



ANALISIS PREDIKSI HARGA STEAM GAME

Gilang Wijanarko

Email : gilangwjnrk@gmail.com

Whatsapp : +6288232217249

STEAM GAME

Steam platform

Steam adalah platform distribusi digital untuk video game yang dikembangkan oleh Valve Corporation. Secara ringkas, Steam memungkinkan pengguna untuk:

- Membeli dan mengunduh game secara online
- Bermain bersama teman lewat fitur multiplayer dan komunitas
- Mengakses fitur tambahan seperti forum diskusi, Steam Workshop, dan update otomatis
- Menyimpan koleksi game dalam satu tempat tanpa perlu CD fisik

Steam menjadi semacam “toko game virtual” terbesar di dunia, dengan ribuan judul dari game indie hingga AAA, dan sering menawarkan diskon besar-besaran.

IKTISAR DATA

Dataset dapat diperoleh pada:
<https://www.kaggle.com/datasets/benjaminlundkvist/steam-sales-historical-dataset/code>

Jumlah kolom : 11

Jumlah baris : 1108 (jumlah game)

Dataset ini berisikan data penjualan steam game yang terdiri dari rating, review, diskon, platform rilis, harga original dan harga setelah diskon

Memahami karakteristik dan pola harga setelah diskon dapat membantu dalam memproyeksikan **tren perubahan harga** berdasarkan fitur yang tersedia.

STRUKTUR DATA

Label

Price (€) : Harga setelah diskon yang dipengaruhi oleh pricing dan reviews (float4).

Fitur

- **Game Name** : Nama gim yang menentukan jumlah baris data (object).
- **Rating** : Rata-rata penilaian oleh pengguna (float64).
- **#Reviews** : Jumlah reviews/penilaian yang diberikan pada game (object).
- **Discount%** : Besarnya potongan harga yang mempengaruhi harga jual (float64).
- **Original Price (€)** : Harga awal yang ditentukan pengembang (float64).
- **Release Date** : Waktu rilis gim (Object).
- **Windows** : Rilis di Windows (Ya/Tidak) (int64).
- **Linux** : Rilis di Linux (Ya/Tidak) (int64).
- **MacOS** : Rilis di MacOS (Ya/Tidak) (int64).
- **Fetches at** : Waktu perekaman data dari API/database steam. Waktu perekaman dilakukan pada 8 Agustus 2025, 11 Agustus 2025, dan 2 September 2025 (object).

TUJUAN ANALISI

Analisis regresi linear adalah jenis analisis yang memperhitungkan perubahan variable X (independent) yang mempengaruhi variable Y (dependen). Analisis ini berguna untuk mengukur tren dan pola yang terbentuk dari model bisnis yang dijalankan.

Analisi ini dilakukan untuk menjawab pertanyaan, seperti:

- **Memprediksi** perubahan pada harga jual gim pada platform steam

Tools dan library yang digunakan dalam analisis ini terdiri dari:



1. HANDLING DATA

Pricing pada steam sangat dipengaruhi oleh region. Hal ini disebabkan perbedaan kurs, daya beli dan kondisi pasar lokal. Perbedaan harga antar region dapat berpengaruh pada analisis data penjualan steam sejenis.

	Game Name	Rating	#Reviews	Discount%	Price (€)	Original Price (€)	Release Date	Windows	Linux	MacOS	Fetches At
0	PEAK	7.0	77,683	-38.0	4.64	7.49	16 Jun, 2025	1	0	0	2025-08-11 00:00
1	Cyberpunk 2077	7.0	772,574	-65.0	20.99	59.99	9 Dec, 2020	1	0	1	2025-08-11 00:00
2	Rust	7.0	1,038,433	-50.0	19.99	39.99	8 Feb, 2018	1	0	1	2025-08-11 00:00
3	Microsoft Flight Simulator (2020) 40th Anniver...	5.0	63,905	-60.0	27.99	69.99	17 Aug, 2020	1	0	0	2025-08-11 00:00
4	Sid Meier's Civilization VII	4.0	34,643	-30.0	48.99	69.99	10 Feb, 2025	1	1	1	2025-08-11 00:00

Rating pada game merupakan angka bulat yang menggambarkan pembulatan terdekat. Rating berada pada rentang 0 – 8 berdasarkan konversi ke skala numerik. Umumnya steam menggunakan rating seperti “Overwhelmingly Positive”, “Mixed”, “Very Negative”, dan lain-lain.

Data uang tersedia pada platform Kaggle.com senantiasa terus diperbarui dengan data tambahan per 2 September 2025.

1. HANDLING DATA

```
[ ] df.duplicated().sum()
```

```
np.int64(0)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1108 entries, 0 to 1107
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Game Name             1108 non-null   object
1   Rating                1108 non-null   float64
2   #Reviews              1108 non-null   object
3   Discount%             1108 non-null   float64
4   Price (€)             1108 non-null   float64
5   Original Price (€)    1108 non-null   float64
6   Release Date          1108 non-null   object
7   Windows               1108 non-null   int64
8   Linux                 1108 non-null   int64
9   MacOS                 1108 non-null   int64
10  Fetched At            1108 non-null   object
dtypes: float64(4), int64(3), object(4)
memory usage: 95.3+ KB
```

Cek Duplikat

Jumlah data duplikat = 0, sehingga tidak ada data yang di-drop.

Hapus Kolom

Kolom yang dihapus dalam analisis ini terdiri dari:

- Game Name = terdiri dari nama gim yang berjumlah 1108 data.
- Release Date = analisis membutuhkan analisis dengan fitur numerik yang berpengaruh langsung pada harga akhir gim.
- Fetched At = tidak berpengaruh secara langsung pada perubahan pricing gim. Fetched At berpengaruh pada banyaknya data yang digunakan dan keterbaruan data.

Format Data

Perubahan pada tipe data:

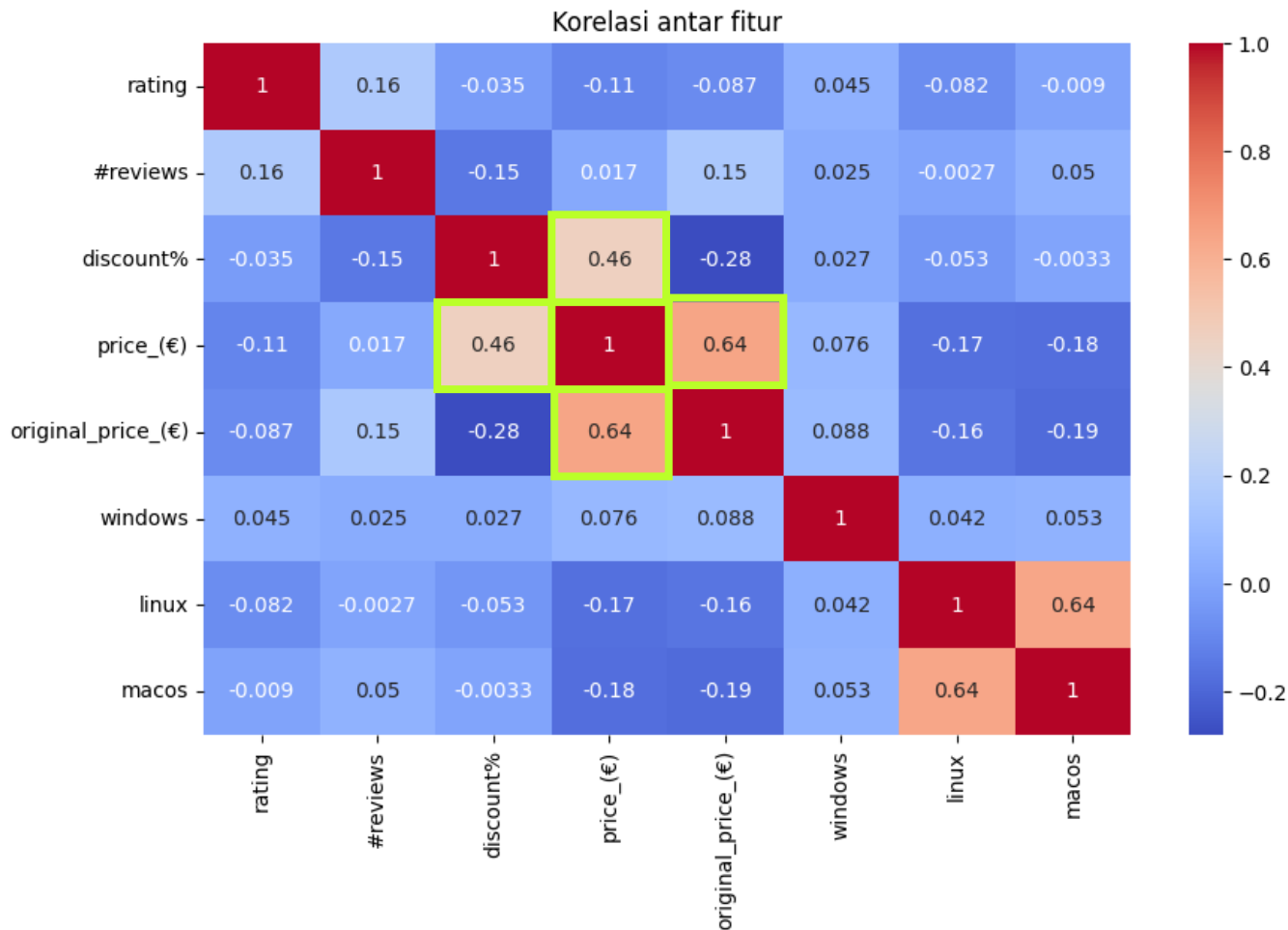
- #Reviews = Data berupa objek yang masih dapat dijumpai spasi " " dan koma "," sehingga perlu untuk diubah terlebih dahulu menjadi int64.

2. EXPLANATORY DATA ANALYSIS

	rating	#reviews	discount%	price_(€)	original_price_(€)	windows	linux	macos
count	1108.000000	1.108000e+03	1108.000000	1108.000000	1108.000000	1108.000000	1108.000000	1108.000000
mean	6.446751	2.536044e+04	-51.344765	12.561715	27.860397	0.994585	0.242780	0.337545
std	1.291994	7.387433e+04	22.493165	8.695701	15.682801	0.073421	0.428957	0.473086
min	3.000000	1.300000e+01	-95.000000	0.990000	2.990000	0.000000	0.000000	0.000000
25%	5.000000	1.291000e+03	-75.000000	5.990000	17.490000	1.000000	0.000000	0.000000
50%	7.000000	4.740500e+03	-50.000000	9.990000	24.990000	1.000000	0.000000	0.000000
75%	7.000000	1.963700e+04	-30.000000	15.990000	39.990000	1.000000	0.000000	1.000000
max	8.000000	1.038433e+06	-10.000000	59.990000	99.990000	1.000000	1.000000	1.000000

Harga akhir (price (€)) memiliki rata-rata €12.561715 dengan median €9.99. Hal ini menandakan banyak gim dengan harga besar yang menggeser nilai rata-rata ke angka yang besar (**skewed right**). Beberapa gim bahkan memiliki harga sebesar €59.99. Harga gim tidak simetris, banyak gim yang memiliki harga murah dan sedikit gim dengan harga yang sangat mahal. Banyak dijumpai outlier gim dengan harga mahal dari rata-rata harga gim.

2. EXPLANATORY DATA ANALYSIS



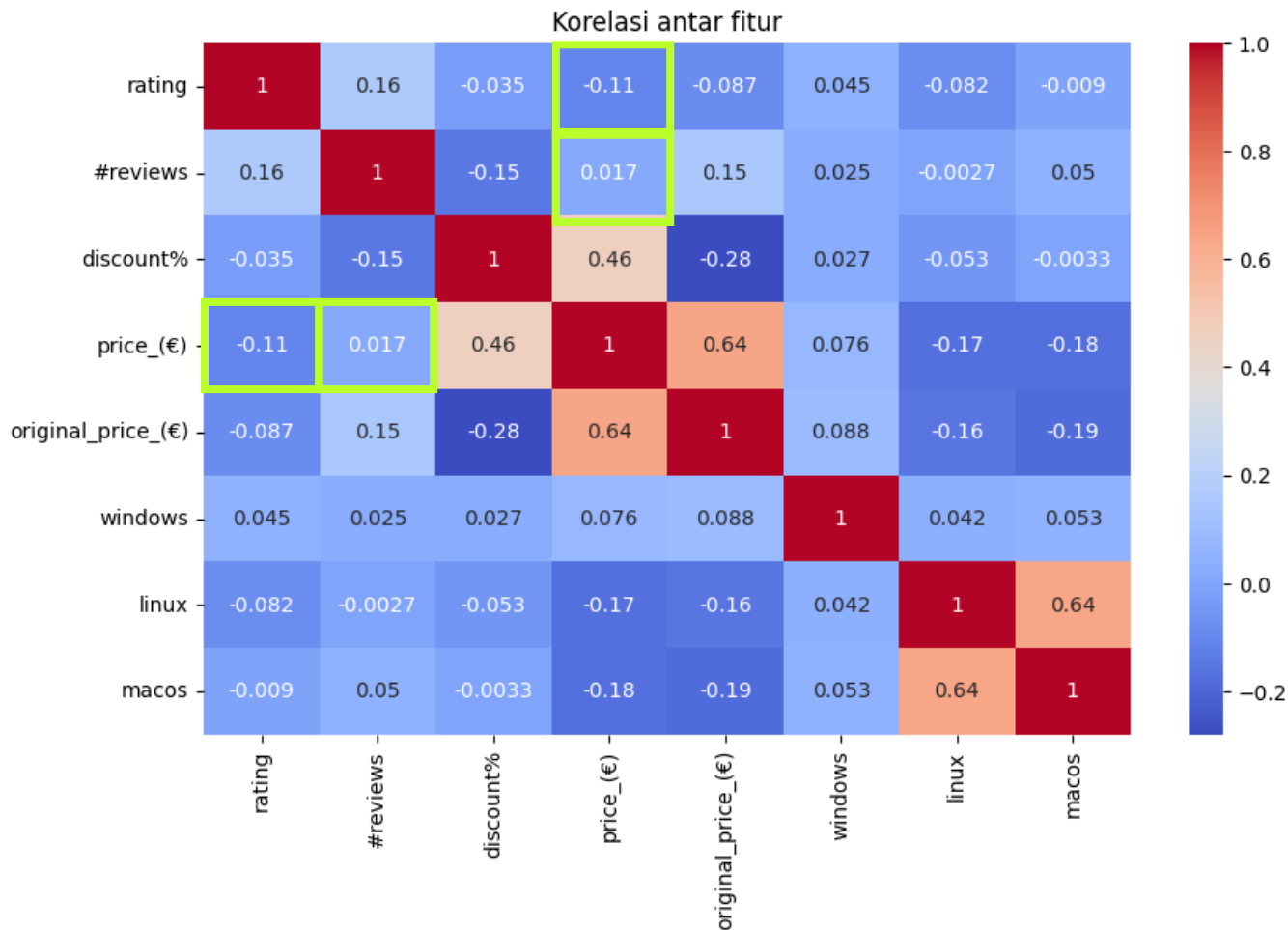
Korelasi Heatmap

Beberapa temuan dari hubungan antar korelasi fitur dan label di antaranya:

Original Price (€) → Price (€) Terdapat korelasi positif yang cukup kuat sebesar 0.64, menunjukkan bahwa **harga awal sangat memengaruhi harga akhir**. Artinya, gim yang awalnya mahal cenderung tetap memiliki harga akhir yang tinggi, meskipun sudah didiskon.

Discount (%) → Price (€) Korelasi negatif sebesar -0.46 mengindikasikan bahwa **semakin besar diskon yang diberikan, semakin rendah harga akhir gim**. Walaupun nilai diskon ditulis sebagai angka positif, arah hubungannya berlawanan dengan harga akhir

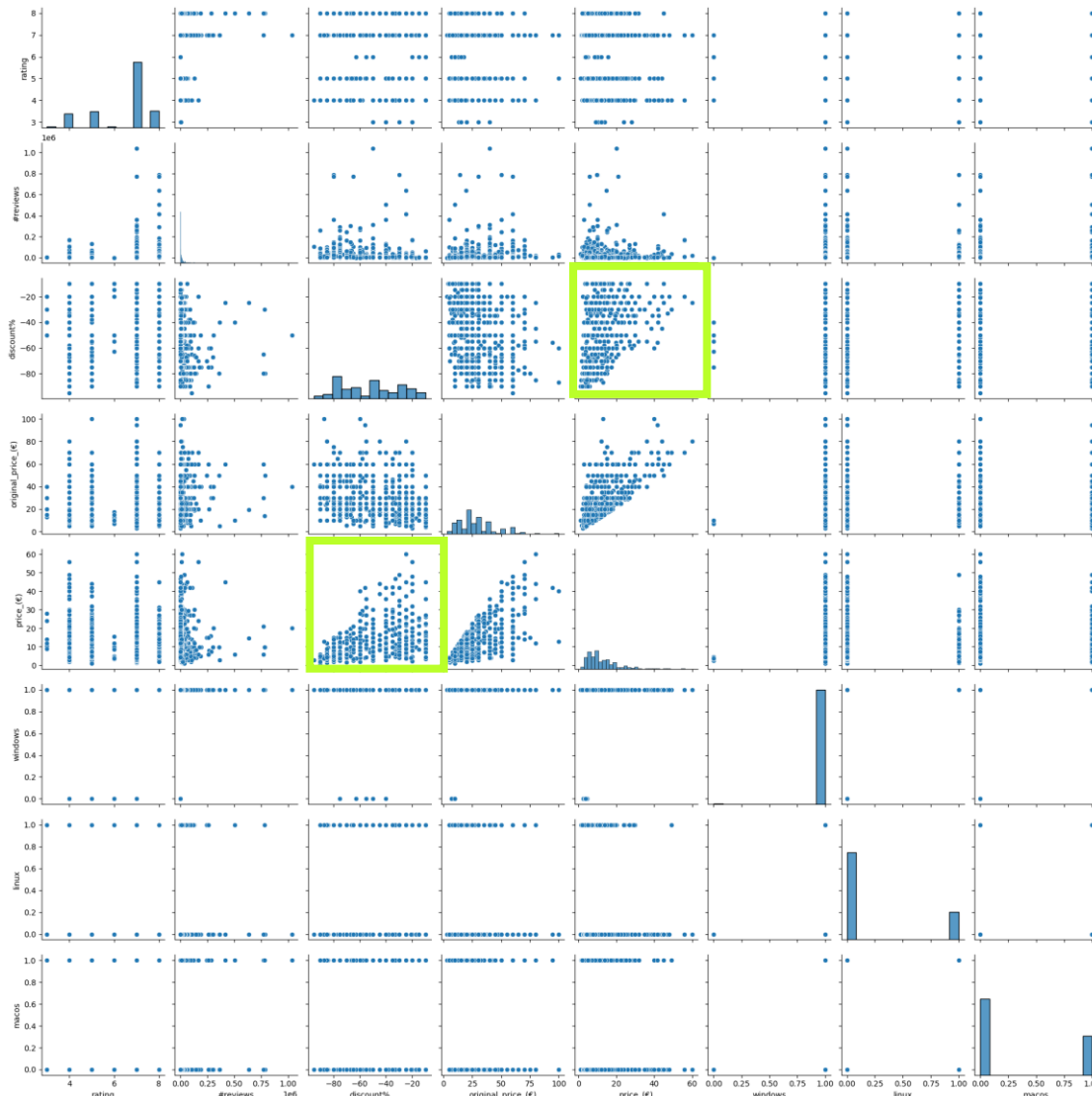
2. EXPLANATORY DATA ANALYSIS



#Reviews → Price (€) Korelasi sangat lemah sebesar 0.017, artinya **jumlah ulasan tidak berpengaruh signifikan terhadap harga gim**. Banyaknya review lebih mencerminkan popularitas dan pengalaman pemain, bukan penentu harga.

Rating → Price (€) Korelasi negatif lemah sebesar -0.11 menunjukkan bahwa **gim dengan rating tinggi belum tentu memiliki harga mahal**. Preferensi pemain terhadap genre, franchise, atau gaya permainan bisa membuat gim berkualitas tetap dijual dengan harga terjangkau.

2. EXPLANATORY DATA ANALYSIS



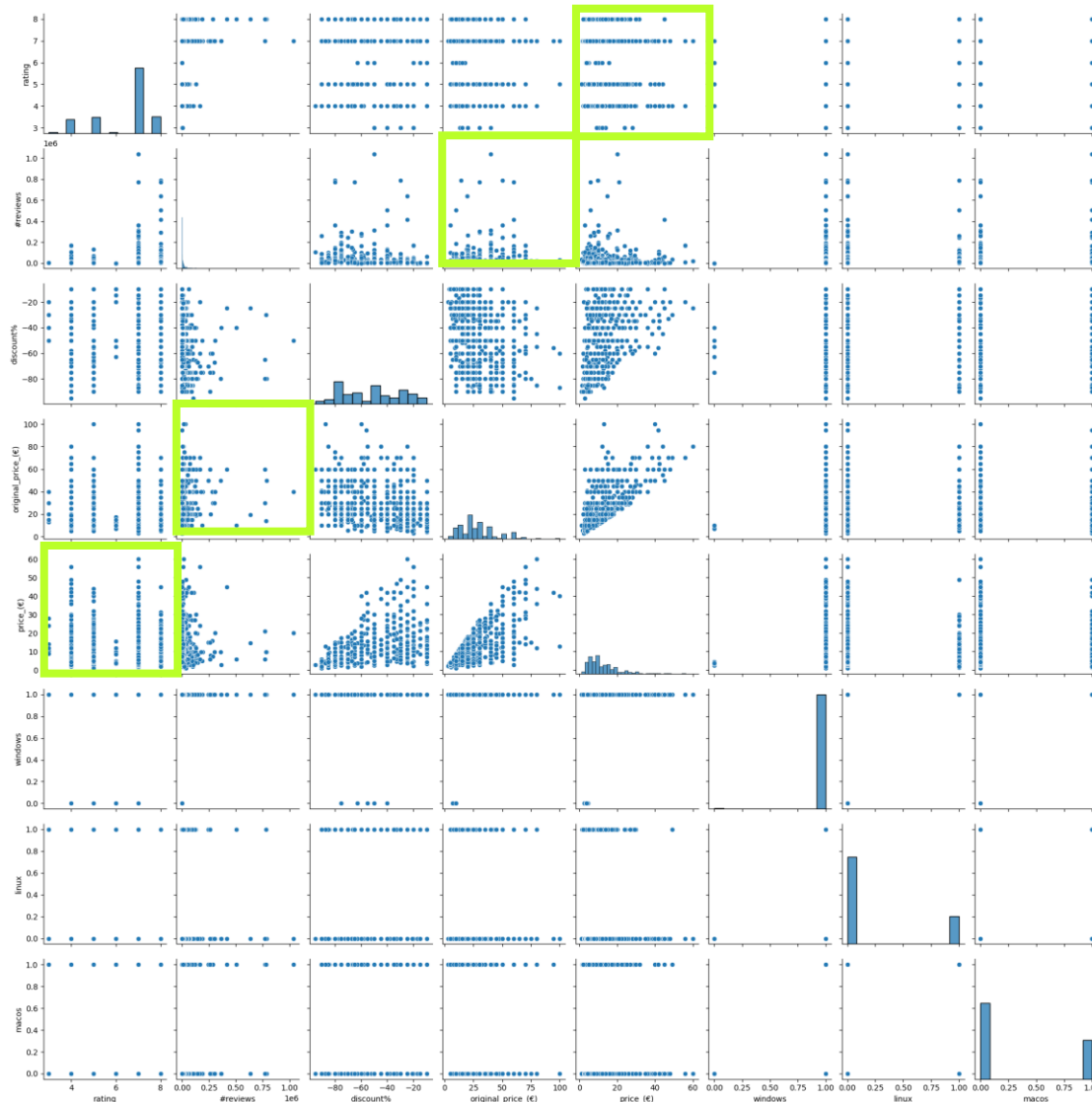
Pairplot

Beberapa temuan dari perbandingan pairplot di antaranya:

Original Price (€) dan Discount (%) menunjukkan **hubungan linear yang jelas** terhadap Price (€).

- Harga akhir gim cenderung meningkat seiring naiknya harga awal.
- Diskon yang lebih besar secara konsisten menurunkan harga akhir. Kedua fitur ini membentuk pola garis miring pada scatter plot, menandakan korelasi linier yang kuat.

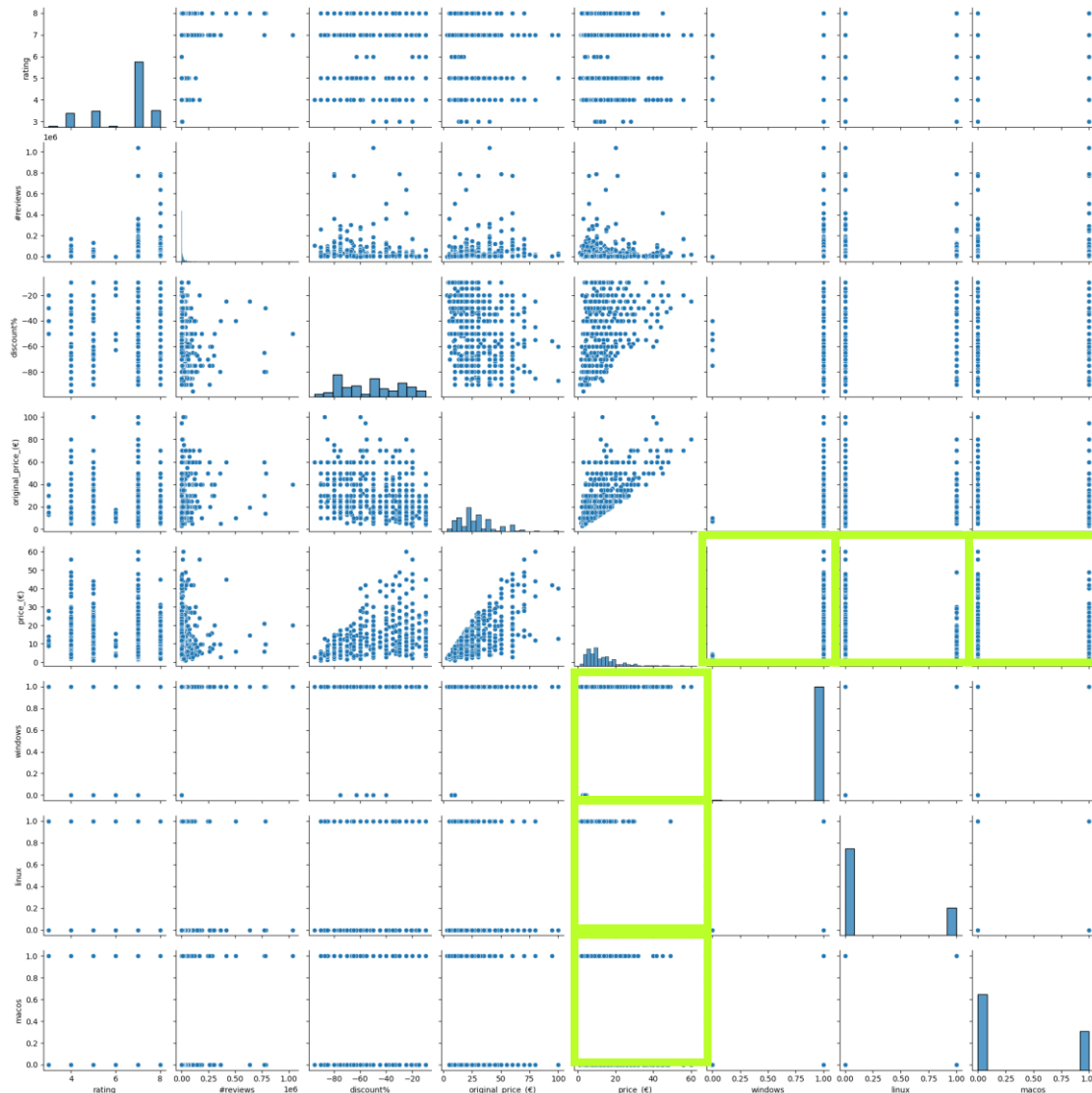
2. EXPLANATORY DATA ANALYSIS



Rating dan Jumlah Reviews **tidak menunjukkan pola linear** yang signifikan terhadap harga.

- Titik-titik pada scatter plot terlihat menyebar acak, tanpa arah yang jelas.
- Hal ini mengindikasikan bahwa kualitas atau popularitas gim (berdasarkan ulasan) tidak selalu menentukan harga.

2. EXPLANATORY DATA ANALYSIS



Platform Rilis (Windows, Linux, MacOS) juga **tidak menunjukkan pola linear** yang konsisten terhadap harga.

- Karena fitur ini bersifat biner (0/1), pola hubungan sulit dibaca secara linier.
- Dukungan platform lebih mencerminkan aspek teknis atau segmentasi pasar, bukan penentu harga langsung.

3. ANALISIS PREDIKTIF

```
[ ] features = ['rating', '#reviews', 'discount%', 'original_price_(€)', 'windows', 'linux', 'macos']  
    target = 'price_(€)'
```

```
[ ] x = df[features]  
    y = df[target]
```

```
[ ] x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=42)
```

```
[ ] model = LinearRegression()  
    model.fit(x_train, y_train)  
    y_pred = model.predict(x_test)
```

Dalam membentuk model prediktif digunakan fitur seperti: rating, reviews, diskon, harga original, dan platform rilis (Windows, Linux, MacOS). Target yang ingin diprediksi adalah harga akhir/price (€).

Split data training set dan test set dilakukan pada perbandingan **80:20**. Jumlah data training cukup untuk data yang tidak terlalu banyak (1108 baris data) dengan test set yang optimal.

Pemodelan dilakukan menggunakan **regresi linear** yang menunjukkan peningkatan harga berdasarkan perubahan pada fitur lain.

3. ANALISIS PREDIKTIF

MSE: 9.79
RMSE: 3.13
MAE: 2.29
MAPE: 29.06%

Mean Absolute Error (MAE)
sebesar €2.29, prediksi cenderung stabil dan tidak terpengaruh secara signifikan oleh outlier.

Root Mean Square Error (RMSE)
sebesar €3.13. Kesalahan prediksi cenderung ada pada dengan harga yang tinggi (sebab data skewed right)

Mean Absolut Percentage Error (MAPE) sebesar 29,06% dari harga asli.

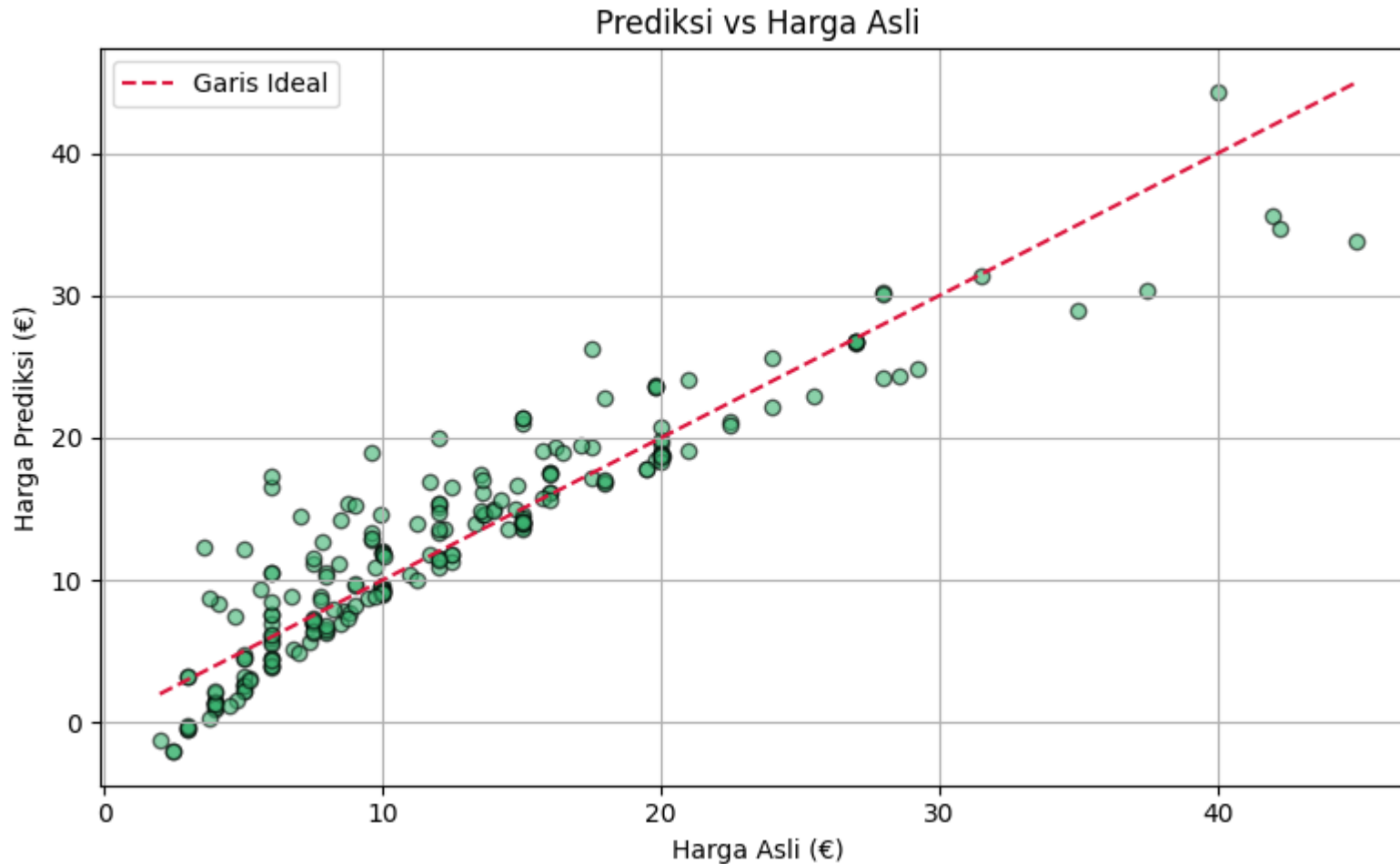


Model lebih stabil pada gim dengan harga yang lebih rendah/dibawah median harga.

Kurang optimal pada gim harga mahal, sebab pada harga besar semakin menjauh dari model linear.

Penggunaan model non linear yang dapat meningkatkan akurasi prediksi.

3. ANALISIS PREDIKTIF



Berdasarkan pemodelan prediksi harga gim pada platform steam, diketahui sebaran harga ada pada gim dengan harga terjangkau. Model membentuk garis linear yang menandakan peningkatan pada fitur lain mempengaruhi prediksi harga.

Model prediksi masih belum dapat memperkirakan gim premium dengan harga yang besar (lebih dari €30).

KESIMPULAN

1. Pengaruh harga awal

Harga awal rilis sangat mempengaruhi besarnya harga akhir yang dapat dibeli pemain. Pemain dapat melihat harga original rilis pada pengumuman launching game ataupun melalui komunitas gim.

2. Diskon Game

Pemberian diskon gim sangat berpengaruh pada harga akhir pada gim yang bisa dibeli. Semakin tinggi diskon maka harga semakin murah (perbandingan terbalik)

3. Platform, rating, dan reviews

Pemilihan platform tidak berdampak secara langsung pada prediksi harga, sebab platform berkaitan pada segmentasi pasar. Sedangkan rating dan reviews berkaitan pada popularitas gim dan pengalaman pemain.



REKOMENDASI

Pemodelan regresi linear mampu menghasilkan prediksi harga gim yang **cukup akurat**, terutama pada rentang harga rendah hingga menengah. Hubungan linier antara fitur seperti `original_price_` (€) dan `discount%` terhadap harga akhir memungkinkan model menangkap pola dasar dengan baik.

Namun, untuk meningkatkan akurasi dan mengurangi kesalahan prediksi, terutama pada gim dengan harga tinggi, pemilihan model alternatif seperti **Random Forest** atau **Decision Tree** dapat dipertimbangkan. Model-model ini lebih fleksibel dalam menangani hubungan non-linier dan lebih tahan terhadap outlier.

Distribusi harga gim yang skewed ke kanan, dengan sebagian kecil gim berharga tinggi, menyebabkan outlier yang dapat memengaruhi performa model. Untuk mengatasi hal ini, pendekatan seperti **transformasi logaritmik** pada harga dapat digunakan agar distribusi lebih normal dan model lebih stabil.





**TERIMA KASIH
ATAS
PERHATIANNYA!**

Hubungi Saya di :

[www.linkedin.com/in/gilang-
wijanarko-1b6b4b16b](https://www.linkedin.com/in/gilang-wijanarko-1b6b4b16b)