

Analisis Multivariat Korelasi Antara Durasi Film, Rating, dan Keuntungan Box Office Di Marvel Cinematic Universe

Wicaksono Yuli Sulistyono^{*1}, Panggah Widiandana², Muhammad Immawan Aulia³,
Muhamad Rosidin⁴

^{1,4}Prodi PJJ Sistem Informasi, Universitas Siber Muhammadiyah

^{2,3}Prodi PJJ Informatika, Universitas Islam Mulia

E-mail: *¹wicaksono@sibermu.ac.id, ²panggah.widiandana@uim-yogya.ac.id,
³muhimmawanaulia16@uim-yogya.ac.id, ⁴rosidin@sibermu.ac.id

Abstrak

Analisis Multivariat merupakan metode statistik yang digunakan untuk menganalisis data yang terdiri dari banyak variabel. Marvel Cinematic Universe (MCU) telah merilis puluhan film yang sukses secara komersial dan mendapatkan rating tinggi dari penonton dan kritikus. Hal tersebut membuat MCU menjadi subjek yang menarik untuk dianalisis menggunakan metode analisis multivariat. Penelitian ini bertujuan untuk menganalisis korelasi antara durasi film, rating, dan keuntungan box office dalam MCU menggunakan metode analisis multivariat. Hasil analisis menunjukkan bahwa durasi film dan rating memiliki korelasi positif yang kuat, yang berarti film dengan durasi yang lebih panjang cenderung memiliki rating yang lebih tinggi. Selain itu, rating dan keuntungan box office juga memiliki korelasi positif yang menunjukkan bahwa film dengan rating yang lebih tinggi cenderung menghasilkan keuntungan box office yang lebih besar. Penelitian ini menegaskan pentingnya kualitas konten dalam menarik audiens dan meningkatkan pendapatan box office, dan dapat digunakan oleh pembuat film dan distributor untuk membuat keputusan yang lebih baik tentang durasi film, rating, dan strategi pemasaran.

Kata kunci—Box Office, Korelasi, Marvel;

1. PENDAHULUAN

Perkembangan teknologi sekarang semakin berkembang maju dan meningkat dari waktu ke waktu [1][2]. Salah satunya adalah kebutuhan akan metode analisis data yang lebih kompleks dan mendalam semakin meningkat. Analisis Multivariat merupakan metode statistik yang digunakan untuk menganalisis data yang terdiri dari banyak variabel serta diduga antar variabel tersebut saling berhubungan satu sama lain. Metode ini telah dikembangkan oleh berbagai ahli sebagai metode analisis data kuantitatif sejak abad ke-19 [3]. Analisis Multivariat dapat digunakan di dunia film yang digunakan untuk menganalisis hubungan antara berbagai variabel seperti durasi film, rating, dan keuntungan box office. Variabel-variabel ini saling terkait dan mempengaruhi satu sama lain, sehingga analisis multivariat dapat memberikan pemahaman yang lebih mendalam tentang dinamika industri film [4].

Marvel Cinematic Universe (MCU) merupakan salah satu *franchise* film terbesar dan paling sukses di dunia. MCU telah merilis puluhan film yang sukses secara komersial dan mendapatkan rating tinggi dari penonton dan kritikus. Hal tersebut membuat MCU menjadi subjek yang menarik untuk dianalisis menggunakan metode analisis multivariat [5].

Durasi film adalah salah satu faktor yang dapat mempengaruhi keberhasilan film. Durasi film dapat mempengaruhi penilaian penonton dan keuntungan *box office*. Namun, hubungan antara durasi film dan keberhasilan film di MCU belum banyak diteliti [6]. Rating film juga merupakan

faktor penting yang dapat mempengaruhi keberhasilan film. Rating film dapat mencerminkan kualitas film dan mempengaruhi keputusan penonton untuk menonton film tersebut². Analisis multivariat dapat digunakan untuk menganalisis hubungan antara rating film dan keberhasilan film di MCU [7]. Keuntungan *Box Office* adalah keuntungan dalam pembuatan film. Besaran keuntungan dapat dipengaruhi oleh berbagai faktor, termasuk durasi film dan pemeran. Analisis multivariat dapat memberikan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi Keuntungan *Box Office* di MCU [8].

Penelitian ini akan menggunakan data primer dan skunder untuk menganalisis hubungan antara durasi film, rating film, dan Keuntungan *Box Office* di MCU. Data ini akan dianalisis menggunakan metode analisis multivariat, yang memungkinkan peneliti untuk menganalisis hubungan antara variabel-variabel ini secara simultan dan serentak. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang dinamika industri film, khususnya MCU. Hasil ini juga dapat digunakan oleh pembuat film dan distributor untuk membuat keputusan yang lebih baik tentang durasi film, rating film, dan strategi pemasaran.

2. METODE PENELITIAN

Tahapan penelitian yang dilakukan menggunakan alur Data Science yang digambar seperti pada Gambar 1.



Gambar 1. Alur Kerja Data Science [9]

Gambar 1 menunjukkan alur kerja data science yang dimulai dari *capture data* dan *clean data* kemudian analisis dan berakhir di bagian *reporting* atau visualisasi.

2.1 Pengumpulan Data (*Capture Data*)

Data untuk penelitian ini dikumpulkan dari berbagai sumber yang dapat diandalkan dan relevan dengan Marvel Cinematic Universe (MCU). Data tersebut meliputi durasi film, rating film, dan Keuntungan *Box Office*. Sumber data didapat melalui Kaggle (<https://www.kaggle.com/datasets/ummalma/mcu-movies-and-tv-shows>)

2.2 Pembersihan Data (*Manage and Data*)

Setelah data dikumpulkan, langkah selanjutnya adalah pembersihan data. Proses ini melibatkan pengecekan dan penanganan data yang hilang, penghapusan data duplikat, dan konversi data ke format yang tepat. Tujuannya adalah untuk memastikan bahwa data siap untuk analisis [10].

2.3 Eksplorasi Data (*Exploratory Analysis*)

Eksplorasi data dilakukan untuk memahami karakteristik dan pola dalam data. Ini melibatkan analisis statistik deskriptif, visualisasi data, dan pengujian asumsi statistik. Hasil dari eksplorasi data akan membantu dalam pemilihan metode analisis yang tepat [11].

2.4 Analisis Akhir (Final Analysis)

Tahap ini, analisis multivariat akan dilakukan untuk menganalisis hubungan antara durasi film, rating film, dan keuntungan *box office*. Metode yang digunakan dapat berupa regresi linier berganda, analisis jalur, atau metode lain yang sesuai. Setelah model dibuat, evaluasi model dilakukan untuk memastikan bahwa model tersebut valid dan dapat dipercaya. Ini melibatkan pengujian hipotesis, pengecekan kebaikan fit model, dan validasi silang.

2.5 Laporan (Reporting)

Hasil dari analisis akan disajikan dalam bentuk laporan atau visualisasi penelitian. Laporan ini akan mencakup metodologi penelitian, hasil analisis, interpretasi hasil, dan rekomendasi untuk penelitian lebih lanjut. Visualisasi data akan sangat menunjang sebuah analisa data yang mana kedua hal tersebut saling bergantung satu sama lain, karena sebuah analisa data akan lebih mudah jika di visualisasikan dalam bentuk grafik atau tabel dari data yang sudah di Analisa [12].

Alur Data Science ini diharapkan dapat memberikan pemahaman yang lebih mendalam tentang hubungan antara durasi film, rating film, dan Keuntungan *Box Office* di MCU. Selain itu, hasil penelitian ini juga diharapkan dapat digunakan oleh pembuat film dan distributor untuk membuat keputusan yang lebih baik tentang durasi film, rating film, dan strategi pemasaran.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Hasil penelitian ini menjelaskan secara detail berdasarkan langkah-langkah dalam analisis Data Science. Tahapan pertama merupakan tahapan pengumpulan data yang digunakan dalam analisis adalah dataset dari Marvel Cinematic Universe (MCU). Dataset ini mencakup 66 entri dengan total 23 kolom, yang berisi informasi seperti judul film, tanggal rilis, pendapatan *box office*, durasi film, rating, dan beberapa variabel lainnya. Setiap kolom dalam dataset memiliki tipe data yang bervariasi, termasuk objek, integer, dan float, yang mencerminkan berbagai jenis informasi yang relevan untuk analisis korelasi antara durasi film, rating, dan keuntungan *box office* seperti pada Gambar 2.

```

movies_data.info() #melihat info data
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 23 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Movie                                     66 non-null     object
1   Release Date                             66 non-null     object
2   Release Month                            66 non-null     object
3   Release Day                              66 non-null     int64
4   Release Year                             66 non-null     int64
5   Ownership                                 66 non-null     object
6   Domestic Box Office                      66 non-null     int64
7   Inflation Adjusted Domestic              66 non-null     int64
8   International Box Office                 66 non-null     int64
9   Inflation Adjusted International         66 non-null     float64
10  Worldwide Box Office                     66 non-null     int64
11  Inflation Adjusted Worldwide             66 non-null     float64
12  Opening Weekend                          66 non-null     int64
13  Inflation Adjusted Opening Weekend       66 non-null     int64
14  Budget                                   66 non-null     int64
15  Inflation Adjusted Budget                66 non-null     int64
16  IMDb Score                               66 non-null     float64
17  Meta Score                               66 non-null     int64
18  Tomatometer                             66 non-null     int64
19  Rotten Tomato Audience Score            66 non-null     int64
20  Run Time In Minutes                     66 non-null     int64
21  Phase                                    33 non-null     object
22  Director                                 66 non-null     object
dtypes: float64(3), int64(14), object(6)
memory usage: 12.0+ KB

```

id	judul	tahun	genre	director	box_office	run_time	imd_b_score
0	Iron Man	2008	Action	Jon Favreau	585171547	126	7.9
1	The Incredible Hulk	2008	Action	Louis Leterrier	265573859	112	6.6
2	Iron Man 2	2010	Action	John Dahl	621156389	124	6.9
3	Thor	2011	Action	Joss Whedon	449326518	115	7.0
4	Captain America: The First Avenger	2011	Action	Joe Johnston	370569776	124	6.9
5	The Avengers	2011	Action	Joss Whedon	1515100211	143	8.0
6	Iron Man 3	2013	Action	Shane Black	1215392272	130	7.1
7	Thor: The Dark World	2013	Action	Joss Whedon	644602516	112	6.8
8	Captain America: The Winter Soldier	2014	Action	Anthony Russo	714401889	132	7.7
9	Guardians of the Galaxy	2014	Action	Xavier執	770882395	121	8.0
10	Avengers: Age of Ultron	2015	Action	Joss Whedon	1395316979	141	7.3
11	Ant-Man	2015	Action	Peyton Reed	518858449	117	7.3
12	Captain America: Civil War	2016	Action	Joe Russo	1151895686	147	7.8
13	Doctor Strange	2016	Action	Scott Derrickson	676343174	115	7.5
14	Guardians of the Galaxy Vol. 2	2017	Action	Xavier執	869087963	136	7.6
15	Spider-Man: Homecoming	2017	Action	Jon Watts	678271291	133	7.4
16	Thor: Ragnarok	2017	Action	Taika Waititi	850482778	130	7.9
17	Black Panther	2018	Action	Ryan Coogler	1336494320	134	7.3
18	Avengers: Infinity War	2018	Action	Anthony Russo	2048359754	149	8.4
19	Ant-Man and the Wasp	2018	Action	Peyton Reed	629144550	118	7.0

Gambar 2 Proses Capture Data

Pembersihan data adalah langkah penting untuk memastikan kualitas analisis. Berdasarkan gambar yang disediakan, langkah pertama dalam pembersihan data adalah mengidentifikasi dan menghapus duplikasi data. Pemeriksaan awal untuk mendeteksi duplikasi di seluruh dataset menggunakan perintah `movies_data.duplicated().sum()` menunjukkan bahwa tidak ada data yang duplikat. Pemeriksaan tambahan berdasarkan kolom "movie" menggunakan perintah `marvel_corr.duplicated().sum()` juga menunjukkan hasil yang sama, yaitu tidak ada duplikasi data. Gambar 3 menjelaskan tentang bagaimana mencari data *duplicate* didalam python.

```
#movies_data.duplicated().sum() #utk mencari ada brp total data yg duplicate
marvel_corr.duplicated().sum() #->contoh mencari data duplicate yg ada yg kolom movie
```

Gambar 3 Proses Clean Data

Setelah data dibersihkan, langkah berikutnya adalah menampilkan hasil filter dari data yang relevan untuk analisis korelasi. Berdasarkan gambar 4, data difilter untuk menyertakan kolom-kolom "Movie", "Worldwide Box Office", "Run Time In Minutes", dan "IMDb Score". Hasil filter ini menampilkan judul film, total pendapatan box office global, durasi film dalam menit, dan skor IMDb untuk masing-masing film MCU. Informasi ini penting untuk analisis lebih lanjut karena menggambarkan variabel utama yang akan dikorelasikan.

	Movie	Worldwide Box Office	Run Time In Minutes	IMDb Score
0	Iron Man	585171547	126	7.9
1	The Incredible Hulk	265573859	112	6.6
2	Iron Man 2	621156389	124	6.9
3	Thor	449326518	115	7.0
4	Captain America: The First Avenger	370569776	124	6.9
5	The Avengers	1515100211	143	8.0
6	Iron Man 3	1215392272	130	7.1
7	Thor: The Dark World	644602516	112	6.8
8	Captain America: The Winter Soldier	714401889	132	7.7
9	Guardians of the Galaxy	770882395	121	8.0
10	Avengers: Age of Ultron	1395316979	141	7.3
11	Ant-Man	518858449	117	7.3
12	Captain America: Civil War	1151895686	147	7.8
13	Doctor Strange	676343174	115	7.5
14	Guardians of the Galaxy Vol. 2	869087963	136	7.6
15	Spider-Man: Homecoming	678271291	133	7.4
16	Thor: Ragnarok	850482778	130	7.9
17	Black Panther	1336494320	134	7.3
18	Avengers Infinity War	2048359754	149	8.4
19	Ant-Man and the Wasp	629144550	118	7.0

Gambar 4 Proses Filter Data

Gambar 4 merupakan tampilan proses filter yang dilakukan dengan mencari judul, *Worldwide Box Office*, Durasi dan Rating. Tahap terakhir adalah menghitung korelasi antara variabel-variabel utama, yaitu "Worldwide Box Office", "Run Time In Minutes", dan "IMDb Score" seperti pada Gambar 5.

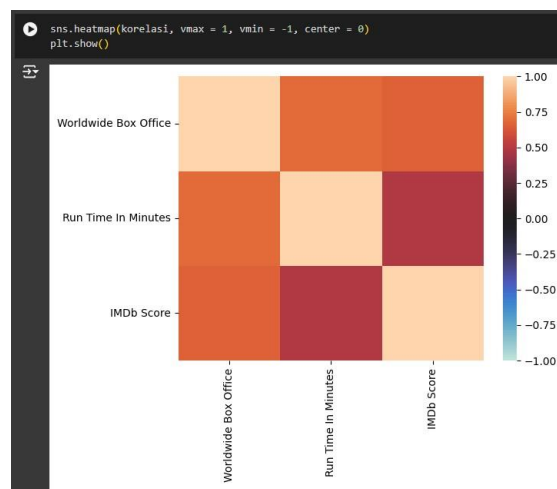
	Worldwide Box Office	Run Time In Minutes	IMDb Score
Worldwide Box Office	1.000000	0.689413	0.659261
Run Time In Minutes	0.689413	1.000000	0.491220
IMDb Score	0.659261	0.491220	1.000000

Gambar 5 Nilai Korelasi Antar Atribut

Berdasarkan Gambar 5, korelasi antara variabel-variabel tersebut ditunjukkan dalam bentuk matriks korelasi. Hasil korelasi menunjukkan bahwa:

1. Korelasi antara "Worldwide Box Office" dan "Run Time In Minutes" adalah 0.689, yang menunjukkan korelasi positif yang cukup kuat.
2. Korelasi antara "Worldwide Box Office" dan "IMDb Score" adalah 0.659, yang juga menunjukkan korelasi positif yang cukup kuat.
3. Korelasi antara "Run Time In Minutes" dan "IMDb Score" adalah 0.491, yang menunjukkan korelasi positif yang moderat.

Dengan demikian, hasil ini menunjukkan adanya hubungan positif antara durasi film, skor IMDb, dan keuntungan box office global di Marvel Cinematic Universe. Film dengan durasi lebih panjang dan skor IMDb yang lebih tinggi cenderung memiliki pendapatan box office global yang lebih tinggi.



Gambar 6 Diagram Heatmap Korelasi

Berdasarkan heatmap yang diberikan pada Gambar 6, kita dapat melihat korelasi antara tiga variabel: *Worldwide Box Office*, *Run Time in Minutes*, dan *IMDb Score* (rating). Korelasi antara *Worldwide Box Office* dan *Run Time in Minutes* (Durasi) terlihat positif, meskipun tidak terlalu kuat, ditunjukkan dengan warna oranye muda. Ini mengindikasikan bahwa film dengan durasi lebih panjang cenderung memiliki pendapatan *box office* yang lebih tinggi, meskipun hubungan ini tidak signifikan. Sebaliknya, korelasi antara *Worldwide Box Office* dan *IMDb Score* lebih kuat, ditunjukkan dengan warna oranye yang lebih gelap, menunjukkan bahwa film dengan rating IMDb yang lebih tinggi cenderung memiliki pendapatan *box office* yang lebih besar. Sementara itu, korelasi antara *Run Time in Minutes* dan *IMDb Score* adalah yang paling lemah, terlihat dari warna merah muda, yang menunjukkan bahwa durasi film tidak memiliki hubungan signifikan dengan rating IMDb. Secara keseluruhan, visualisasi ini menunjukkan bahwa dalam Marvel Cinematic Universe, pendapatan *box office* lebih dipengaruhi oleh rating IMDb daripada durasi film. Durasi film memiliki pengaruh moderat terhadap pendapatan *box office*, namun tidak signifikan terhadap rating IMDb. Hal ini menunjukkan bahwa kualitas film yang tercermin dalam rating IMDb lebih penting untuk pendapatan *box office* daripada durasi film.

3.2 Pembahasan

Hasil ini menunjukkan bahwa durasi film dan rating memainkan peran penting dalam keberhasilan film di *box office*. Analisis multivariat yang dilakukan menunjukkan temuan bahwa

durasi film, rating, dan keuntungan *box office* di Marvel Cinematic Universe memiliki hubungan yang signifikan. Misalnya, durasi film dan rating memiliki korelasi positif yang kuat, yang berarti film dengan durasi yang lebih panjang cenderung memiliki rating yang lebih tinggi. Selain itu, rating dan keuntungan *box office* juga memiliki korelasi positif, yang berarti film dengan rating yang lebih tinggi cenderung menghasilkan keuntungan *box office* yang lebih besar. Namun, penting untuk dicatat bahwa hasil ini tidak menunjukkan hubungan sebab akibat. Misalnya, meskipun film dengan durasi yang lebih panjang cenderung memiliki rating yang lebih tinggi, ini tidak berarti bahwa meningkatkan durasi film akan secara otomatis meningkatkan ratingnya. Faktor lain, seperti kualitas cerita dan penampilan aktor, juga dapat mempengaruhi rating film. Selain itu, meskipun penelitian ini menunjukkan adanya hubungan antara durasi film, rating, dan keuntungan *box office*, hubungan ini mungkin tidak berlaku untuk semua film. Misalnya, ada film pendek yang mendapatkan rating tinggi dan menghasilkan keuntungan *box office* yang besar, dan ada film panjang yang mendapatkan rating rendah dan tidak sukses di *box office*.

4. KESIMPULAN

Berdasarkan analisis korelasi antara durasi film, rating, dan pendapatan *box office* dalam Marvel Cinematic Universe, dapat disimpulkan bahwa rating memiliki pengaruh yang lebih signifikan terhadap pendapatan *box office* dibandingkan dengan durasi film. Meskipun film dengan durasi lebih panjang cenderung meraih pendapatan *box office* yang lebih tinggi, pengaruhnya tidak sekuat rating. Durasi film tidak memiliki korelasi yang signifikan dengan rating IMDb, menegaskan bahwa kualitas film yang tercermin dalam rating lebih krusial untuk kesuksesan komersial dibandingkan dengan panjang durasi film. Hasil ini menggarisbawahi pentingnya kualitas konten dalam menarik audiens dan meningkatkan pendapatan *box office*.

5. SARAN

Proyek ini menganalisis hubungan antara durasi film, rating IMDb, dan pendapatan *box office* dalam film-film Marvel Cinematic Universe (MCU) menggunakan analisis multivariat untuk memahami faktor-faktor yang memengaruhi keberhasilan komersial film. Hasil menunjukkan bahwa rating film memiliki pengaruh lebih signifikan terhadap pendapatan dibandingkan durasi, meskipun film dengan durasi lebih panjang cenderung juga menghasilkan pendapatan lebih tinggi. Untuk penelitian lebih lanjut, dapat ditambahkan data dari film di luar MCU dan variabel lain seperti biaya produksi atau pemeran utama untuk memperluas wawasan, serta visualisasi tambahan untuk memperjelas pola-pola yang muncul.

DAFTAR PUSTAKA

- [1] W. Y. Sulistyono, I. Riadi, and A. Yudhana, "Penerapan Teknik SURF pada Forensik Citra untuk Analisa Rekayasa Foto Digital," *JUITA J. Inform.*, vol. 8, no. 2, p. 179, 2020, doi: 10.30595/juita.v8i2.6602.
- [2] I. Riadi, A. Yudhana, and W. Y. Sulistyono, "Analisis Perbandingan Nilai Kualitas Citra pada Metode Deteksi Tepi," *Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 2, pp. 345–351, 2020.
- [3] D. U. Wustqa, E. Listyani, R. Subekti, R. Kusumawati, M. Susanti, and K. Kismiantini, "Analisis Data Multivariat Dengan Program R," *J. Pengabd. Masy. MIPA dan Pendidik. MIPA*, vol. 2, no. 2, pp. 83–86, 2018, doi: 10.21831/jpmmp.v2i2.21913.
- [4] M. dos S. F. Madson, "KNN algorithm and multivariate analysis to select and classify starch films," *Food Packag. Shelf Life*, no. 100976, 2022.
- [5] R. Haryadi, F. Fariyal, and S. Sanjaya, "How does Marvel Cinematic Universe (MCU) imply the counseling practice?," *ProGCouns J. Prof. Guid. Couns.*, vol. 3, no. 2, pp. 59–68, 2022, doi:

-
- 10.21831/progcouns.v3i2.53917.
- [6] and S.-G. L. Kim, Ahyun, Silvana Trimi, “Exploring the key success factors of films: a survival analysis approach,” *Serv. Bus.*, vol. 15, pp. 613–638, 2021.
- [7] N. A. Mayangky, D. N. Kholifah, I. Balla, and I. J. Thira, “Pengaruh Rating Film Terhadap Jumlah Audience Yang Menonton Film,” *Indones. J. Softw. Eng.*, vol. 5, no. 2, pp. 113–120, 2019, doi: 10.31294/ijse.v5i2.6963.
- [8] M. S. Erstiawan and M. P. Wangi, “Peran Anggaran dan Realisasi Anggaran Produksi Film Independen dalam Perspektif Akuntansi,” *Efektor*, vol. 10, no. 1, pp. 132–147, 2023, doi: 10.29407/e.v10i1.18667.
- [9] I. Permata Sari, J. Rawamangun Muka Raya, and J. Timur, “Implementasi Data Science dalam Ritel Online: Analisis Customer Retention dan Clustering Customer dengan Metode K-Means,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 1, pp. 417–425, 2021.
- [10] N. P. A. Widiari, I. M. A. D. Suarjaya, and D. P. Githa, “Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali,” *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 8, no. 2, p. 137, 2020, doi: 10.24843/jim.2020.v08.i02.p07.
- [11] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, “Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 23–27, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475.
- [12] R. Al Ghivary, M. Mawar, N. Wulandari, N. Srikandi, and A. N. M. F, “Peran Visualisasi Data Untuk Menunjang Analisa Data Kependudukan Di Indonesia,” *Pentahelix*, vol. 1, no. 1, p. 57, 2023, doi: 10.24853/penta.1.1.57-62.
-