

**Tugas Datamining**  
**Analisis Sentimen Google Playstore Terhadap**  
**Genshin Impact**



**Hugo Faiq Riseva A11.2022.14329**

## Deskripsi Singkat

Genshin Impact adalah salah satu aplikasi permainan terpopuler di Google Play Store, dengan jutaan unduhan dan berbagai ulasan pengguna. Analisis sentimen terhadap ulasan-ulasan ini dapat memberikan wawasan berharga mengenai persepsi pengguna terhadap aplikasi ini, terutama terkait kepuasan, ketidakpuasan, dan ekspektasi pengguna. Dengan menggunakan metode pembelajaran mesin, penelitian ini bertujuan untuk mengidentifikasi sentimen utama pengguna dan mengelompokkan ulasan menjadi sentimen positif, negatif, dan netral.

## Masalah dan Tujuan yang Ingin Dicapai

**Masalah :** Bagaimana persepsi pengguna terhadap aplikasi Genshin Impact di Google Play Store? Faktor-faktor apa yang menonjol dalam ulasan pengguna, dan bagaimana pola sentimen tersebut dapat digunakan untuk meningkatkan pengalaman pengguna?

### Tujuan :

1. Mengidentifikasi sentimen utama dalam ulasan pengguna Genshin Impact di Google Play Store.
2. Mengelompokkan ulasan berdasarkan sentimen positif, negatif, dan netral.
3. Mengidentifikasi tema-tema atau isu yang sering muncul dalam ulasan negatif dan memberikan rekomendasi untuk pengembang aplikasi.

## Alur / Tahapan / Kerangka Eksperimen

1. Pengumpulan Data : Mengumpulkan data ulasan pengguna aplikasi Genshin Impact dari Google Play Store menggunakan scraping atau dataset publik.
2. Pre Processing Data : Melakukan pembersihan data, seperti menghapus kata-kata umum (stop words), menghapus tanda baca, dan normalisasi teks.
3. Tokenisasi dan Transformasi Teks : Memecah ulasan menjadi token (kata atau frasa) serta menggunakan teknik seperti TF-IDF atau Word2Vec untuk merepresentasikan teks.
4. Pelabelan Data : Melabeli data secara manual untuk membangun model awal (jika dataset tidak berlabel) atau menggunakan data yang sudah berlabel.
5. Pengembangan Model Sentimen : Membangun model klasifikasi sentimen (contoh: Naive Bayes, SVM, atau LSTM) untuk mengelompokkan ulasan ke dalam kategori positif, negatif, dan netral.

6. Evaluasi Model : Menggunakan metrik seperti akurasi, precision, recall, dan F1-score untuk mengevaluasi performa model sentimen.
7. Analisis Sentimen dan Topik : Menggunakan model untuk menganalisis sentimen dan mengidentifikasi tema-tema utama dari ulasan pengguna.
8. Rekomendasi : Menyusun rekomendasi untuk pengembang Genshin Impact berdasarkan hasil analisis sentimen dan tema negatif yang muncul.

## **Penjelasan Datasets**

### **Sumber Data :**

Data berasal dari Komentar dan ulasan yang tersedia di aplikasi Google Playstore. Dengan menggunakan metode Crawling Data .

### **Penjelasan Atribut :**

1. Review\_ID : Identitas unik untuk setiap ulasan.
2. User\_ID : Identitas unik pengguna yang memberikan ulasan.
3. Review\_Text : Teks ulasan yang diberikan oleh pengguna.
4. Rating : Peringkat aplikasi dari pengguna (dari 1 hingga 5 bintang).
5. Review\_Date : Tanggal ulasan ditulis.
6. Sentiment\_Label : Kategori sentimen ulasan (positif, negatif, atau netral) - dihasilkan setelah pelabelan atau dari dataset berlabel.
7. Issue\_Type (opsional) : Kategori permasalahan atau topik yang sering muncul (seperti kinerja aplikasi, pembelian dalam aplikasi, pengalaman bermain, dll).

## **Crawling Data**

Crawling merupakan proses menggali jauh ke dalam world wide web atau target tertentu untuk mengambil data. Crawling data dilakukan dengan menggunakan bahasa pemrograman Python pada tools Google Colaboratory. Crawl

ing data dilakukan pada game Genshin Impact yang terdapat pada Google Play Store pada bulan Mei 2024, dengan ulasan yang diambil sebanyak 398 ulasan.

## **Labeling**

Labeling yaitu proses memberikan nilai score sentimen positif atau negatif dengan nilai score, jika score bernilai tiga sampai dengan lima maka akan menghasilkan nilai yaitu satu yang artinya positif, sedangkan score bernilai satu sampai dengan dua maka akan menghasilkan nilai yaitu nol yang artinya negative



```
df = pd.read_csv('/content/data_real.csv', index_col=0)
df.head()
```

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedat	appVersion	label
e82290bf-ac79-457f-a724-71b7d2e589fb	Agis Sayid Abdillah	lh.googleusercontent.com/a-/ALV-U...	Wah ini sih game nya keren abisssss Saya suka ...	5	593	4.6.0_22174900_22378599	2024-05-11 03:23:30	NaN	NaN	4.6.0_22174900_22378599	1
570138d7-1e47-40c2-af2b-97fcb1ff5848	Luvina Irene	lh.googleusercontent.com/a-/ALV-U...	Hii, aku suka banget sama game ini, dari dulu ...	4	12	4.6.0_22174900_22378599	2024-05-25 15:09:21	NaN	NaN	4.6.0_22174900_22378599	1
59e80afe-c076-4dec-8826-0b27f3493643	Rikaellie	lh.googleusercontent.com/a-/ALV-U...	Dari segi apapun menurut saya bagus.saya suka ...	5	72	4.6.0_22174900_22378599	2024-05-28 04:38:48	NaN	NaN	4.6.0_22174900_22378599	1
7d14da6a-504d-4e60-8e59-e8ba09c77b40	Humam	lh.googleusercontent.com/a-/ALV-U...	Cuma satu minus di game ini, kurang kompatibel...	4	26	4.6.0_22174900_22378599	2024-05-17 05:50:43	NaN	NaN	4.6.0_22174900_22378599	1
b3787f21-f5b0-4687-a53b-5f2c352226fd	Channel Gaming	lh.googleusercontent.com/a-/ALV-U...	Ping jaringan di in game tidak jelas udh insta ...	3	0	4.6.0_22174900_22378599	2024-05-30 15:23:27	NaN	NaN	4.6.0_22174900_22378599	1

## Preprocessing Text

### 1. Normalisasi

Merupakan tahapan reprocessing yang dilakukan untuk menubah atau kalimat pada suatu kata yang singkat untuk memenuhi tujuan tersendiri yaitu “dgn” menjadi “dengan”, “yg” menjadi “yang”. Preprocessing Text 1. Normalisasi Merupakan tahapan reprocessing yang dilakukan untuk menubah atau kalimat pada suatu kata yang singkat untuk memenuhi tujuan tersendiri yaitu “dgn” menjadi “dengan”, “yg” menjadi “yang”.

### 2. Stopwords

Digunakan untuk menghapus kata yang memiliki informasi yang rendah seperti yang, dan, dari, di, dll.

### 3. Tokenize

Memisahkan kalimat, paragraf atau dokumen menjadi kata-kata atau bagian-bagian tertentu.

### 4. Stemming

Digunakan untuk memperkecil jumlah indeks yang berbeda dari satu data sehingga sebuah kata yang memiliki suffix maupun prefik akan kembali ke bentuk dasarnya.

# Preprocessing Teks

## 1. Normalisasi

```
norm = {'dgn': 'dengan', 'gue': 'saya', 'bgmn': 'bagaimana', 'tdk': 'tidak', 'blum': 'belum', 'fast': 'cepat', 'trims': 'terima kasih', 'brg': 'barang', 'gx': 'tidak', 'recommended': 'rekomendasi', 'good': 'bagus', 'gw': 'saya', 'yg': 'yang', 'gk': 'tidak', 'bgs': 'bagus', 'gak': 'tidak', 'bgt': 'banget', 'skrg': 'sekarang', 'tp': 'tapi', 'bejir': 'astaga', 'donlaud': 'download', 'bansos': 'bantuan', 'banyak': 'banyak', 'udah': 'sudah', 'kejelekanya': 'buruk', 'bgitu': 'begitu', 'tetep': 'tetap', 'mnit': 'menit', 'njing': 'astaga', 'bgt': 'banget', 'dll': 'dan lain lain', 'dri': 'dari', 'anjay': 'wah', 'bgus': 'bagus', 'benr': 'benar', 'ttp': 'tetap', 'tpi': 'tapi', 'dri': 'dari', 'kureng': 'kurang', 'udh': 'sudah', 'dkit': 'sedikit', 'mntap': 'bagus', 'kgk': 'tidak', 'bnget': 'banget', 'rada': 'agak', 'yt': 'YouTube', 'worth it': 'layak', 'biar': 'agar', 'p2w': 'game menguntungkan hanya untuk pemain berbayar', 'gmn': 'bagaimana', 'cman': 'cuma', 'udah': 'sudah', 'gpp': 'tidak masalah', 'lgi': 'lagi', 'cakep': 'bagus', 'org': 'orang', 'kalo': 'jika', 'ngotak': 'tidak berfikir', 'krn': 'karena', 'apik': 'bagus', 'system': 'sistem', 'adain': 'sediakan', 'tpi': 'tapi', 'pencinya': 'tekan', 'amat': 'sekali', 'bgst': 'astaga', 'chr': 'chara', 'ajah': 'saja', 'banyak': 'banyak', 'bnget': 'banget', 'apah': 'apa', 'bansos': 'bantuan', 'mntap': 'bagus', 'bkan': 'bukan', 'tvat': 'tempat', 'fzp': 'pemain dapat mengakses dan menikmati game tanpa biaya apa pun di muka', 'krn': 'karena', 'malahan': 'bahkan', 'best': 'terbaik', 'terus': 'lalu', 'udh': 'sudah', 'bgt': 'banget', 'bgs': 'bagus', 'smpe': 'sampai', 'yutub': 'YouTube', 'thn': 'tahun', 'apk': 'aplikasi', 'plot twins': 'putar balik alur', 'plss': 'tolong', 'agak': 'perkiraan', 'ajah': 'saja', 'sampein': 'sampaikan', 'btw': 'ngomong-ngomong', 'krna': 'karena', 'sm': 'sama', 'enjoy': 'senang', 'bodoamat': 'tidak peduli', 'nice': 'bagus', 'trus': 'terus'}
```

```
def normalisasi(str_text):  
    for i in norm:  
        str_text = str_text.replace(i, norm[i])  
    return str_text
```

```
df_copy['content'] = df_copy['content'].apply(lambda x: normalisasi(x))
```

reviewId	userName	userImage	content	score	thumbsupCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion	label
e82290bf-ac79-457f-a724-71b7d2e589fb	Agis Sayid Abdillah	lh.googleusercontent.com/a-/ALV-U...	wah ini sih game nya keren abissss saya suka ...	5	593	4.6.0_22174900_22378599	2024-05-11 03:23:30	NaN	NaN	4.6.0_22174900_22378599	1
570138d7-1e47-40c2-af2b-97fcb1f5848	Luvina Irene	lh.googleusercontent.com/a-/ALV-U...	hii, aku suka banget sama game ini, dari dulu ...	4	12	4.6.0_22174900_22378599	2024-05-25 15:09:21	NaN	NaN	4.6.0_22174900_22378599	1
59e80a1e-c076-4dec-8826-0b27f3493643	Rikaellie	lh.googleusercontent.com/a-/ALV-U...	dari segi apapun menurut saya bagus.saya suka ...	5	72	4.6.0_22174900_22378599	2024-05-28 04:38:48	NaN	NaN	4.6.0_22174900_22378599	1
7d14da6a-504d-4e60-8e59-c8ba08c77b40	Humam	lh.googleusercontent.com/a-/ALV-U...	cuma satu minus di game ini, kurang kompatibel...	4	26	4.6.0_22174900_22378599	2024-05-17 05:50:43	NaN	NaN	4.6.0_22174900_22378599	1
b3787f21-f5b0-4687-a53b-5f2c352226fd	Channel Gaming	lh.googleusercontent.com/a-/ALV-U...	ping jaringan di in game tidak jelas sudah ins...	3	0	4.6.0_22174900_22378599	2024-05-30 15:23:27	NaN	NaN	4.6.0_22174900_22378599	1

## 2. Stopwords

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory, StopWordRemover, ArrayDictionary  
more_stop_words = []  
  
stop_words = StopWordRemoverFactory().get_stop_words()  
new_array = ArrayDictionary(stop_words)  
stop_words_remover_new = StopWordRemover(new_array)  
  
def stopword(str_text):  
    str_text = stop_words_remover_new.remove(str_text)  
    return str_text
```

```
df_copy['content'] = df_copy['content'].apply(lambda x: stopword(x))  
df_copy.head()
```

	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion	label
reviewId											
e82290bf-ac79-457f-a724-71b7d2e589fb	Agis Sayid Abdillah	lh.googleusercontent.com/a-/ALV-U...	wah sih game nya keren abisssss suka design ka...	5	593	4.6.0_22174900_22378599	2024-05-11 03:23:30	NaN	NaN	4.6.0_22174900_22378599	1
570138d7-1e47-40c2-af2b-97fcb1ff5848	Luvina Irene	lh.googleusercontent.com/a-/ALV-U...	hii, aku suka banget sama game ini, dulu penge...	4	12	4.6.0_22174900_22378599	2024-05-25 15:09:21	NaN	NaN	4.6.0_22174900_22378599	1
59e80a1e-c076-4dec-8826-0b27f3493643	Rikaellie	lh.googleusercontent.com/a-/ALV-U...	segi apapun bagus,saya suka banget malah,cuma ...	5	72	4.6.0_22174900_22378599	2024-05-28 04:38:48	NaN	NaN	4.6.0_22174900_22378599	1
7d14da6a-504d-4e60-8e59-c8ba08c77b40	Humam	lh.googleusercontent.com/a-/ALV-U...	cuma satu minus game ini, kurang kompatibel be...	4	26	4.6.0_22174900_22378599	2024-05-17 05:50:43	NaN	NaN	4.6.0_22174900_22378599	1
b3787f21-f5b0-4687-a53b-5f2c352226fd	Channel Gaming	lh.googleusercontent.com/a-/ALV-U...	ping jaringan in game jelas instal semingguan....	3	0	4.6.0_22174900_22378599	2024-05-30 15:23:27	NaN	NaN	4.6.0_22174900_22378599	1

### 3. Tokenize

```

Tokenize

tokenized = df_copy['content'].apply(lambda x:x.split())
tokenized

reviewId
e82290bf-ac79-457f-a724-71b7d2e589fb    [wah, sih, game, nya, keren, abisssss, suka, d...
570138d7-1e47-40c2-af2b-97fcb1ff5848    [hii,, aku, suka, banget, sama, game, ini,, du...
59e80a1e-c076-4dec-8826-0b27f3493643    [segi, apapun, bagus,saya, suka, banget, malah...
7d14da6a-504d-4e60-8e59-c8ba08c77b40    [cuma, satu, minus, game, ini,, kurang, kompat...
b3787f21-f5b0-4687-a53b-5f2c352226fd    [ping, jaringan, in, game, jelas, instal, semi...
...
f6f0b6b6-bd64-4771-9767-95d68cf9aea4    [overall, storynya, bagus,, grafiknya, bagus, ...
f445f6a7-3d73-40be-8ea1-e03c61b3ec24    [overall, bagus, gameplay,story,, event,, kala...
266ed51f-50be-4a4e-9edd-a0380d1a7510    [game, bagus, cukup, suka, sama, ada, genshin,...
5abb923d-bc92-45bc-8e43-2fd05861e178    [game, nya, bagus, segi, cerita, grafik, kesel...
1883bbfb-f750-4c5d-a621-79c53d13d97a    [segi, gameplay,, grafik,, multiplayer, nya, s...
Name: content, Length: 398, dtype: object

```

### 4. Stemming

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

def stemming(content):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    do = []
    for w in content:
        dt = stemmer.stem(w)
        do.append(dt)
    d_clean = []
    d_clean = " ".join(do)
    print(d_clean)
    return d_clean

tokenized = tokenized.apply(stemming)

tokenized.to_csv('data_bersih2.csv', index=False)
data_clean = pd.read_csv('data_bersih2.csv', encoding='latin1')
data_clean.head()

```

nah sih game nya keren abisssss suka desain karakter animasi tarung power full banget untuk saran sih jadi basahan gacha donor persudakan ngar para free player dapat seru hii aku suka banget sama game ini dulu ken download hp ga kuat akhir sekarang bisa gamenya seru banget satu masalah story quest raiden dua pas suruh ikut jejak musuh malah ngecih segitiga apa bagus saya suka banget malah cuma saya harap dev denger player dan baik soal story walaupun story npc mc woga depan mc paimon npc quest dialog bicara dan dialog suara cuma satu minus game ini kurang kompatibel beberapa device mid-range tentu makin paksa buat main lama makin panas frame drop sekali grafik low-medium 60fps mau gimana nama k ping jaring in game jelas instal minggu ping jaring dalam game pernah ubah stabil 100 ms buat kadang basic attack nya delay padahal game ping jaring bagus bagus saja kasih bintang 3 keren ghaca nya sudah cukup sebal kenapa bersusah payah ghaca 30 bola karakter satu buat saya tidak main beberapa minggu saya benar suka grafik dan cara game konsep baik story bagus sekali jaring oke gameplay lumayan bagus meski batas fitur lumayan meski batas saja grafik fps bisa full hight padahal spesifikasi handphone sgggen opt quest nya seru map nya lumayan seru dieksplor ketika ar sudah tinggi semua nya jadi atidak bosenin hari login cuma farming buat karakter aja itu artefak bagus susah didapetin sangat2 suka semua game buat hoyoverse game buat hoyoverse selalu saji story game fun fresh seru sedih pokok story nya keren abiss keluh soal teka-teki puzzle nya sulit turut story nya sih bagus banget lalu untuk environment area nya behhh indah banget cuyy yaa sudah nyentuh ar tinggi pasti bakal ngerasa bosen end game content nya bisa bilang mini segi mana game sudah bagus mau sebut game kikir apa game tetap bagus karna desain karakter dunia ide storynya keren semua game downloadnya gratis cuman 1 jadi keluh saya masal segi positif game tentu eksplorasi menyenangkan atidak ribet sih kalau sumeru karakter desain waifuable husbuable alur cerita kalah bagus adik visual bagus meski tidak bagus masalah wish lumayan ganti artefak sistem upgrade nya gimana sih sudah dapat bagus peluang cuman 5 di upgrade naik substatnya fokus banget def stat tinggi mulai selera lanjut game nya overall bagus segi grafik story nya bagus apa pas sudah masuk ark fontaine baik banget story nya lebih hidup sayang kan kurang konten game mode khusus hanya spiral 2 gameplay nya memang tarik story questnya bikin boring selesai quest kepala selalu frustrasi karna dialog berfikir panjang mana bisa skip itu developer selalu pelit beri hadiah game nya padahal bagus bagus aja kata nya harap ke 90 aku dapat kan karakter bintang 5 bahkan buku bintang 4 developer harus nya itu tepat janji dong jangan buat rugi player 4 game nya bagus grafik main lore cerita tempur cutscene nya region nya karakter nya alam pandang nya semua sempurna game baik pernah aku main jauh ini game bakal tetap di hati sering adakan event kasih primogem baik bug ada game sama waktu isi resin persintidakat 1 resin 8 menit lama resin sangat butuh main baru gamenya bagus ngomong-ngomong tidak tau mula ku jelasin ya genshin game open world macam game gacha gitu jadi gacha dapet ampas tuh wajarin aja nama gacha genshin bukan game 5 saran tambah menu baru dual koneksi atur bisa dapat jaring stabil grafik patah2 war terima kasih sukses selalu download genshin impact jadi tagih main nya hari main genshin impact walaupun tantang nya yang susah banget tapi aku suka banget sama grafik nya suara karakter nya elemental 4 genshin impact versi sekarang sudah mulai lebih baik ya sudah banyak fitur cukup bantu player hapus data quest dahulu sangat ngebantu loh buat 128gb story nya sudah terlalu bagus banget si alur cerita sama grafik nya dan waktu ngunduh data awal game sambil keluar apalagi character nya cantik sama ganteng lagi deh game paling bagus main versi 1.8 turut bagus sih dari grafik gameplay sama desain karakter cuma kurang hadiah di banditidakan sama game belah di bilang sedikit banget game nya bagus sekali grafik ok story quest keren ada event selalu keren tarik quest archonnya bagus tarik kalau tambah lebih banyak story quest nya ya note kasih untung daya penintidakan karakter lama farming artefak lama dapat bagus kadang resin tak guna guna banyak cerita npc karakter utama hadiah event sering sesuai main event hadiah anniver jauh sih bagus kalau animasi waktu quest archon bisa buka dari bikin anime nanti muntidakin buruk game nya balik saja dulu mohon maaf atas saran ga jelas saya hai hoyo suka banget aku main genshin momentum 2020 seru banget alur cerita tarik banyak quest juga ngegacha karakter walau atidak ngeselin kalah rate off dulu tetap seru kok