

Analisis Faktor-Faktor yang Mempengaruhi Popularitas Aplikasi Pada Play Store

Andhika Wibawa, Azhar Mukhlisin, Tri Aji Bagaskara

Informatics Engineering, Department of Production Technology and Industry, Sumatera Institute of Technology

Abstract

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi popularitas aplikasi di Play Store. Play Store merupakan toko/repositori aplikasi Android yang memiliki beberapa *metrics* data aplikasi seperti rating, jumlah pengguna, kategori aplikasi, jenis aplikasi (berbayar atau gratis), genre aplikasi, dan lain-lain. Dari data ini, dilakukan pengolahan informasi dalam bentuk visualisasi yang diharapkan dapat berguna bagi para pengembang aplikasi di Android nantinya. Terdapat beberapa kesimpulan yang dihasilkan yaitu dari 7026 sampel aplikasi, rating aplikasi rata-rata adalah 4.16, didominasi oleh kategori keluarga (21.5%), dengan rating konten untuk semua umur (80.87%). Lalu, terdapat kecocokan pada 50 aplikasi terpopuler yaitu memiliki ukuran kurang dari 100 MB dengan rating umur rata-rata didesain untuk semua umur, dimana semuanya adalah bertipe gratis.

This is an open access article under the [CC BY-NC](#) license



Keywords:

play store;
popularitas;
aplikasi;
visualisasi

Article History:

Received: XXX
Revised: XXX
Accepted: XXX
Published: XXX

Corresponding Author:

Andhika Wibawa, Azhar Mukhlisin, Tri Aji Bagaskara

Informatics Engineering,
Sumatera Institute of Technology,
Indonesia

Email:

andhika.119140218@student.itera.ac.id
azhar.119140180@student.itera.ac.id
tri.119140214@student.itera.ac.id

INTRODUKSI

Play Store merupakan sebuah toko aplikasi *online* yang memuat berbagai jenis aplikasi berbayar maupun gratis. Toko ini merupakan aplikasi bawaan yang berjalan pada sistem operasi Android [1]. Play Store memiliki sistem ulasan/sentimen terhadap aplikasi yang ada yaitu berupa *rating* bintang dari 1 hingga 5 yang dapat disertai dengan teks penjelasan [2]. Sentimen yang diberikan biasanya berupa opini masyarakat mengenai (kualitas) suatu aplikasi [3]. Aplikasi yang mempunyai *rating* tinggi akan direkomendasikan dan muncul pada halaman depan atau timbul dalam filter terbaik bila diurutkan melalui *rating* [4]. Beberapa data *metrics* aplikasi yang terdapat pada Play Store antara lain meliputi rating, jumlah pengunduh, jenis aplikasi (gratis atau berbayar), kategori aplikasi, dan lain-lain [5].

Dari banyaknya *metrics* aplikasi yang terdapat pada Play Store tersebut, para pengembang di Android dapat mengambil dan mengolahnya agar dapat mengetahui tren perkembangan aplikasi yang saat ini banyak

diadopsi oleh pengembang dari aplikasi-aplikasi populer. Beberapa penelitian yang telah dilakukan sebelumnya antara lain mencakup analisis sentimen [6], pengguna (*userbase*) berdasarkan catatan perubahan aplikasi (*release note*) [7], korelasi antara jumlah unduhan aplikasi dengan rating [8], dan lain-lain.

Pada penelitian ini, penulis mencoba menganalisis variabel apa saja yang dapat mempengaruhi popularitas suatu aplikasi pada Play Store. Lalu, hasilnya ditampilkan dalam bentuk poster dan bantuan visualisasi data menggunakan bahasa pemrograman Python. Python dipilih karena memiliki jumlah library yang besar, sintaks kode yang jelas, dan didukung oleh komunitas besar serta integrasi yang kuat dengan bahasa dan alat lainnya [9][10].

METODE DAN PEMBAHASAN

Secara garis besar, terdapat 6 tahapan utama yang dilakukan pada penelitian ini untuk menghasilkan visualisasi data dan kesimpulan akhir nantinya. Beberapa tahapan ini yaitu:



Gambar 1. Alur Metode Penelitian

Adapun detail lebih lanjut mengenai tiap-tiap tahapan dapat dilihat pada bagian di bawah ini:

A. Menentukan tujuan visualisasi data

Dalam hal ini, tujuan yang ingin dicapai adalah melihat faktor-faktor yang mempengaruhi popularitas aplikasi di Play Store. Faktor-faktor yang terdapat pada kolom dataset antara lain kategori, rating, ukuran, harga, dan konten rating (batas umur) aplikasi. Sebelum faktor-faktor tersebut diolah menjadi informasi, dilakukan beberapa proses *data preprocessing* terlebih dahulu.

B. Menentukan target visualisasi data

Target utama dari penelitian dan visualisasi data ini adalah pengembang pemula yang ingin membuat aplikasi di *platform* Android dan merilisnya di Play Store. Meski begitu, informasi yang dihasilkan juga harus bersifat singkat (*to-the-point*) dan mudah dipahami oleh semua golongan. Oleh karena itu, hasil dari penelitian ini juga akan diatur ulang sedemikian rupa (dalam bentuk

poster) sehingga menarik untuk dibaca dan dieksplorasi oleh pembaca.

C. Mencari dan mengeksplorasi dataset

Dari beberapa dataset yang tersedia di internet, penulis memutuskan untuk menggunakan dataset dari Kaggle dengan link sebagai berikut: <https://www.kaggle.com/datasets/lava18/google-play-store-apps>. Data tidak diambil langsung dari Play Store karena situsnya bersifat dinamis serta membutuhkan waktu yang lama untuk mengambil dan mengolahnya (*data scraping*).

D. Membersihkan data (*data cleaning*)

Data cleaning merupakan proses untuk membersihkan struktur data dari berbagai kesalahan/error yang mempengaruhi integritas dan akurasi data [11]. Proses data cleaning dapat dilakukan melalui bantuan visualisasi ataupun eksplorasi data secara manual [12]. Proses ini umumnya selalu dilakukan sebelum melakukan pengambilan informasi karena dataset yang diolah pastinya memiliki tingkat *error* dan *outlier* tertentu yang cenderung mengganggu [13]. Pada penelitian ini, proses *cleaning* yang dilakukan meliputi penghapusan aplikasi dengan nama yang sama (didominasi oleh aplikasi *clone*/kurang berkualitas), pengubahan tipe data dari *string* menjadi *float/integer*, serta penghapusan beberapa *outlier* dalam data melalui pengamatan manual.

E. Reduksi data (*data reduction*)

Reduksi data merupakan sebuah proses untuk menyaring, memilih, menghapus data yang tidak diperlukan [14]. Dari banyaknya fitur serta jumlah data yang ada, tidak semua fitur bersifat penting untuk diolah menjadi suatu informasi. Oleh karena itu, kami mereduksi lagi beberapa fitur dari data seperti versi aplikasi, genre (mirip seperti kategori), serta versi Android yang dibutuhkan aplikasi. Lalu, kategori aplikasi serta kategori umur juga dilakukan penyusutan untuk nilai-nilai yang saling berkaitan, misalnya kategori bisnis dan keuangan digabung menjadi satu, dsb.

F. Mengolah informasi dari data

Setelah melakukan beberapa proses *data cleaning* dan *data reduction* pada tahapan sebelumnya, didapatkan data bersih yang sudah cukup layak digunakan untuk menghasilkan informasi dan visualisasi data nantinya. Visualisasi ini dapat digambarkan ke dalam beberapa diagram seperti diagram lingkaran, garis, dan juga diagram batang [15].

HASIL DAN PEMBAHASAN

A. Ringkasan data

Dataset yang diolah berasal dari situs Kaggle dengan judul "Google Play Store Apps". Dataset (mentahan) ini terdiri dari 10841 baris dengan 13 fitur/kolom yaitu App, Category, Rating, Reviews, Size, Installs, Type, Price, Content Rating, Genres, Last Updated, Current Ver, dan Android Ver. Adapun gambaran dari dataset yang belum diolah tersebut adalah sebagai berikut:

App	Category	Rating	...
Photo Editor & Candy ...	ART_AND_DESIGN	4.1	...
Coloring book moana	ART_AND_DESIGN	3.9	...
U Launcher Lite – FREE Live ...	ART_AND_DESIGN	4.7	...
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	...
...

Tabel 1. Data mentah aplikasi Play Store

Dari data mentah tersebut, kemudian dilakukan proses seperti *data cleaning* dan *data reduction*. Salah satu proses *data cleaning* yang dilakukan pada dataset ini yaitu:

```
# Buang nilai duplikat, NaN, dan outlier (observasi manual)
df.dropna(how = "any", inplace = True)
df.drop_duplicates(subset = ["App"], inplace = True)
df = df[df["Category"] != "1.9"]
df = df[df["Content Rating"] != "Unrated"]
df = df[df["Size"] != "Varies with device"]
df = df[df["Type"] != "0"]
```

Gambar 2. Proses penghapusan duplikat, NaN, dan *outlier* pada data

Dimana data yang dibuang didasarkan pada hasil observasi manual pada dataset. Lalu, untuk memproses berbagai perhitungan *rating* dan ukuran aplikasi nantinya, dilakukan juga *cleaning* pada data angka yang masih mengandung *string*.

```
# Hapus tanda koma dan plus
df["Installs"].replace({
    ",": "",
    "\+": ""
}, regex = True, inplace = True)

# Ubah format juta dan ribuan ke digit nol
df["Size"].replace({
    "M": "000",
    "k": "00",
    "\.": ""
}, regex = True, inplace = True)

# Hapus dollar di depan angka
df["Price"].replace({
    "$": ""
}, regex = True, inplace = True)
# Ubah dan perjelas nama kolom harga
df.columns.values[7] = "Dollar Price"
```

Gambar 3. Proses pengolahan *string* menjadi angka (*integer/float*)

Terakhir, proses *data reduction* berupa penghapusan kolom yang tidak digunakan penggabungan beberapa kategori aplikasi yang masih saling berkaitan. Cuplikan kode penggabungan kategori aplikasi dapat dilihat seperti di bawah ini:

```
# Translasi dan gabungkan beberapa kategori
# yang saling berhubungan

df["Category"].replace({
    "FAMILY": "Keluarga",
    "GAME": "Permainan",
    "TOOLS": "Utilitas",
    "MEDICAL": "Kedokteran",
    "BUSINESS": "Bisnis dan Keuangan",
    "FINANCE": "Bisnis dan Keuangan",
    # Dan seterusnya ...
}, inplace = True)
```

Gambar 4. Proses translasi dan penggabungan kategori aplikasi

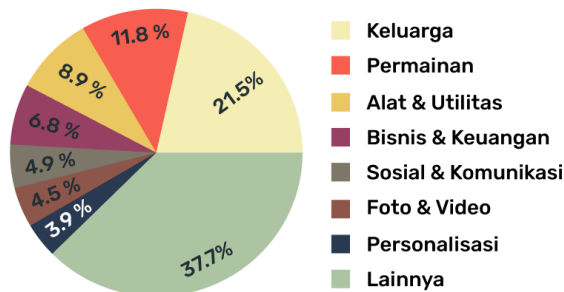
Lalu, setelah dilakukan beberapa proses *data cleaning* dan *data reduction*, dataset tersebut menjadi berjumlah 7026 baris dengan jumlah 10 kolom yaitu App, Category, Rating, Reviews, Size, Installs, Type, Dollar Price, Content Rating, dan Last Updated. Gambaran dari data setelah dilakukan *data preprocessing* tersebut yaitu:

App	Category	Rating	...
Photo Editor & Candy ...	Seni dan Desain	4.1	...
Coloring book moana	Seni dan Desain	3.9	...

U Launcher Lite – FREE Live ...	Seni dan Desain	4.7	...
Sketch - Draw & Paint	Seni dan Desain	4.5	...
...

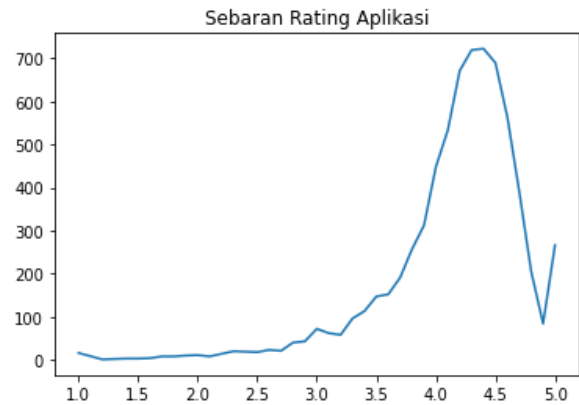
Tabel 2. Data final aplikasi Play Store**B. Pengolahan informasi**

Setelah dilakukan berbagai proses pada tahapan sebelumnya, data kini siap diolah menjadi informasi yang bermanfaat. Langkah pertama yang dilakukan adalah melihat sebaran kategori aplikasi yang ada pada sampel dataset.

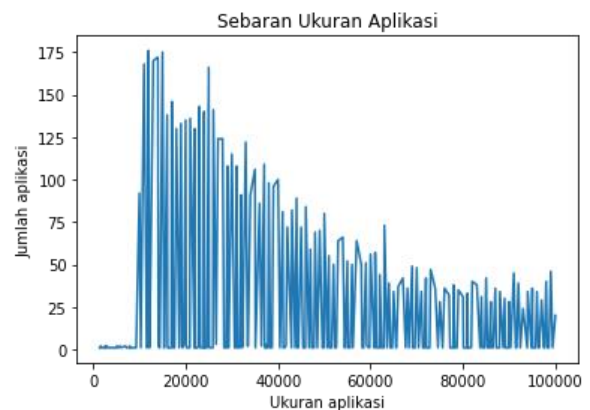
Kategori Aplikasi**Gambar 5.** Sebaran kategori aplikasi

Dari gambar tersebut, terlihat bahwa sebagian besar aplikasi memiliki kategori keluarga (21,5% / 1511 aplikasi) dan permainan (11,8% / 832 aplikasi). Lalu, beberapa kategori sisanya yang masing-masing bernilai kurang dari 3,9% disusun oleh 20 kategori yang berbeda, seperti olahraga, gaya hidup, dan lain-lain.

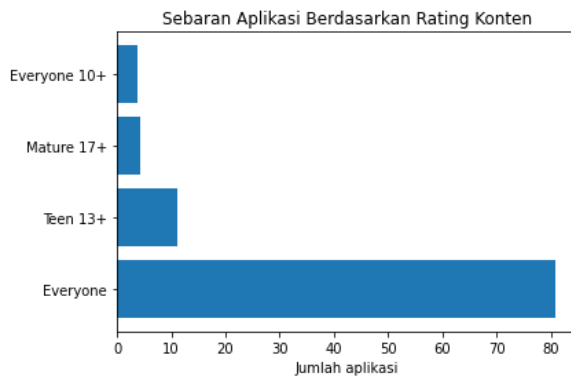
Selanjutnya yaitu melihat sebaran skor (*rating*) pada tiap-tiap aplikasi yang ada di Play Store. Didapatkan nilai *rating* rata-rata yaitu 4,16 dengan *rating* terbanyak yaitu 4.4 (722 aplikasi) dan *rating* tersedikit yaitu 1.2 (1 aplikasi). Grafik *rating* aplikasi dapat dilihat pada gambar berikut:

**Gambar 6.** Sebaran rating aplikasi

Setelah melihat *rating* aplikasi, dilihat pula ukuran dari tiap-tiap aplikasi yang ada di Play Store. Grafik menunjukkan ukuran aplikasi terbesar berada pada rentang 100 MB dengan jumlah 20 aplikasi dan terendah 1,4 MB dengan jumlah 1 aplikasi. Lalu, rata-rata aplikasi yang ada pada sampel memiliki ukuran 41 MB.

**Gambar 7.** Sebaran ukuran aplikasi

Langkah selanjutnya yaitu melihat sebaran aplikasi berdasarkan *rating* konten (umur) pengguna. Didapat hasil tertinggi yaitu *rating* “everyone” dengan nilai 80,85% (5676 aplikasi), disusul oleh “teen” (11,09% / 779 aplikasi), “mature” (4,35% / 306 aplikasi), dan “everyone 10+” (3,68% / 259 aplikasi). Hasil grafik dari proses ini dapat dilihat pada gambar di bawah.



Gambar 8. Sebaran *rating* konten aplikasi

Dengan mengalikan *rating* dan jumlah pengguna aplikasi, didapatkan *list* aplikasi dengan jumlah pengguna dan *rating* terbanyak. Adapun peringkat 5 terbesar yang dihasilkan dari *list* tersebut yaitu:

Nama	Kategori	Rating	Pengguna
Subway Surfers	Permainan	4.5	100 Juta
Google News	Berita dan Majalah	3.9	100 Juta
SHAREit	Utilitas	4.6	50 Juta
UC Browser	Sosial dan Komunikasi	4.5	50 Juta
Dropbox	Produktivitas	4.4	50 Juta

Tabel 3. 5 Aplikasi terpopuler di Play Store

Jika dilihat dari 50 *list* aplikasi dengan jumlah pengguna dan rating terbanyak tersebut, aplikasi-aplikasi ini juga memiliki beberapa karakteristik, yaitu:

1. Berukuran kurang dari 100 MB, dimana yang terbesar berukuran 89 MB dan terkecil 17 MB
2. Sebagian besar memiliki rating untuk semua umur (36 aplikasi), dengan sisanya untuk remaja sebanyak 8 aplikasi dan diatas 10 tahun sebanyak 6 aplikasi
3. Semua bersifat gratis untuk diunduh

KESIMPULAN

Pada hasil dan analisis pembahasan aplikasi yang telah dilakukan sebelumnya, dapat ditarik banyak kesimpulan, antara lain:

1. Dari 7026 aplikasi terdapat 541 aplikasi berbayar dan selebihnya adalah aplikasi gratis.

2. Kategori aplikasi terbanyak sesuai dengan jumlah pengguna adalah kategori keluarga sebesar 21,5%.
3. Untuk kategori aplikasi berdasarkan umur, terdapat 5682 aplikasi untuk semua umur yang mana setara dengan 80,87% dari total sampel aplikasi dan didominasi oleh kategori keluarga dan permainan.
4. Aplikasi tertinggi berdasarkan jumlah pengguna dan *rating* adalah aplikasi Subway Surfers, Google News, SHAREit, UC Browser, dan Dropbox.
5. Aplikasi terpopuler umumnya berukuran kurang dari 100 MB, didesain untuk semua umur, dan bersifat gratis untuk diunduh.

REFERENSI

- [1] Rafael. S. O, "ANALISIS DATASET GOOGLE PLAYSTORE MENGGUNAKAN METODE EXPLORATORY DATA ANALYSIS," *Jurnal Strategi*, vol. 2, 2020.F.
- [2] Maylani, "Implementasi Metode Data Mining Untuk Memprediksi Warna Anak Kucing Pada Proses Pengembangbiakan Kucing Ras Menggunakan Algoritma Support Vector Machine (SVM)," p. 114, 2021.
- [3] M. F. Al-shufi, "SENTIMEN ANALISIS MENGENAI APLIKASI STREAMING FILM MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DI PLAY STORE," *SISMATIK*, 2021.
- [4] B. Wahyudi, "Prediksi Peringkat Aplikasi di Google Play Menggunakan Metode Random Forest," *Jurnal Nasional Teknologi Komputer*, vol. 2, 2022.
- [5] E. D. P. Daulay, "Sentimen Analisis pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes," *e-Proceeding of Engineering*, vol. 7, 2020.
- [6] M. I. Ahmadi, "SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)," vol. 1, 2020.
- [7] A. Z.H, "An Empirical Study on Release Notes Patterns of Popular Apps in the Google Play Store," *Queen's University*, 2019.
- [8] A. A. Sunasara, "Play Store App Analysis," *International Research Journal of Engineering and Technology*, vol. 8, no. 04, 2021
- [9] A. N. Syahrudin, "INPUT DAN OUTPUT PADA BAHASA PEMROGRAMAN PYTHON," *Jurnal Dasar Pemrograman Python STMIK*, 2018.

- [10] Fitri. dkk, "Penggunaan Bahasa Pemrograman Python Sebagai Pusat Kendali Pada Robot 10-D," *Jurnal Ilmiah Universitas Pendidikan Indonesia*, vol. 3, 2017.
- [11] K. H. Tae, "Data Cleaning for Accurate, Fair, and Robust Models:A Big Data - AI Integration Approach," 2019.
- [12] Conder. L. S, "Arachnid: Generalized Visual Data Cleaning," *SIGMOD*, 2019.
- [13] N. P. A. Widiari, "Teknik Data Cleaning Menggunakan Snowflake untuk Studi Kasus Objek Pariwisata di Bali," *JURNAL ILMIAH MERPATI*, vol. 8, no. 2, 2020.
- [14] D. P. Ayu, "Evaluasi Layanan Kepegawaian Pada Aplikasi E-Master Untuk Tenaga Kependidikan," 2021.
- [15] K. M. Sukiakhy, "KLASIFIKASI DAN VISUALISASI PARIWISATA ACEH DENGAN MENGGUNAKAN DATA PADA TWITTER," (*Jurnal Sistem Informasi Musirawas*, vol. 5, 2020.