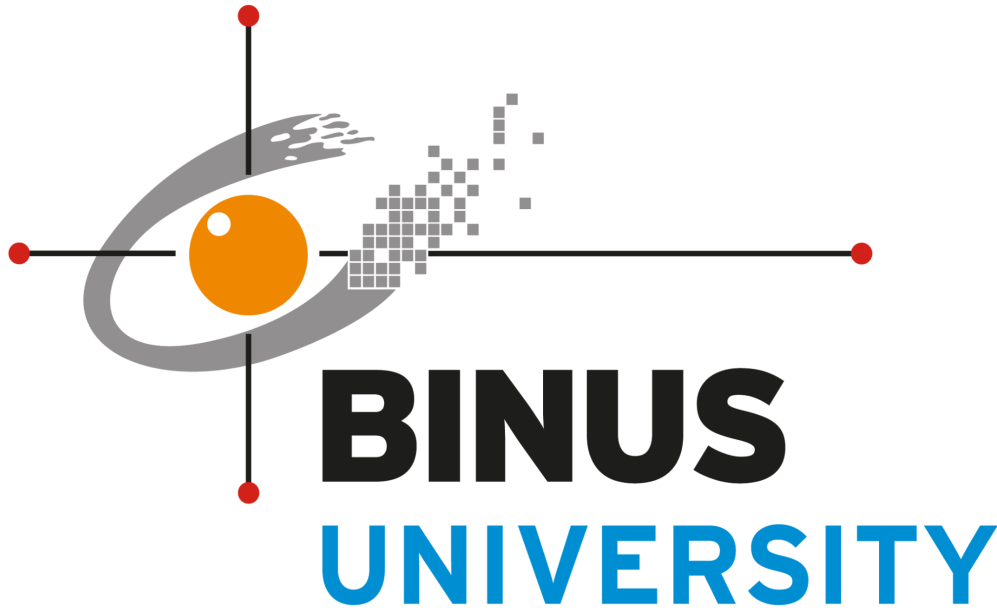


Laporan Data Analisis
Deteksi Berita Palsu atau Hoaks di Media Sosial
dengan Natural Language Processing (NLP)



Disusun oleh :

Adriel Jonathan Santoso	2440026231
Eykman Omassy	2440041882
Novia Sari	2440051076
Fidelia Lavani	2401954905

May 30, 2022

DAFTAR ISI

DAFTAR ISI	1
BAB I. PENDAHULUAN	2
1.1 Latar Belakang Masalah	2
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II. TINJAUAN PUSTAKA	4
2.1 Landasan Teori	4
2.1.1 Berita Bohong atau Hoax	4
2.1.2 Media Sosial	5
2.1.3 Natural Language Processing (NLP)	6
BAB III. METODE PENELITIAN	8
3.1 Lokasi dan Waktu Penelitian	8
3.2 Teknik Pengumpulan Data	8
3.3 Teknik Analisis Data	10
3.4 Analisis Sistem	11
BAB IV. KESIMPULAN DAN SARAN	19
4.1 Kesimpulan	19
4.2 Saran	19
REFERENSI	20

BAB I PENDAHULUAN

1. Pendahuluan

1.1. Latar Belakang Masalah

Sosial media adalah teknologi perangkat lunak sebagai penghubung antara satu orang dengan orang lainnya untuk berkomunikasi dan berbagi konten serta informasi. Dengan adanya sosial media, berbagai informasi seperti berita, foto, video, artikel, dan lain-lainnya dapat diakses dengan mudah dari mana saja selama pengguna memiliki koneksi internet.

Saat ini, sosial media bagaikan dua sisi bagi masyarakat terutama dalam penyebaran berita dengan adanya berita hoax/palsu. Berita hoax/palsu adalah artikel berita yang diterbitkan bukan berdasarkan fakta. 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.

Pencegahan penyebaran berita hoax/palsu dimulai dari masyarakat yang bijak dalam seleksi berita yang tersebar di media sosial. Masyarakat perlu mengetahui berita yang terpercaya dari segala macam berita yang ada. Masyarakat perlu pengarahan terhadap cara memilih informasi yang benar dari berita yang beredar. Dengan adanya data analysis berita palsu, masyarakat dapat berpikir kritis dalam membaca dan memahami sebuah berita.

1.2. Rumusan Masalah

Adapun rumusan masalah untuk proposal ini adalah sebagai berikut.

- Apa pengertian dari berita palsu ?
- Apa saja topik berita yang umumnya ada di Indonesia?
- Apa saja faktor-faktor yang mendefinisikan berita sebagai berita palsu?
- Siapa yang umumnya terbawa oleh arus berita palsu di Indonesia?
- Bagaimana pengaruh berita palsu terhadap masyarakat Indonesia?
- Mengapa masyarakat dapat mempercayai berita palsu?
- Di mana saja berita palsu dipublikasikan atau disebarkan ke masyarakat umum?
- Bagaimana solusi dalam menangani arus berita palsu?

1.3. Batasan Masalah

Batasan-batasan masalah dalam penelitian ini adalah sebagai berikut.

- Penelitian ini melakukan analisis terhadap berita palsu di Indonesia.

- Analisis berita palsu dilakukan terhadap dataset pemberitaan dan monitoring di media massa tahun 2019 dan 2020 di Indonesia.
- Penelitian ini menggunakan metode analisis Natural Language Processing (NLP).
- Penelitian ini akan menggunakan tiga model klasifikasi data, yaitu MultiNomial Naive Bayes, Support Vector Machine, dan Passive Aggressive Classifier lalu melakukan perbandingan performa untuk ketiganya.

1.4. **Tujuan Penelitian**

Berdasarkan latar belakang dan rumusan masalah diatas adapun tujuan umum dan khusus dari penelitian ini, sebagai berikut :

1. Tujuan Umum

Tujuan utama dari penulisan proposal penelitian ini yaitu diharapkan agar dapat membantu pembaca dalam mendeteksi informasi yang asli atau palsu dalam kehidupan sehari-harinya.

2. Tujuan Khusus

- a. Mampu membedakan penulisan berita yang benar ataupun tidak
- b. Menganalisis validitas isi dari berita yang beredar
- c. Mendorong masyarakat untuk melakukan penulisan dan penyampaian berita secara baik dan benar
- d. Mendorong masyarakat agar dapat berpikir secara aktif dan kritis saat membaca dan memahami sebuah berita
- e. Mendorong masyarakat untuk turut melaporkan setiap berita palsu yang beredar

1.5. **Manfaat Penelitian**

Dari setiap penelitian pasti terdapatnya manfaat yang memberikan hal positif maupun hal negatif. Maka dari itu, manfaat dari penelitian ini adalah kita dapat mengetahui berita yang terpercaya dari segala macam berita yang ada. Dimana pembaca mendapatkan pengarahannya terhadap cara memilih informasi yang benar dari berita yang beredar. Dari berita yang beredar juga pembaca dapat menganalisa berita yang beredar dan berpikir kritis apa yang telah dibaca. Oleh karena itu, dari analisa Fake News ini dapat meninjau berita yang didapatkan oleh masyarakat yang membaca berita.

BAB II TINJAUAN PUSTAKA

2. Tinjauan Pustaka

2.1. Landasan Teori

2.1.1. Berita Bohong atau Hoax

2.1.1.1. Definisi Hoax

Menurut KBBI Daring, kata *hoax* merupakan informasi, kabar, atau berita bohong. [1]

Menurut Santoso dkk., *hoax* merupakan informasi palsu yang dianggap benar dan dapat membawa persepsi manusia ke arah yang salah. Penyebaran informasi hoax umumnya memiliki beberapa tujuan, dengan tujuan utama untuk menggiring dan memanipulasi opini masyarakat. [2]

Dalam kamus Cambridge, kata *hoax* berarti tipuan perbuatan, perkataan yang tidak jujur, palsu atau bohong dengan tujuan menyesatkan atau lelucon belaka. [3]

2.1.1.2. Sejarah Hoax

Kata Hoax berasal dari bahasa Latin “*Hoc est corpus*” yang pada zaman dahulu digunakan oleh para penyihir sebagai mantra untuk memberdayakan atau mengecoh orang lain dari ucapan mereka. Penggunaan kata hoax ditemukan pertama kali pada tahun 1808 dalam bahasa Inggris. Kata itu ditemukan dalam sebuah buku karya Linda Walsh yang berjudul *Sins Against Science*. Setelah itu, istilah hoax mulai naik daun sejak perilisan film yang berjudul “*The Hoax*” yang disutradarai oleh Lasse Hallstrom pada tahun 2006. [4]

2.1.1.3. Jenis-Jenis Hoax

Menurut Hidayat, dkk., ada beberapa jenis informasi hoax meliputi: [5]

- a. *Fake news* atau berita bohong, yaitu berita yang berusaha menggantikan berita asli dengan memasukkan unsur ketidakbenaran.
- b. *Clickbait* atau autan jebakan, yaitu autan yang diletakkan di dalam suatu situs dengan tujuan agar orang dapat mengunjungi situs lainnya. Umumnya *clickbait* memiliki judul yang berlebihan dan foto yang menarik.

- c. *Confirmation bias*, yaitu kecenderungan untuk menginterpretasi kejadian yang baru saja terjadi sebagai bukti dari kepercayaan yang sudah ada.
- d. *Misinformation*, yaitu informasi yang salah dan tidak akurat, terutama yang digunakan untuk menipu.
- e. *Satire*, yaitu sebuah tulisan humor atau ironi yang digunakan untuk mengkritisi kejadian yang sedang naik daun.
- f. *Post-truth*, yaitu kejadian yang mengedepankan emosi dibandingkan fakta sebenarnya dalam membuat opini publik.
- g. Propaganda, yaitu aktivitas yang menyebarluaskan informasi, fakta, rumor, atau setengah-kebenaran untuk mempengaruhi opini masyarakat.

2.1.2. Media Sosial

2.1.2.1. Definisi Media Sosial

Media sosial merupakan kumpulan *Internet-based* dan *mobile services* yang mengizinkan penggunaanya untuk berpartisipasi dalam pertukaran online, berkontribusi dalam konten yang dibuat user, atau mengikuti komunitas online. [6]

2.1.2.2. Jenis-Jenis Media Sosial

Jenis Internet service yang dikategorikan sebagai media sosial adalah sebagai berikut. [6]

- a. Blog
Blog merupakan suatu jurnal online yang mana halaman yang ditunjukkan berdasarkan urutan kronologis terbalik (mulai dari yang terakhir dimasukkan).
- b. Wiki
Wiki merupakan suatu website yang mana tiap pengunjunnya diperbolehkan untuk memodifikasi halaman manapun dan membentuk halaman baru menggunakan browsernya.
- c. *Social bookmarking*
Situs social bookmarking memperbolehkan penggunaanya untuk mengumpulkan, merapikan, dan menyebarkan link ke website tertentu.
- d. *Social network sites*
Situs social network merupakan web-based service yang memperbolehkan tiap individu untuk (1) membentuk suatu

profil publik atau semi-publik, (2) memperlihatkan daftar pengguna yang berhubungan dengannya, dan (3) menunjukkan dan mengunjungi daftar koneksi dari teman dalam sistem

e. *Status-update services*

Status-update services merupakan microblogging services yang memperbolehkan penggunanya untuk membagikan update singkat mengenai seseorang atau event yang sedang berlangsung dan melihat update yang dibuat oleh orang lain.

f. *Virtual world content*

Virtual world content merupakan situs yang menyediakan lingkungan virtual menyerupai dunia game dimana para pengguna dapat berinteraksi antara satu sama lain di dalamnya.

g. *Media-sharing sites*

Media-sharing sites merupakan situs yang memperbolehkan penggunanya untuk mengunggah video atau foto.

2.1.3. Natural Language Processing (NLP)

2.1.3.1. Definisi Natural Language Processing

Natural Language Processing (NLP) merupakan cabang dari bidang artificial intelligence yang memberikan komputer kemampuan untuk membaca, memahami, dan menginterpretasi bahasa manusia. [7]

2.1.3.2. Lima Tugas Utama NLP

Tugas NLP secara garis besar adalah mengubah data teks dan suara dari manusia menjadi informasi yang dapat dipahami oleh komputer. Tugas tersebut dijelaskan secara detail sebagai berikut. [7]

- a. *Speech recognition* atau *speech-to-text*, yaitu tahap yang bertugas untuk mengkonversi data suara menjadi teks.
- b. *Part of speech tagging* atau *grammatical tagging*, yaitu proses yang menentukan fungsi dan konteks dari suatu bagian kalimat berdasarkan letaknya.
- c. *Word sense disambiguation*, yaitu proses seleksi definisi suatu kata yang memiliki banyak arti dengan menggunakan

analisis semantic untuk menentukan arti yang paling sesuai dengan konteks kalimat.

- d. *Named entity recognition (NEM)*, yaitu proses mengidentifikasi kata atau frasa menjadi suatu entitas. Contoh : kata 'Jakarta' dikategorikan sebagai Lokasi dan kata 'Andi' dikategorikan sebagai Nama.
- e. *Co-reference resolution*, yaitu proses yang mengidentifikasi apakah dua kata merujuk pada entitas yang sama. Contoh : pengecekan kata 'dia' dan 'Susi' menunjuk ke objek yang sama atau tidak.
- f. *Sentiment analysis*, yaitu proses yang bertugas untuk menentukan kualitas subjektif dari teks, seperti emosi, tingkah lakunya, kecurigaan pengirim, dan lain-lain.
- g. *Natural language generation*, yaitu proses menyusun informasi dari komputer menjadi bahasa manusia.

BAB III METODE PENELITIAN

3. Metode Penelitian

3.1. 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.

3.2. Teknik Pengumpulan Data

Pada penelitian kali ini, kami memutuskan untuk melakukan metode penelitian dengan cara kuantitatif setelah itu merangkum hasil data yang didapat ke dalam bentuk kuantitatif yang diolah menjadi sebuah tabel ataupun grafik agar pembaca mampu memahami penelitian kami lebih jelas dan mengerti maksudnya lebih dalam. Data - data yang kami kumpulkan memiliki beberapa kriteria di dalamnya yaitu :

1. Judul berita dan konten berita yang diambil sebagai data. Di dapat melalui website portal berita yang resmi ataupun tidak resmi.
2. Berita yang diambil menggunakan bahasa Indonesia
3. Berita harus memiliki tanggal penulisan, penulis berita dan narasumber berita

Adapun kami juga menambahkan beberapa kriteria yang tidak kami masukkan dalam proses pengumpulan berita yaitu:

1. Tidak menggunakan bahasa Indonesia
2. Menghindari beberapa website yang memang sudah terindikasi pembuat berita palsu

Dalam kriteria ini akan ada kemungkinan sebuah perubahan tambahan ataupun pengurangan di masa yang mendatang sesuai dengan bagaimana nantinya format sebuah berita yang disajikan ke masyarakat umum. Dalam tekniknya juga akan sewaktu - waktu berubah, teknik ini kemungkinan juga akan berbeda sesuai dengan format berita dari setiap negara ataupun waktu pembuatan berita tahun ke tahun.

3.2.1 Sumber Data

Data yang digunakan dalam penelitian ini diambil dari Jakarta open data yang sudah mengumpulkan berita dengan melabeli berita yang terindikasi hoax, non-hoax ataupun netral. Dengan berjumlah 308 artikel selama 3 bulan dari

Oktober hingga Desember pada tahun 2020. Informasi yang didapatkan adalah tanggal, isu, *tonality*, judul berita, dan *link* dari berita seperti pada Tabel 1 yang menunjukkan sample data dari analisa dan Monitoring Pemberitaan di Media Massa Bulan November Tahun 2020.

tanggal	isu	tonality	berita_jakarta	link
2020-11-02	DKI Jakarta Raih Penghargaan Sustainable Transport Award (STA) 2021	POSITIF	Jakarta Raih STA 2021 Transjakarta Terus Kembangkan Sistem Integrasi	https://www.beritajakarta.id/read/84340/jakarta-raih-sta-2021-transjakarta-terus-kembangkan-sistem-integrasi#.X57XtlMzbdc
2020-11-03	Halte Transjakarta Rusak akan Didesain Ulang Kekinian	POSITIF	Transjakarta Siapkan Halte dengan Konsep Modern dan Kekinian	https://www.beritajakarta.id/read/84352/transjakarta-siapkan-halte-dengan-konsep-modern-dan-kekinian#.X6N34JmzbUo
2020-11-13	Strategi Pemprov DKI Jakarta Mengantisipasi Banjir	POSITIF	150 Petugas Gabungan Kerja Bakti di Pancoran	https://www.beritajakarta.id/read/84743/150-petugas-gabungan-kerja-bakti-di-pancoran#.X7IbO9Mzbdc
2020-11-06	Target Gubernur Anies Banjir Surut dalam Enam Jam	NETRAL	Sudinsos Jaktim Gelar Simulasi Evakuasi Korban Banjir	https://www.beritajakarta.id/read/84577/sudinsos-jaktim-gelar-simulasi-evakuasi-korban-banjir#.X6oAfZMzbUo
2020-11-10	PSBB Transisi Kembali Diperpanjang	NETRAL	Tingkatkan Perilaku 3M Pemprov DKI Perpanjang PSBB Masa Transisi Hingga 22 November	https://www.beritajakarta.id/read/84540/tingkatkan-perilaku-3m-pemprov-dki-perpanjang-psbb-masa-transisi-hingga-22-november#.X6oA6JMzbUo

2020-11-13	Peningkatan Kasus DBD	NETRAL	Gubernur Resmikan Tiga Aplikasi Inovasi Layanan Kesehatan	https://www.beritajakarta.id/read/65834/gubernur-resmikan-tiga-aplikasi-inovasi-layanan-kesehatan#.XFFe4VUzaHs
2020-10-17	Sorotan Pelaksanaan Perda Penanggulangan COVID	NETRAL		
2020-11-04	Keputusan Menaikan UMP 2021 Dikritik	NEGATIF	80 Pelaku Usaha Perhotelan Usulkan Penyesuaian UMP untuk 2021	https://www.beritajakarta.id/read/84459/80-pelaku-usaha-perhotelan-usulkan-penyesuaian-ump-untuk-2021#.X6N3fZMzbUo
2020-11-12	Commitment Fee Formula E	NEGATIF		
2020-11-16	Kerumunan Massa Habib Rizieq	NEGATIF		

Tabel 1. November tahun 2020

3.3. Teknik Analisis Data

Dengan penelitian ini, beberapa tahapan dalam teknik analisis data dibawah ini:

- Pengumpulan berita dari website Jakarta Open Data serta melalui situs berita online yaitu beritajakarta.id.
- Berita atau dataset yang telah dikumpulkan kemudian melakukan uji coba terhadap algoritma pendeteksi berita hoax dari peneliti lain. Untuk metode kami sendiri menganalisisnya dari korelasi antara judul dengan berita yang disampaikan
- 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.

- d. Masuk ke dalam tahap *screening* yang bertujuan untuk menghapuskan karakter ataupun variabel diluar dari huruf ataupun angka. Misalnya simbol teks.
- e. 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.
- f. Ketika hasil dari algoritma tersebut sudah berhasil dilakukan, maka tahap berikut yaitu mengklasifikasi *tonality* dari data tersebut menjadi tiga yaitu positif, negatif dan netral. Netral disini merupakan berita yang terindikasi berpotensi menjadi berita palsu maupun asli sesuai dengan beberapa faktor luar yang kami perkirakan
- g. 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.

3.4. Analisis Sistem

Mendefinisikan 2 variabel untuk membaca dua dataset.

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

Fig. 1 Dua variable

Setiap dataset harus diberi satu kolom bernama label / *news category* berupa 1 dan 0. Nilai 1 untuk dataset berita aktual / *realnews* dan 0 berita palsu / *fakenews*.

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

Fig 2 Kategori Dataset

Melakukan pengecekan untuk memastikan bahwa data memiliki keseimbangan dalam jumlah datanya. Berdasarkan hasil pengecekan (Fig. 4) dapat dikatakan bahwa dataset sudah cukup seimbang karena sedikitnya perbedaan jumlah dataset.

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

Fig. 3 Cek keseimbangan dataset

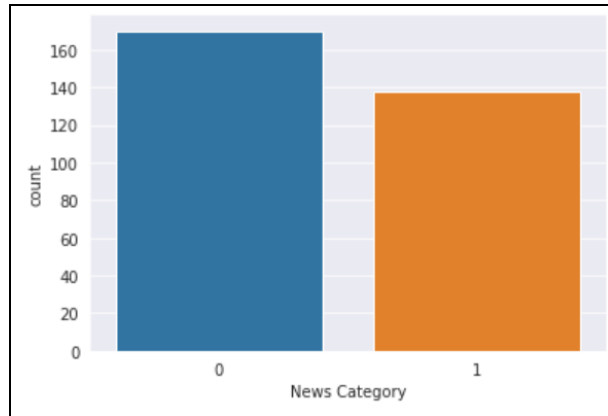


Fig. 4 Hasil pengecek keseimbangan dataset

Melakukan pengecekan untuk data kosong pada dataset. Pengecekan dilakukan untuk setiap kolom pada dataset. Hasil pengecekan (Fig. 6) menunjukkan bahwa setiap kolom dalam dataset tidak memiliki data yang kosong.

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

Fig. 5 Pengecekan data kosong

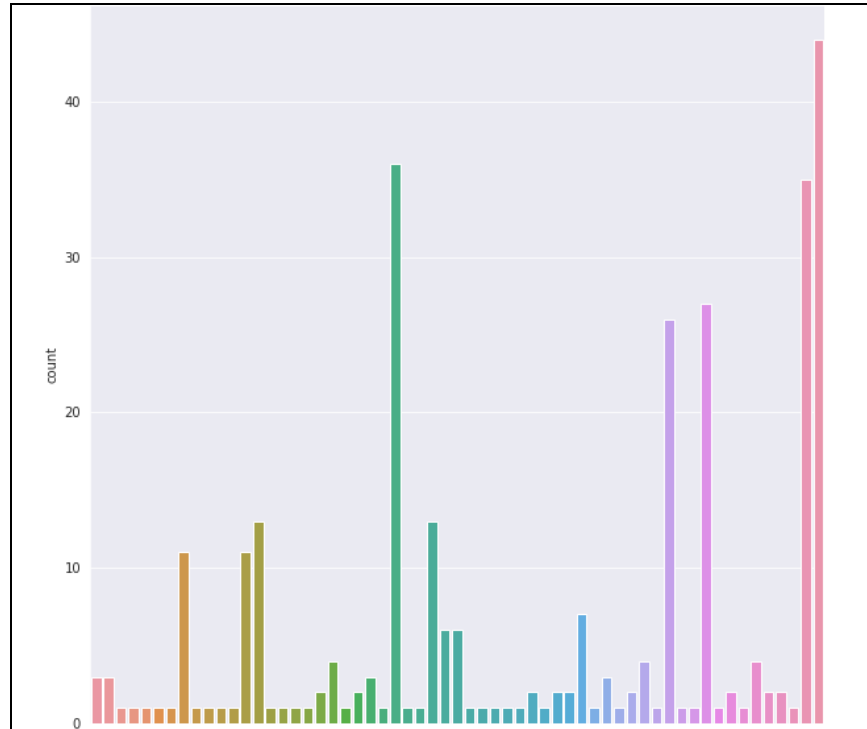
tanggal	0
isu	0
tonality	0
berita_jakarta	0
link/page	0
News Category	0

Fig. 6 Hasil pengecekan data kosong

Dari data yang diatas dapat dilihat pengecekan data yang bernilai NULL pada tabel. Pada semua kolom dataset memiliki 6 kolom yang terdiri tanggal, isu, tonality, berita_jakarta, link/page, dan News Category yang digunakan untuk pengecekan data nantinya.

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

Fig. 7 Detail jumlah per isu



Dapat dilihat bahwa pada hasil dataset yang dimiliki pada kolom isu “Berbagai Upaya DKI Jakarta Hadapi La Nina” paling banyak melainkan isu lainnya.

```
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)
```

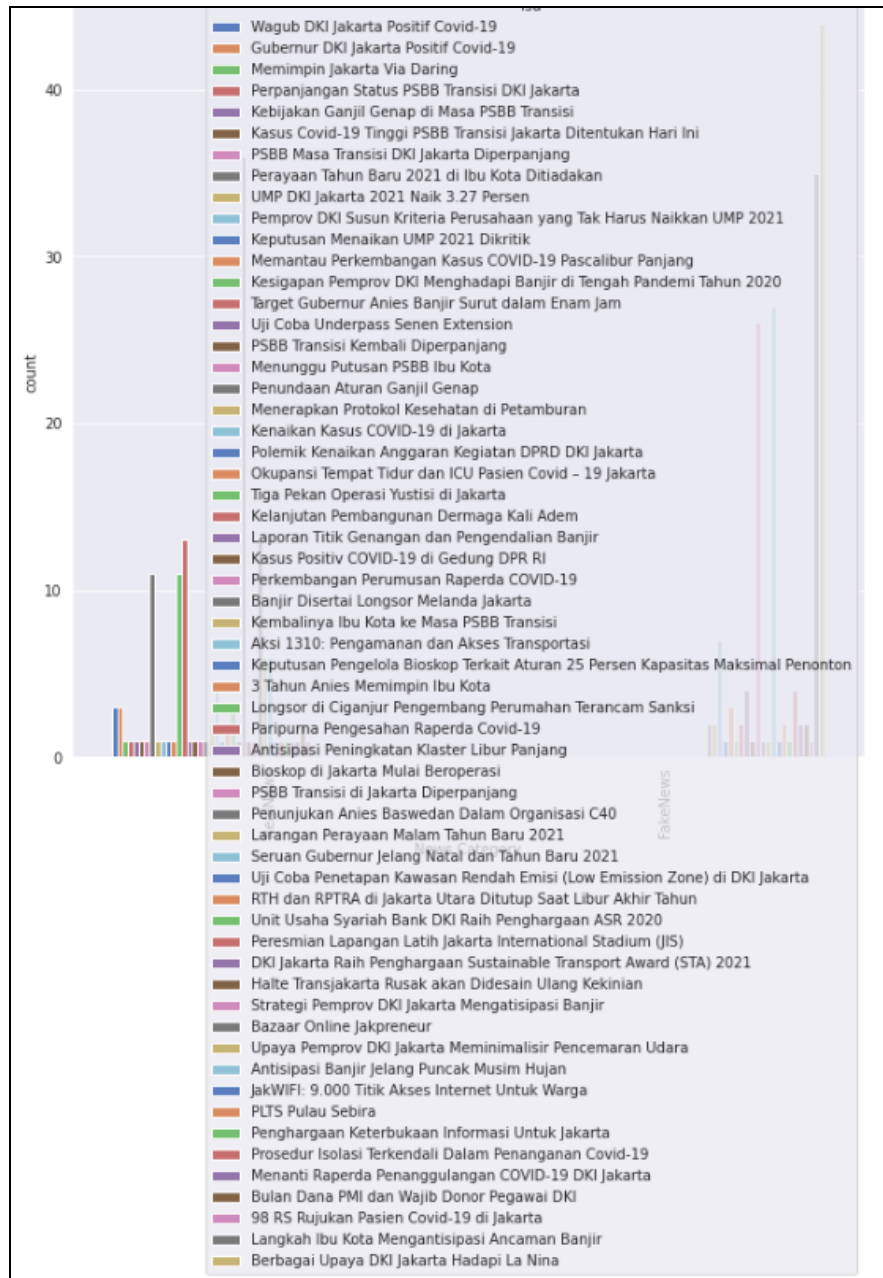


Fig. 10 Hasil detail jumlah per isu setiap dataset

Dari Fig. 10 dapat dilihat bahwa isu pada tiap dataset dan jumlah pada tiap datasetnya. Pada grafik diatas juga terbagi menjadi 2 bagian yaitu RealNews yang disebelah kiri dan FakeNews sebelah kanan. Tahap selanjutnya yaitu pembuatan *wordcloud*, maka Kolom isu dan berita_jakarta digabungkan menjadi satu dan kolom lainnya yang tidak dipakai dalam pembuatan *wordcloud* dihilangkan (Fig. 11).

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

Fig. 11 Penggabungan dan penghapusan kolom

```
from wordcloud import WordCloud

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

Fig. 12 Import library wordcloud

Pembuatan cloud kata (Fig. 13 dan Fig. 14) untuk melakukan analisis kata per kata yang paling sering muncul dalam data. Maka dibuatnya stopwords yang akan dimasukkan kedalam suatu bahasa. Cara kerja mesin ini untuk mencari yang mengabaikan kata yang berhenti saat melakukan indeksing data serta mengambil hasil data yang ingin dicari.

```
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()
```

Fig. 13 Wordcloud dari berita palsu



Fig. 14 Hasil wordcloud berita palsu


```
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()
```

Fig. 15 Word Cloud dari berita aktual



Fig. 16 Hasil wordcloud berita aktual

Berdasarkan dataset yang kita yang kita miliki, kami perlu memasukkan beberapa algoritma processing. NLP, yang menjadi kunci utama dalam melakukan pendeteksian sebuah berita asli maupun palsu. Namun algoritma NLP sendiri tidak bisa menyelesaikan masalahnya, kita memasukkan lagi algoritma pendukung untuk mendapatkan hasil yang sesuai. Model algoritma tersebut terdiri dari :

a. Multinomial Naive bayes

- Naive bayes ini sangat sering digunakan dalam NLP. Naive bayes memiliki relasi yang sama dengan Bayes teorema.

$$P(c|x) = P(x|c) * P(c) / P(x)$$

Dengan menggunakan naive bayes, kita berasumsi bahwa semua fitur yang tersedia bersifat independen terhadap satu sama lain. Berdasarkan data yang kami punya kami mengklasifikasi fake news dan real news.

Diberikan berita x , kami akan menghitung $P(\text{berita benar}|x)$ serta $P(\text{berita palsu}|x)$. Jika $P(\text{berita benar}|x) > P(\text{berita palsu}|x)$, algoritma memprediksi itu adalah berita benar. Jika tidak, berita tersebut akan diprediksi sebagai palsu

```
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)
```

Fig. 17 Pemisahan data

```
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
```

Fig. 18 Import library Multinomial Naive Bayes

```
pipe = Pipeline([
    ('vect', CountVectorizer()),
    ('tfidf', TfidfTransformer()),
    ('clf', MultinomialNB())
])

model = pipe.fit(x_train, y_train)
prediction = model.predict(x_test)

score = metrics.accuracy_score(y_test, prediction)
print("accuracy:  %0.3f" % (score*100))

fig, ax = plot_confusion_matrix(conf_mat=confusion_matrix(y_test, prediction),
                                show_absolute=True,
                                show_normed=True,
                                colorbar=True)

plt.show()
```

Fig. 19 Multinomial Naive Bayes

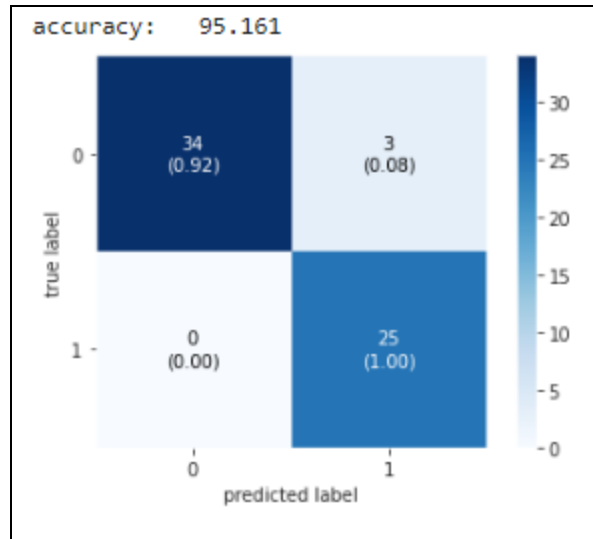


Fig. 20 Hasil Multinomial Naive Bayes

b. Count Vectorizer

- Count vectorizer ini umumnya digunakan untuk memecah sebuah kumpulan dokumen atau kalimat dan membangun sekumpulan kosakata dari character kata yang unik. Dokumen baru juga bisa dibuat berdasarkan kosakata tersebut.

BAB IV. KESIMPULAN DAN SARAN

4.1. 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.

Melalui penelitian ini kami juga merasa terbantu oleh algoritma yang ada dengan tingkat akurasi tertinggi 95% dalam pengecekan keaslian sebuah berita. Dan kami merekomendasikan algoritma NLP dan multinomial naive bayes untuk kasus ini karena terbilang cocok dan efisien. Algoritma ini mampu handle ribuan data berita yang tersebar dan mengkategorisasi tiap berita dengan cepat dan akurat. Namun kami melihat bahwa algoritma ini memiliki sedikit kekurangan, yaitu kecepatan prosesing yang terbilang sedikit lambat apabila data yang dimasukkan sangat besar seperti contohnya 500.000 data. Tapi NLP ini bisa diakali dengan beberapa cara supaya bisa lebih cepat seperti contohnya menggunakan loop Cython yang memanfaatkan objek Cython C. Objek Cython C adalah objek pada bahasa pemrograman C atau C++ yaitu *int*, *float*, *double*, dan *struct* yang dapat dikompilasi dalam bentuk kode cepat namun tingkat rendah.

4.2. Saran

4.2.1 Saran untuk pembaca

Kami menyarankan pembaca untuk melakukan validasi ulang setiap berita yang dibaca apakah berita tersebut bersifat aktual maupun palsu. Bisa dengan mengkonfirmasi sumber multimedia yang tercantum di berita tersebut seperti gambar, audio, maupun video. Pembaca juga diharapkan tidak mudah untuk terprovokasi oleh berita yang mengandung propaganda ataupun konspirasi yang memicu kemarahan sosial.

4.2.2 Saran untuk penelitian berikutnya

Kedepannya, dataset yang digunakan dapat memiliki jumlah data yang luas terutama tidak terbatas pada satu portal berita saja. Semakin luas data yang dimiliki, maka akan semakin banyak analisis data, akan semakin akurat hasil kata kunci untuk membantu masyarakat memilah berita. Selain itu, tentunya algoritma yang kami terapkan tidaklah sempurna dan dapat menerapkan algoritma lain yang dapat bekerja lebih efisien dan cepat.

REFERENSI

- [1] Kementerian Pendidikan Kebudayaan, Riset, dan Teknologi. (n.d.). Hoax. Pada KBBI. Diakses pada 10 Juni 2022, dari <https://kbbi.kemdikbud.go.id/entri/hoaks>
- [2] Santoso, D. N., Rachmawanto, E. H., Nugraha, A., Nugroho, A. A., Setiadi, D. R. I. M. & Basuki, R. S. (2020). Hoax classification and sentiment analysis of Indonesian news using Naive Bayes optimization. *Telecommunication Computing Electronics and Control*, 18(2), 799-806. 10.12928/TELKOMNIKA.v18i2.14744
- [3] Cambridge University. (n.d.). Hoax. Pada Cambridge Dictionary. Diakses pada 10 Juni 2022, dari <https://dictionary.cambridge.org/dictionary/english/hoax>.
- [4] Assidik, G. K. (2018). Kajian Identifikasi dan Upaya Penangkalan Pemberitaan Palsu (Hoax) pada Pembelajaran Bahasa Indonesia. Kementerian Pendidikan Kebudayaan, Riset, dan Teknologi. <http://repositori.kemdikbud.go.id/10238/1/KAJIAN%20IDENTIFIKASI%20DAN%20UPAYA%20PENANGKALAN.pdf>
- [5] Hidayat, N., Qalby, N., Alaydrus, S. S., Darmayanti, A., & Salsabila, A. P. (2019). Pengaruh Media Sosial Terhadap Penyebaran Hoax oleh Digital Native. *Research Gate*. https://www.researchgate.net/publication/330135150_PENGARUH_MEDIA_SOSIAL_TERHADAP_PENYEBARAN_HOAX_OLEH_DIGITAL_NATIVE
- [6] Dewing, M. (2012). Social Media: An Introduction. *Library of Parliament*. <https://bdp.parl.ca/staticfiles/PublicWebsite/Home/ResearchPublications/InBriefs/PDF/2010-03-e.pdf>
- [7] IBM Cloud Education. (2020). *Natural Language Processing (NLP)*. Diakses 10 Juni 2022, dari [https://www.ibm.com/cloud/learn/natural-language-processing#:~:text=Natural%20language%20processing%20\(NLP\)%20refers,same%20way%20human%20beings%20can](https://www.ibm.com/cloud/learn/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can).
- Dina, S. (2017). *Cara Cerdas Mencegah Penyebaran Hoaks di Medsos*. Diakses 7 Juni 2022, dari https://kominfo.go.id/content/detail/11347/cara-cerdas-mencegah-penyebaran-hoaks-di-medsos/0/sorotan_media
- George, J, A. (2020). *Fake News Detection using NLP Techniques*. Diakses 15 Juni 2022, dari <https://medium.com/analytics-vidhya/fake-news-detection-using-nlp-techniques-c2dc4be05f99>

Umam. *Pengertian Media Sosial, Sejarah, Fungsi, Jenis, Manfaat, dan Perkembangannya*. Diakses 7 Juni 2022, dari

https://www.gramedia.com/literasi/pengertian-media-sosial/#A_Pengertian_Media_Sosial

Vidi A. (2020). *5 Alasan Orang Sebarkan Hoaks, Nomor 2 Paling Sering Dilakukan*. Diakses 7 Juni 2022, dari

<https://www.liputan6.com/cek-fakta/read/4342173/5-alasan-orang-sebarkan-hoaks-nomor-2-paling-sering-dilakukan>

Yuanita. (2017). *Ini Cara Mengatasi Berita “Hoax” di Dunia Maya*. Diakses 15 Juni 2022, dari https://kominfo.go.id/content/detail/8949/ini-cara-mengatasi-berita-hoax-di-dunia-maya/0/sorotan_media

Yunita. (2017). *Penyebaran Informasi Hoax Menimbulkan Keresahan di Masyarakat*. Diakses 7 Juni 2022, dari

https://kominfo.go.id/content/detail/9058/penyebaran-informasi-hoax-menimbulkan-keresahan-di-masyarakat/0/sorotan_media

Wolf, T. (2018). *100 Times Faster Natural Language Processing in Python*. Diakses 15 Juni 2022, dari

<https://medium.com/huggingface/100-times-faster-natural-language-processing-in-python-ee32033bdced>