahwa sisa galat yaitu jarak antara nilai aktual dan nilai prediksi, berdistribusi secara normal. Nilai residual sebagian besar berada diantar -2 dan +2 dengan puncak distribusi hampir nol. Pola ini menunjukkan bahwa kesalahan prediksi tersebar secara merata dan *model* tidak memiliki kecenderungan prediksi yang signifikan terhadap nilai tertentu. Hasil menunjukkan bahwa asumsi homoskedastisitas terpenuhi – kondisi dimana varian residual tetap konstan pada seluruh rentang nilai prediksi.

Selain itu, distribusi residual menunjukkan stabilitas *model*. Jika sebaran residual terlalu lebar, itu menunjukkan bahwa *model* belum mampu menangkap kompleksitas data secara optimal. Namun, dalam penelitian ini, sebaran residual yang relatif simetris menunjukkan bahwa *model* memiliki tingkat stabilitas yang baik dan tidak menunjukkan pola sistematis tertentu, sehingga layak untuk interpretasi prediktif.

Akibatnya, kedua grafik yang menunjukkan perbandingan antara nilai aktual dan prediksi serta distribusi residual, secara visual mendukung validitas *model* regresi linier berganda yang digunakan. Meskipun akurasi *model* belum sepenuhnya sempurna, pola hubungan linier yang konsisten dan distribusi residual yang hampir normal menunjukkan bahwa *model* sudah cukup representatif untuk menjelaskan hubungan antar variabel penelitian.



Gambar 4. Evaluasi Model

Sumber Gambar: Penulis

Nilai rata – rata kuadrat kesalahan (MSE) menunjukkan seberapa baik kemampuan *model* untuk menghasilkan prediksi yang lebih dekat dengan kondisi sebenarnya. Nilai MSE yang lebih rendah menunjukkan seberapa baik hubungan antara nilai aktual dan nilai prediksi. Kesalahan prediksi rata – rata *model* masih moderat, hasil MSE sebesar 4,86 dalam penelitian ini. Ini adalah nilai yang wajar mengingat data ulasan *IMDb* yang sangat beragam dan subjektif, dengan *rating* mulai dari 1 hingga 10. Karena elemen manusia seperti emosi, preferensi dan pengalaman dimodelkan secara deterministik, nilai MSE di digital seperti ulasan film dapat dikategorikan cukup baik.

Sementara itu, hasil nilai koefisien determinasi (R^2) sebesar 0,3486 mempengaruhi tiga variabel independen dalam *model – helpful vote*, *total vote* dan sentimen *review* – dapat menjelaskan sekitar 34,86% variasi nilai penilaian film atau hampir sepertiga dari variasi penilaian. Sisanya yaitu 65,14% dipengaruhi oleh variabel di luar *model*, seperti kualitas produksi film, popularitas aktor, strategi pemasaran dan faktor lainnya.

Hasil penelitian ini dapat dianggap cukup baik dan konsisten dengan temuan empiris sebelumnya. Ini karena meskipun nilai R^2 belum tergolong tinggi, temuan ini masih sejalan dengan karakteristik *model* linier pada data sosial yang tidak deterministik. Adanya perbedaan data yang belum sepenuhnya dapat dijelaskan oleh *model* ditunjukkan oleh perbedaan yang cukup besar antara nilai MSE dan R^2 seperti yang ditunjukkan pada **Gambar.4** karena ketimpangan pengaruh antar variabel. variabel sentimen *review* yang memiliki koefisien tinggi (6,8066) kemungkinan besar bertanggung jawab atas kondisi ini.

V. KESIMPULAN

Penelitian ini menemukan bahwa sebagai variabel dependen, *rating* film Spider-Man: Brand New Day dipengaruhi secara signifikan oleh variabel *helpful vote* (X_1), *total vote* (X_2) dan sentimen *review* (X_3). Data ini diperoleh melalui analisis regresi linier berganda dari data ulasan

film yang dikumpulkan dari platform *Kaggle*. Persamaan regresi berikut dihasilkan dari perhitungan *model*:

$$\gamma = 6.0887 + 0.0644x_1 - 0.0407x_2 + 6.8066x_3$$

Persamaan tersebut menunjukkan bahwa variabel sentimen *review* (X_3) memiliki koefisien sebesar 6,8066 yang menunjukkan bahwa ulasan dengan nada positif memiliki pengaruh terbesar terhadap peningkatan *rating*, dengan peningkatan satu satuan pada variabel *helpful vote* (X_1) meningkatkan *rating* film sebesar 0,0644 poin, sedangkan peningkatan satu satuan pada *total vote* (X_2) menurunkan *rating* sebesar 0,0407 poin. Hubungan ini menunjukkan bahwa persepsi emosional penonton memainkan peran utama dalam menentukan penerimaan publik terhadap film. Di sisi lain, tingkat partisipasi pengguna yang ditunjukkan dengan suara dan penilaian juga memperkuat validitas sosial dari penilaian tersebut.

Sekitar 34,86% variasi nilai *rating* film dapat dijelaskan oleh *model*, menurut nilai *Mean Squared Error* (MSE) sebesar 4,86 dan nilai *R-squared* (R_2) sebesar 0,3486. Faktor – faktor eksternal seperti kualitas produksi, daya tarik aktor, strategi pemasaran dan nilai nostalgia terhadap karakter membentuk 65,14% variasi. Meskipun akurasi *model* masih dianggap moderat, nilai MSE yang rendah menunjukkan bahwa tingkat kesalahan prediksi berada dalam batas yang wajar untuk data sosial yang subjektif.

Hasil visualisasi menunjukkan bahwa ada hubungan linier positif antara nilai aktual dan nilai prediksi; sebagian besar titik data terkonsentrasi di sekitar garis regresi. Penemuan ini mendukung hasil sebelumnya bahwa *model* dapat secara konsisten mengidentifikasi pola hubungan antara variabel dependen dan independen. Selain itu, distribusi residual yang mirip dengan kurva normal menunjukkan bahwa asumsi klasik regresi telah terpenuhi dan *model* tidak menunjukkan bias yang signifikan.

Secara keseluruhan, penelitian ini menegaskan bahwa peringkat atau *rating* film sebagai indikator keberhasilan dipengaruhi oleh interaksi antara faktor kognitif (*helpful vote* dan *total vote*) dan faktor afektif (sentimen *review*) yang tercermin dalam perilaku pengguna internet. Menurut *model* yang dibuat, persepsi publik terhadap kualitas film dapat diprediksi dengan menggabungkan aspek partisipasi dan emosi. Jadi penelitian ini membantu membangun literatur tentang analitik prediktif di industri hiburan dan memberi pelaku industri dasar empiris untuk membuat strategi promosi dan distribusi berbasis data yang lebih baik.

VII. REFERENSI

- Bigne, E., Ruiz, C., Cuenca, A., Perez, C., & Garcia, A. (2021). What drives the helpfulness of online reviews? A deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. *Journal of Destination Marketing and Management*, 20, 100570. <https://doi.org/10.1016/j.jdmm.2021.100570>
- Cholilawati, & Suliyanthini, D. (2022). Perubahan Perilaku Konsumen Selama Pandemi Covid-19. *Equilibrium: Jurnal Pendidikan*, 9(1), 18–24. <https://journal.unismuh.ac.id/index.php/equilibrium/index>
- Dau Bere, N., Laura Dekafrio, A., Kartika Kusmayati, N., Kurniawati, Y., Ekonomi dan Program Studi Manajemen, F., & Mahardhika, S. (2024). Pengaruh Media Sosial Terhadap Perilaku Konsumen Dikalangan Generasi Milenial Di Surabaya. *TEKNOBIS: Teknologi, Bisnis Dan Pendidikan*, 2(3), 430–438. <https://jurnalmahasiswa.com/index.php/teknobis>

- ISKANDI, I., Yassir, L., & Badia, E. H. (2023). The Symbol of the Spider in Spider-Man Film: Cultural Implications and Ideological Dimensions. *MEDIOVA: Journal of Islamic Media Studies*, 3(1), 85–101. <https://doi.org/10.32923/medio.v3i1.3325>
- Kim, A., Trimi, S., & Lee, S. G. (2021). Exploring the key success factors of films: a survival analysis approach. *Service Business*, 15(4), 613–638. <https://doi.org/10.1007/s11628-021-00460-x>
- Ramos, J., Apóstolo, D., & Oliveira, H. G. (2022). Movie rating prediction using sentiment features. *2nd Workshop on Sentiment Analysis and Linguistic Linked Data, SALLD 2022 - Held in Conjunction with the International Conference on Language Resources and Evaluation, LREC 2022 - Proceedings, June, 9–18*. <https://aclanthology.org/2022.salld-1.3/>
- Saumya, S., Roy, P. K., & Singh, J. P. (2023). Review helpfulness prediction on e-commerce websites: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 126, 107075. <https://doi.org/10.1016/J.ENGAPPAL.2023.107075>
- Steinke, I., Wier, J., Simon, L., & Seetan, R. (2022). Sentiment Analysis of Online Movie Reviews using Machine Learning. *International Journal of Advanced Computer Science and Applications*, 13(9), 618–624. <https://doi.org/10.14569/IJACSA.2022.0130973>
- Uwase N., M. (2025). Consumer Behavior in the Digital Age. *Iaa Journal of Art and Humanities*, 12(1), 77–83. <https://doi.org/10.59298/iaajah/2025/1217783>
- Wang, X., Guo, P., Li, X., Gangopadhyay, A., Busart, C. E., Freeman, J., & Wang, J. (2023). Reproducible and Portable Big Data Analytics in the Cloud. *IEEE Transactions on Cloud Computing*, 11(3), 2966–2982. <https://doi.org/10.1109/TCC.2023.3245081>
- Wratsari, E. T. A. W., Junaedi, M. S., & Mahestu Noviandra Krisjanti. (2022). Online Streaming Services Uses During The COVID-19 Pandemic in Indonesia. *Kinerja*, 26(2), 306–320. <https://doi.org/10.24002/kinerja.v26i2.6068>
- Yves Gambier, H. J. (2024). *Chinese Films Abroad Distribution and Translation* (I. I. J. Rifka Alkhilyatul Ma'rifat, I Made Suraharta (ed.); Vol. 2). https://www.routledge.com/Chinese-Films-Abroad-Distribution-and-Translation/Gambier-Jin/p/book/9781032739410?srsId=AfmBOoqqhWJfAstQLmt-bzpJzOul4ISNBxEdQwzLZSw90dFsZR8n_6KK