

Platinum Challenge

Team 4

By: Ivan, Ridzan, Prabowo & Dish

Data Source

data_preprocessed

Python

```
['warung ini dimiliki oleh pengusaha pabrik tahu yang sudah puluhan tahun terkenal membuat tahu putih di bandung tahu berkualitas dipadu keahlian memasak dipadu kretivitas jadilah warung yang menyajikan menu 'mohon ulama lurus dan k212 mmbri hujjah partai apa yang harus diwlh agar suara islam tidak pecah pecah', 'lokasi strategis di jalan sumatera bandung tempat nya nyaman terutama sofa di lantai 2 paella nya enak sangat pas dimakan dengan minum bir dingin appetiser nya juga enak enak ', 'betapa bahagia nya diri ini saat unboxing paket dan barang nya bagus menetapkan beli lagi ', 'duh jadi mahasiswa jangan sombong dong kasih kartu kuning segala belajar dulu yang baik tidak usahlah ikut ikut politik nanti sudah selesai kuliah nya mau ikut politik juga tidak telat dasar mahasiswa 'makanan beragam harga makanan di food stall akan ditambahkan 10 lagi di kasir suasana ramai dan perlu perhatian untuk mendapatkan parkir dan tempat duduk ', 'pakai kartu kredit bca tidak untung malah rugi besar', 'tempat unik bagus buat foto makanan enak pegawai ramah bersih dan luas wifi kencang harga standar dan sesuai dengan tempat nya ada menu masakan makanan barat dan indonesia menu favorit lychee mojito 'saya bersama keluarga baru saja menikmati pengalaman kuliner yang menyenangkan di rm sari sunda di jalan setiabudhi bandung tapi karena tidak ada hubungan untuk rm sari sunda di jalan setiabudhi maka sa 'bersyukur', 'simcard indosat inaktiv gara gara lupa isi pulsa dan kabar nya aktif jika pinda ke pasca bayar ribet banget', 'sifat ini sering muncul pada orang orang yang tidak punya tujuan hidup', 'sekadar menceritakan pengalaman saya pesan steak grilled beef 200 tidak disajikanlah steak 200 tidak tersebut dalam 2 potongan 100 tidak 100 tidak rasanya tidak buruk namun kualitas daging nya yang membuat saya 'pengalaman bersama indosat hari ini semoga tidak terjadi pada pelanggan lain sempat sampai marah marah dengan pelayanan pelanggan indosat', 'anak sekarang sulit untuk dinasehati', 'hanya mengatasi masalah kayak gini saja anies sandi tidak becus ke mana saja pasukan kebersihan yang begitu banyak jumlah nya ', 'setiap bumn dibentuk dengan uu bukan dibentuk dengan uu ', 'sehabis puas bermain di trans studio bandung saya bersama teman sepakat untuk makan ayam kakek di sini ayam nya selalu enak dan gurih ', 'rasa bakso cuanki dan batagor cukup selalu ramai pengunjung buka sekitar jam 10 pagi dari jalan riau menuju jalan a yani ada sisi kiri jalan sekitar jalan anggrek harga per porsi sekitar rp 14 000 datangl 'sgwr 2018 beri dampak positif terhadap hobi anak anak', 'berada di lembang pemandangan indah udara sejuk penataan resto yang bagus dilengkapi dengan makanan yang sangat nikmat pilihan menu beragam dan minuman yang lengkap lingkungan yang cocok untuk makan ber 'berbuka puasa di sini pada minggu lepas pak supir yang mencadangkan tempat makan ini makanan nya di sini sedap dan harga berpatutan tempat nya luas dan cantik rekomendasi too others to dine in here to rasa 'hahaha suka banget nonton kartun indonesia keluarga pak tomat', 'bagi teman teman yang sedang berkunjung ke bandung bisa mampir ke kopi anjis untuk makan makanan yang tidak terlalu berat ada cilok rakyat ditemankan dengan teh sereh dan lekker ', 'di restoran ini saya dan keluarga makan malam yang terakhir di bandung kami berjumlah 8 orang dan bila nya hanya sekitar 400 an benar benar murah meriah sampai makan double double lagi makanan nya juga en ... 'dari segi tempat mungkin sangat sederhana tapi rasa tidak perlu diragukan lagi harga nya pun cukup terjangkau kekurangan nya harus sabar antri dan panas panasan menunggu giliran ', 'bongkar saja pak sby sekalian toh ini negri juga sudah hancur kok', 'rumah itu memiliki pagar besar berwarna hitam ', 'tidak pernah bosan datang ke sini baik siang maupun malam hari untuk sekadar menikmati keindahan tempat yang ada di kampung daun ini konsep yang sangat pas sekali dengan nuansa jawa barat yang asri dan tenang mew ...']
```

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#).

Problem Statement

Rumusan:

Pendalaman Machine Learning untuk menguji nilai statement pada penggunaan social media Netizen indonesia.

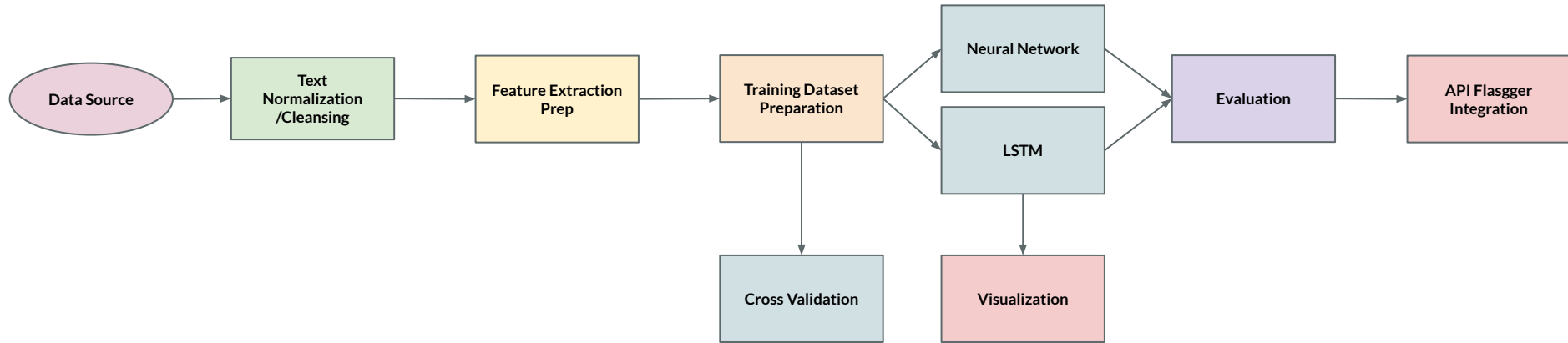
Tujuan:

Melakukan test pada machine menggunakan process LSTM dan Neural Network.

Hypothesis:

- > Tingginya accuracy dapat tercapai pada process Neural Network*
- > Hasil goodfit dapat tercapai pada process Training Validation & Loss*

Challenge Flow Chart



Pre Processing

Team melakukan komparasi pre-processing guna untuk melihat sejauh mana perbedaan yang dihasilkan oleh data.

Hasil ditemukan bahwa ada nya perbedaan di beberapa testing namun perbedaan yang tidak terlalu signifikan ini justru memberikan hasil akhir yang menarik!



Feature Extraction Process (TF-IDF)

Kita melakukan eksperimen TF-IDF yang dimana untuk output sebelum cleansing sangat baik, oleh karena itu kita menggunakan metode TF-IDF dengan hasil seperti dibawah:

Training Output	Precision	Recall	F1-Score	Support
Negative	0.74	0.69	0.81	687
Neutral	0.90	0.24	0.76	230
Positive	0.82	0.95	0.90	1283
Accuracy			0.86	2200
Macro AVG	0.82	0.63	0.82	2200
Weighted AVG	0.80	0.80	0.86	2200

Feature Extraction Process (BOW)

Kita melakukan eksperimen terhadap BOW sebelum Cleansing dan kita mendapatkan hasil sebelum Cleansing dengan hasil output seperti dibawah:

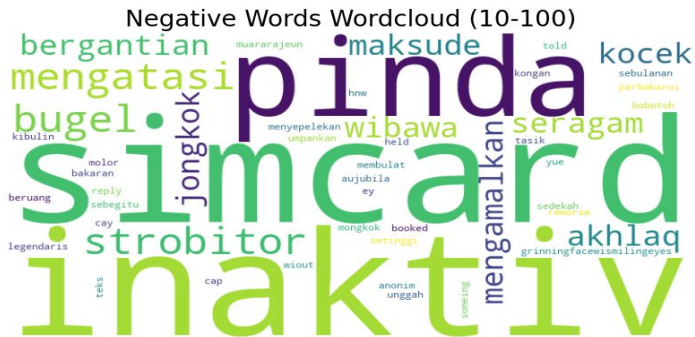
Training Output	Precision	Recall	F1-Score	Support
Negative	0.77	0.80	0.78	706
Neutral	0.85	0.65	0.74	220
Positive	0.88	0.90	0.89	1274
Accuracy			0.84	2200
Macro AVG	0.83	0.78	0.80	2200
Weighted AVG	0.84	0.84	0.84	2200

Cleansing Process

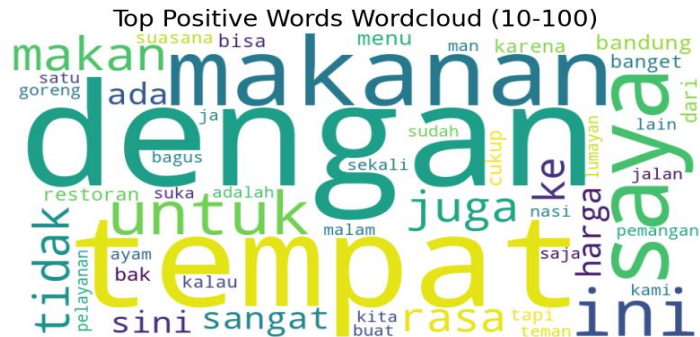
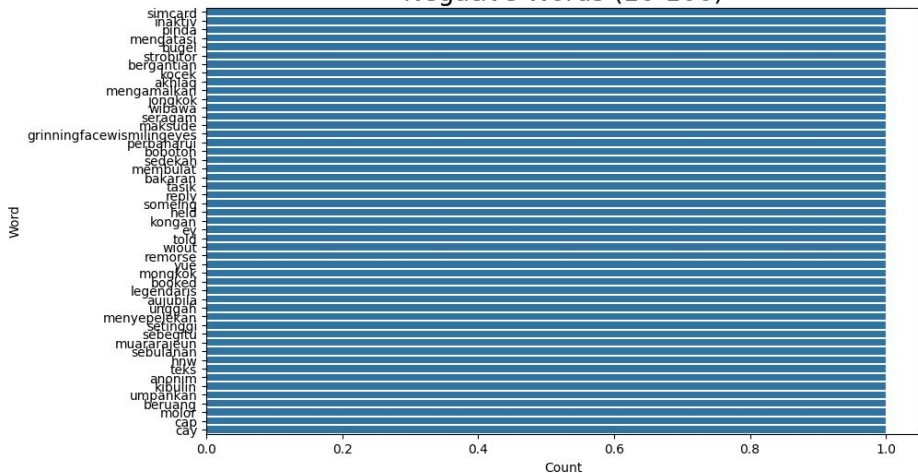
Berikut adalah semua komparasi cleansing yang Team lakukan menggunakan F1 Score!

Nama	Tanpa Cleansing	Stopword Binar	Stopwords kami	Steaming Sastrawi	NLTK	KamusAlay Binar	KamusAlay Kami	Remove Chars Kami
Negative	0.81	0.78	0.79	0.81	0.77	0.78	0.79	0.79
Neutral	0.76	0.76	0.75	0.78	0.76	0.74	0.76	0.75
Positive	0.90	0.89	0.90	0.91	0.89	0.88	0.89	0.90
Accuracy	0.86	0.84	0.85	0.87	0.84	0.84	0.84	0.85

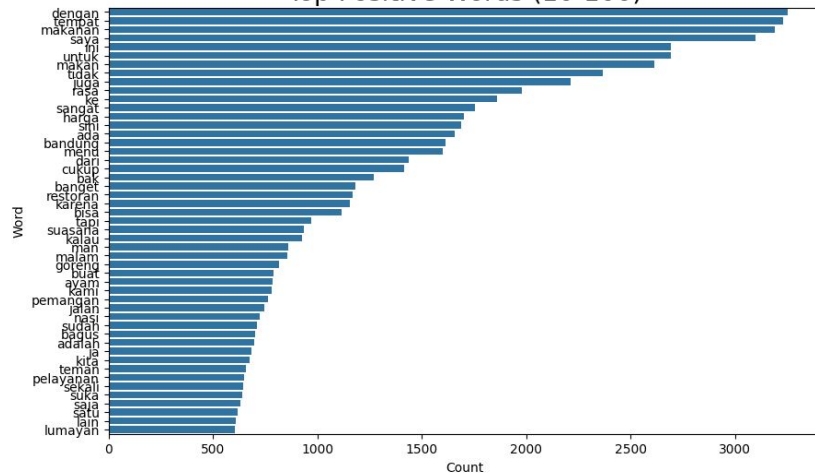
EDA Stopwords and New Dictionary



Negative Words (10-100)



Top Positive Words (10-100)



Training and Model Neural Network

Menggunakan TF-IDF, Cleansing versi kami. Team Melakukan pengujian dengan mengganti fungsi Activation dan Solver.

Dimana didapatkan beberapa point:

- Menggunakan Activation logistic dan solver sgd akan mendapatkan F1 Score pada Negative dan Neutral menjadi nihil = 0
- Menggunakan Activation tanh dan solver sgd akan mendapatkan F1 Score pada Neutral menjadi nihil = 0

Logistic	Adam
Negative	0.77
Neutral	0.71
Positive	0.87
Accuracy	0.83

Tanh	Adam
Negative	0.75
Neutral	0.69
Positive	0.87
Accuracy	0.81

Relu	Adam
Negative	0.79
Neutral	0.76
Positive	0.89
Accuracy	0.83

Neural Network Cross - K - Fold Validation

Training 1

F1 Score	
Negative	0.79
Neutral	0.79
Positive	0.89
Accuracy	0.85

Training 2

F1 Score	
Negative	0.77
Neutral	0.74
Positive	0.89
Accuracy	0.84

Training 3

F1 Score	
Negative	0.79
Neutral	0.81
Positive	0.90
Accuracy	0.85

Training 4

F1 Score	
Negative	0.78
Neutral	0.74
Positive	0.89
Accuracy	0.84

Training 5

F1 Score	
Negative	0.78
Neutral	0.73
Positive	0.89
Accuracy	0.84

rata-rata akurasi 0.8456363636363637

FEATURE - EXTRACTION - LSTM

Dalam model LSTM ini kami menggunakan 4 Layer dan 256 neuron karena keterbatasan dataset yang hanya 11.000 rows.

Tokenizer: tokenizer berguna untuk memecah teks menjadi token

Contoh: saya(1) dapat(2) hadiah(3) —> [1, 2, 3]

Pad Sequences: proses untuk menyelaraskan panjang urutan token agar sesuai dengan panjang yang diharapkan oleh model. Proses ini melibatkan menambahkan nilai nol (0) di akhir urutan. Contoh jika maxlength adalah 4

Contoh: saya(1) dapat(2) hadiah(3) —> [1,2,3,0]

LSTM - EVALUATION

Experiment 1

f-1 Score	
Negative	0.83
Neutral	0.78
Positive	0.91
Accuracy	0.88

Experiment 2

f-1 Score	
Negative	0.82
Neutral	0.80
Positive	0.91
Accuracy	0.87

Experiment 3

f-1 Score	
Negative	0.81
Neutral	0.77
Positive	0.91
Accuracy	0.87

Experiment 4

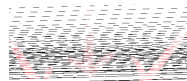
f-1 Score	
Negative	0.83
Neutral	0.80
Positive	0.92
Accuracy	0.88

Experiment 5

f-1 Score	
Negative	0.83
Neutral	0.74
Positive	0.91
Accuracy	0.87

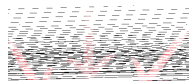
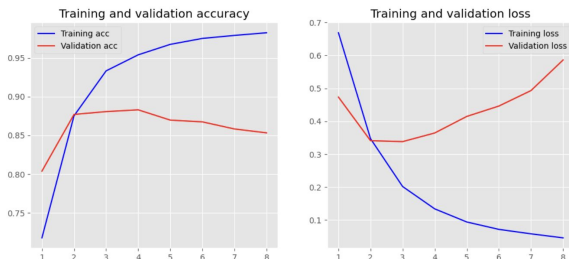
rata-rata akurasi 0.87

LSTM - EXPERIMENT



Experiment 1

Hyperparameters	
Embed Dim	100
Units	128
Dropout	0.2
Epochs	10
Batch Size	128



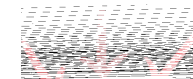
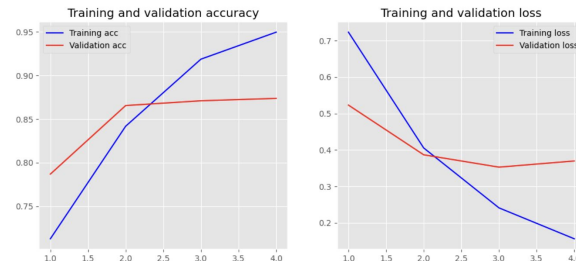
Experiment 3

Hyperparameters	
Embed Dim	100
Units	256
Spatial Dropout 1D	0.4
Epochs	20
Batch Size	128



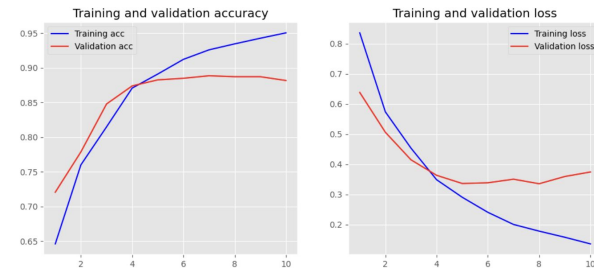
Experiment 2

Hyperparameters	
Embed Dim	100
Units	256
Dropout	0.3
Epochs	10
Batch Size	128



Experiment 4

Hyperparameters	
Embed Dim	128
Units	256
Spatial Dropout 1D	0.6
Epochs	10
Batch Size	256

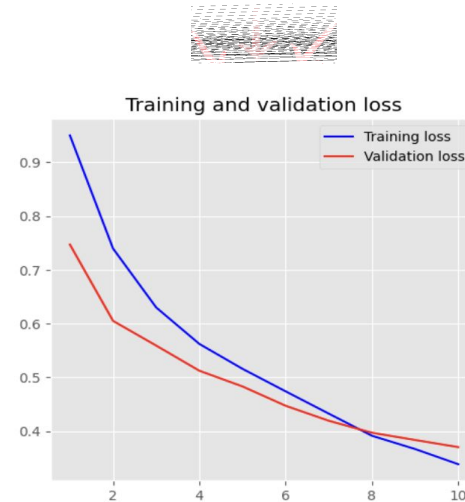
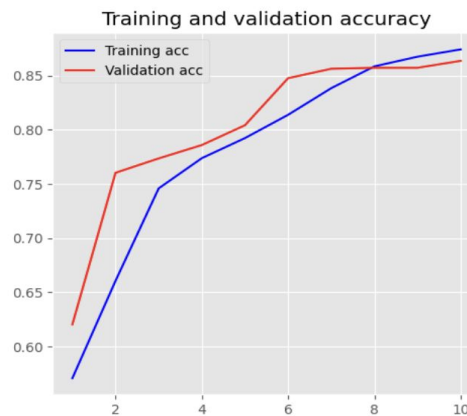
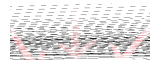


LSTM

Experiment 5

f-1 Score	
Negative	0.83
Neutral	0.74
Positive	0.91
Accuracy	0.87

Hyperparameters	
Embed Dim	128
Units	256
SpatialDrop out1D	0.7
Epochs	10
Batch Size	256



Final Experiment

DEMO

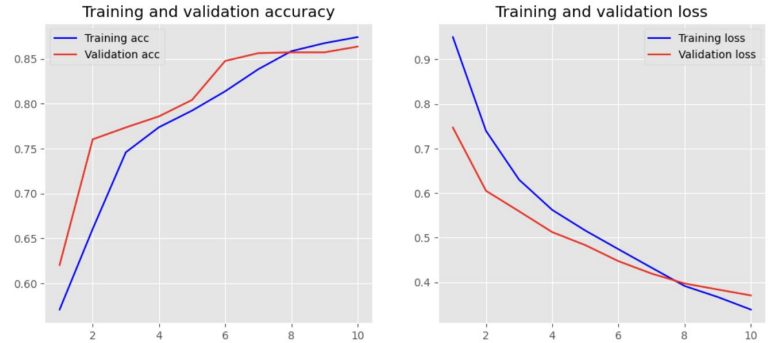


Click [Demo API](#) to showcase

Hypothesis & Conclusion

Terlihat pada chart di samping bahwa tujuan hypothesis tercapai yaitu:

- > Hasil goodfit dapat tercapai pada process Training Validation & Loss
- > Tingginya accuracy dapat tercapai pada process Neural Network



F1 Score

