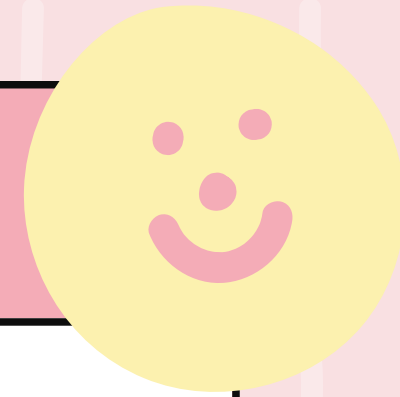


KELOMPOK BIG DATA PROJECT



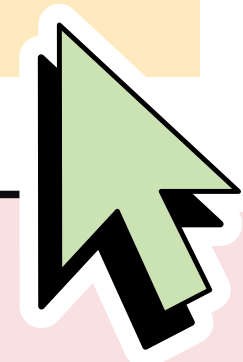
Deteksi Berita Palsu atau Hoaks di Media Sosial dengan Natural Language Processing (NLP)

- ADRIEL JONATHAN SANTOSO (2440026231)
- EYKMAN OMASSY (2440041882)
- NOVIA SARI (2440051076)
- FIDELIA LAVANI (2401954905)

UNIVERSITAS BINA NUSANTARA

WELL DONE!

GOALS





Isi Presentasi Kami

1

Latar Belakang

2

tujuan

3

Metode Penelitian

4

Kesimpulan

ALMOST
THERE!





Latar Belakang

have
a
nice
day

EXERCISE

Saat ini, sosial media bagaikan dua sisi bagi masyarakat terutama dalam penyebaran berita dengan adanya berita hoax/palsu. Terdapat banyak penyebaran berita hoax/palsu melalui sosial media dengan berbagai tujuan seperti provokasi, fanatisme, pengakuan, dan lain-lain. Berita hoax/palsu dapat menyebabkan keresahan, ketidakamanan, ketidaknyamanan bagi masyarakat. Selain itu, berita hoax dapat berdampak pada persatuan negara jika berita hoax/palsu tersebut digunakan sebagai bentuk provokasi suatu topik antara satu kelompok dengan kelompok lainnya.

Tujuan

1

tujuan utama

tujuan utama dari penelitian ini yaitu diharapkan agar dapat membantu masyarakat dalam mendeteksi informasi yang asli atau palsu dalam kehidupan sehari-harinya.

MAKE IT
HAPPEN

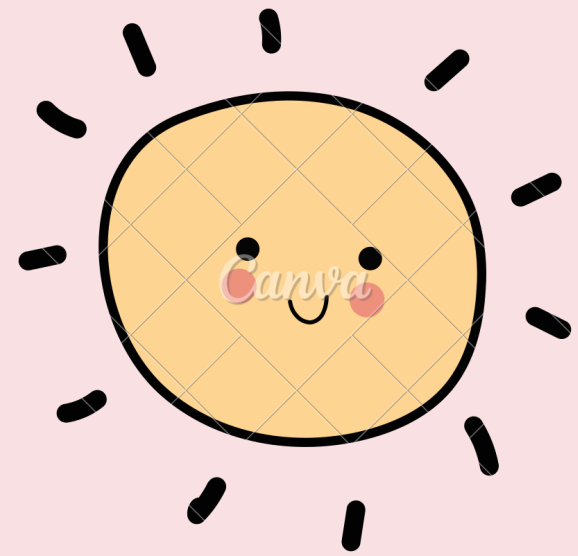
Tujuan

2

tujuan khusus

1. Mampu membedakan penulisan berita yang benar ataupun tidak
2. Menganalisis validitas isi dari berita yang beredar
3. Mendorong masyarakat untuk melakukan penulisan dan penyampaian berita secara baik dan benar
4. Mendorong masyarakat agar dapat berpikir secara aktif dan kritis saat membaca dan memahami sebuah berita
5. Mendorong masyarakat untuk turut melaporkan setiap berita palsu yang beredar

MAKE IT
HAPPEN



METODE PENELITIAN



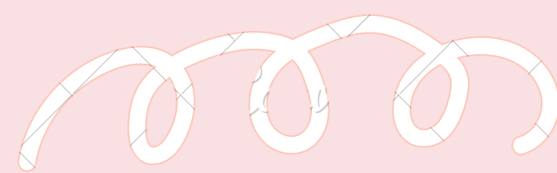


1

Lokasi dan Waktu Penelitian



Dalam proses pengerjaan penelitian, kami lakukan secara remote work yang melalui online platform dari rumah masing - masing tanpa perlu melakukan pertemuan secara tatap muka. Penelitian dimulai dari bulan Juni tanggal 1 tahun 2022 yang memiliki timeline check setiap seminggu sekali dengan waktu sesuai diskusi via chat



2

Metode Pengumpulan Data

Pada analisis ini, kami memutuskan untuk melakukan metode kuantitatif setelah itu merangkum hasil data yang didapat ke dalam bentuk kuantitatif yang diolah menjadi sebuah tabel ataupun grafik.

3

Tahap Penelitian

IMPORTANT!

1. Pengumpulan berita dari website Jakarta Open Data serta melalui situs berita online

2. Berita atau dataset yang telah dikumpulkan kemudian melakukan uji coba terhadap algoritma pendeteksi berita hoax dari peneliti lain

3. Melakukan sebuah tahapan converting, yang mengubah semua menjadi huruf kecil. Selain itu, kata-kata yang memiliki arti variabel yang sama akan dianggap sebagai satuan kata yang sama.

NOTES

IMPORTANT!


5. Selanjutnya, data yang sudah didapatkan dibagi menjadi dua segmen yaitu Real news data dan Fake news data yang dipilih melalui pihak ketiga secara acak yaitu algoritma.

4. Masuk ke dalam tahap screening yang bertujuan untuk menghapuskan karakter ataupun variabel diluar dari huruf ataupun angka.

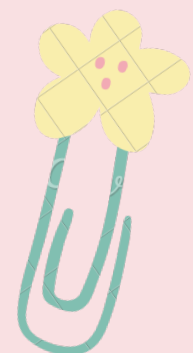
6. Ketika hasil dari algoritma tersebut sudah berhasil dilakukan, maka tahap berikut yaitu mengklasifikasi tonality dari data tersebut menjadi tiga yaitu positif, negatif dan netral.



IMPORTANT!



8. Data tersebut kemudian dibuat analisis dengan melakukan membandingkan antara rumusan masalah dan hasil yang didapatkan. Proses ini nantinya akan menimbulkan insight baru terhadap berita yang sudah dipublikasikan khalayak umum.



4

Analisis Data



```
import pandas as tu
```

```
[ ] real = tu.read_csv("realnews.csv")  
    fake = tu.read_csv("fakenews.csv")
```



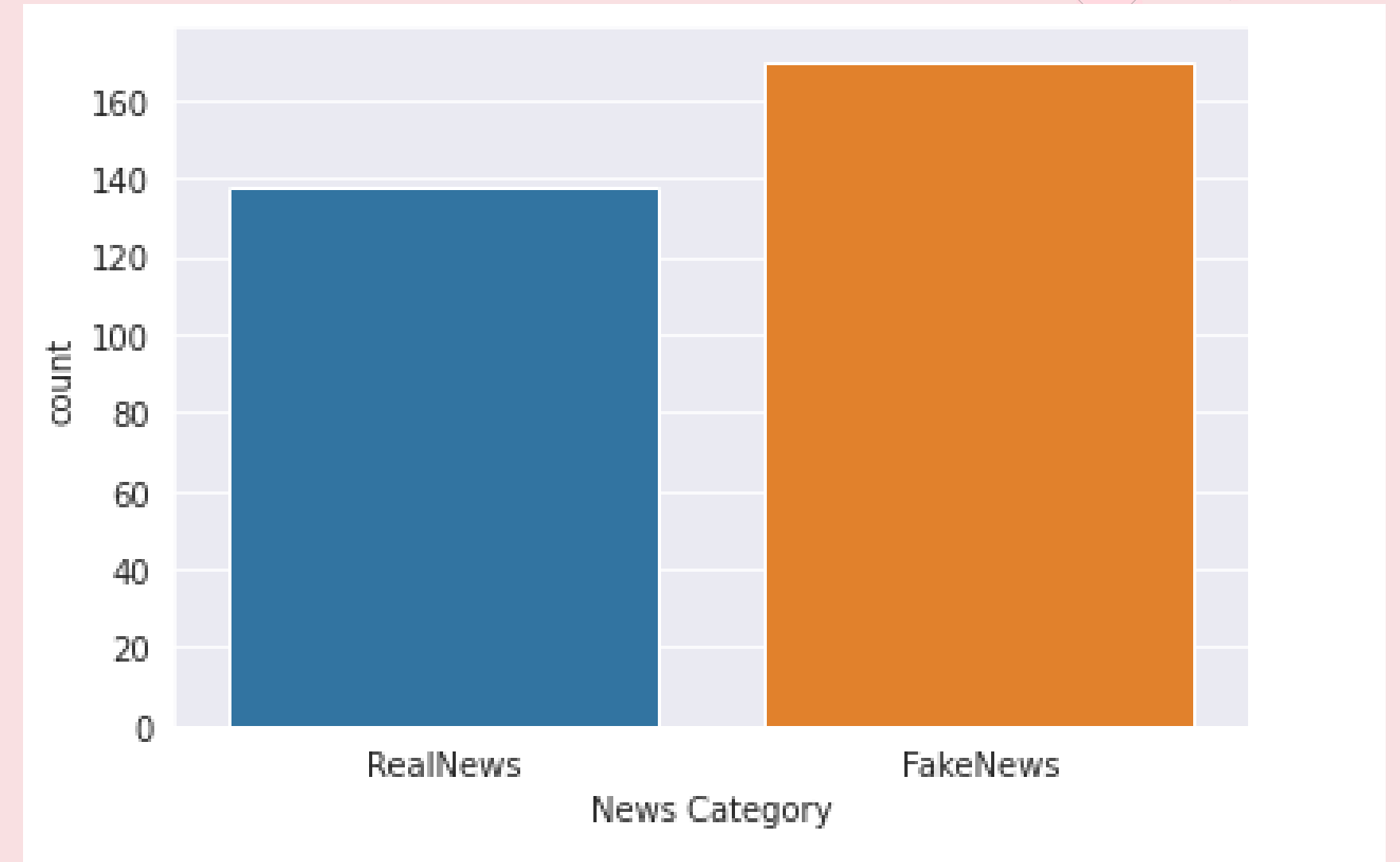
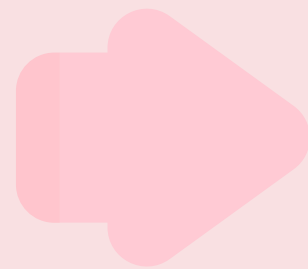
```
real['News Category'] = "RealNews"  
fake['News Category'] = "FakeNews"  
Table = tu.concat([real, fake])
```



```
import seaborn as sea  
sea.countplot(Table['News Category']);
```

YAY!

Hasil Grafik Keseimbangan Data Berita Palsu dan Asli



4

Analisis Data (Continued)

```
Table.isnull().sum()
```

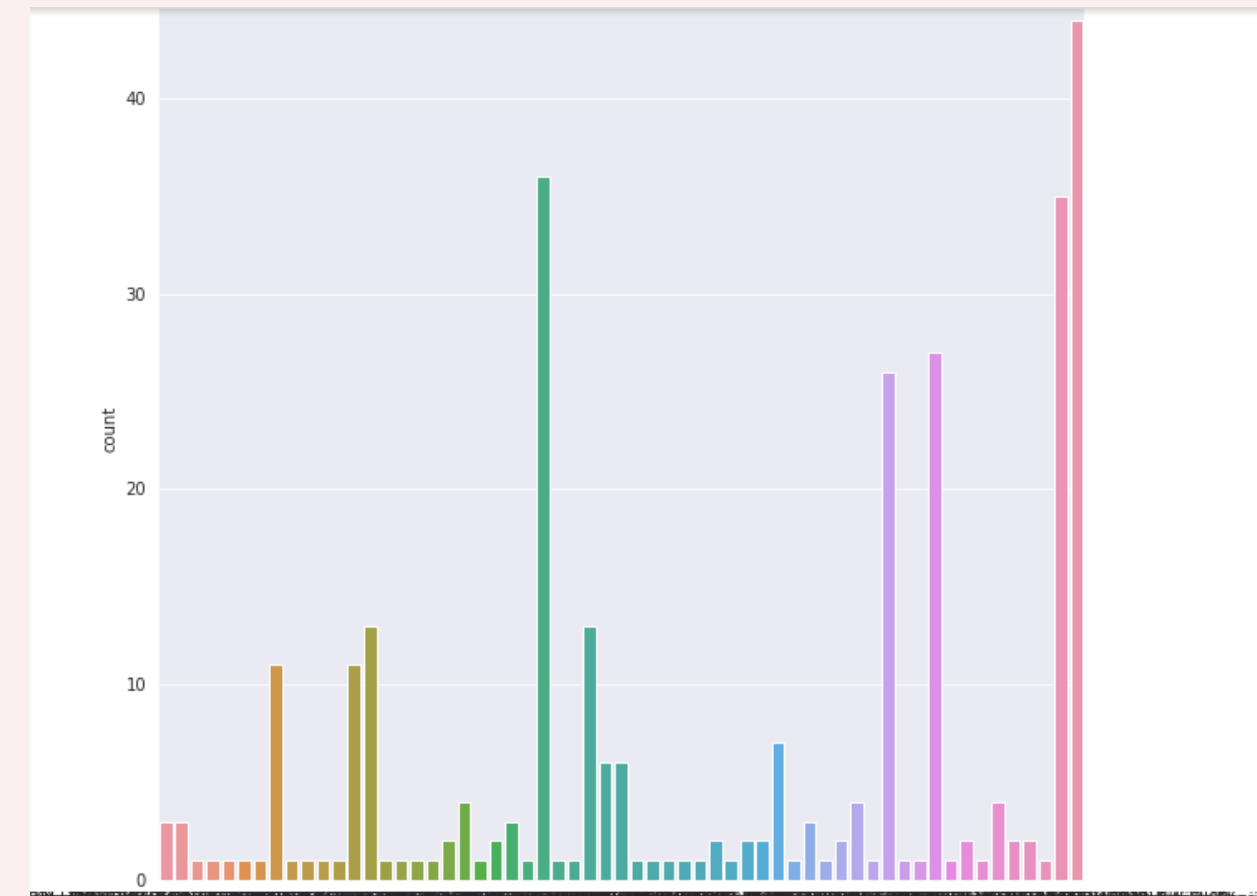
Result :

```
tanggal      0
isu           0
tonality      0
berita_jakarta  0
link/page     0
News Category  0
dtype: int64
```


4

Analisis Data (Continued)

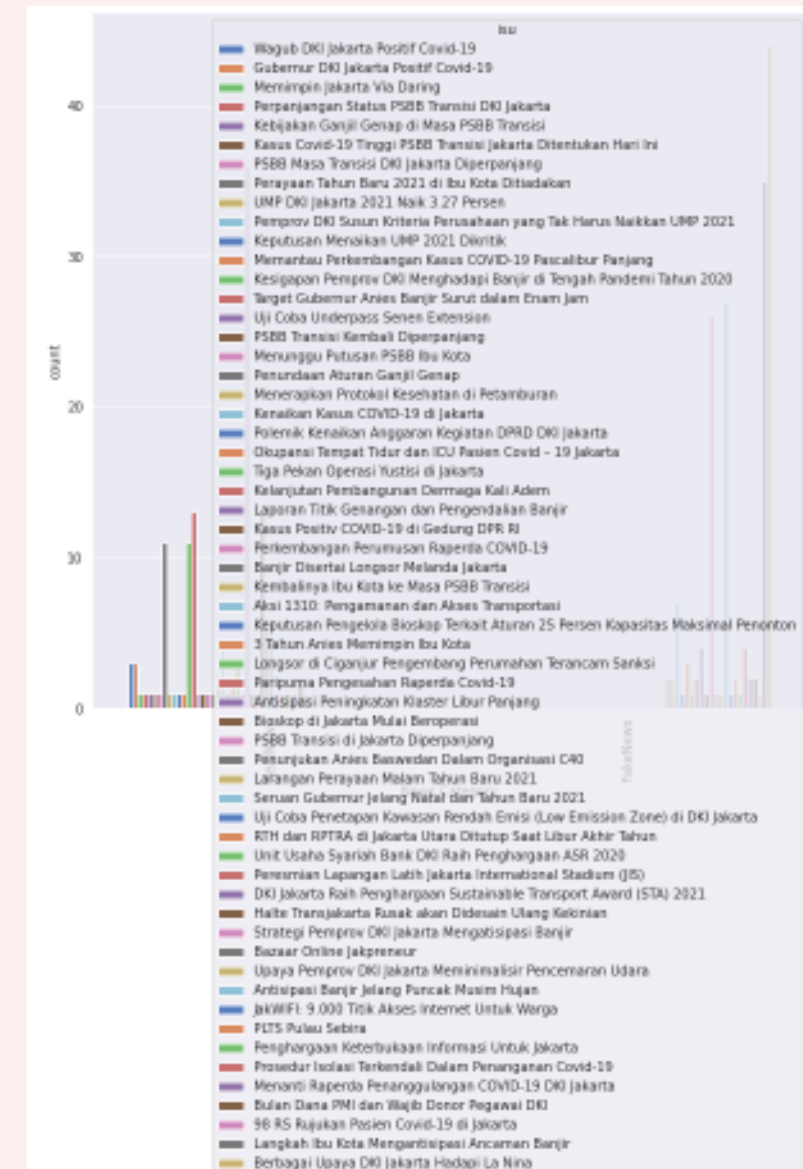
```
▶ import matplotlib.pyplot as plt  
data['isu'].value_counts()  
plt.figure(figsize = (10,10))  
sea.countplot(data['isu']);
```



4

Analisis Data (Continued)

```
plt.figure(figsize = (10,10))
chart = sea.countplot(x = "News Category",hue ="isu", data = Table , palette = 'muted')
sea.set_style("darkgrid")
chart.set_xticklabels(chart.get_xticklabels(),rotation=90)
```



4

Analisis Data (Continued)

```
[ ]
    Table['isu'] = Table['berita_jakarta'] + " " + Table['isu']
    Table = Table.drop(['berita_jakarta', 'link/page', 'tanggal'], axis=1)
```



```
from wordcloud import WordCloud
```

```
import nltk
nltk.download('stopwords')
```



```
wordcloud = WordCloud(width = 800, height = 800,
                        background_color = 'white',
                        stopwords = stopwords.words('english')
                        , min_font_size = 10).generate(" ".join(Table[Table['News Category'] == "FakeNews"].isu))
```

```
plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
plt.show()
```


4

Analisis Data (Continued)

```
print(Table[Table['News Category'] == 0])
```

Empty DataFrame

Columns: [isu, tonality, News Category]

Index: []

▶ print(Table)

	isu	tonality	News Category
0	Terkonfirmasi Positif COVID-19 Wagub Ariza Sam...	NETRAL	RealNews
1	Besok JIC Adakan Doa Bersama Virtual untuk Kes...	NETRAL	RealNews
2	437 Orang Kontak Erat Gubernur dan Wagub Mengi...	NETRAL	RealNews
3	Gubernur Konfirmasi Positif COVID-19 Gubernur ...	NETRAL	RealNews
4	Besok JIC Adakan Doa Bersama Virtual untuk Kes...	NETRAL	RealNews
..
165	Kelurahan Karet Semanggi Tambah Sumur Resapan...	POSITIF	FakeNews
166	Petugas Bersihkan Saluran Air Jl Balam Kramat ...	POSITIF	FakeNews
167	Dinas SDA DKI Keruk Lumpur Kali Sekretaris di ...	POSITIF	FakeNews
168	Posko Siaga Bencana Kampung Melayu Mulai Aktif...	POSITIF	FakeNews
169	Saluran Air Jl Utan Kayu Raya Dibersihkan Berb...	POSITIF	FakeNews

[308 rows x 3 columns]

4

Analisis Data (Continued)

```
wordcloud = WordCloud(width = 800, height = 800,  
                      background_color = 'white',  
                      stopwords = stopwords.words('english')  
                      , min_font_size = 10).generate(" ".join(Table[Table['News Category'] == "RealNews"].isu))  
  
plt.figure(figsize = (8, 8), facecolor = None)  
plt.imshow(wordcloud)  
plt.axis("off")  
plt.tight_layout(pad = 0)  
plt.show()
```


Canva

4

Analisis Data (Continued)

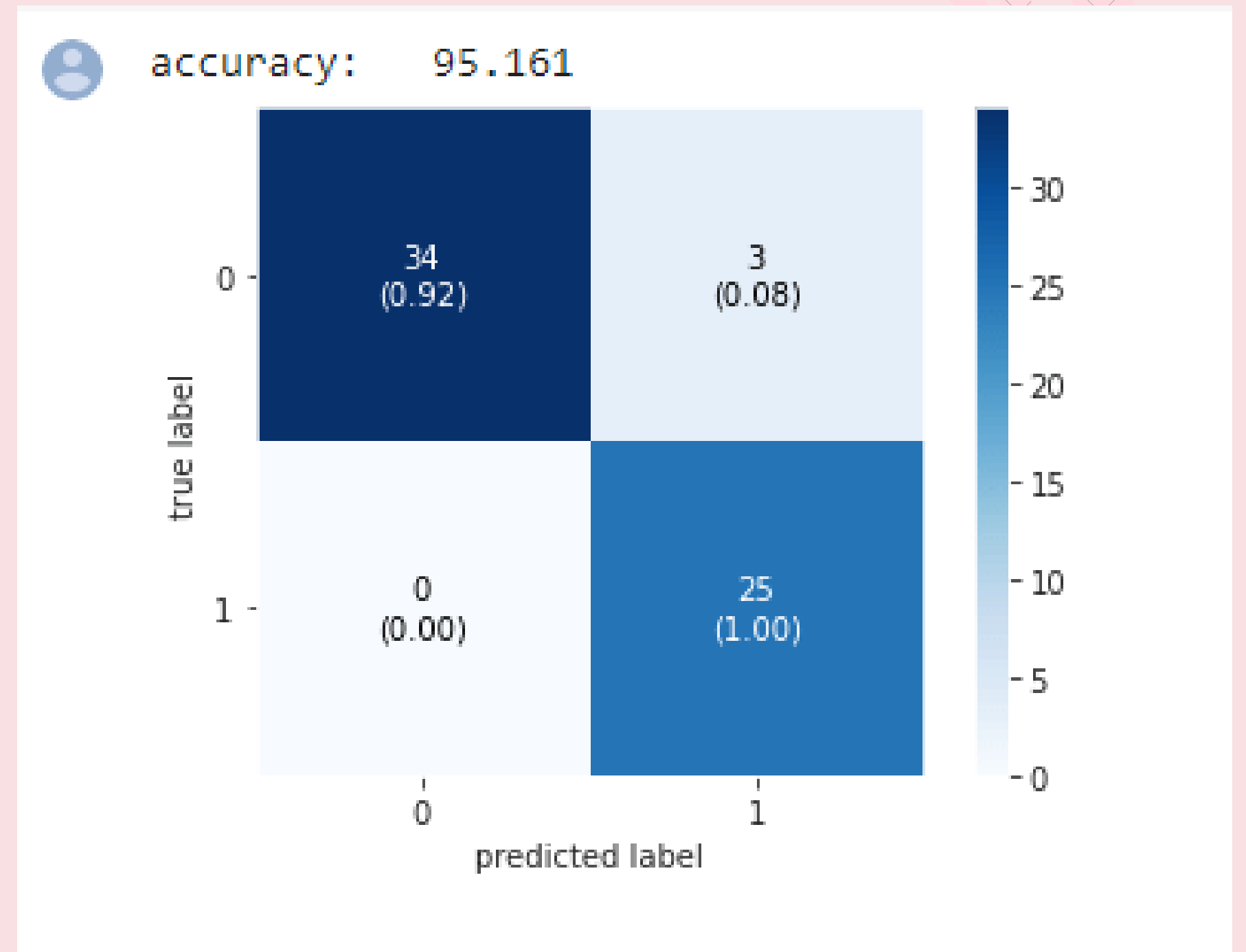
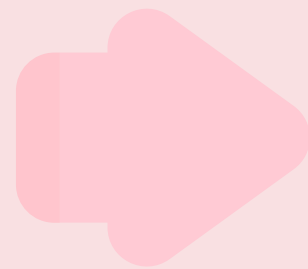
```
[ ] import sklearn
    from sklearn.model_selection import train_test_split
    x_train,x_test,y_train,y_test = train_test_split(data['isu'],data['News Category'],test_size=0.2, random_state = 1)
```

```
[ ]
    from sklearn.feature_extraction.text import TfidfTransformer
    from sklearn.feature_extraction.text import CountVectorizer
    from sklearn.feature_extraction.text import TfidfVectorizer
    from sklearn.pipeline import Pipeline
    from sklearn.naive_bayes import MultinomialNB
    from sklearn.metrics import accuracy_score
    import sklearn.metrics as metrics
    from mlxtend.plotting import plot_confusion_matrix
    from sklearn.metrics import confusion_matrix
```

Analisis Data (Continued)

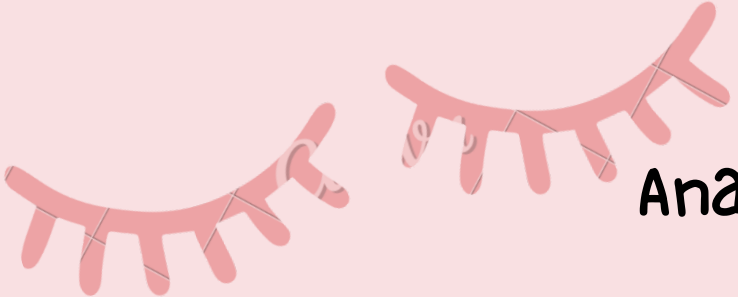
YAY!

Confusion Matrix (MultiNomial Naive Bayes Model)

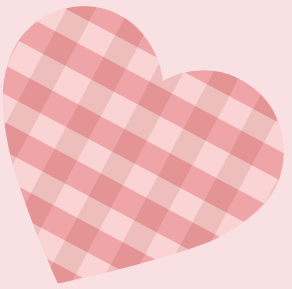
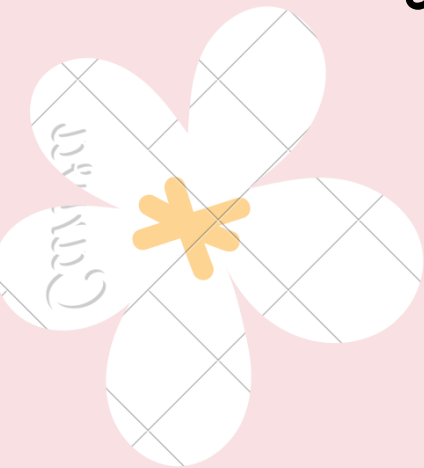

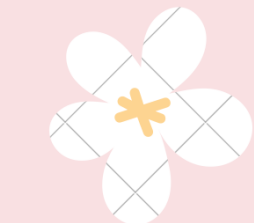
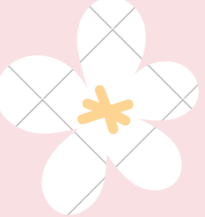
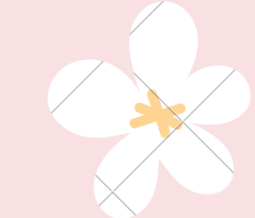
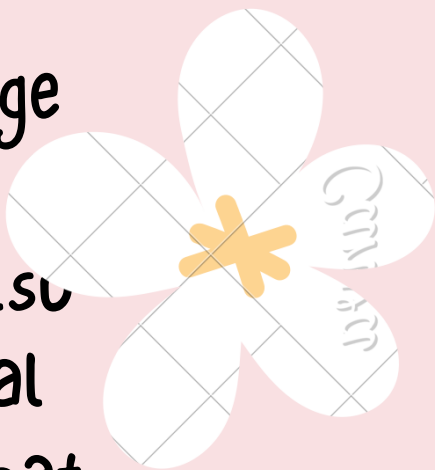
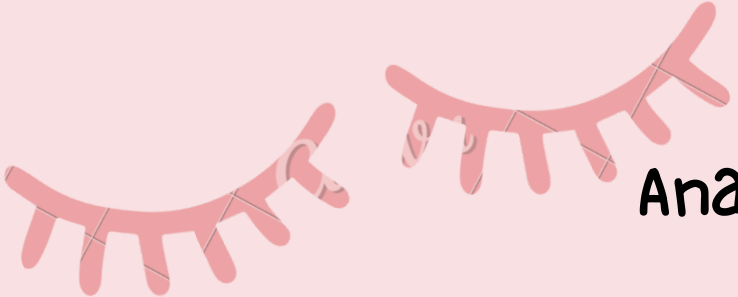




Kesimpulan



Analisis deteksi berita palsu/hoax menggunakan NLP (Natural Language Processing) berhasil dilakukan dengan tingkat akurasi 95.161%. Berdasarkan hasil analisis wordcloud, terdapat kata dari isu berita palsu yang meragukan atau provokatif sedangkan kata dari isu berita aktual dengan kata yang bermakna. Dengan demikian kata kunci tersebut dapat digunakan patokan bagi masyarakat untuk lebih berhati-hati dan seleksi berita. Masyarakat juga perlu menghimbau lebih dalam terhadap penulisan judul berita dengan apa yang disampaikan oleh berita.





THANKYOU

