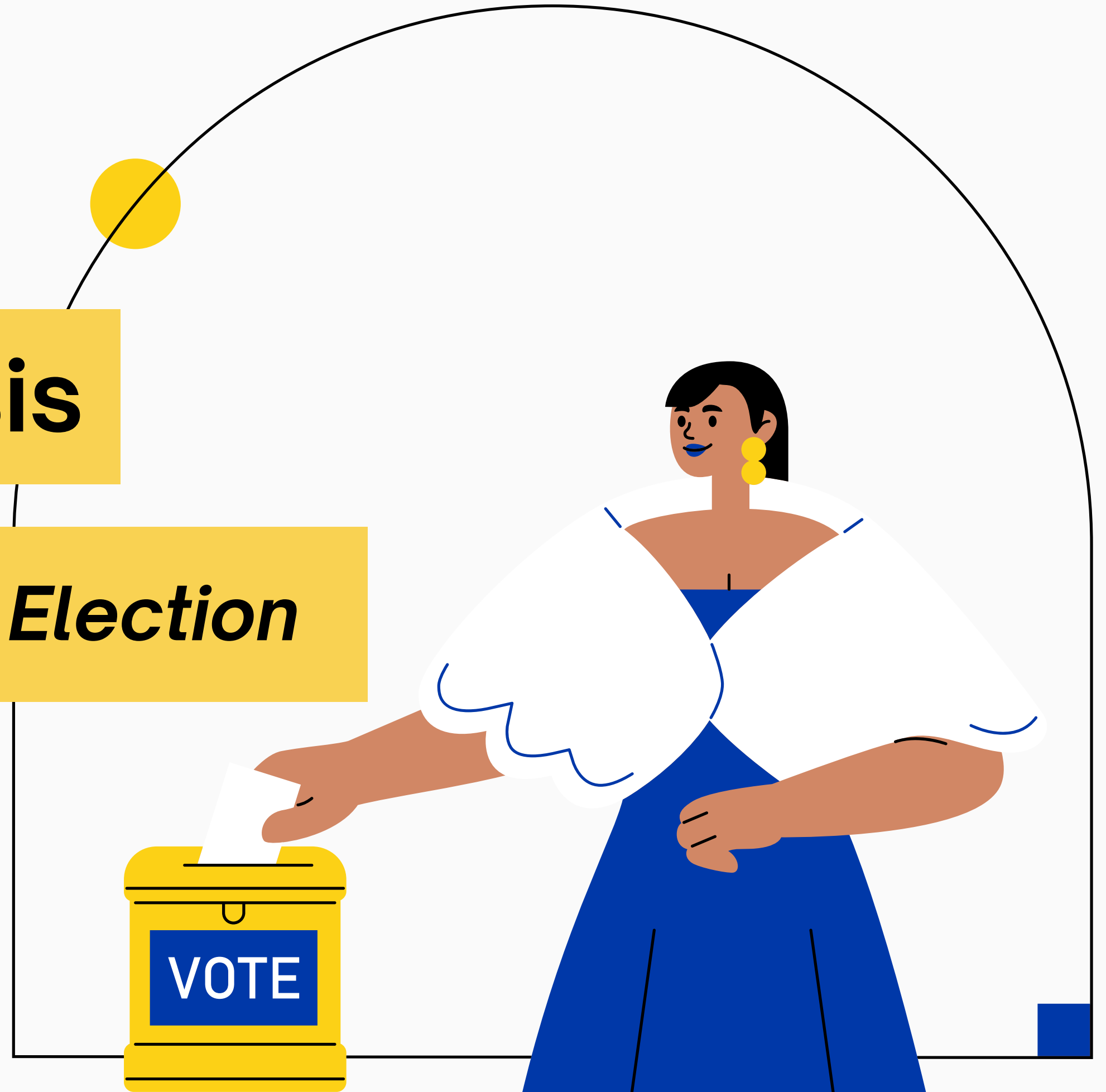
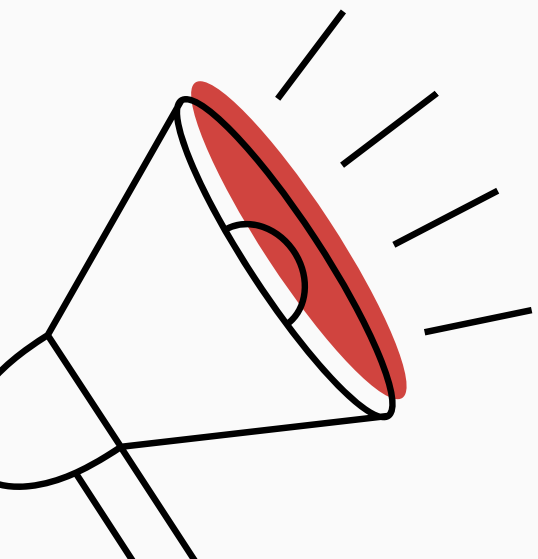


Sentiment Analysis

on 2019 Indonesia Election

BY : ALAN TURING



TOPIC OUTLINE •

BACKGROUND

DATA UNDERSTANDING

TEXT PREPROCESSING

EDA

MODELING

EVALUATION



BACKGROUND



INTRODUCTION



Election in the country of Indonesia

Pemilihan umum di Indonesia berlangsung tiap lima tahun, warga memilih presiden dan wakil presiden. Pasangan calon akan bersaing, pemenangnya memimpin negara selama lima tahun ke depan. Ini adalah momen penting dalam kehidupan politik Indonesia, menentukan arah negara berdasarkan pilihan rakyat.

BACKGROUND

Sentiment Analysis on Indonesia Election

Di era digital, Twitter telah menjadi platform utama bagi masyarakat untuk berbagi pandangan, opini, dan reaksi terhadap peristiwa politik. Analisis terhadap percakapan yang terjadi di Twitter selama masa kampanye dan pemilihan dapat memberikan pemahaman yang mendalam tentang pola sentimen yang muncul.



GOALS

Developed Sentiment Analysis Model

Membangun model untuk memahami sentimen masyarakat tentang pemilihan umum Indonesia tahun 2019 dari data Twitter.



DATA UNDERSTANDING



DATASET

Dataset terdiri dari 1815 baris dan 2 kolom, dengan setiap baris mewakili satu tweet yang mencerminkan opini masyarakat Indonesia selama pemilihan umum 2019.

Jenis Sentimen

- positif : 612 data
- netral : 607 data
- negatif : 596 data

Feature	Description
tweet	Berisi teks tweet yang menjadi subjek analisis, memuat beragam pendapat dan respons dari masyarakat terkait pemilihan umum di Indonesia pada tahun 2019.
sentimen	Berisi label sentimen tweet yang dapat berupa positif, negatif, atau netral, menggambarkan pandangan masyarakat terhadap pemilihan umum 2019 di Indonesia.

TEXT PREPROCESSING



CLEANING PIPELINE

1. Split Hashtag
2. Lowercasing
3. Remove URL
4. Remove Numeric
5. Remove String Emoticon
6. Remove Punctuation
7. Extract Emoji
8. Remove Special Character
9. Remove Repeating Character
10. Remove Single Alphabeth
11. Checking Informal Words
12. Stemming



SPLIT HASHTAG

Memproses teks yang berisi hashtag dengan memisahkan kata-kata yang ditulis dalam format camel case menjadi kata-kata terpisah, sehingga memudahkan pemrosesan lebih lanjut.

Before	After
#PilihOrangBaik #PilihJelasIslamnya #PilihBajuPutih	Pilih Orang Baik Pilih Jelas Islamnya Pilih Baju Putih
#17AprilCoblosJokowiAmin #17AprilCoblosBajuPutih	17April Coblos Jokowi Amin 17April Coblos Baju Putih

- Terdapat 556 perubahan data

LOWER CASING

Memproses teks menjadi huruf kecil bertujuan untuk memastikan konsistensi dalam data teks

Before	After
Kata @prabowo Indonesia tidak dihargai bangsa asing!	kata @prabowo indonesia tidak dihargai bangsa asing!
Iya bener Aa, kita MANTAP kan pilihan ke Pemimpin yg bs memberi Solusi @jokowi bukan yg menyalahkan pemimpin sebelumnya	iya bener aa, kita mantap kan pilihan ke pemimpin yg bs memberi solusi @jokowi bukan yg menyalahkan pemimpin sebelumnya

- Terdapat 1754 perubahan data

REMOVE URL

Membersihkan teks dari URL dengan menggantikan URL yang cocok dengan string kosong, sehingga menghasilkan teks yang tidak mengandung tautan.

Before	After
batuan langka, tasbih jokowi hadiah dari habib luthfi seharga mercy? http://dlvr.it/r2pvzvâ	batuan langka, tasbih jokowi hadiah dari habib luthfi seharga mercy?
populer b1 4: jokowi: naikkan tax ratio drastis bisa guncang ekonomi http://brt.st/6b4kâ	populer b1 4: jokowi: naikkan tax ratio drastis bisa guncang ekonomi

- Terdapat 626 perubahan data

REMOVE NUMERIC

Menghapus angka dari teks selain angka (1, 2, 01, 02) dan menggantikan angka tertentu dengan kata-kata terkait, sehingga menghasilkan teks yang tidak mengandung representasi numerik.

Before	After
dengan memilih 02 kita sudah bersodakoh	dengan memilih dua kita sudah bersodakoh
no. 5 donk mang. dan sayah percaya, semua saudara seperjuangan 02	no. donk mang. dan sayah percaya, semua saudara seperjuangan dua

- Terdapat 757 perubahan data

REMOVE STRING EMOTICON

Menghapus string emotikon dari teks dengan menggantikan emotikon tersebut dengan string kosong, sehingga menghasilkan teks yang tidak mengandung representasi visual emotikon.

Before	After
leres pakdhe :) salah satu aspek penting dari faktor ekonomi yg bertumbuh adalah stabilitas politik	leres pakdhe salah satu aspek penting dari faktor ekonomi yg bertumbuh adalah stabilitas politik
akutuh kasihan ngelihat pak jokowi udah berumur, tapi masih ttp trsnyum disaat org ngebenci dia:(diluar dia jadi capres	akutuh kasihan ngelihat pak jokowi udah berumur, tapi masih ttp trsnyum disaat org ngebenci dia diluar dia jadi capres

- Terdapat 12 perubahan data

REMOVE PUNCTUATION

Membersihkan tanda baca dari teks dengan menggantikan tanda baca tersebut dengan spasi kosong, dan menghapus spasi yang berlebihan, sehingga menghasilkan teks yang bebas dari tanda baca dan spasi yang tidak diperlukan.

Before	After
dengan memilih dua kita sudah bersodakoh, karna prabowo sandi menyumbangkan gaji mereka untuk negara,fakir miskin, yatim piatu, masjid.	dengan memilih dua kita sudah bersodakoh karna prabowo sandi menyumbangkan gaji mereka untuk negara fakir miskin yatim piatu masjid
assalamu'alaikum. saya tahu satu suara itu sgt berharga dlm upaya memenangkan dua @prabowo & @sandiuno saya hrs merelakan suara	assalamualaikum saya tahu satu suara itu sgt berharga dlm upaya memenangkan dua prabowo sandiuno saya hrs merelakan suara

- Terdapat 1742 perubahan data

EXTRACT EMOJI

Mengekstraksi emoji dari teks dan menggantinya dengan makna teks terkait, menjaga teks non-emoji tidak berubah, sehingga menghasilkan teks yang terstandarisasi tanpa emoji.

sumber :
<https://github.com/jolicode/emoji-search/blob/main/synonyms/cldr-emoji-annotation-synonyms-id.txt>

Before	After
ayo milih jokowi 😊	ayo milih jokowi bahagia
habis gini presiden akan ganti nih 😍	habis gini presiden akan ganti nih jatuh cinta

- Terdapat 48 perubahan data

REMOVE SPECIAL CHAR

Menghapus karakter-karakter khusus dari teks, mempertahankan hanya kata-kata yang terdiri dari alphanumeric, sehingga menghasilkan teks yang hanya berisi kata-kata yang tersusun atas alphanumeric.

Before	After
kata prabowo indonesia tidak dihargai bangsa asing berita ini ðŸ‘† pasti hoax	kata prabowo indonesia tidak dihargai bangsa asing berita ini pasti hoax
prabowo sandi sepakat tak ambil gaji karena negara sedang susah â€¦ go aceh	prabowo sandi sepakat tak ambil gaji karena negara sedang susah go aceh

- Terdapat 1111 perubahan data

REMOVE REPEATING CHAR

Menghapus karakter berulang lebih dari dua kali dalam teks, sehingga menghasilkan teks yang bebas dari karakter berulang.

Before	After
yg diserang ekonomi makro woooyyy lu liat pengangkut sembako dulu belum ada tol dia dijalan mikir jam sekarang tinggal bayar waktu terganti gw pelakunya	yg diserang ekonomi makro woy lu liat pengangkut sembako dulu belum ada tol dia dijalan mikir jam sekarang tinggal bayar waktu terganti gw pelakunya
prabowo sudah berbuat apa untuk rakyat kecil sedang gaji karyawannya aja belum dibayar sampai sekarang heheheheeee	prabowo sudah berbuat apa untuk rakyat kecil sedang gaji karyawannya aja belum dibayar sampai sekarang hehehehe

- Terdapat 84 perubahan data

REMOVE SINGLE WORD

Membersihkan karakter tunggal yang berdiri sendiri dalam teks, menghasilkan teks yang tidak mengandung karakter tunggal yang tidak berarti.

Before	After
populer b jokowi naikkan tax ratio drastis bisa guncang ekonomi	populer jokowi naikkan tax ratio drastis bisa guncang ekonomi
y memang lebih kecil tp pngetahuan tata kelola ekonomi negara jauh kelasnya	memang lebih kecil tp pngetahuan tata kelola ekonomi negara jauh kelasnya

CHECKING INFORMAL WORD

Mencari kata-kata informal dalam kalimat dengan mencocokkan kata formal pada KBBI

Terdapat 6964 Kata Informal yang dapat dipertimbangan untuk diproses pada penelitian lebih lanjut.

Informal Words
abg, abis, afk, ajja, akhlaq, baper, bljr, cukong, dableg, wkwkwkw

STEMMING

Melakukan stemming pada teks, menghasilkan bentuk dasar kata tersebut.

Before	After
presiden jokowi dan raja salman sepakat tingkatkan kerja sama ekonomi via metroball	presiden jokowi dan raja salman sepakat tingkat kerja sama ekonomi via metroball
berita terbaru prabowo saya tak ingin naikkan pajak persen dalam setahun telah ada di	berita baru prabowo saya tak ingin naik pajak persen dalam tahun telah ada di

- Terdapat 1690 perubahan data

STOPWORD

Pada project ini stopword tidak digunakan karena terdapat beberapa kata yang akan mengubah makna sentimen apabila di hapus.

Stopword
tidak, bukan, belum, enggak, entah, mungkin, seharusnya, sebaiknya, sebelumnya

DATASET PIPELINE

Frequency Based Vectorizing

- Count Vector
- TF-IDF
- Co-Occurrence Matrix
- N-Gram (unigram)
- N-Gram (unigram - bigram)
- N-Gram (unigram - trigram)

Prediction Based Vectorizing

- CBOW
- SkipGram

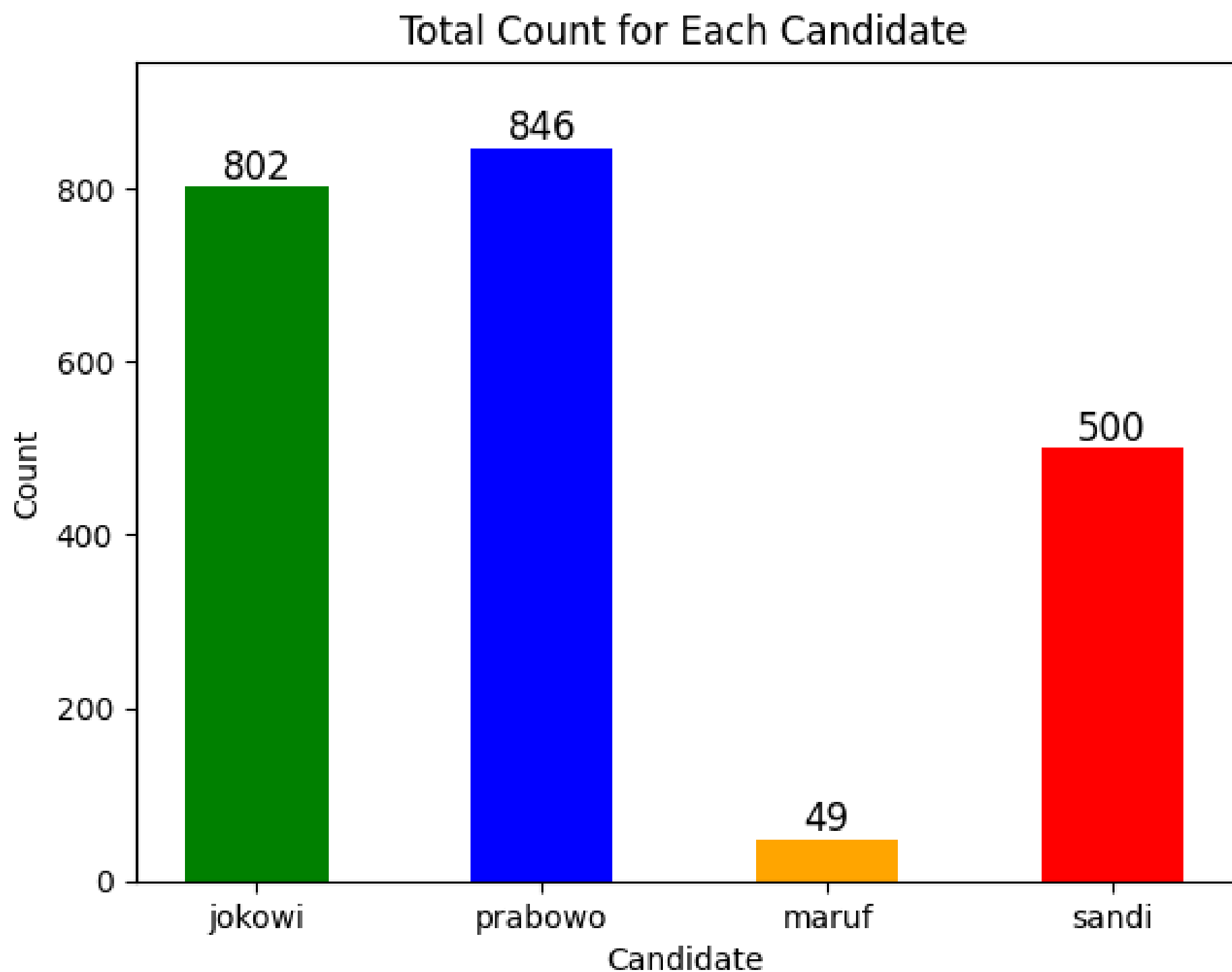


EDA



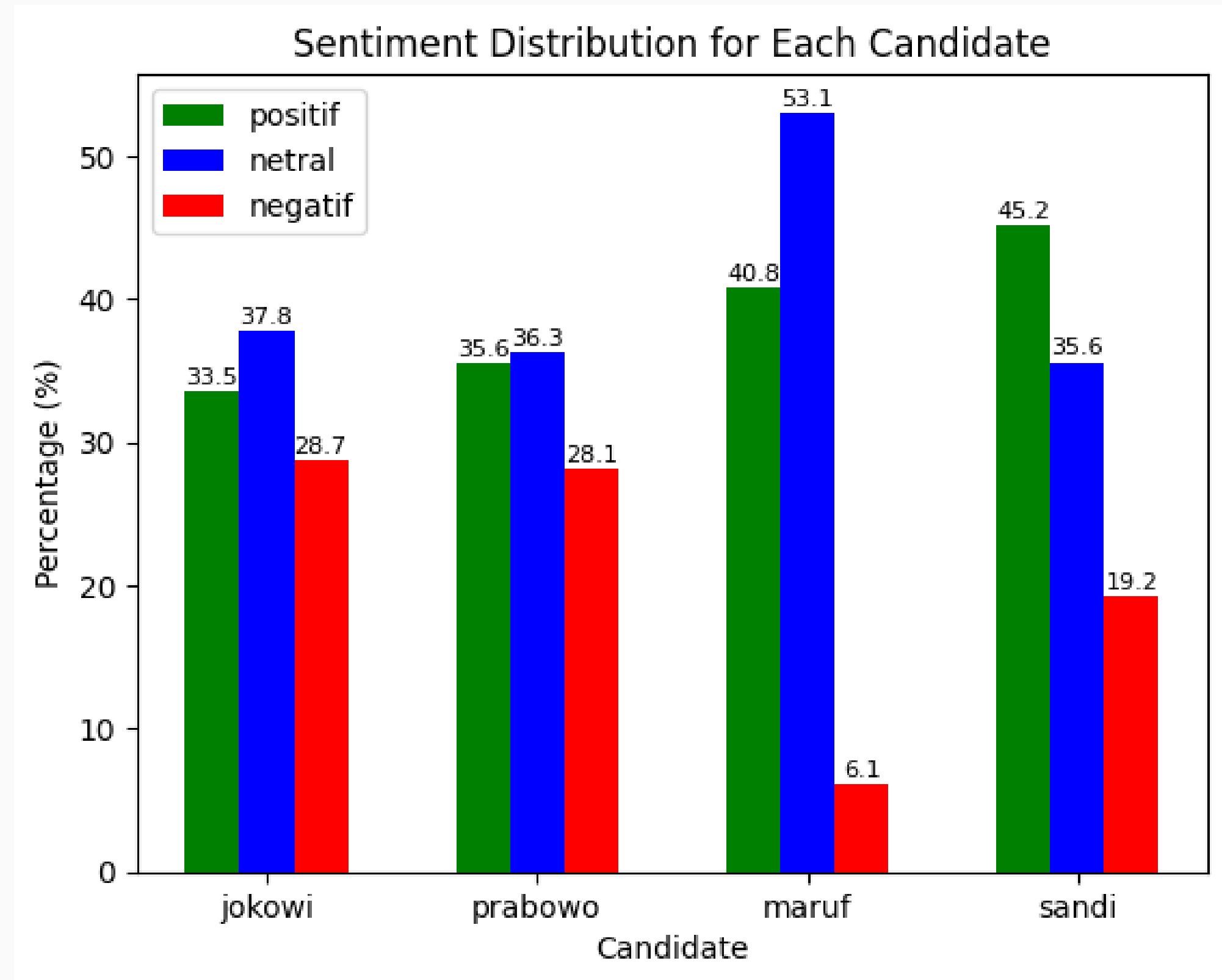
Total Count

Analisis data total penyebutan kandidat menunjukkan dominasi Prabowo, diikuti oleh Jokowi, Sandi, dan Maruf. Pola ini mencerminkan tidak hanya visibilitas kandidat, tetapi juga potensi pengaruhnya terhadap opini publik, menyoroti pentingnya pemahaman tren dalam dinamika pemilihan.



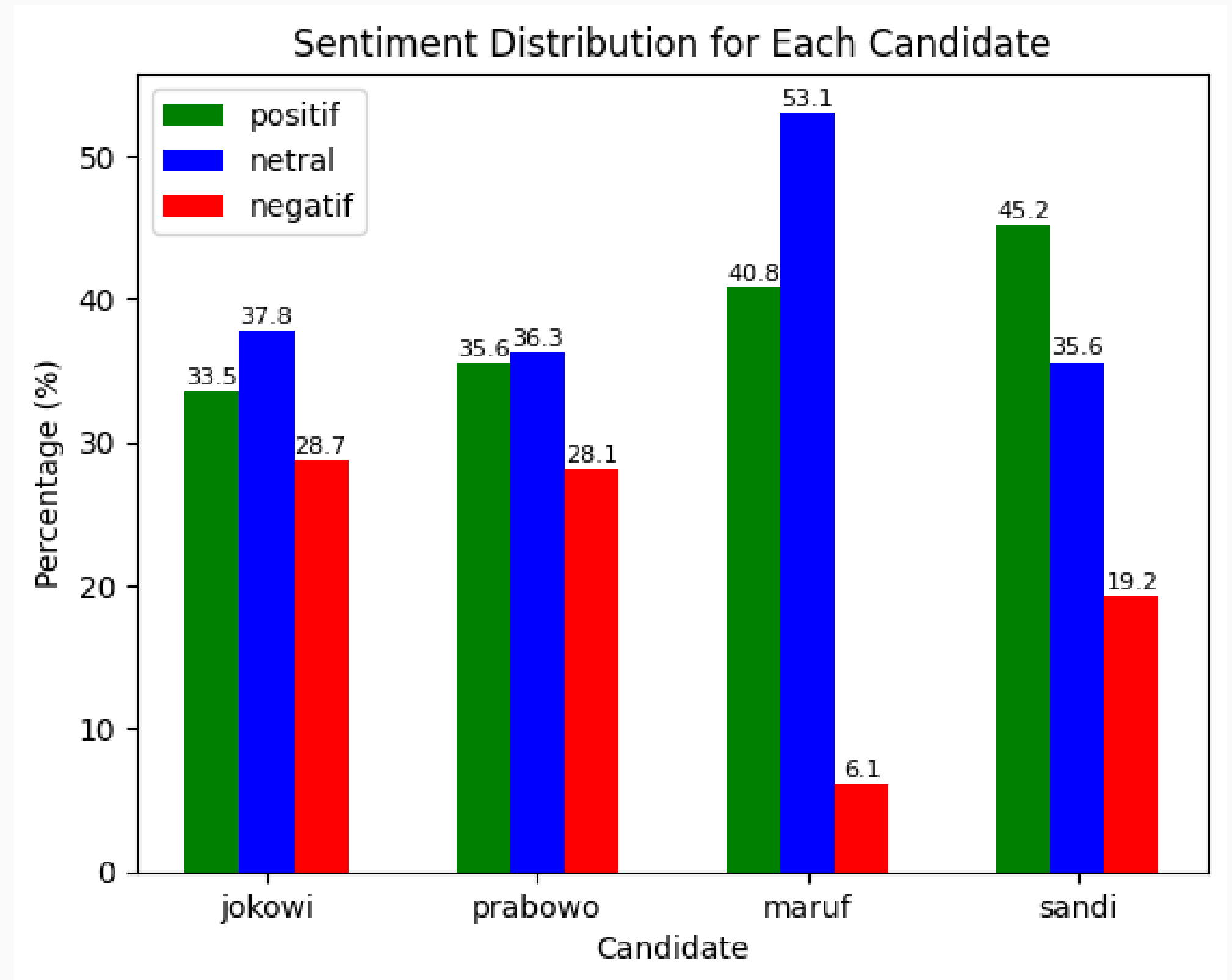
Sentiment Distribution

- **Jokowi** : Dukungan stabil, ditandai dengan tingkat sentimen netral yang tinggi, tingkat sentimen positif yang hampir setara dengan Prabowo menunjukkan basis dukungan yang kuat.
- **Prabowo** : Polarisasi opini yang seimbang, dengan distribusi merata antara sentimen positif, netral, dan negatif, menunjukkan divergensi pendapat dalam masyarakat, meskipun proporsi sentimen negatifnya lebih kecil.



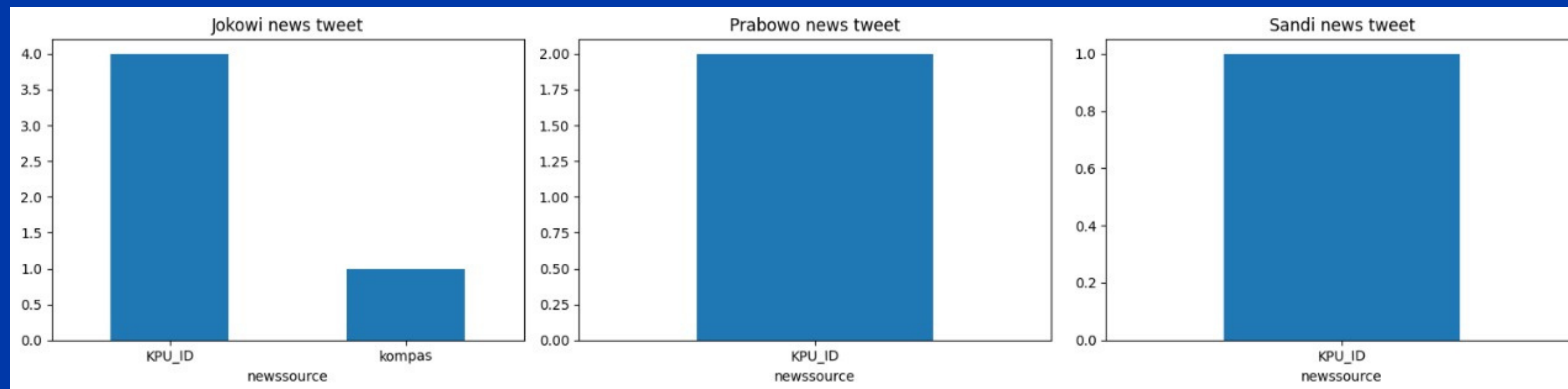
Sentiment Distribution

- **Maaruf** : Meskipun kurang populer dibandingkan kandidat lainnya, tingkat sentimen positif yang signifikan menunjukkan adanya basis pendukung yang solid.
- **Sandi** : Tingkat dukungan yang tinggi tercermin dari tingkat sentimen positif yang tertinggi, dan proporsi sentimen negatifnya yang rendah menunjukkan adanya kepuasan atau konsistensi opini yang baik dalam masyarakat



News

- nytimes, jurnal_IDN, tempoteve, KPU_ID, CNNIndonesia, Kompas, Metro_TV, Berisatu, tvonetv, PaaISocmed
- Kami melakukan eksplorasi dari dataset untuk menemukan apakah sumber media tersebut menyebutkan kandidat tertentu
- Dari data ini dapat disimpulkan bahwa KPU menyebutkan kandidat lebih banyak daripada berita lain



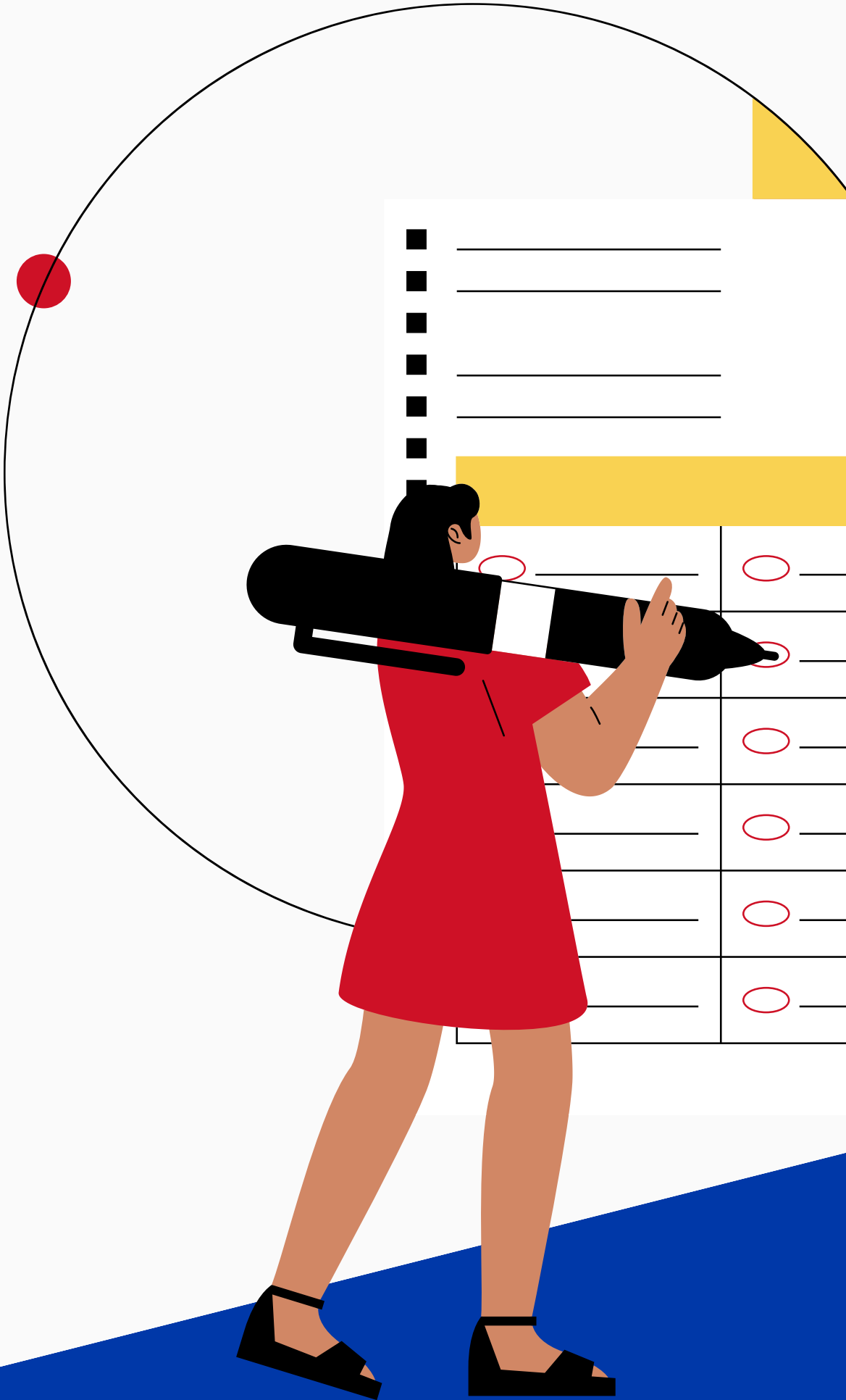
MODELING



LSTM

ARSITEKTUR

Layer	
Embedding	Input Dim=1000, Output_dim=50
LSTM	Dense = 64
Hidden layer	Dense = 256, Activation = Relu
Dropout	0.5
Output	3, Activation=Softmax



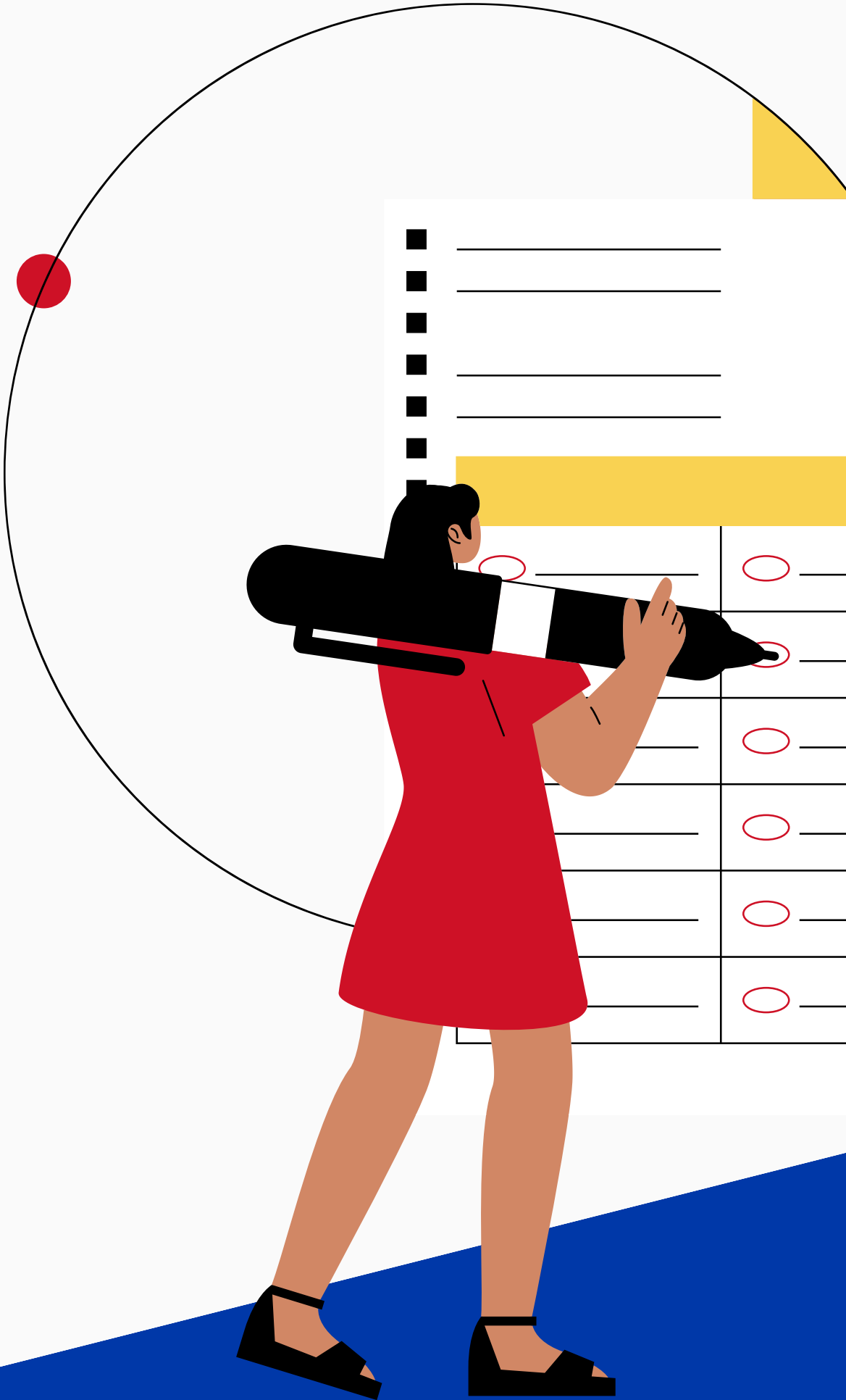
EVALUATION



RANDOM FOREST

COUNT VECTORIZER

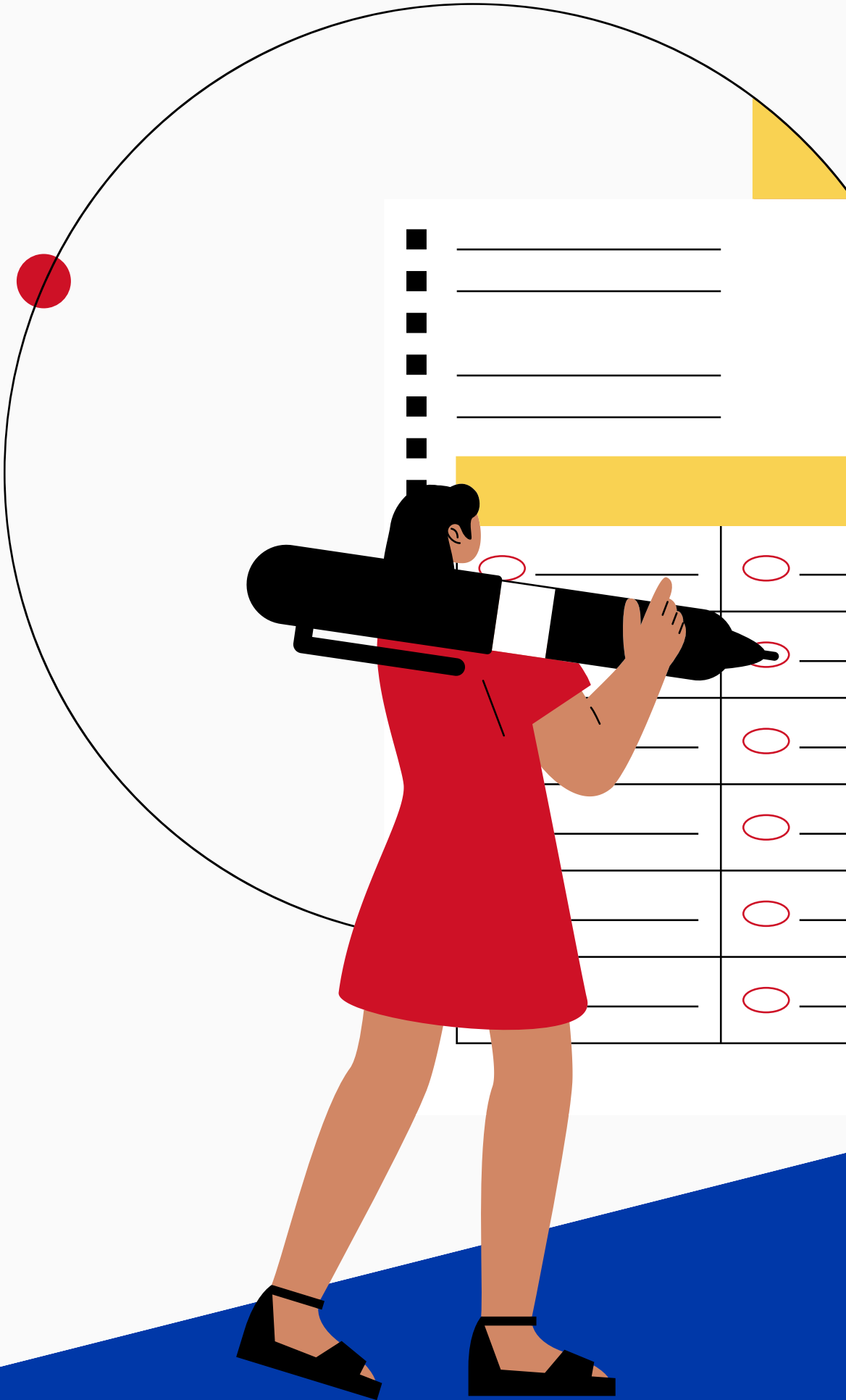
Ngram	Test Size	Accuracy	F1
Unigram	10	61%	61%
Unigram	20	59%	59%
Unigram	30	59%	59%
Unigram - Bigram	10	59%	59%
Unigram - Bigram	20	58%	58%



RANDOM FOREST

TF-IDF

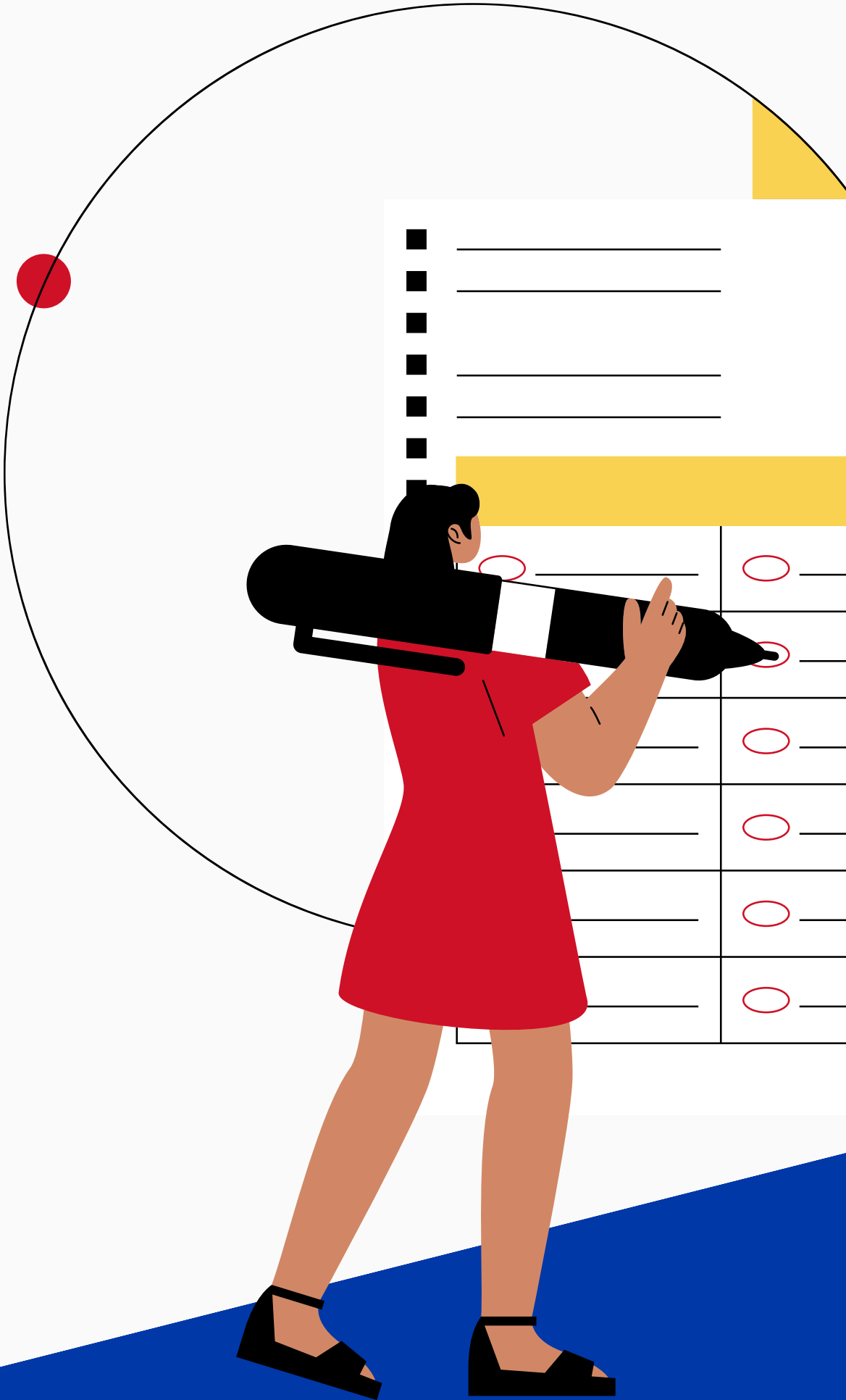
Ngram	Test Size	Accuracy	F1
Unigram - Bigram	10	66%	66%
Unigram	20	61%	61%
Unigram - Bigram	20	61%	61%
Unigram - Bigram	30	58%	58%
Unigram	10	58%	58%



LSTM

COUNT VECTORIZER

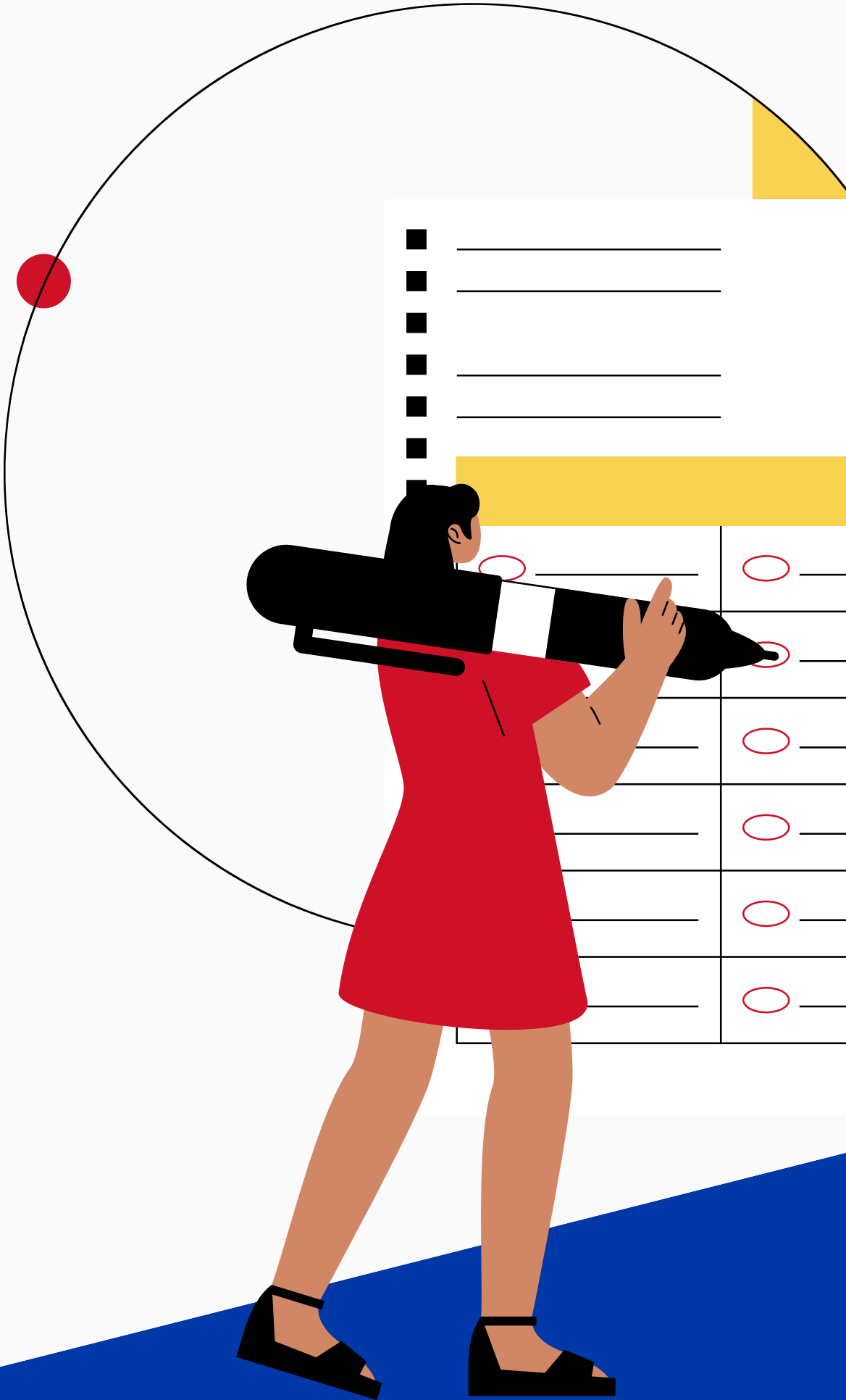
Ngram	Accuracy	F1
Unigram	36%	19%
Unigram - Bigram	36%	19%
Trigram	36%	19%
Bigram	32%	15%

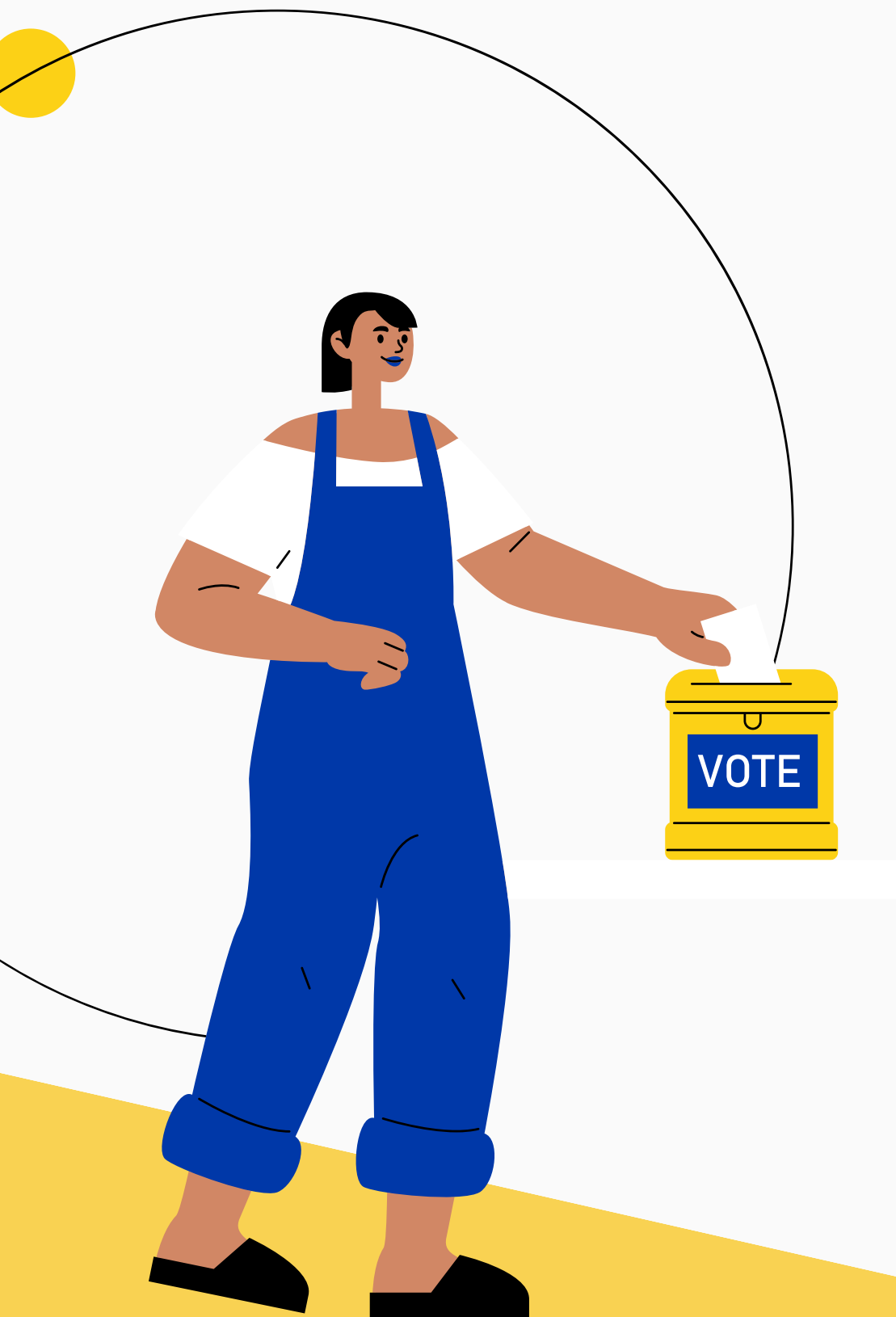


LSTM

TF-IDF

Ngram	Accuracy	F1
Bigram	36%	19%
Unigram - Bigram	33%	19%
Unigram	33%	19%
Trigram	33%	15%





CONCLUSION

Dalam proyek ini, kami menganalisis sentimen pada data tweet terkait Pemilihan Umum 2019 di Indonesia. Menggunakan vectorizer TF-IDF Unigram - Bigram dan model Random Forest, kami mencapai tingkat akurasi 66% dan skor F1 61%. Hasil evaluasi menunjukkan efektivitas model dalam memprediksi sentimen dari tweet, memberikan wawasan penting terhadap opini dan sikap masyarakat terhadap peristiwa politik ini.

Your Vote Is Your Voice. Use It Wisely.

