

SENTIMENT ANALYSIS

IKN by Andi Setianto

SLIDE NOW





Business Understanding

- Analisis sentimen terhadap IKN penting karena memberikan wawasan tentang pandangan dan perasaan masyarakat terhadap ibu kota tersebut.
- Melalui analisis sentimen, dapat diketahui apakah penduduk mendukung atau tidak mendukung keputusan pemerintah terkait pemindahan ibu kota. Informasi ini dapat membantu pemerintah dalam pengambilan keputusan yang lebih baik, memperbaiki kebijakan, dan merespons kebutuhan dan kekhawatiran masyarakat secara lebih efektif.



Data Understanding



Collecting

- Data text berasal dari Tweets 7 hari terakhir (28 Mei 2023 s.d. 4 Juni 2023)
- Sebanyak 1000 tweets
- Dengan kata kunci 'IKN'

1

Cleaning

- Data dilakukan cleaning, seperti slang words, hapus retweet, hapus data duplicate, dan processing untuk special character
- Tersisa 484 data saja

2

Preprocessing

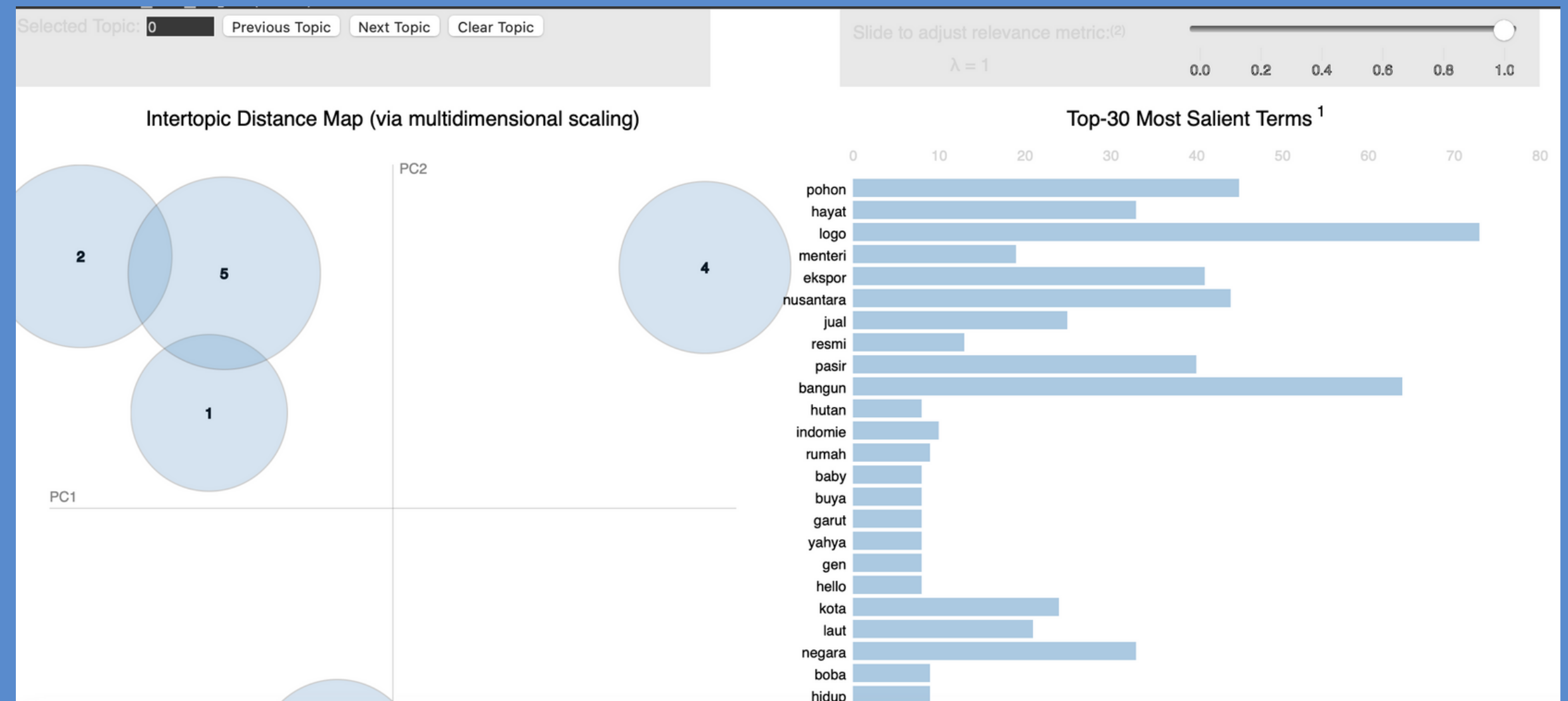
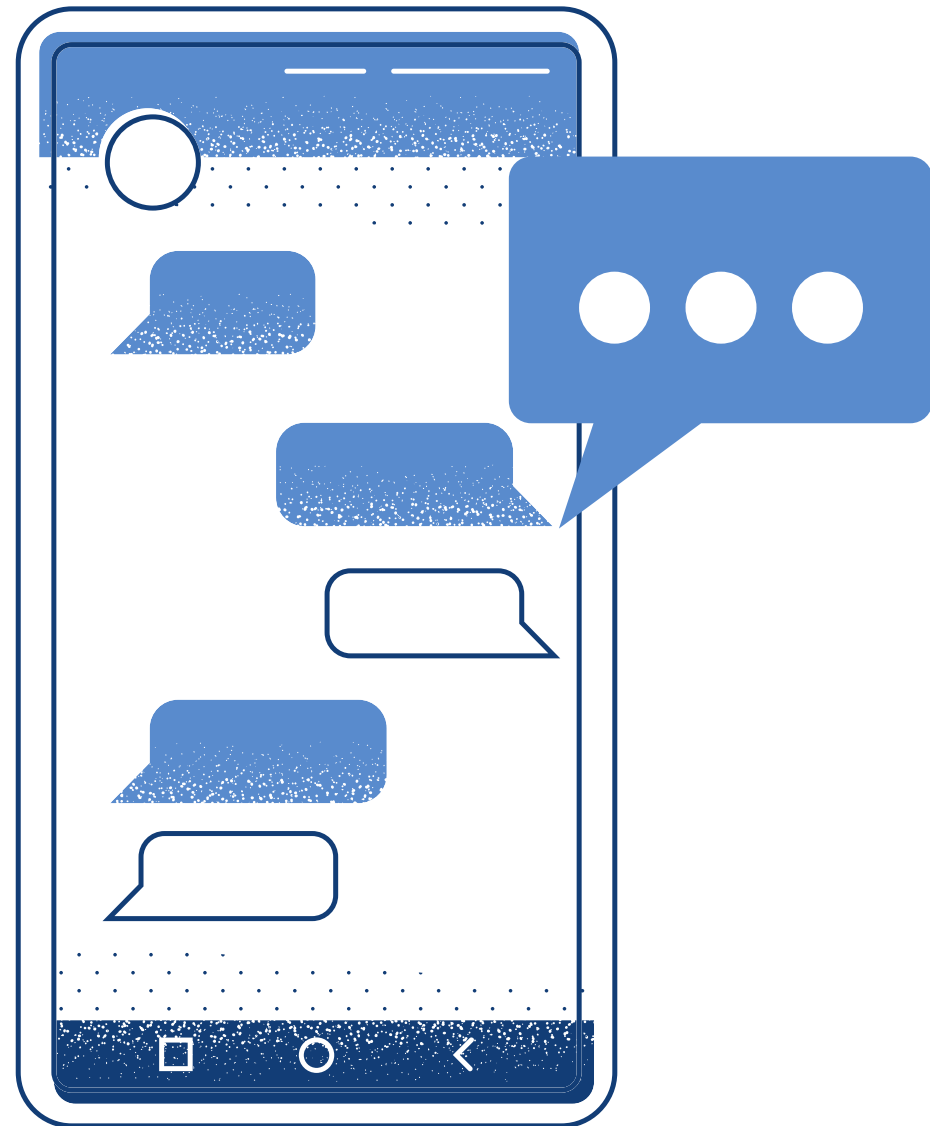
- Proses preprocessing berupa stemming, tokenizing, dll. (akan dijelaskan pada halaman selanjutnya)

3

Preprocessing ::

- ``casefoldingText`` : mengubah semua huruf dalam teks menjadi huruf kecil.
- ``tokenizingText`` : Fungsi ini digunakan untuk membagi teks menjadi token-token atau kata-kata individu. Ini dilakukan menggunakan fungsi ``word_tokenize`` dari pustaka ``nltk``. Hasilnya berupa daftar kata-kata.
- ``filteringText`` : Fungsi ini digunakan untuk menyaring kata-kata yang tidak relevan atau stopwords dalam bahasa Indonesia. Stopwords adalah kata-kata umum yang sering muncul dalam teks namun tidak memberikan makna yang signifikan. Fungsi ini membandingkan setiap kata dalam teks dengan daftar stopwords yang telah ditentukan dan menghilangkan kata-kata tersebut dari teks.
- ``stemmingText`` : Proses menghapus imbuhan atau akhiran kata sehingga hanya menyisakan kata dasar. Dalam code tersebut, metode stemming yang digunakan adalah metode stemming Bahasa Indonesia yang disediakan oleh pustaka Sastrawi.
- ``toSentence`` : Fungsi ini digunakan untuk menggabungkan kembali token-token atau kata-kata menjadi kalimat yang utuh. Setiap kata dalam daftar kata-kata diubah menjadi string, dan kemudian digabungkan kembali menggunakan spasi sebagai pemisah antara kata-kata.

LDA / Topic Extraction



Dengan menggunakan LDA, code tersebut dapat membantu mengidentifikasi topik utama yang muncul dalam teks. Kita dapatkan bahwa belakangan sedang sering membahas tentang pohon, hayat, dan logo. Hal tersebut dimungkinkan karena baru saja ada peresmian logo IKN

EXTRACT NOW

Polaritas Sentimen

Untuk polaritas sentimen, saya menggunakan dictionary dari Indonesian Sentiment Lexicon. Yaitu berisi dictionary kata kata yang memiliki arah negative, dan arah positive, dengan bobot 1 hingga 5 (ataupun negasinya)

link:

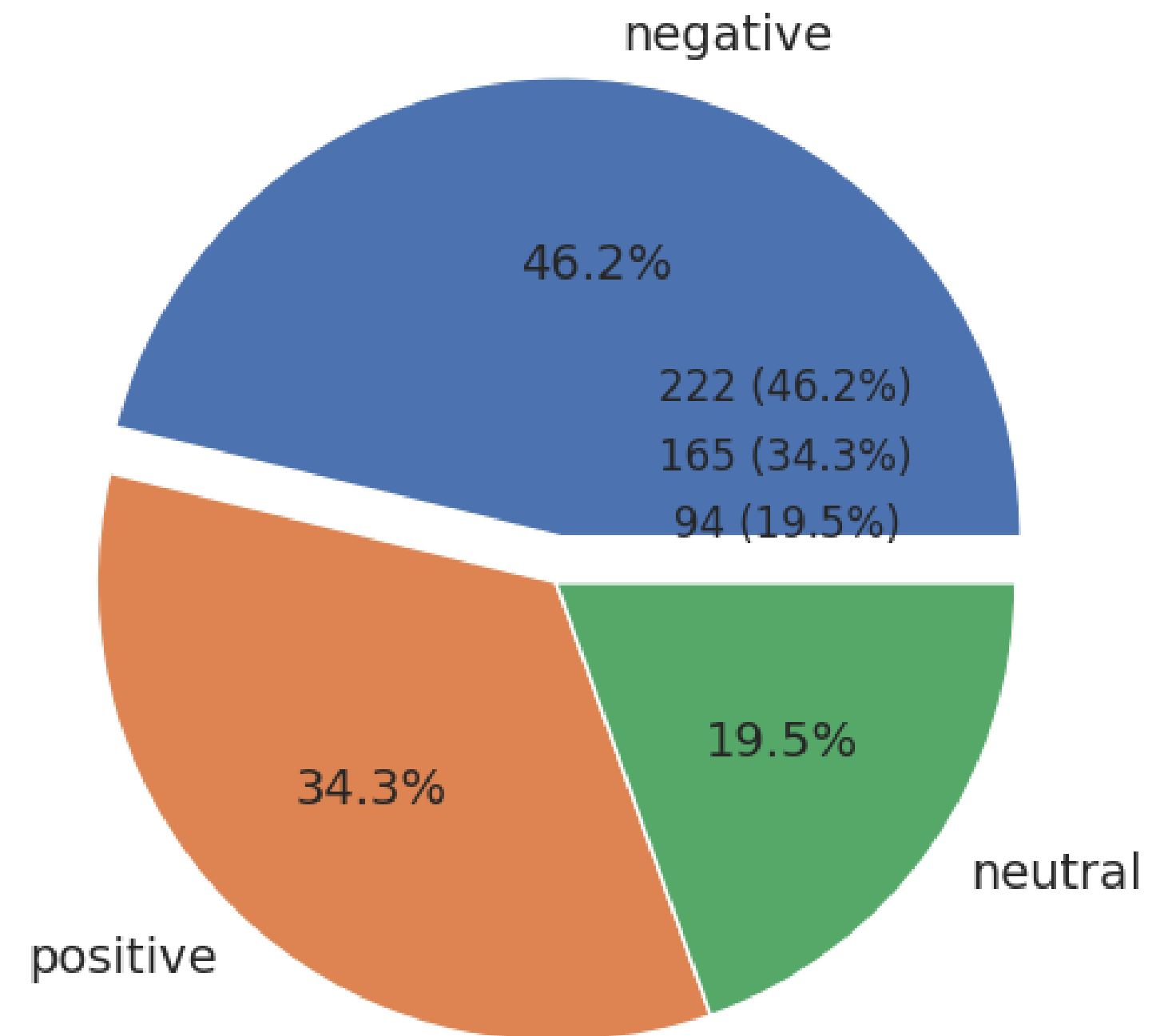
https://www.researchgate.net/publication/321757985_InSet_Lexicon_Evaluation_of_a_Word_List_for_Indonesian_Sentiment_Analysis_in_Microblogs

1	word	weig
2	putus tali gantung	-2
3	gelebah	-2
4	gobar hati	-2
5	tersentuh (perasaan)	-1
6	isak	-5
7	larat hati	-3
8	nelangsa	-3
9	remuk redam	-5
10	tidak segan	-2
11	gemar	-1
12	tak segan	-1
13	sesal	-4
14	pengen	-2

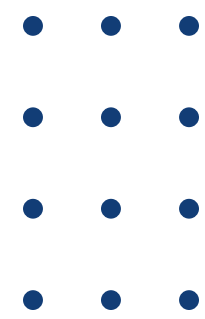
Visualisasi

Dari diagram tersebut terlihat bahwa mayoritas (46%) pengguna tweet masih memiliki sentimen yang negatif terhadap IKN, selisih 10 % dengan yang bersentimen positif. Namun di sini terdapat 19.5% opini yang netral, sehingga ranah yang masih swing seperti ini bisa menjadi fokus bagi pemerintah, agar mereka puas / yakin dengan kebijakan pemerintah terhadap IKN

Sentiment Polarity on Tweets Data



Contoh Sentiment Positive dan Negative

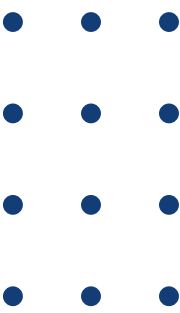


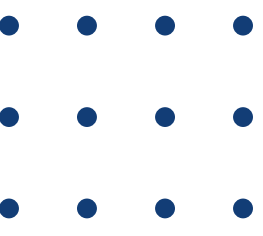
positive

	selamat pagi indonesia semangat pagi untuk semangat indonesia baru selamat datang logo ikn 🇮🇩	16	positive
	presiden joko widodo secara resmi meluncurkan logo ibu kota nusantara ikn di istana negara jakarta pada selasa logo yang nantinya akan menjadi identitas visual bagi ikn tersebut memiliki tema pohon hayat dengan sejumlah makna di dalamnya namun di media sosial...	14	positive
	ikn adalah jalan menuju peradaban dunia yang maju dan ramah lingkungan kojta masa depan tapi bicara peradaban itu sulit diterima oleh k drone karena memang tidak punya adab maunya ke jaman batu bisa pilih siapapun perempuan semauanya	12	positive
	kontribusi millenial untuk ikn 🇮🇩💪 rotiboy jisoo dating kim namjoon raffi kurnia meiga pelan toothless tiger jk awal juni libur semangat adel happy th we are one andara super bowl kim seungmin	12	positive

negative

1	bahrawi itulah sebabnya ikn ngotot di wujudkan ujung ujungnya para siluman siluman baru beli tanah di seluruh kalimantan siluman baru dari sungai sungai kuning	-28	negative
2	gimana kalo dibalik begini baru tau kalo april nanti harus pindah ke ikn kasian asn yang terpaksa ninggalin jakarta krn keputusan peninggalan jokowi karena sudah kontrak terpaksa harus pindah dgn setengah hati	-26	negative
3	memang apa berguna itu ikn yg baru sudah kesenjangan kaya miskin makin lebar gizi buruk pada anak selalu meningkat ekonomi di rezim petruk gak pernah bagus penegak hukum macam sampah korupsi dimanamana utang makin gak ngotak malah bikin masalah terus blok	-25	negative
4	negara isinya makelar nggak makelar motor listrik mobil listrik tanah kereta cepat investasi ikn apalagi dah tu sampe pasir ikutan di makelarin	-20	negative

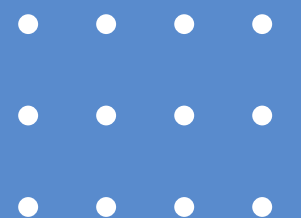
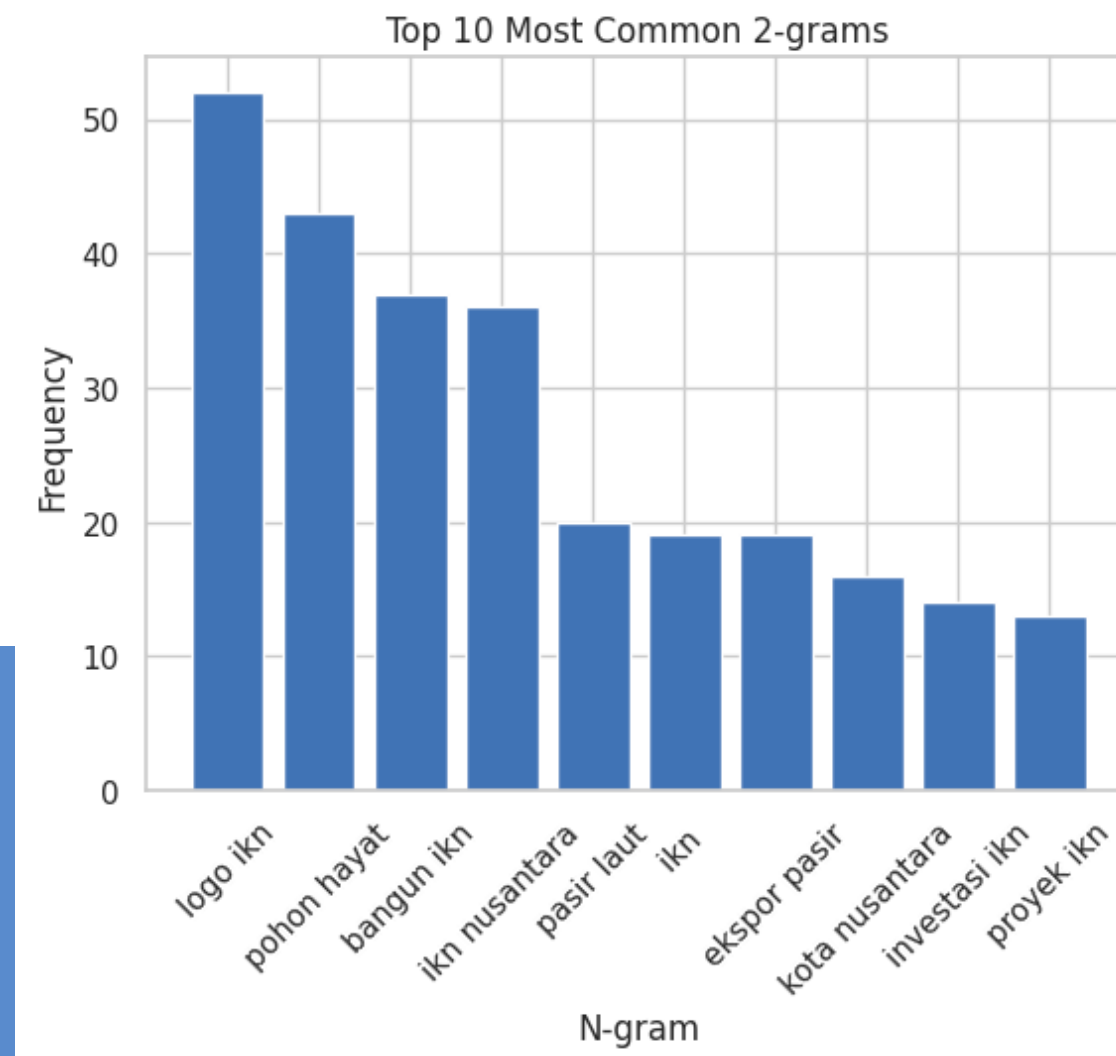
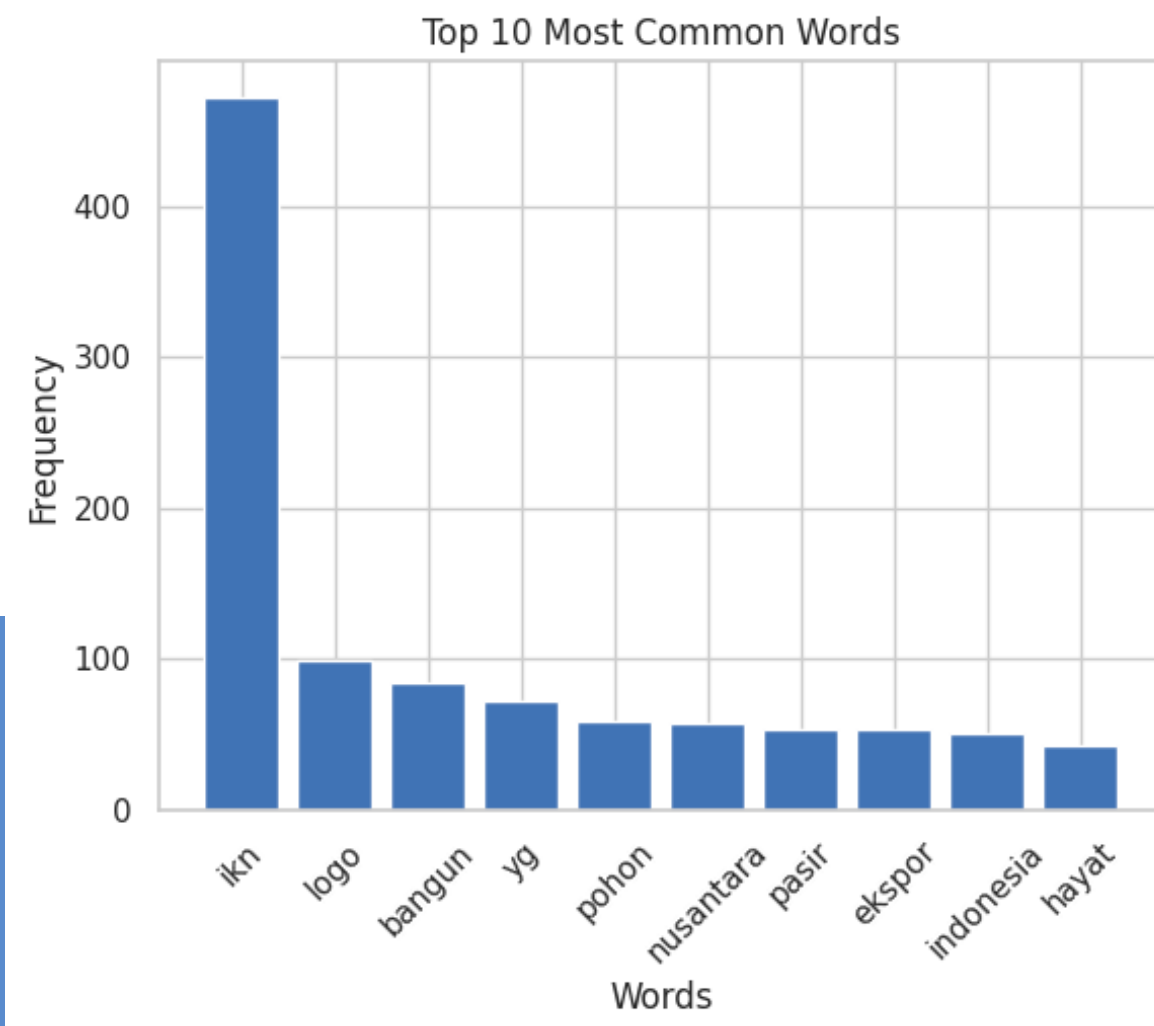




Common Words

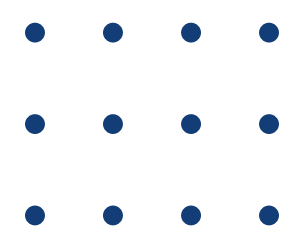
Top 10 kata yang paling sering muncul yaitu IKN, logo, dan bangun. Hal itu relevan dengan keywords kita yaitu terkait IKN, secara konteks juga relevan karena kita sedang melaksanakan 'pembangunan', kemudian masih berhubungan dengan momen belakangan ini, yaitu peresmian logo IKN

Tampak juga pada diagram 2-grams, bahwa 'logo + ikn' sedang ramai dibicarakan, begitu juga dengan 'pohon hayat' sebagai deskripsi dari logo tersebut

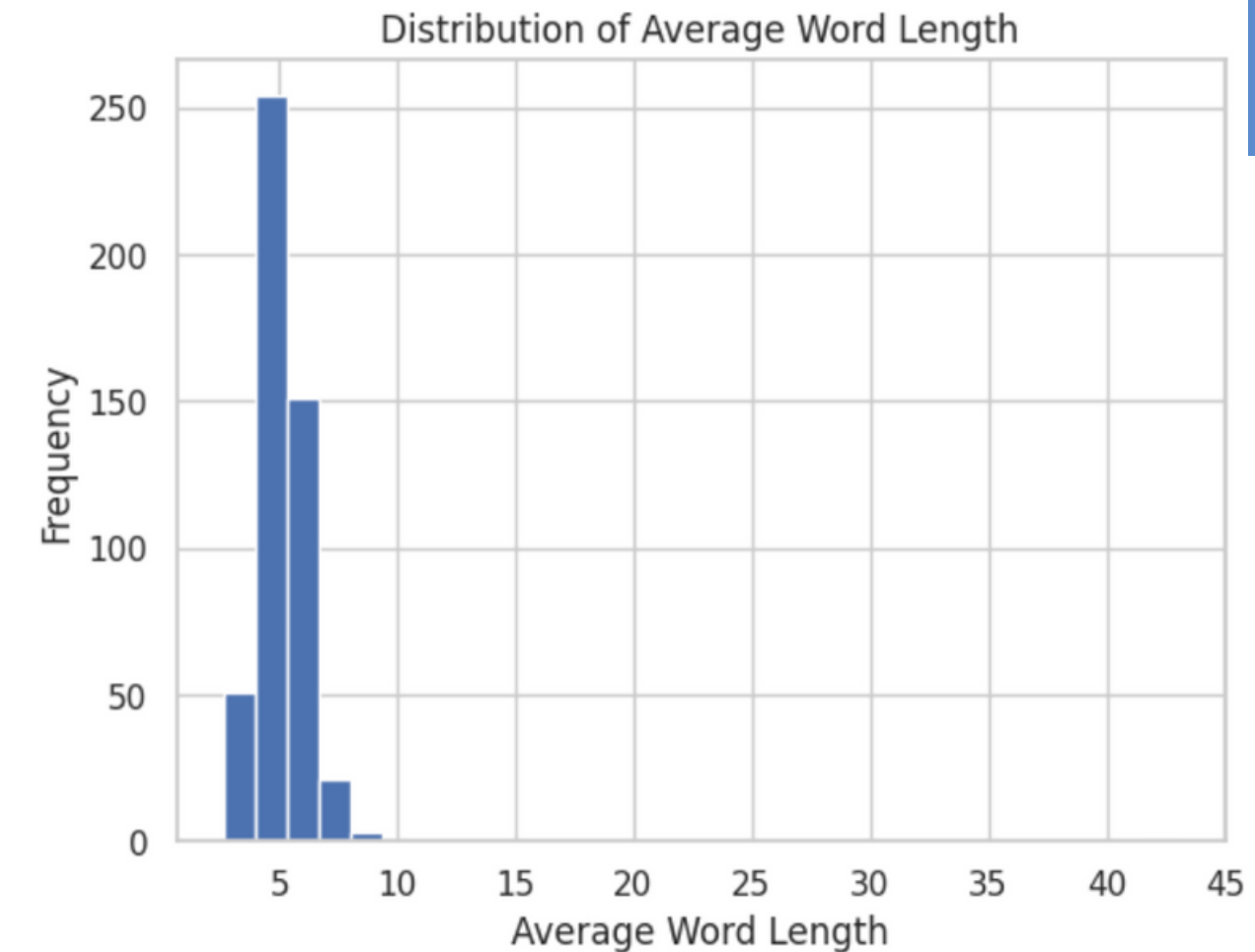
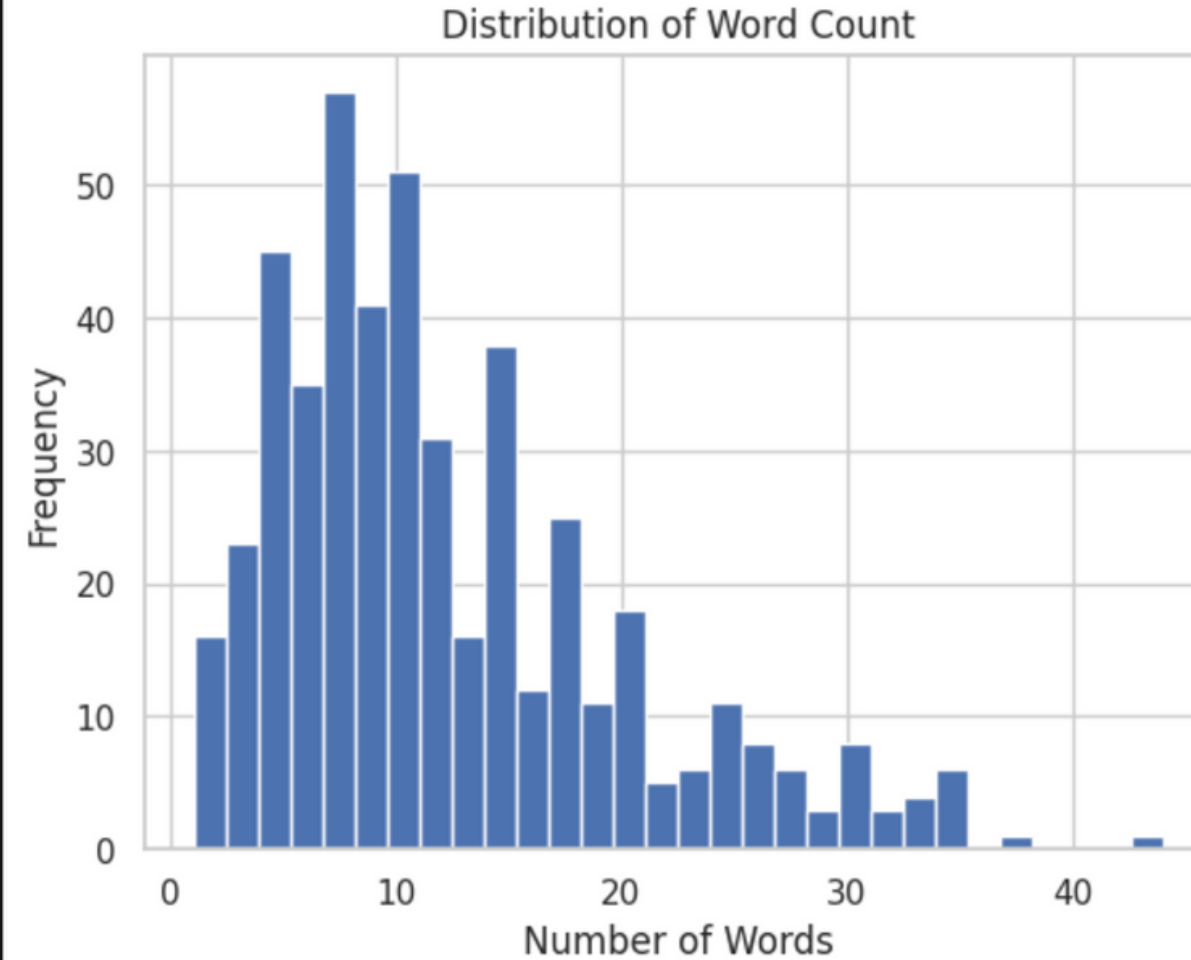
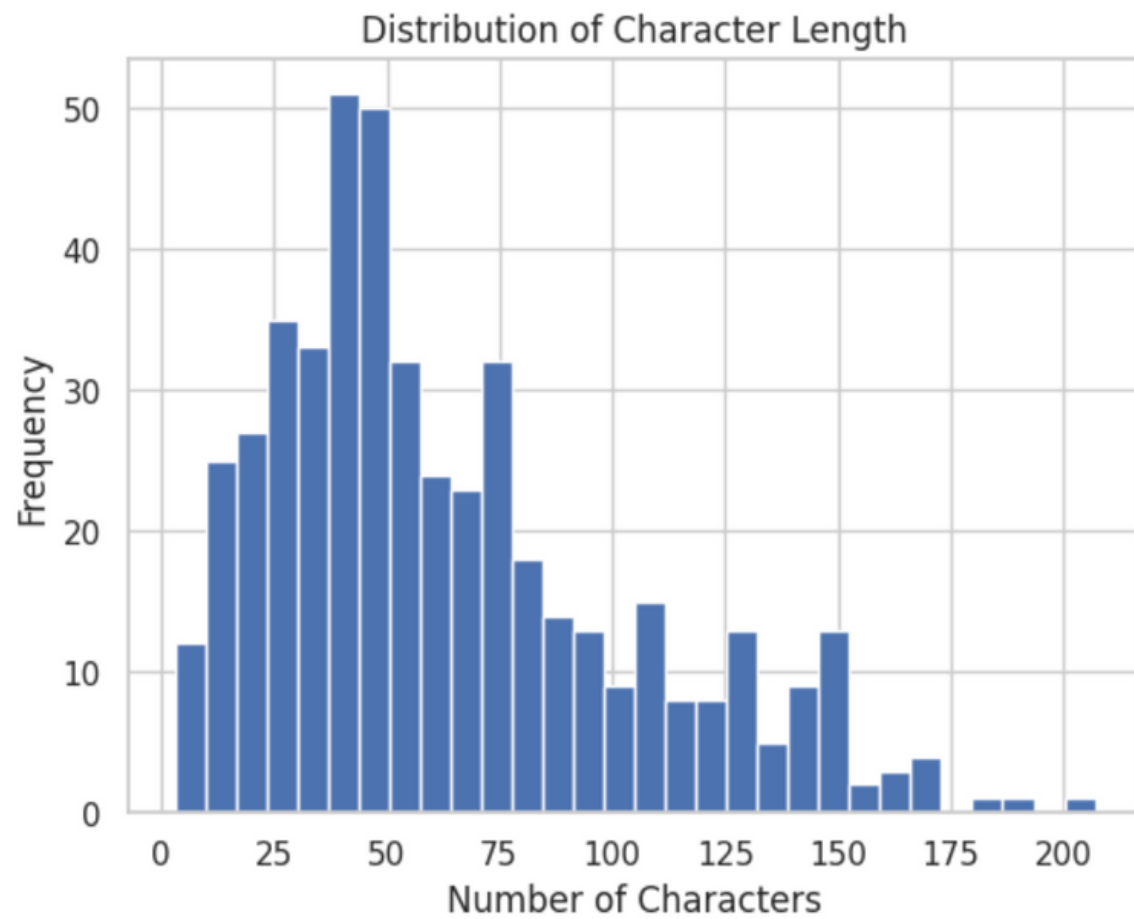


[illegible][illegible]

Kemudian pada Words Cloud of Negative Word, kita melihat kata 'hutang' beberapa kali, hal ini karena kekhawatiran netizen terhadap IKN, yaitu hutang Indonesia yang semakin membengkak



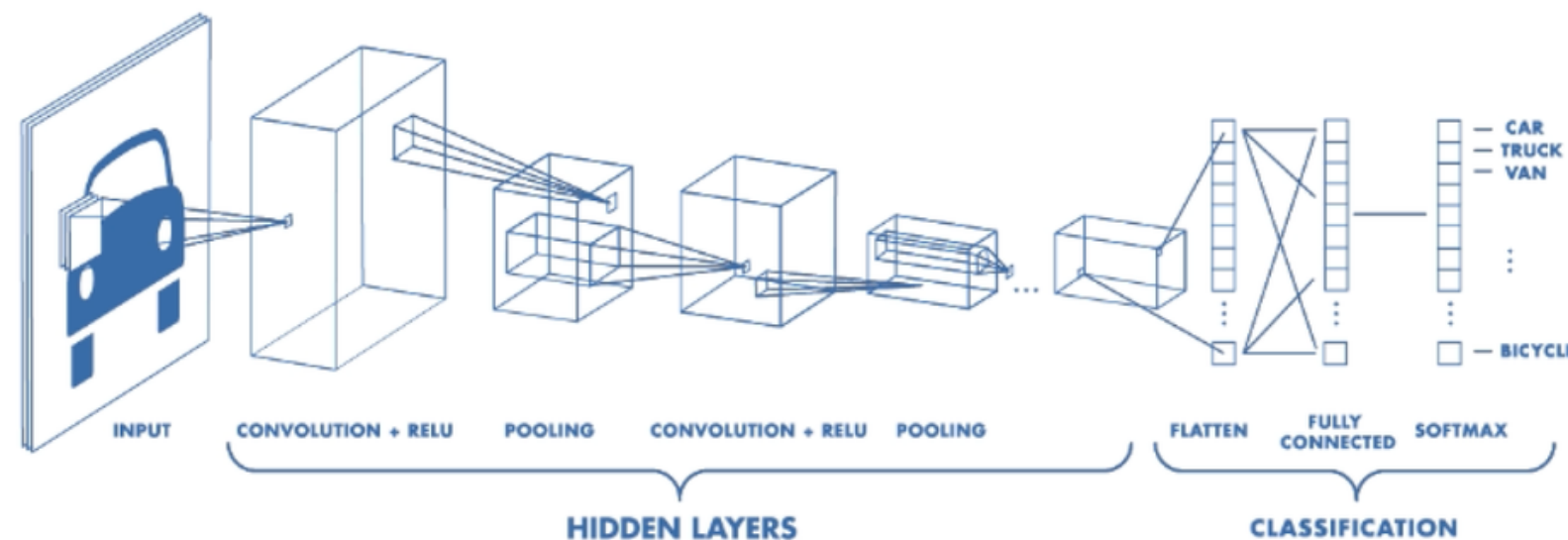
Distribution of Words & Characters



Puncak yang lebih tinggi menunjukkan adanya panjang karakter/words yang lebih umum atau sering muncul dalam tweet. Sebaliknya, lembah yang lebih dalam menunjukkan panjang karakter yang lebih jarang terjadi. Puncak-puncak dan lembah-lembah tersebut dapat memberikan wawasan tentang kecenderungan panjang karakter/words yang paling umum atau khas dalam dataset tweet

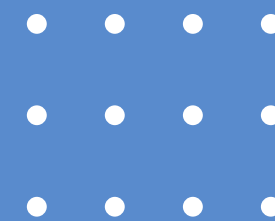
distribusi condong ke salah satu sisi, hal ini menunjukkan adanya kecenderungan panjang karakter/words yang lebih umum di salah satu sisi.

Modelling with CNN



Convolutional Neural Network (CNN)

CNN (Convolutional Neural Network) adalah sebuah jenis arsitektur jaringan saraf yang digunakan untuk tugas-tugas pengenalan pola pada data berstruktur seperti gambar, teks, dan suara. CNN terdiri dari lapisan konvolusional yang melakukan konvolusi antara filter dengan input data, lapisan pooling yang mengurangi dimensi spasial data, fungsi aktivasi untuk memperkenalkan non-linearitas, dan lapisan fully connected untuk klasifikasi akhir. Melalui proses pelatihan dengan optimisasi, CNN dapat mempelajari fitur-fitur hierarkis dari data input dan mencapai kinerja yang sangat baik dalam tugas-tugas pengenalan pola yang kompleks.



Important Point (CNN)

- `create_model()`: Merupakan fungsi untuk membuat model CNN dengan arsitektur yang ditentukan. Model ini terdiri dari lapisan embedding, konvolusi, pooling, dropout, dan dense layers. Model juga didefinisikan dengan menggunakan fungsi `compile()` yang menentukan loss function, optimizer, dan metrik evaluasi yang akan digunakan. Fungsi ini juga mencetak ringkasan (summary) dari model.
- Hyperparameters: Menentukan kumpulan hyperparameters yang akan dicari dalam proses hyperparameter tuning. Hyperparameters ini mencakup berbagai fitur arsitektur CNN, seperti dimensi embedding, jumlah filter, learning rate, jumlah epoch, dan ukuran batch.
- `random_search`: Mengevaluasi model menggunakan `RandomizedSearchCV` untuk menemukan kombinasi hyperparameters terbaik.
- `model_prediction`: Melatih model menggunakan hyperparameters terbaik yang ditemukan melalui `RandomizedSearchCV`. Model ini menggunakan validation split sebesar 0.1 untuk validasi setiap epoch selama pelatihan.

maka kita dapat mencari kombinasi hyperparameters terbaik untuk model CNN dalam klasifikasi teks dan melatih model dengan menggunakan hyperparameters tersebut.


Model Accuracy

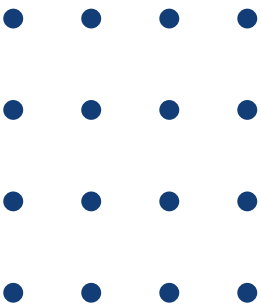


Plot ini memberikan gambaran tentang bagaimana performa model dalam hal akurasi selama proses pelatihan, dan dapat membantu dalam menganalisis kinerja model serta mengidentifikasi overfitting atau underfitting.

Jika dilihat dari plot tersebut, pola tersebut menunjukkan adanya overfitting pada model. Train accuracy meningkat hingga mendekati 1, sedangkan validation accuracy cenderung stagnan di sekitar 0.5 hingga 0.6. Hal ini mengindikasikan bahwa model terlalu "menghafal" data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data validasi atau data baru. Perlu dilakukan upaya untuk mengendalikan overfitting, seperti menggunakan regularisasi, dropout, mengurangi kompleksitas model, atau mengumpulkan lebih banyak data training.

Contoh Hasil Polarity

	text	polarity
0	profil aulia akbar sosok yang mendesain logo ikn pohon hayat	Negative
1	bisa jadi disharing dulu yg penting seperti buat bahan kaca dan tinta laser ampasnyaaa dijual lagi buat ikn	Negative
2	ikn bisa jadi mangkrak	Neutral
3	reklamasi ikn	Neutral
4	gaet investor singapura jokowi bakal jelaskan konsep kota hijau ikn	Neutral
...
92	logo resmi ikn 	Positive
93	online mereka akan berpikir sesuatu hal lain yang terdapat di balik pohon lalu mengimajinasikan alasan dipilihnya gambar pohon unt...	Negative
94	kami warga sofifi akan mengawal pembangunan ikn	Positive
95	pelototi terus jkw rezim tolak periode penundaan pemilu tolak ikn batalkan kenaikan ppn pecat luh...	Negative
96	ikn hanya angan mukidiudahlah sejahterakan rakyat buat lapangan pekerjaan sebanyak banyaknya bayar hutang dlu agar tidak menggununghukum mati para koruptor yg mencuri uang rakyat jangan sibuk ikn ikn ikn aja	Negative

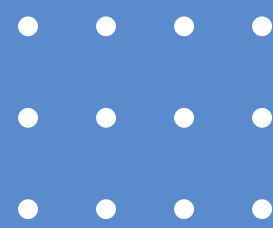


Conclusion & Recommendation



- Kesimpulan dari data sentiment analisis terkait ibu kota Nusantara adalah mayoritas sentimen negatif (46.2%), menandakan adanya ketidakpuasan yang signifikan. Meskipun ada pendapat positif (34.3%), sentimen netral (19.5%) cukup rendah. Perlu perhatian terhadap masalah yang mempengaruhi persepsi dan keadaan ibu kota Nusantara.
- Jika ingin spesifik melihat sentimen terkait IKN, mungkin topik terkait 'logo' perlu dikecualikan
- Tingkatkan data cleaning & proses slang word, karena masih banyak slang word yang muncul
- Model perlu ditingkatkan, mungkin dengan memperbaiki data training/ ditingkatkan, atau mengganti dengan model lainnya seperti LSTM ataupun naive bayes
- Perlu mencari metode manual polarization yang lebih baik lagi, karena terkadang sistem salah mengklasifikasikan (yang seharusnya dimasukan pada kelas 'positif', malah masuk ke kelas 'negatif')





THANK YOU FOR YOUR ATTENTION

andisetianto123@gmail.com

