

BAB III

HASIL ANALISIS DAN PEMBAHASAN

3.1 Data Selection

Pada rentang waktu 15 September 2023 hingga 1 Mei 2024, data ulasan untuk permainan Stumble Guys dikumpulkan menggunakan teknik scraping. Menggunakan library google play scraper, berhasil terkumpul sebanyak 1500 data. Data yang terkumpul memiliki lima atribut, yaitu reviewId, userName, at, content, dan score. Selanjutnya, hanya atribut content dan score yang dipilih untuk digunakan.



reviewId	userName	userImage	content	score	timestamp	reviewMetadata	at	reviewContent	reviewId	timestamp
7940114190-444f02771-04a18704a48	Ferd Rudy	https://lh3.googleusercontent.com/...	game ini sangat seru dan seru banget	5	1644360000		2024-01-15 10:00:00	None	5	1644360000
4022819190-444f02771-04a18704a48	2021 B	https://lh3.googleusercontent.com/...	game ini sangat seru dan seru banget	5	1644360000		2024-01-15 10:00:00	None	5	1644360000
7940114190-444f02771-04a18704a48	Adnan Yusuf	https://lh3.googleusercontent.com/...	game ini sangat seru dan seru banget	5	1644360000		2024-01-15 10:00:00	None	5	1644360000
7940114190-444f02771-04a18704a48	Muhammad Rizki Akbar	https://lh3.googleusercontent.com/...	game ini sangat seru dan seru banget	5	1644360000		2024-01-15 10:00:00	None	5	1644360000
7940114190-444f02771-04a18704a48	Ahmad Wicaksono	https://lh3.googleusercontent.com/...	game ini sangat seru dan seru banget	5	1644360000		2024-01-15 10:00:00	None	5	1644360000

Gambar 3. 1 Hasil data selection

3.2 Labeling

labeling digunakan untuk mengkategorikan ulasan pengguna game Stumble Guys. Jika rating berada pada rentang 1 sampai 2, sentimen akan diberi label negatif. Sebaliknya, jika rating berada pada rentang 4 dan 5, sentimen akan diberi label positif. Proses ini dilakukan dengan menerapkan fungsi pelabelan pada kolom score Dapat dilihat pada gambar 3.2

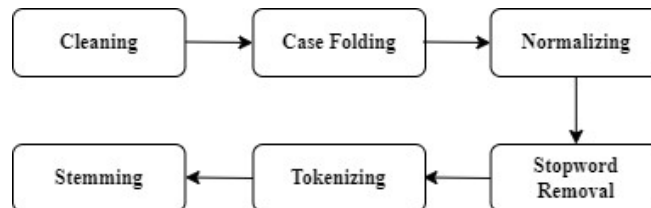
	content	score	Label
113	Di map super slide terkadang ada bug yaitu tom...	3	None
698	Saran saya untuk iklan nya mohon di hilangkan ...	5	Positif
0	Gamenya bagus,seru,dan banyak pilihan map dan ...	5	Positif
1374	game apaan si main block dash malah bag jump k...	3	None
1003	Bang jangan di update nya terus-terusan saya m...	3	None
...
807	Scopely tolong dong Benerin stumble guysnya bi...	4	Positif
914	Aku kasih 5,tapi saranku ke developer mohon il...	5	Positif
909	Gamenya bagus banget banyak skin baru map baru...	5	Positif
896	1.Knp ya pingnya tuh tiba-tiba nambah ke 987 p...	1	Negatif
589	Baru aja download ulang dan developer nya bene...	4	Positif

1500 rows x 3 columns

Gambar 3. 2 Hasil labeling

3.3 Preprocessing

Dalam tahap Preprocessing, penulis melakukan serangkaian langkah yang diperlukan karena dateset tidak terstruktur. Penulis mengambil ulasan memakai scraping untuk memaksimalkan hasil Analisis Sentimen. Diagram berikut menggambarkan urutan langkah pra-pemrosesan.



Gambar 3. 3 Langkah preprocessing

3.3.1 Cleaning

Pada gambar 5 ditunjukkan hasil dari proses text cleaning di mana dilakukan penghapusan karakter khusus dan tanda baca, yang tidak diperlukan dalam teks

	content	score	Label
0	sep	5	Positif
1	Aplikasi nya berat	1	Negatif
2	Gamernya bagus	5	Positif
3	HAPUS BUG SERVER NYA DONG MASA GUA MAU MAIN HA...	1	Negatif
4	Game nya seru tetapi suka banget ngeleg banget	5	Positif
...
1495	Mewing	5	Positif
1496	Saat Map Mr Beast, Ada banyak sekali bug, Saat...	1	Negatif
1497	saya kasih bintang 1 karena saat nak ngebuka ga...	1	Negatif
1498	game ini haram anj	1	Negatif
1499	Banyak bug	1	Negatif

1500 rows x 3 columns

Gambar 3.4 Hasil cleaning

3.3.2 Case folding

Dalam tahap Case Folding, semua kata atau kalimat diubah menjadi huruf kecil. Ini merupakan hasil dari proses tersebut.

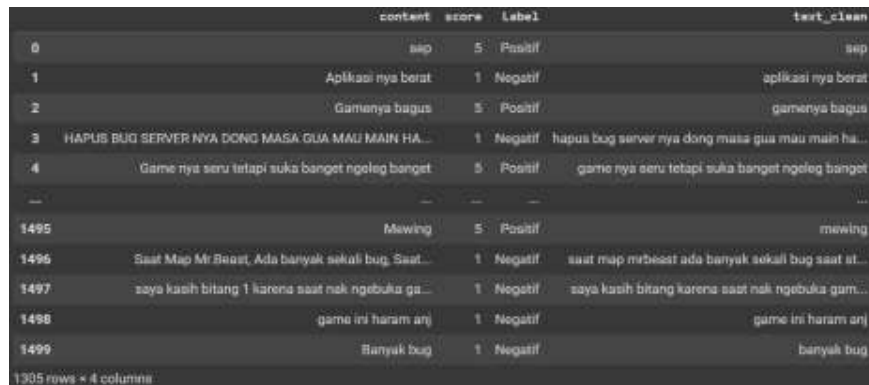
	content	score	Label	text_clean
0	sep	5	Positif	sep
1	Aplikasi nya berat	1	Negatif	aplikasi nya berat
2	Gamernya bagus	5	Positif	gamernya bagus
3	HAPUS BUG SERVER NYA DONG MASA GUA MAU MAIN HA...	1	Negatif	hapus bug server nya dong masa gua mau main ha...
4	Game nya seru tetapi suka banget ngeleg banget	5	Positif	game nya seru tetapi suka banget ngeleg banget
...
1495	Mewing	5	Positif	mewing
1496	Saat Map Mr Beast, Ada banyak sekali bug, Saat...	1	Negatif	saat map mrbeast ada banyak sekali bug saat et...
1497	saya kasih bintang 1 karena saat nak ngebuka ga...	1	Negatif	saya kasih bintang karena saat nak ngebuka gam...
1498	game ini haram anj	1	Negatif	game ini haram anj
1499	Banyak bug	1	Negatif	banyak bug

1500 rows x 4 columns

Gambar 3.5 Hasil case folding

3.3.3 Normalizing

Pada tahap Normalizing, dilakukan penyesuaian atau transformasi teks ke dalam bentuk standar atau format yang lebih terstruktur.



	content	score	Label	text_clean
0	sep	5	Positif	sep
1	Aplikasi nya berat	1	Negatif	aplikasi nya berat
2	Gamenya bagus	5	Positif	gamenya bagus
3	HAPUS BUG SERVER NYA DONG MASA GUA MAU MAIN HA...	1	Negatif	hapus bug server nya dong masa gua mau main ha...
4	Game nya seru tetapi suka banget ngeleg banget	5	Positif	game nya seru tetapi suka banget ngeleg banget
...
1495	Mewing	5	Positif	mewing
1496	Saat Map Mr Beast, Ada banyak sekali bug. Saat...	1	Negatif	saat map mrbeast ada banyak sekali bug saat it...
1497	saya kasih bintang 1 karena saat nak ngobrol ga...	1	Negatif	saya kasih bintang karena saat nak ngobrol gam...
1498	game ini haram anj	1	Negatif	game ini haram anj
1499	Banyak bug	1	Negatif	banyak bug

Gambar 3. 6 Hasil normalizing

3.3.4 Stopword removal

Pada tahap Stopword Removal, dilakukan penghapusan kata-kata yang tidak berarti atau tidak penting dalam teks, Pada tahap penghapusan stopwords, tujuannya adalah untuk hanya memfokuskan pada kata-kata yang penting.



	content	score	Label	text_clean	text_stopword
0	sep	5	Positif	sep	sep
1	Aplikasi nya berat	1	Negatif	aplikasi nya berat	aplikasi nya berat
2	Gamenya bagus	5	Positif	gamenya bagus	gamenya bagus
3	HAPUS BUG SERVER NYA DONG MASA GUA MAU MAIN HA...	1	Negatif	hapus bug server nya dong masa gua mau main ha...	hapus bug server nya gua mau main game server pas...
4	Game nya seru tetapi suka banget ngeleg banget	5	Positif	game nya seru tetapi suka banget ngeleg banget	game nya seru suka banget ngeleg banget
...
1495	Mewing	5	Positif	mewing	mewing
1496	Saat Map Mr Beast, Ada banyak sekali bug. Saat...	1	Negatif	saat map mrbeast ada banyak sekali bug saat it...	map mrbeast bug stage landing screen melompat ...
1497	saya kasih bintang 1 karena saat nak ngobrol ga...	1	Negatif	saya kasih bintang karena saat nak ngobrol gam...	kasih bintang nak ngobrol gamenya nolongin pem...
1498	game ini haram anj	1	Negatif	game ini haram anj	game haram anj
1499	Banyak bug	1	Negatif	banyak bug	bug

Gambar 3. 7 Hasil stopword removal

3.3.5 Tokenizing

Langkah Tokenizing adalah proses berikutnya yang memungkinkan dokumen untuk dibagi menjadi kata-kata terpisah.

konten	scene	Label	text_clean	text_tokenized	text_tukana
0	0	Positif	hey	hey	[hey]
1	1	Negatif	aplikasi nya berat	aplikasi nya berat	[aplikasi, nya, berat]
2	0	Positif	permainannya bagus	permainannya bagus	[permainannya, bagus]
3	1	Negatif	hapus bug server nya dong masa gua main main itu...	hapus bug server nya dong masa gua main main itu...	[hapus, bug, server, nya, dong, masa, gua, main, main, itu, ...]
4	0	Positif	game nya seru banget udah banget ngelung banget	game nya seru banget udah banget ngelung banget	[game, nya, seru, banget, udah, banget, ngelung, banget]
1485	0	Positif	menyukai	menyukai	[menyukai]
1486	1	Negatif	masi mau rebawati ada banyak enkali bag saat...	masi mau rebawati ada banyak enkali bag saat...	[masi, mau, rebawati, ada, banyak, enkali, bag, saat, ...]
1487	1	Negatif	siya baahh bilang i kerena saat nak ngabuka ga...	siya baahh bilang i kerena saat nak ngabuka ga...	[siya, baahh, bilang, i, kerena, saat, nak, ngabuka, ga, ...]
1490	1	Negatif	game ini funam anj	game ini funam anj	[game, funam, anj]
1499	1	Negatif	banyak buat	buat	[buat]

Gambar 3. 8 Hasil tokenizing

3.3.6 Stemming

Tahap Stemming adalah lanjutan dari tahap sebelumnya. Tahapan ini dapat mengubah kata menjadi bentuk dasarnya. Jadi pada tahap ini kata-kata yang berlebihan akan disaring menjadi bentuk standarnya.

Tabel 3. 1 Hasil stemming

Kata asli	Hasil
Menyenangkan	Senang
Permainan	Main
Tantangan	Tantang
Karakter	Karakter

3.4 Transformation

Tahapan ini dimulai dengan membagi data menjadi dua bagian, yaitu data training dan data testing, dalam tiga skenario berbeda. Proses pembagian data ini dilakukan dengan menggunakan fungsi train test split, dengan pengaturan parameter random_state=0. Tabel 3.2, menampilkan hasil dari pembagian data dalam ketiga skenario tersebut.

Tabel 3. 2 Hasil pembagian data

Skenario rasio perbandingan	Data training	Data testing
90:10	1173	131
80:20	1043	261
70:30	912	392

3.5 Evaluasi

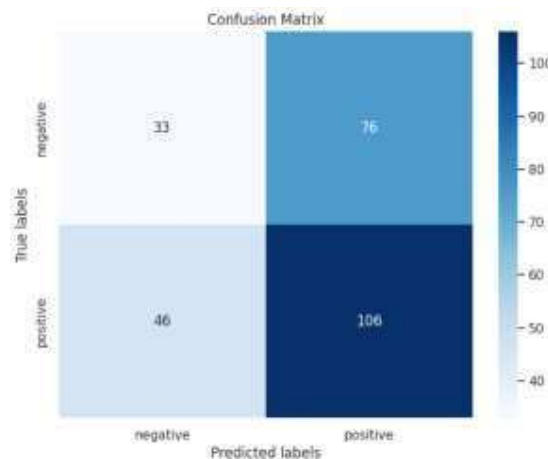
Tahap ini dilakukan evaluasi dengan confusion matrix. Tabel 3.3 memperlihatkan hasil perbandingan dari seluruh evaluasi confusion matrix terhadap setiap skenario.

Tabel 3. 3 Hasil perbandingan akurasi

Training	Testing	Akurasi
80	20	86%
70	30	85%
90	10	81%

Di Tabel 3.3 terdapat hasil akurasi dari berbagai metode split data, seperti 80- 20 dengan akurasi 88%, 70-30 dengan akurasi 85%, 90-10 dengan akurasi 81%, Gambar menunjukkan visualisasi hasil analisis sentimen untuk klasifikasi sentimen positif dan negatif.

Menurut tabel 3,3 pemodelan dengan menggunakan algoritma naive bayes pada skenario 80:20 menghasilkan nilai akurasi 86%. Berikut adalah contoh matrix confusion dari pemodelan:



Gambar 3. 9 Confusion matrix


```

MultinomialNB Accuracy: 0.8620689655172413
MultinomialNB Precision: 0.8461538461538461
MultinomialNB Recall: 0.7777777777777778
MultinomialNB f1_score: 0.8105263157894737
confusion_matrix:
[[ 77 22]
 [ 14 148]]
=====
              precision    recall  f1-score   support

   Negatif      0.85      0.78      0.81        99
   Positif      0.87      0.91      0.89       162

 accuracy              0.86       261
 macro avg      0.86      0.85      0.85       261
 weighted avg   0.86      0.86      0.86       261

 True Positive : 148
 True Negative : 77
 False Positive : 22
 False Negative : 14

```

Gambar 3. 11 Hasil accuracy

3.8 confusion matrix

Confusion matrix adalah alat evaluasi yang digunakan untuk mengukur performa metode klasifikasi. Ini memberikan informasi tentang seberapa baik sistem mampu mengklasifikasikan data ke dalam kategori yang benar. Confusion matrix juga berguna sebagai alat visualisasi untuk memahami hasil pembelajaran dari sistem, khususnya dalam konteks klasifikasi dengan dua kategori. Tabel di bawah ini adalah contoh hasil confusion matrix yang menunjukkan prediksi untuk dua kelas.

Tabel 3. 4 Sample data

Content	Score	Label
Game nya agak pay to win masa emote bisa nampol orang :/	4	Positif
sangat bagus sekali ya	5	Positif
GK jelas udh nungguin malahan GK bisa masuk	2	Negatif
Game tai	1	Negatif
Seru	5	Positif

Tabel 3. 5 Hasil confusion matrix

Predict Positif	Predict Negatif
77	22
14	148
TP Positif	TP Negatif
77	148
225	

Tabel di atas adalah hasil dari sebuah model klasifikasi yang diterapkan pada ulasan game 'Stmble Guys. Model ini melakukan prediksi terhadap ulasan sebagai positif atau negatif. Dari total 261 prediksi yang dilakukan, 77 di antaranya diprediksi positif dan tepat, sedangkan 148 diprediksi negatif dan tepat. Namun, model juga salah memprediksi 14 ulasan positif sebagai negatif dan 22 ulasan negatif sebagai positif.

Keterangan:

True Positives : Ada 77 ulasan yang sebenarnya positif dan berhasil diklasifikasikan sebagai positif oleh model.

True Negatives : Ada 148 ulasan yang sebenarnya negatif dan berhasil diklasifikasikan sebagai negatif oleh model.

False Positives : Ada 22 ulasan yang sebenarnya negatif, tetapi salah diklasifikasikan sebagai positif oleh model.

False Negatives : Ada 14 ulasan yang sebenarnya positif, tetapi salah diklasifikasikan sebagai negatif oleh model.

Akurasi

$$Akurasi = \frac{TP\ Positif + TP\ Negatif}{Total\ Prediksi}$$

$$Akurasi = \frac{77 + 148}{261}$$

$$Akurasi = \frac{225}{261}$$

$$Akurasi = \frac{225}{261} = 0,8621 \text{ atau } 86,21\%$$

$$Akurasi = \frac{77 + 148}{261} = \frac{225}{261} = 86,21\%$$

$$Akurasi = \frac{TP \text{ Positif} + Tp \text{ Negetif}}{\text{Total Prediksi}} = \frac{77 + 148}{261} = \frac{225}{261} = 86,21\%$$

Precision

Negatif

$$Presisi \text{ Negatif} = \frac{TP \text{ Negatif}}{\text{Prediksi Negatif}} = \frac{77}{77 + 14} = \frac{77}{91} = 0,85\%$$

Positif

$$Presisi \text{ Positif} = \frac{TP \text{ Positif}}{\text{Prediksi Positif}} = \frac{148}{148 + 22} = \frac{148}{170} = 0,87$$

Recall

Negatif

$$Recall \text{ Negatif} = \frac{TP \text{ Negatif}}{TP \text{ Negatif} + FN \text{ Negatif}} = \frac{77}{77 + 22} = \frac{77}{99} = 0,78\%$$

Positif

$$Recall \text{ Positif} = \frac{TP \text{ Positif}}{TP \text{ Positif} + FN \text{ Positif}} = \frac{148}{148 + 14} = \frac{148}{162} = 0,91\%$$

F1-Score

Negatif

$$F1 \text{ Score} = 2 \cdot \frac{\text{Presisi} \cdot \text{Recall}}{\text{Presisi} + \text{Recall}}$$

$$F1 \text{ Score Negatif} = 2 \cdot \frac{0,85 \cdot 0,77}{0,85 + 0,77} = 2 \cdot \frac{0,6545}{1,62} = 0,4034 = 0,8068 = 0,81\%$$

Positif

$$F1 \text{ Score positif} = 2 \cdot \frac{\text{Presisi positif} \cdot \text{Recall Positif}}{\text{Presisi positif} + \text{Recall Positif}}$$

$$F1 \text{ Score Positif} = 2 \cdot \frac{0,87 \cdot 0,91}{0,87 + 0,91}$$

$$F1 \text{ Score Positif} = 2 \cdot \frac{0,796}{1,78}$$

$$F1 \text{ Score Positif} = 2 \cdot 0,4446$$

$$F1 \text{ Score Positif} = 0,89\%$$

Dengan menggunakan metode klasifikasi yang diimplementasikan, model mencapai tingkat akurasi sebesar 86,21%. Artinya, dari seluruh prediksi yang dilakukan, model berhasil memprediksi dengan tepat 86,21% kasus positif maupun negatif. Presisi negatif sebesar 85% menunjukkan bahwa sebagian besar dari prediksi yang diklasifikasikan sebagai negatif oleh model memang benar-benar negatif, sementara presisi positif sebesar 87% menunjukkan hal serupa untuk prediksi positif. Dari sisi recall, model mampu mengenali 78% kasus negatif secara keseluruhan dan 91% kasus positif, mencerminkan kemampuan model dalam mengidentifikasi ulasan yang sesuai dengan kategori masing-masing. F1-score positif sebesar 0,89% menggambarkan harmonisasi antara presisi dan recall positif, memberikan gambaran tentang performa model dalam memprediksi ulasan game Stumble Guys berdasarkan evaluasi yang dilakukan.