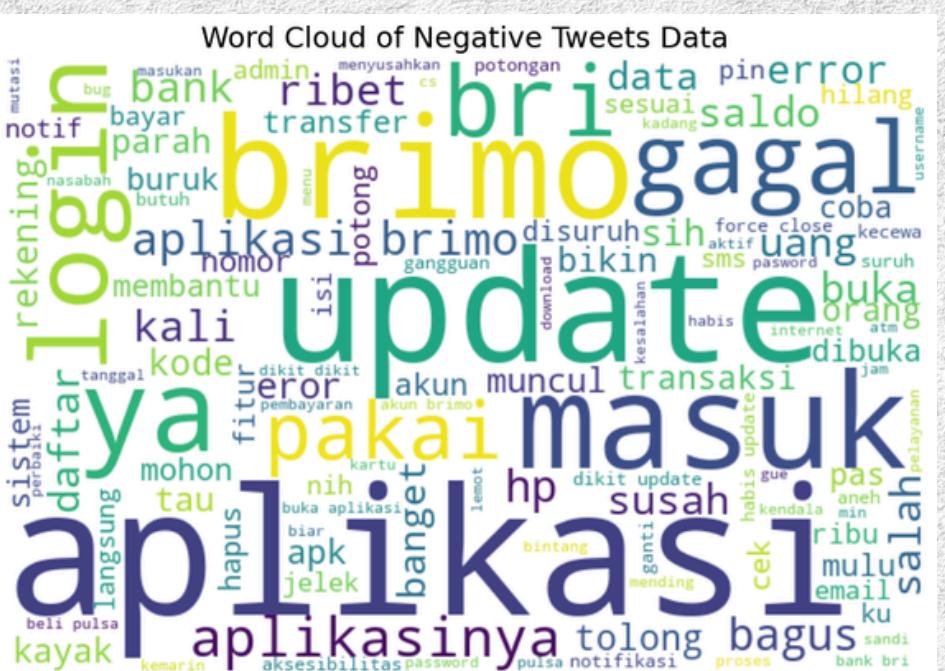


CODING CAMP POWERED BY DBS FOUNDATION

MOBILE BRIMO APP

Review Sentiment Analysis dan prediction from Scraping to Deep Learning Insights



APA KATA PENGGUNA TENTANG BRIMO?



#LifeAtCodingCamp

#LearnAtDicoding

PROJECT OVERVIEW



Proyek ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi BRImo yang tersedia di PlayStore, guna memahami persepsi dan pengalaman pengguna secara mendalam.



Analisis sentimen ini sangat penting karena dapat membantu meningkatkan kualitas layanan, mendeteksi masalah sejak dini, dan menyusun strategi pemasaran yang lebih efektif.



Metodologi utama yang digunakan meliputi pengumpulan data melalui web scraping, pembersihan dan transformasi data teks, serta penerapan model deep learning (CNN, GRU, dan LSTM) untuk mengklasifikasikan sentimen ulasan.

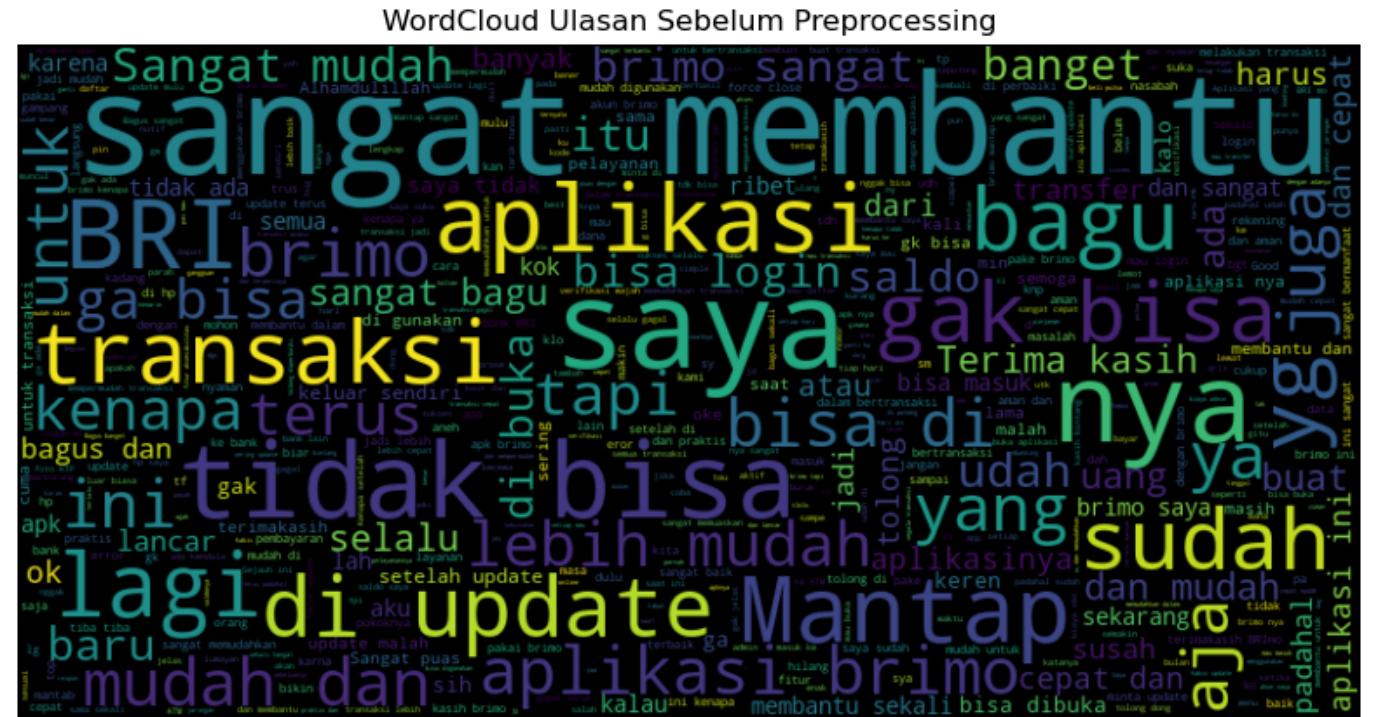


DATA COLLECTION & SCRAPING

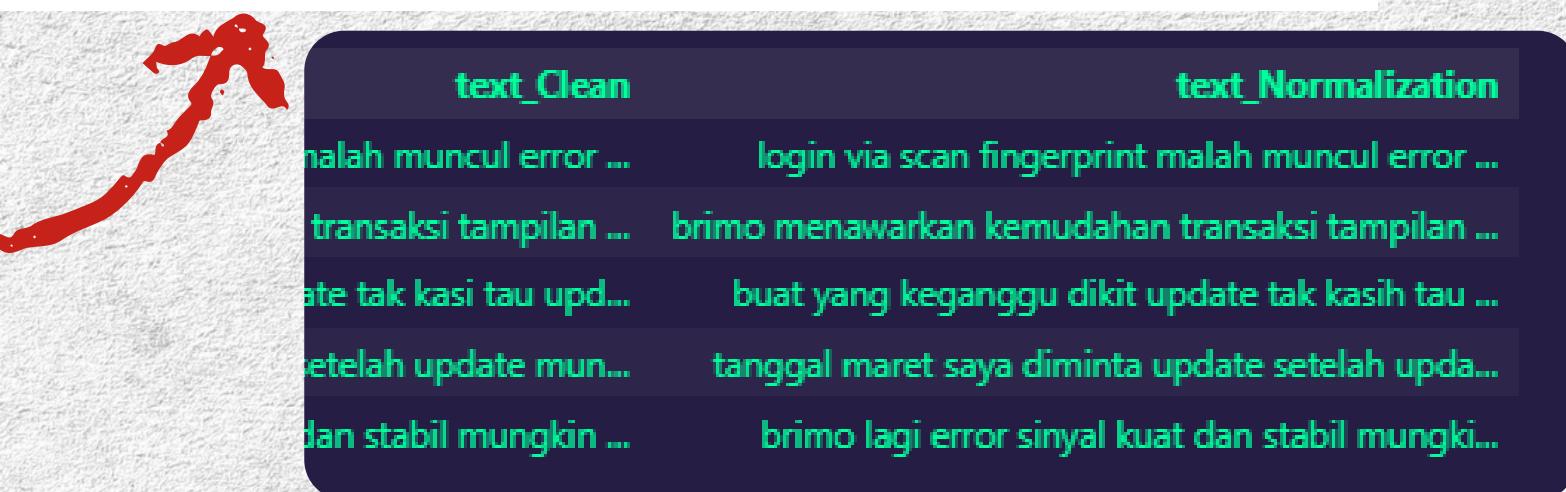
Manfaatkan library google-play-scraper untuk secara otomatis mengumpulkan 30.000 ulasan dari PlayStore. Proses ini mencakup pengambilan data, validasi, dan penyimpanan, yang divisualisasikan melalui diagram alur untuk memudahkan pemahaman metode kerja yang digunakan.



DATA PREPROCESSING & EXPLORATORY ANALYSIS



PREPROCESSING TEKS - CLEANING & NORMALIZATION



text Clean	text Normalization
malah muncul error ...	login via scan fingerprint malah muncul error ...
transaksi tampilan ...	brimo menawarkan kemudahan transaksi tampilan ...
ate tak kasi tau upd...	buat yang keganggu dikit update tak kasih tau ...
setelah update mun...	tanggal maret saya diminta update setelah upda...
dan stabil mungkin ...	brimo lagi error sinyal kuat dan stabil mungkin...

WordCloud "Sebelum Preprocessing" untuk identifikasi noise

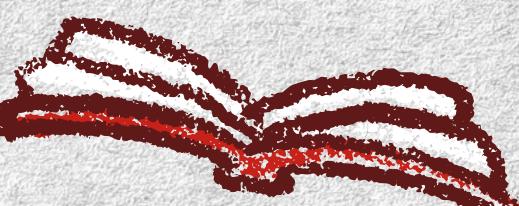
Pembersihan Teks

Hapus URL, mention, hashtag, angka, karakter spesial, dan tanda baca

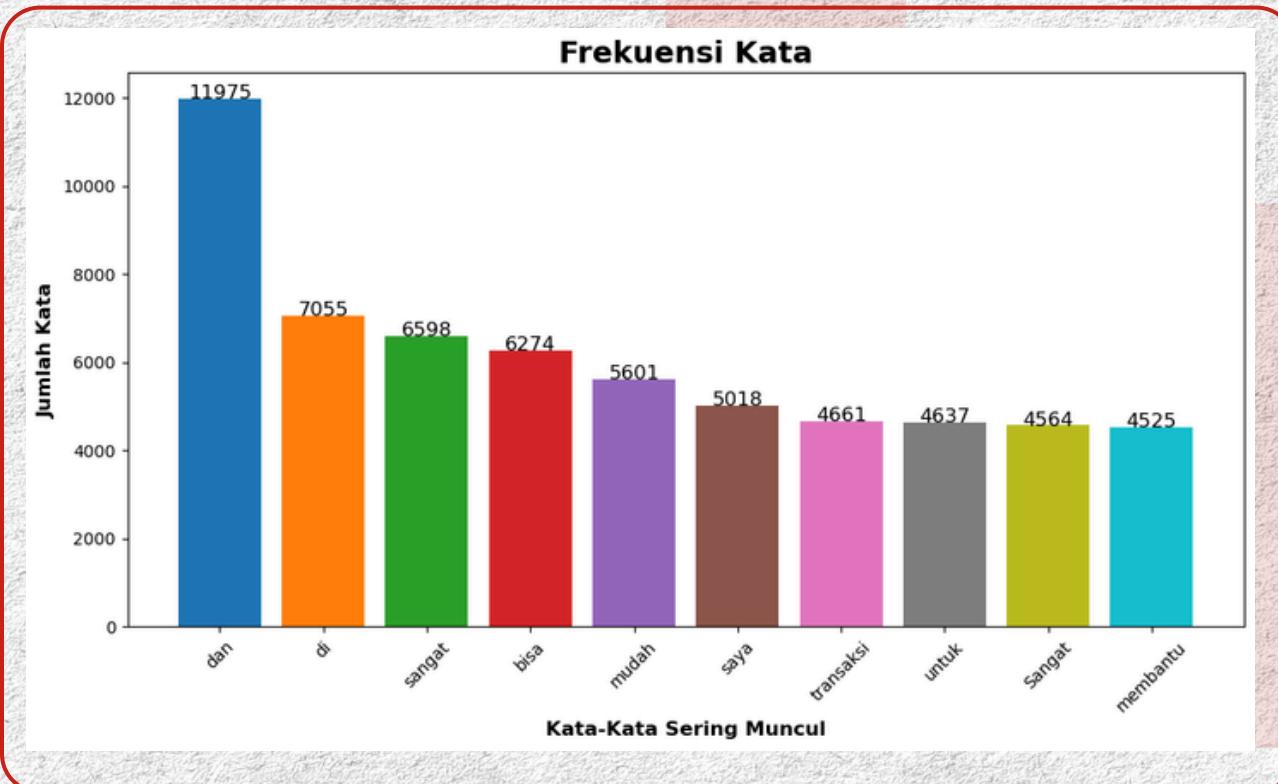
Ubah teks menjadi lowercase

Normalisasi & Slang Replacement

Ganti kata-kata slang ke bentuk baku menggunakan kamus slang



PREPROCESSING TEKS – TOKENISASI, FILTERING & STEMMING



grafik batang frekuensi kata

Penghapusan Stopwords

text_StopWord

login via scan fingerprint muncul error
transa...

brimo menawarkan kemudahan transaksi
tampilan ...

keganggu dikit update kasih tau update
perbaik...

Stemming

text_Stemmindo

[login, via, scan, fingerprint,
muncul, error,...

[brimo, tawar, mudah, transaksi,
tampil, intui...

[ganggu, dikit, update, kasih,
tau, update, ba...

Tokenisasi

text_Tokenization

[login, via, scan, fingerprint, muncul,
error,...

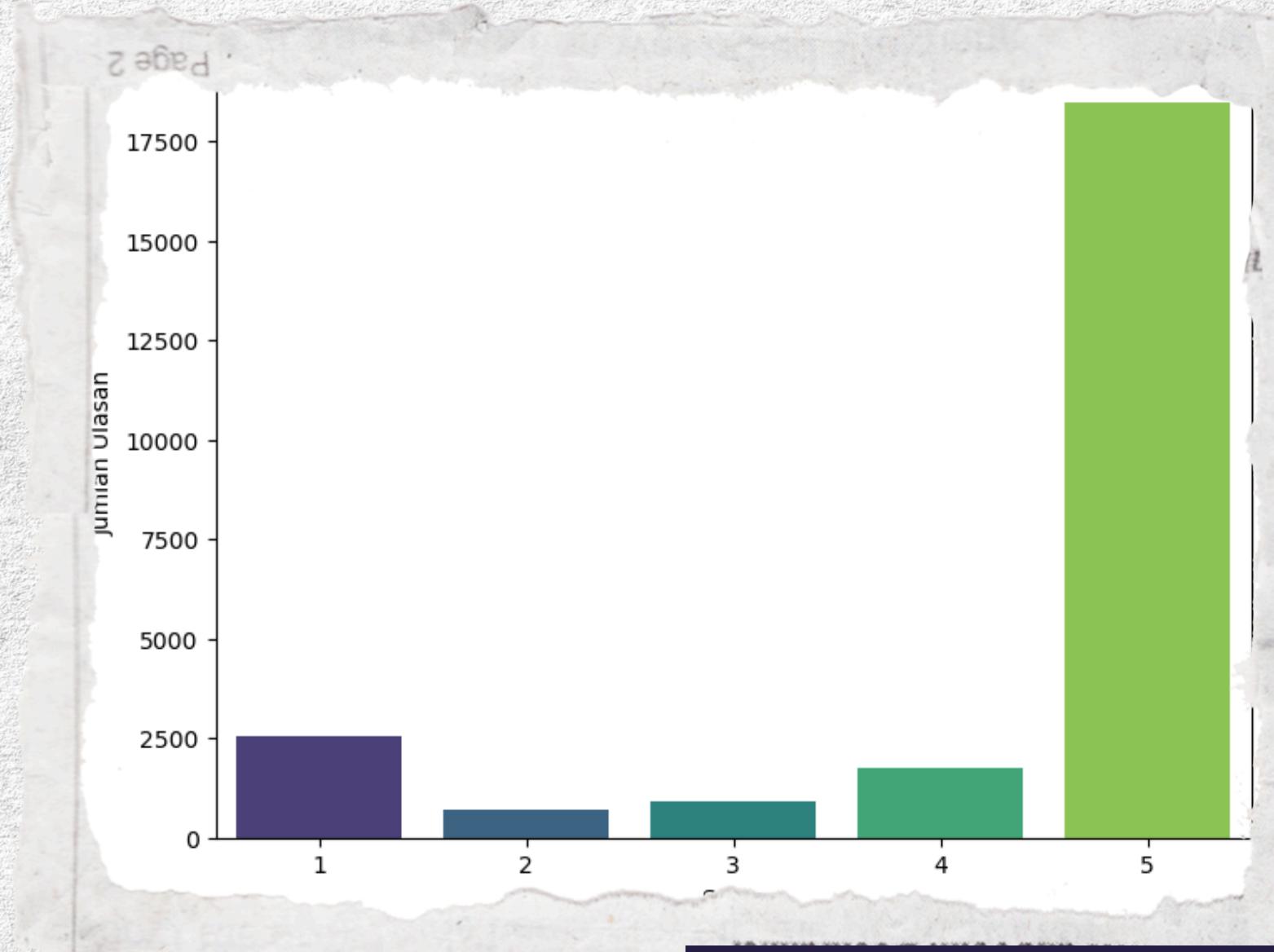
[brimo, menawarkan, kemudahan,
transaksi, tamp...

[keganggu, dikit, update, kasih, tau,
update, ...

[tanggal, maret, update, update, muncul,
fitur...

- Tokenisasi:
Pecah teks menjadi token menggunakan RegexpTokenizer/word_tokenize
- Penghapusan Stopwords:
Hapus stopwords Bahasa Indonesia & Inggris untuk mengurangi noise
- Stemming:
Ubah token ke bentuk dasar menggunakan Sastrawi (proses stemming dalam bahasa Indonesia)

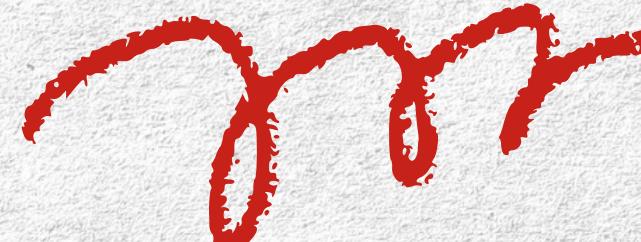
PELABELAN DATA & POLARITAS SENTIMEN



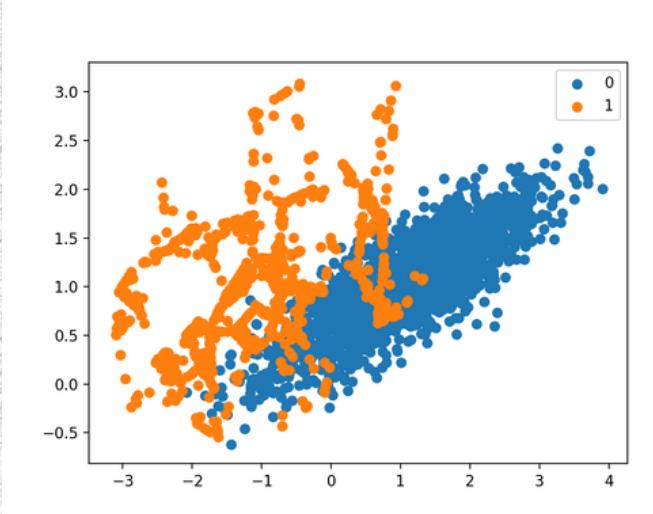
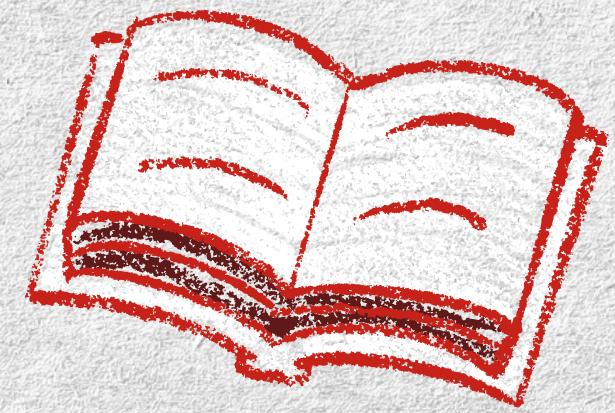
* Menghitung Polaritas Sentimen

Dalam tahap ini, lexicon positif dan negatif diunduh dari GitHub untuk menghitung skor sentimen tiap ulasan. Fungsi `sentiment_analysis_lexicon_indonesia` mengakumulasi skor berdasarkan kata-kata positif dan negatif, kemudian menetapkan label: positif (skor ≥ 0), netral, atau negatif (skor ≤ -7). Hasil pelabelan menunjukkan distribusi 16.868 ulasan positif, 6.324 netral, dan 1.200 negatif, memberikan gambaran mendalam mengenai persepsi pengguna terhadap aplikasi.

```
polarity
positive    16868
neutral      6324
negative     1200
Name: count, dtype: int64
```



MENANGANI DATA IMBALANCE & REPRESENTASI TEKS



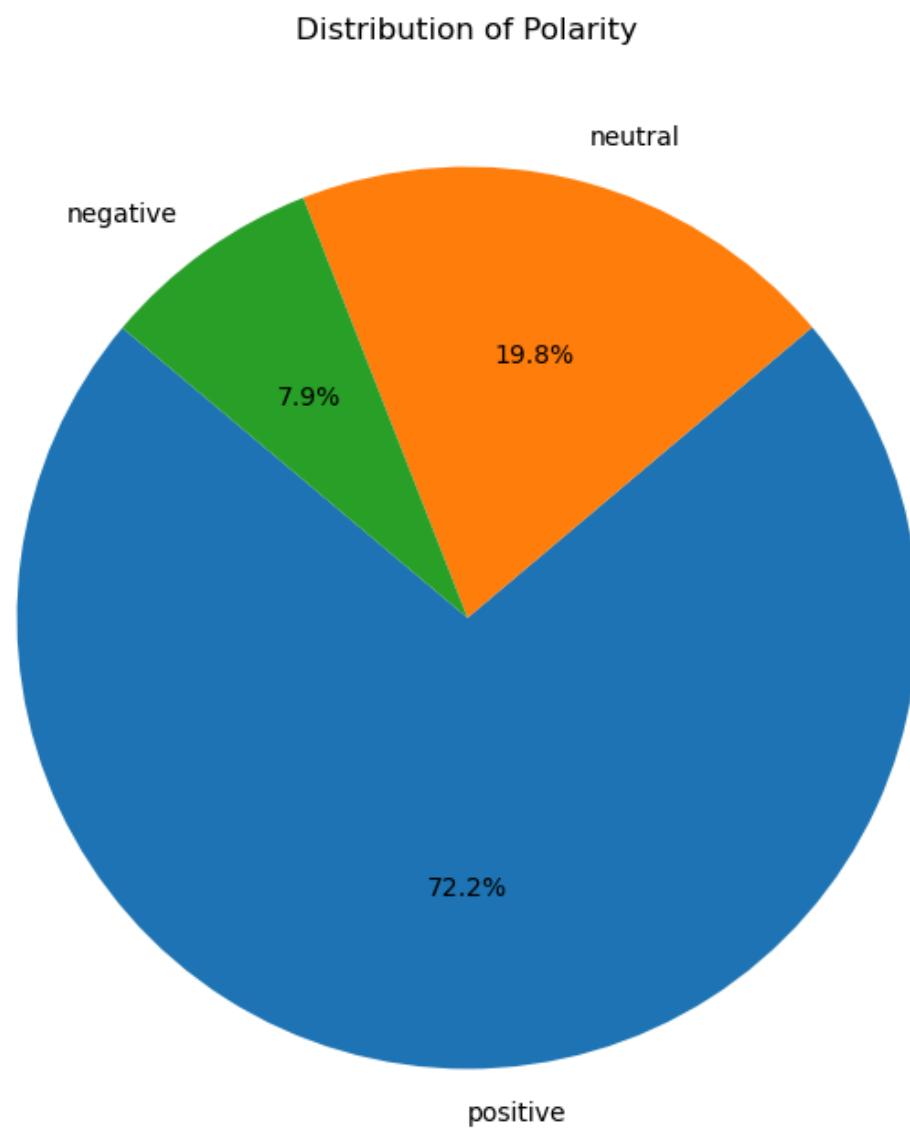
Synthetic Minority Oversampling
Technique

Untuk mengatasi ketidakseimbangan data, kami melakukan oversampling pada kategori minoritas menggunakan teknik seperti SMOTE atau Random Oversampling serta undersampling pada kategori mayoritas guna mencapai keseimbangan. Selain itu, representasi teks dioptimalkan dengan memanfaatkan Word Embeddings (seperti GloVe, Word2Vec, FastText, atau model contextual seperti ELMo, BERT, dan GPT) yang mampu menangkap hubungan semantik kata secara mendalam.

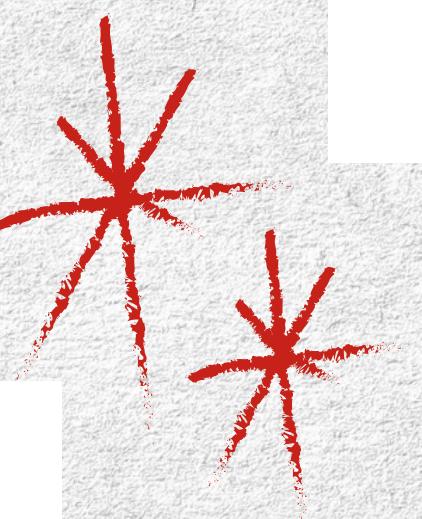
text_Clean	text_Normalization	text_StopWord	text_Tokenization	text_Stemindo	text_String
login via scan fingerprint malah muncul error ...	login via scan fingerprint malah muncul error ...	login via scan fingerprint muncul error transa...	[login, via, scan, fingerprint, muncul, error, ...]	[login, via, scan, fingerprint, muncul, error, ...]	login scan fingerprint muncul error transaksi ...
brimo menawarkan kemudahan transaksi tampilan ...	brimo menawarkan kemudahan transaksi tampilan ...	brimo menawarkan kemudahan transaksi tampilan ...	[brimo, menawarkan, kemudahan, transaksi, tampil, intui...	[brimo, tawar, mudah, transaksi, tampil, intui...	brimo tawar mudah transaksi tampil intuitif ak...
buat yg keganggu dikit update tak kasi tau upd...	buat yang keganggu dikit update tak kasi tau ...	keganggu dikit update kasih tau update perbaik...	[keganggu, dikit, update, kasih, tau, update, ...]	[ganggu, dikit, update, kasih, tau, update, ba...	ganggu dikit update kasih update baik celah le...
tgl maret sy diminta update setelah update mun...	tanggal maret saya diminta update setelah upda...	tanggal maret update update muncul fitur menon...	[tanggal, maret, update, update, muncul, fitur...]	[tanggal, maret, update, update, muncul, fitur...]	tanggal maret update update muncul fitur nonak...
brimo lg error sinyal kuat dan stabil mungkin ...	brimo lagi error sinyal kuat dan stabil mungkin...	brimo error sinyal kuat stabil pengaruh cuaca ...	[brimo, error, sinyal, kuat, stabil, pengaruh, ...]	[brimo, error, sinyal, kuat, stabil, pengaruh, ...]	brimo error sinyal kuat stabil pengaruh cuaca ...
di perangkat saya terkadang brimo tidak meresp...	di perangkat saya terkadang brimo tidak meresp...	perangkat terkadang brimo merespons aplikasi m...	[perangkat, terkadang, brimo, merespons, aplik...	[perangkat, terkadang, brimo, respons, aplikas...	perangkat terkadang brimo respons aplikasi mod...

DATA PREPROCESSING & EXPLORATORY ANALYSIS

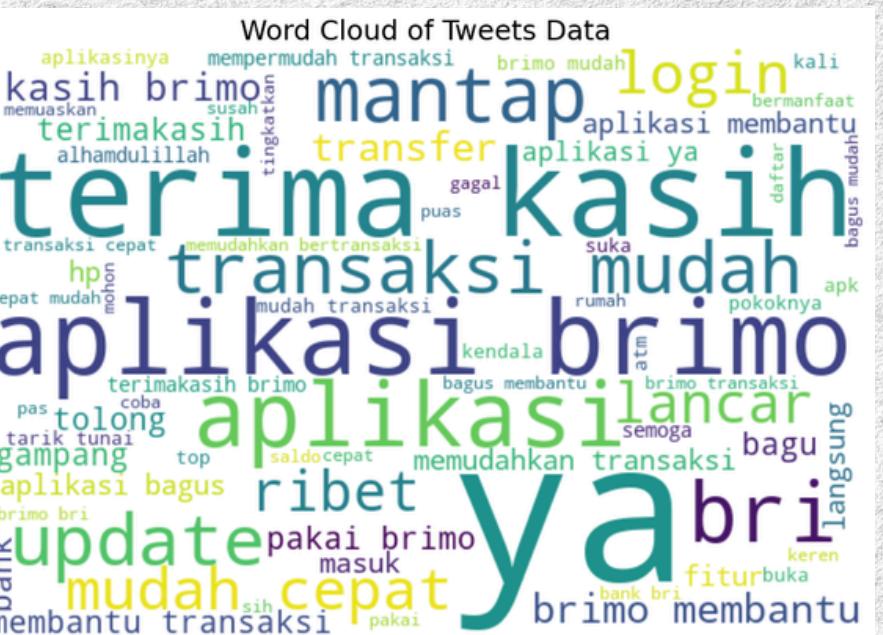
EXPLORATORY ANALYSIS & HYPERPARAMETER TUNING



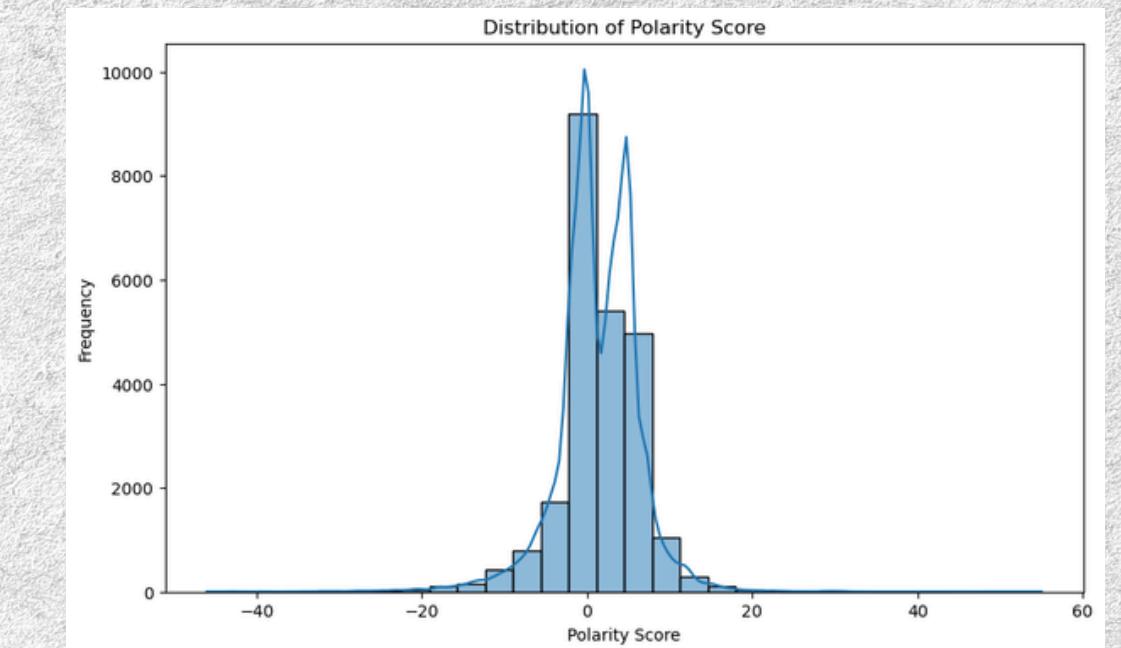
pie chart visualization



Visualisasi eksploratori dilakukan dengan WordCloud, grafik distribusi sentimen, dan grafik batang frekuensi kata, yang mengungkap bahwa mayoritas ulasan menunjukkan sentimen positif, mengindikasikan tingkat kepuasan pengguna yang tinggi. Untuk mengoptimalkan performa model, kami menerapkan teknik pencarian hyperparameter (Grid Search dan Random Search) serta mengeksplorasi model tambahan seperti Bidirectional LSTM, Attention, Transformer, atau BERT guna meningkatkan akurasi dan generalisasi model.



Word Cloud

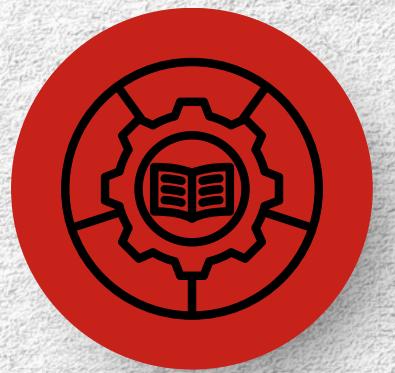


grafik distribusi sentimen



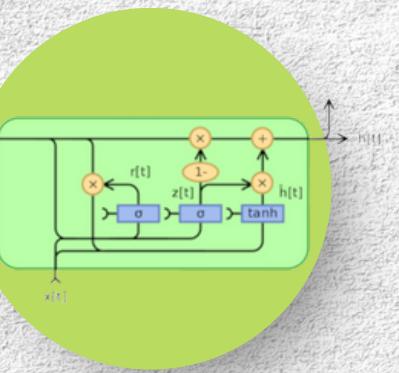
MODEL BUILDING

(PEMBANGUNAN MODEL MACHINE LEARNING)



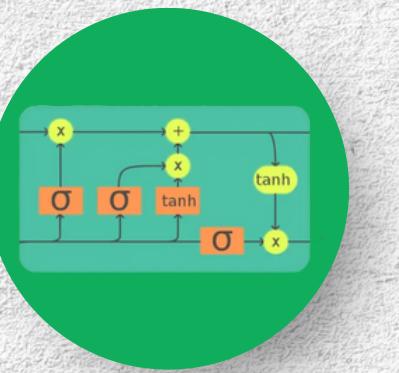
CNN

Menangkap fitur lokal
dari teks



GRU

Efisien dalam
menangani dependensi
sekuensial

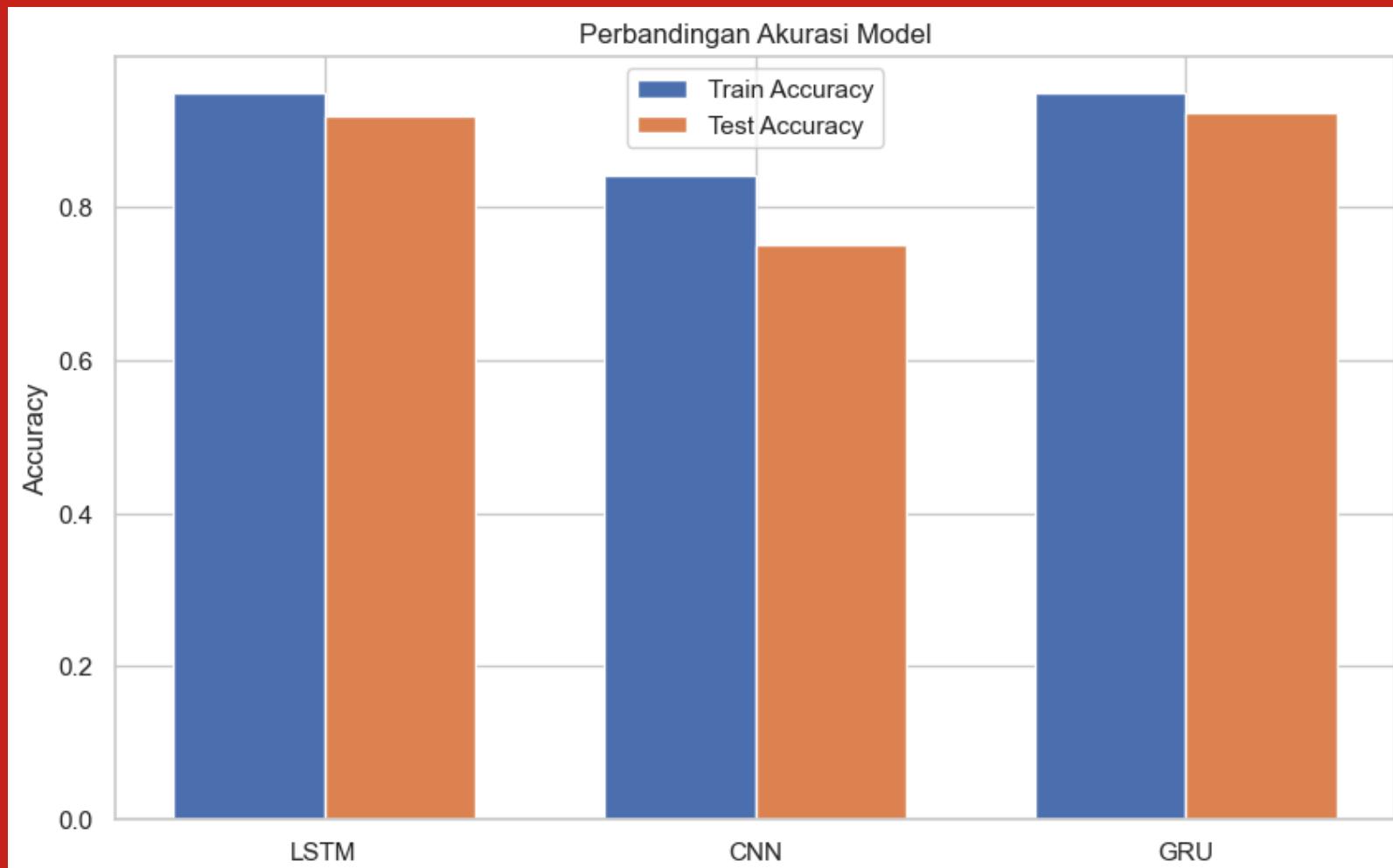


LSTM

Mampu memahami konteks
jangka panjang dalam ulasan
Diagram arsitektur
sederhana masing-masing
model (opsional) membantu
memvisualisasikan peran
unik dan kontribusi tiap
model dalam proses
ekstraksi fitur dan
pemahaman konteks.



MODEL EVALUATION & PERFORMANCE



01

LSTM

Akurasi training sebesar 95,01% dan akurasi testing mencapai 91,93%.

02

CNN

Akurasi training mencapai 84,19%, sedangkan akurasi testing hanya sebesar 75,08%.

03

GRU

Memperoleh akurasi training sebesar 95,02% dan akurasi testing tertinggi sebesar 92,46%.

Model	Accuracy Train	Accuracy Test
0 LSTM	0.950097	0.919262
1 CNN	0.841943	0.750820
2 GRU	0.950166	0.924590

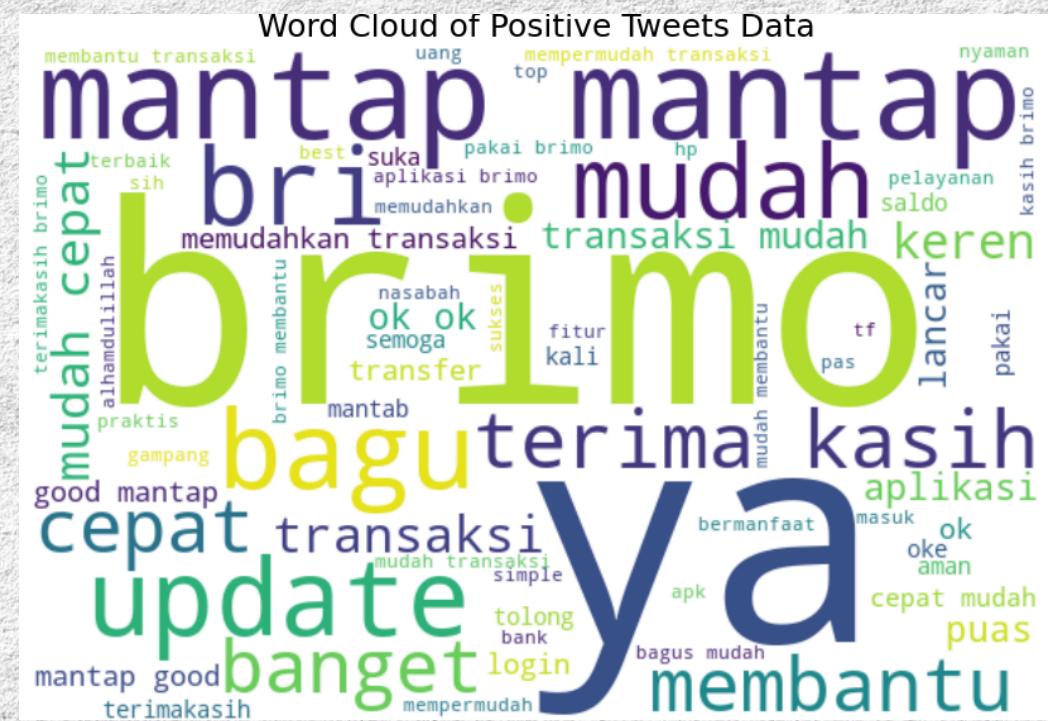
Mengukur performa model dengan membandingkan akurasi training dan testing dari tiga model deep learning: LSTM, CNN, dan GRU. Berdasarkan hasil evaluasi, model LSTM dan GRU menunjukkan performa yang sangat baik dengan akurasi training masing-masing sekitar 95% dan akurasi testing mendekati 92%, sedangkan model CNN hanya mencapai akurasi training 84,19% dan akurasi testing 75,08%. Diagram batang yang divisualisasikan secara jelas menampilkan perbandingan kinerja, sehingga model LSTM dan GRU dapat dianggap sebagai pilihan terbaik untuk klasifikasi sentimen.

KEY RESULTS & INSIGHTS

Temuan utama dari analisis sentimen menunjukkan bahwa mayoritas ulasan pengguna bersentimen positif, mengindikasikan tingkat kepuasan yang tinggi. Namun, ulasan negatif juga mengungkap beberapa area perbaikan, seperti antarmuka pengguna dan kecepatan aplikasi, yang dapat dijadikan dasar untuk peningkatan fitur dan performa aplikasi ke depan. Visualisasi berupa infografis dengan ikon-ikon akan memudahkan pemahaman insight ini secara menyeluruh.

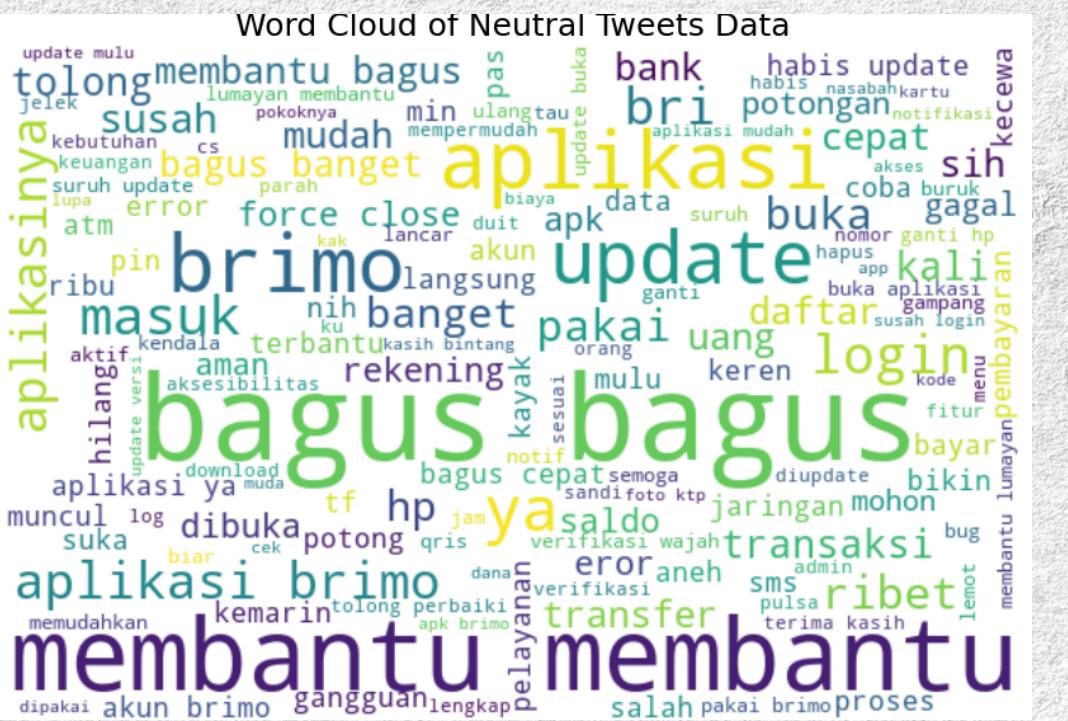
* Positive

Menonjolkan kata-kata yang menunjukkan kepuasan dan pujian pengguna.



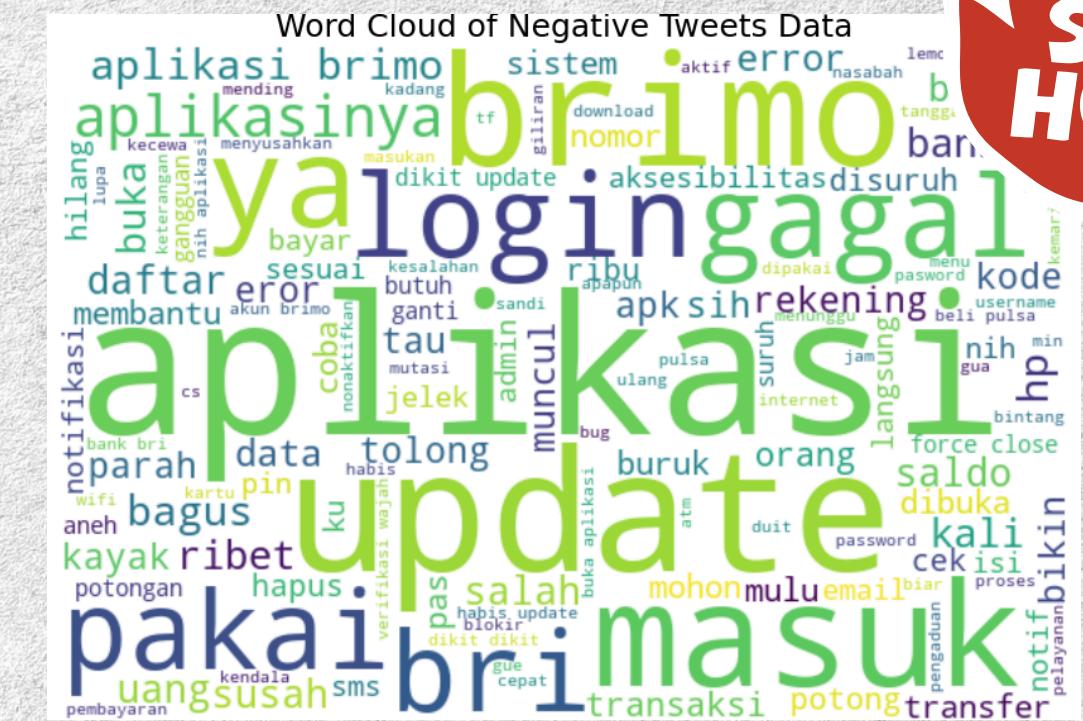
* Netral

Mengungkap kata-kata umum yang tidak mengandung emosi kuat.



* Negative

enampilan kata-kata dominan yang mencerminkan kritik dan ketidakpuasan.



MEMPREDIKSI LABEL PADA DATA BARU

PREDICTION

```
1/1 ----- 0s 46ms/step
```

```
1/1 ----- 0s 46ms/step
```

```
1/1 ----- 0s 54ms/step
```

Text: keren dan menyenangkan

True Label: positive

Predicted Label (LSTM): positive

Predicted Label (CNN): positive

Predicted Label (GRU): positive

Text: Sayangnya, aplikasi BRImo sering mengalami error dan loading yang lama,

True Label: negative

Predicted Label (LSTM): negative

Predicted Label (CNN): negative

Predicted Label (GRU): negative

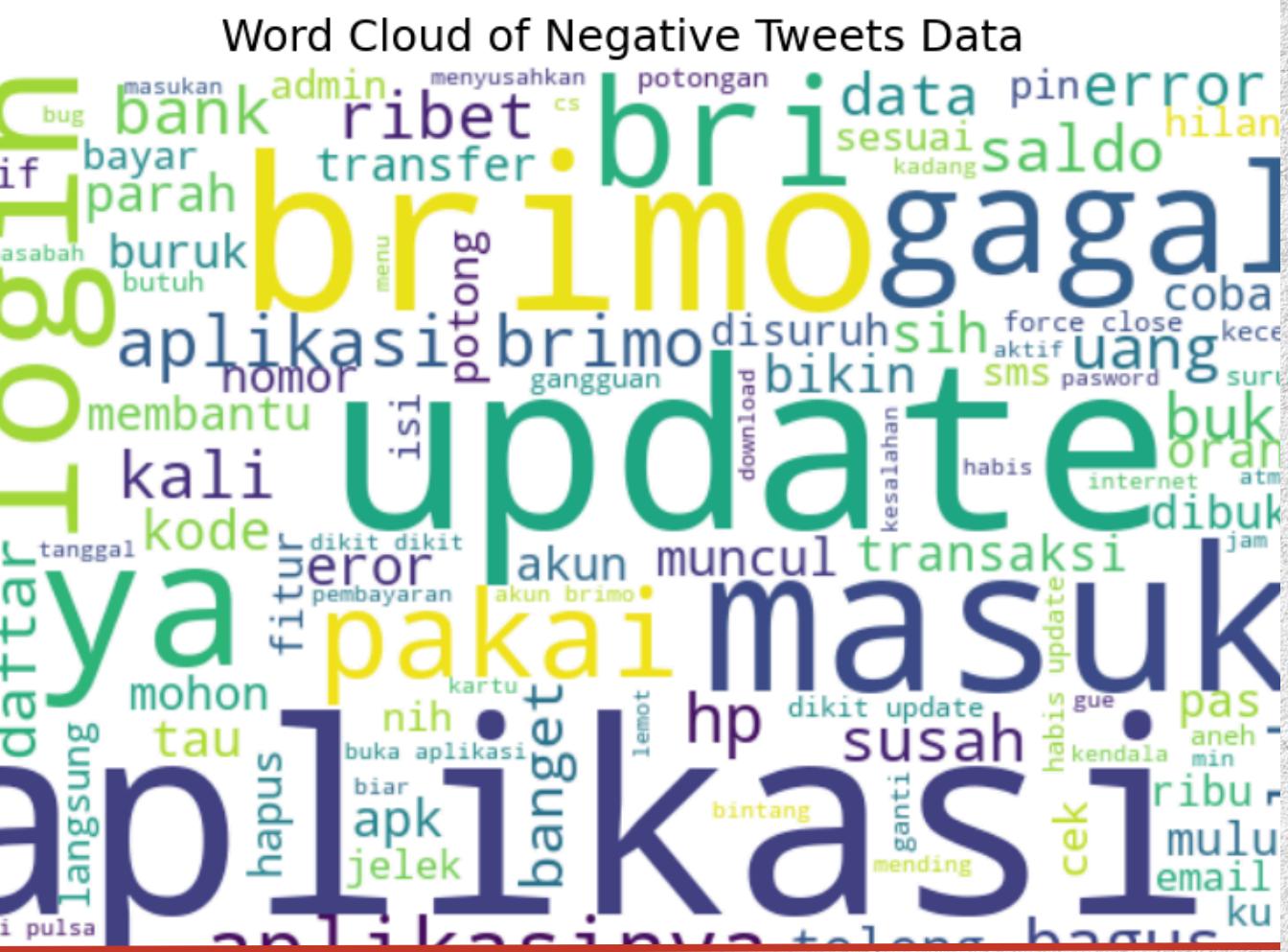
Text: belum optimal tetapi lumayan bagus

True Label: neutral

Predicted Label (LSTM): neutral

Predicted Label (CNN): neutral

Predicted Label (GRU): neutral



* Hasil

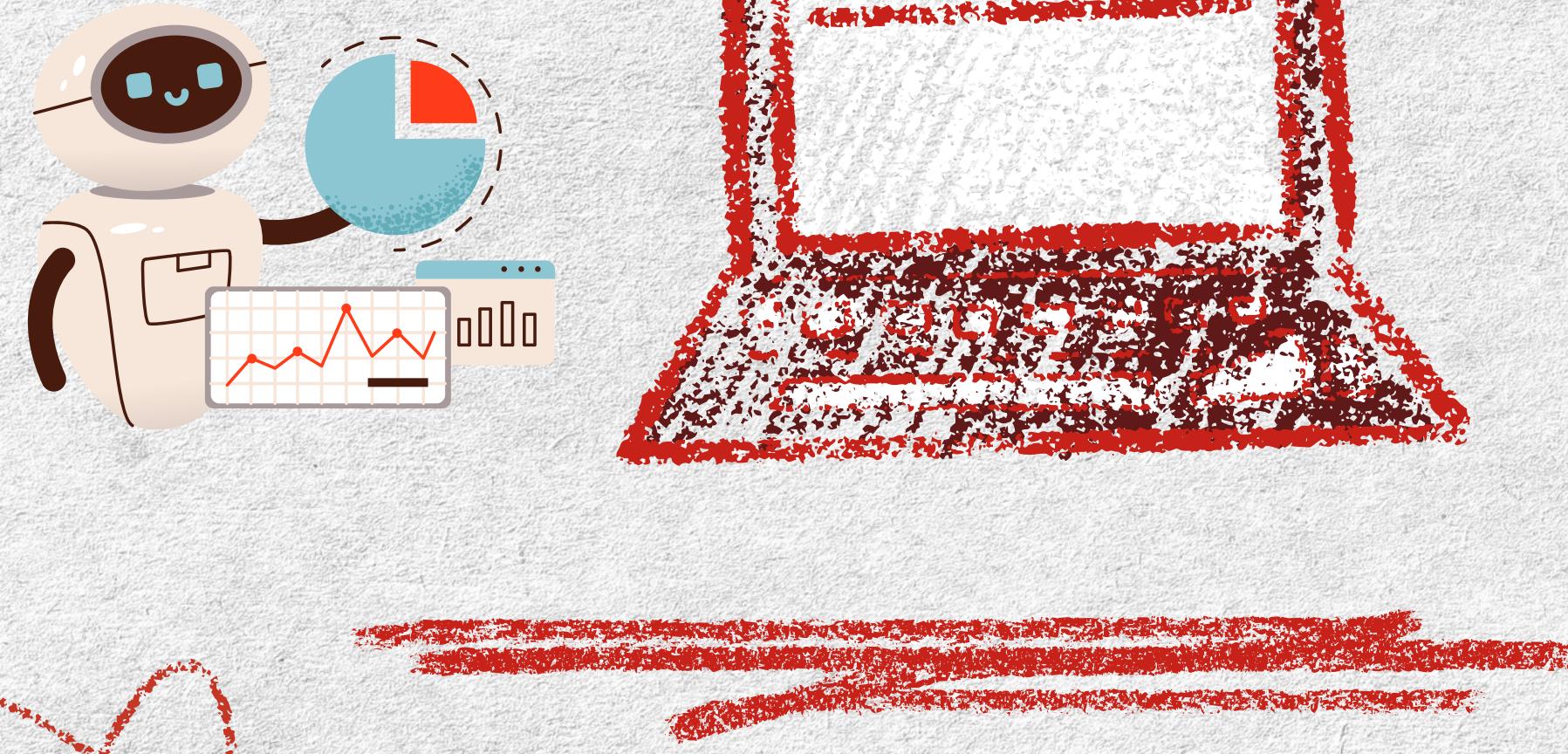
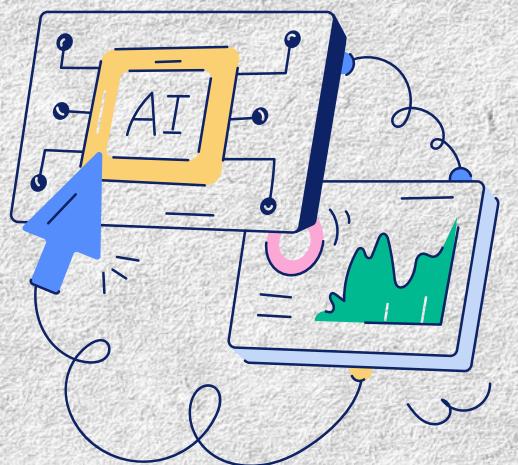
Pada tahap ini, kami menggunakan model LSTM, CNN, dan GRU yang telah dilatih untuk memprediksi sentimen dari ulasan baru. Data baru, seperti "keren dan menyenangkan", "Sayangnya, aplikasi BRImo sering mengalami error dan loading yang lama, tidak bisa login", dan "belum optimal tetapi lumayan bagus", ditokenisasi dan dipadding sesuai panjang input model. Selanjutnya, prediksi dari ketiga model dikonversi ke label melalui argmax dan labelencoder. Hasil prediksi menunjukkan bahwa ketiga model berhasil mengklasifikasikan ulasan dengan benar, sesuai dengan label asli masing-masing.

THANK YOU



Contact & Call-to-Action

- 🌐 <https://github.com/sionpardosi>
- ✉️ spardosi12@gmail.com



#LifeAtCodingCamp
#LearnAtDicoding