

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Studi Kasus

Klasifikasi Berita Hoax/Fake News Menggunakan Metode Naive Bayes Classification

Dwi Welly Sukma Nirad ^a, Aina Hubby Aziira ^b, Febby Apri Wenando ^c, Lutfia Aulia ^d, Ariq Abdurrahman Hakim ^e, Husna Afiqah Yossyafra ^f, Khairin Nisa ^g

a,b,c,d,e,f,g Departemen Sistem Informasi Universitas Andalas, Limau Manih, Kota Padang 25163, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 00 Februari 0000

Revisi Akhir: 00 Maret 0000

Diterbitkan Online: 00 April 0000

KATA KUNCI

Berita Hoax/Fake,

Naive Bayes Classifier,

Data Mining,

Python

KORESPONDENSI

E-mail: *

ABSTRACT

Peningkatan akses mudah terhadap informasi melalui internet dan media sosial telah merubah gaya perolehan informasi oleh masyarakat. Namun, dampak negatif dari kebebasan ini adalah munculnya berita hoax atau fake news yang sulit dibedakan dengan berita yang benar. Berita hoaks memiliki potensi merusak kepercayaan publik dan mempengaruhi keputusan yang diambil oleh individu atau masyarakat. Dalam menghadapi masalah ini, Algoritma Naive Bayes Classifier menjadi salah satu metode yang digunakan untuk membantu mendeteksi kredibilitas berita hoaks dengan mengklasifikasikan teks berdasarkan probabilitas maksimumnya. Oleh karena itu, penelitian ini bertujuan untuk mengaplikasikan metode tersebut dalam mengidentifikasi berita hoax dan meminimalkan dampak negatif yang ditimbulkan.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi mengalami peningkatan yang drastis. Banyak orang sekarang dapat dengan mudah mengakses berbagai informasi melalui media sosial seperti Facebook, Twitter, Instagram, dan juga melalui pesan elektronik singkat seperti WhatsApp dan Telegram. Akses mudah terhadap informasi melalui internet dan media sosial telah mengubah cara masyarakat memperoleh berita. Namun, keleluasaan ini juga membawa konsekuensi negatif, diantaranya adalah munculnya berita hoax atau fake news sehingga sulit masyarakat dalam membedakan antara berita baik dan buruk

yang mereka konsumsi. Berita bohong atau hoaks telah menjadi permasalahan yang besar di Indonesia.

Penyebaran berita hoaks menyebabkan kerugian bagi salah satu pihak dan dapat merusak reputasi seseorang. Berita hoaks memiliki potensi merusak kepercayaan publik, menyebabkan kebingungan, dan mempengaruhi keputusan yang diambil oleh individu atau masyarakat. Pemerintah, melalui Kepolisian Republik Indonesia, telah mengeluarkan ancaman untuk menghukum pihak-pihak yang menyebarkan berita palsu. Namun, tindakan ini kadang dianggap sebagai pembatasan kebebasan berpendapat. Berita hoax yang menyebar dari orang ke orang atau melalui media pers telah menjadi hal yang umum

di media sosial dan platform elektronik lainnya, dan hal ini dapat mempengaruhi masyarakat dengan menimbulkan kegelisahan, ketidaknyamanan, dan rasa amarah yang berlebihan jika berita yang mereka baca mengandung provokasi dan informasi yang negatif [6]. Oleh karena itu, perlu adanya upaya untuk mengklasifikasikan berita hoaks secara efektif guna mengurangi dampak negatif yang ditimbulkan.

Dalam menghadapi masalah ini, penting untuk memudahkan pendeteksian kredibilitas berita hoaks. Salah satu metode yang digunakan adalah Algoritma Naive Bayes Classifier, yang merupakan metode klasifikasi sederhana yang menerapkan teorema Bayes dengan menganggap bahwa semua fitur tidak saling berhubungan. Algoritma ini menggunakan probabilitas keseluruhan, yaitu probabilitas dokumen terhadap kategori (prior). Dengan menggunakan algoritma ini, teks dapat diklasifikasikan berdasarkan probabilitas maksimumnya (posterior).

2. TINJAUAN PUSTAKA

2.1. Berita

Berita adalah informasi baru atau informasi mengenai sesuatu yang sedang terjadi, disajikan dalam bentuk cetak, siaran, Internet, atau dari mulut ke mulut kepada orang ketiga atau orang banyak [11]. Berita harus berisi informasi-informasi yang autentik dan mengungkapkan fakta dari sebuah peristiwa yang terjadi. Pendapat dapat ditambahkan ke dalam sebuah berita, tetapi hanya jika itu adalah pendapat penulis berita atau pendapat orang lain tentang peristiwa tersebut dan tidak mengubah isi dari berita tersebut [4].

2.2. Hoax

Hoax merupakan informasi, kabar, berita yang palsu atau bohong. Pada Kamus Besar Bahasa Indonesia (KBBI) hoax diartikan juga sebagai berita yang bohong. Secara umum hoax dapat diartikan sebagai informasi yang dibuat-buat atau direkayasa untuk menutupi informasi yang sebenarnya [2]. Hoax dalam sebuah berita merupakan hal yang umum kita temukan pada saat sekarang ini. Informasi yang kita terima dari berbagai media bisa saja merupakan sebuah informasi yang direkayasa ataupun tidak. Penyebaran berita hoax ini sangat mengkhawatirkan karena jika ada masyarakat percaya dengan berita tersebut hal ini dapat menjadi sumber perpecahan antar masyarakat. Sehingga dalam memperoleh informasi tersebut kita butuh selektif dan mencari tahu terlebih dahulu terhadap berita yang ada, apakah berita tersebut dapat dipercaya atau tidak.

2.3. Data Mining

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi Artificial Intelligence (AI). Data mining sering juga disebut sebagai Knowledge Discovery in Database (KDD) [10].

Pada penelitian ini kami menggunakan metode data mining classification untuk memproses dataset sekumpulan berita. Metode ini cocok untuk mengelompokkan sebuah dataset berita yang akan digolongkan apakah berita tersebut hoax atau tidak. Melalui metode ini data-data berita yang ingin kita uji dan masih belum memiliki label apakah berita tersebut hoax atau tidak menjadi lebih mudah untuk diberi label atau dikelompokkan.

2.4. Algoritma Naive Bayes

Naive bayes merupakan metode pengklasifikasian berdasarkan probabilitas sederhana dan dirancang agar dapat dipergunakan dengan asumsi antar variabel penjelas saling bebas (independen). Pada algoritma ini pembelajaran lebih ditekankan pada pengestimasi probabilitas. Keuntungan algoritma naive bayes adalah tingkat nilai error yang didapat lebih rendah ketika dataset berjumlah besar, selain itu akurasi naive bayes dan kecepatannya lebih tinggi pada saat diaplikasikan ke dalam dataset yang jumlahnya lebih besar [13].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} = P(X|H) \times P(H) / P(X)$$

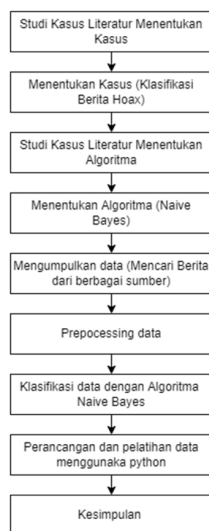
Keterangan:

- X adalah data sampel dengan kelas (label) yang tidak diketahui.
- H adalah hipotesa bahwa X adalah data dengan kelas (label) C. P(H) adalah peluang dari hipotesa H.
- P(X) adalah peluang data sampel yang akan diamati.
- P(X|H) merupakan peluang data sampel X, bila diasumsikan bahwa hipotesa benar (valid).
- Untuk masalah klasifikasi, yang dihitung adalah P(H|X), yaitu peluang bahwa hipotesa benar (valid) untuk data sampel X [14].

3. METODOLOGI

3.1. Tahap Penelitian

Metodologi penelitian dalam tugas besar ini digunakan sebagai acuan dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan. Berikut alur penelitian yang dilakukan.



Penelitian ini dimulai dengan melakukan studi literatur terhadap beberapa artikel yang membahas tentang data mining. Studi awal ini bertujuan untuk menentukan kasus yang akan diangkat berdasarkan beberapa kasus yang terdapat dalam artikel yang telah dibaca. Setelah melakukan studi tersebut, dipilih sebuah kasus klasifikasi berita hoax untuk diangkat dalam penelitian ini. Selanjutnya, dilakukan studi literatur lebih lanjut guna menentukan algoritma yang cocok untuk masalah yang diangkat. Dalam sepuluh artikel yang dipelajari, yang mengusung kasus yang sama namun menggunakan algoritma yang berbeda-beda, diputuskan untuk menggunakan algoritma Naïve Bayes pada penelitian kali ini. Setelah menentukan kasus yang akan diteliti dan algoritma yang akan digunakan, langkah selanjutnya adalah mengumpulkan data dari berbagai situs berita yang ada. Data yang telah dikumpulkan kemudian dilakukan preprocessing, dan dilanjutkan dengan melakukan pemodelan manual menggunakan algoritma Naïve Bayes. Setelah pemodelan manual selesai, langkah berikutnya adalah melakukan pemodelan sistem menggunakan Python.

3.2. Studi Kasus

Pada tahap ini, dilakukan studi kasus melalui studi literatur untuk mengumpulkan informasi terkait kasus-kasus yang telah diteliti sebelumnya dan mendapatkan ide kasus dari studi tersebut. Studi literatur dilakukan dengan menggunakan sumber yang relevan, seperti jurnal dan skripsi-skripsi yang berkaitan dengan data mining. Melalui studi literatur ini, ditemukan sebuah kasus yang akan diteliti, yaitu Klasifikasi Berita Hoax.

Selanjutnya, dilakukan studi literatur kembali untuk menentukan algoritma yang sesuai dengan kasus yang akan diangkat. Studi literatur yang digunakan adalah sumber yang relevan sehingga informasi yang diperoleh dapat dipercaya. Studi literatur ini menghasilkan keputusan untuk menggunakan algoritma Naïve Bayes setelah dilakukan analisis terhadap sumber-sumber yang ada.

Setelah itu dilakukan analisis terhadap kasus yang diangkat dan menemukan berbagai masalah dari penelitian, membuat rumusan masalah, mencari teori-teori pendukung terkait penelitian, dan menentukan tujuan dari penelitian agar lebih terarah dan jelas dalam pelaksanaannya.

3.3. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data untuk mendapatkan informasi dan data-data yang dibutuhkan dalam melakukan penelitian. Data yang digunakan pada penelitian ini adalah sekumpulan berita yang diambil dari beberapa situs berita yang ada. Data-data yang didapat kemudian diolah berdasarkan kebutuhan dalam penelitian agar nantinya dapat dijadikan inputan dalam pelatihan sistem yang akan dibangun untuk menentukan klasifikasi dari sebuah berita.

3.4. Pengumpulan Data

Klasifikasi berita yang dilakukan menggunakan Naïve Bayes dengan mencari nilai dari masing-masing kriteria dan dihasilkan status dari kriteria tersebut hoax atau tidak. Hal tersebut dilakukan terhadap setiap kolom yang ada dengan masing-masing kriteria didalamnya. Nilai yang diperoleh tersebut menunjukkan perbandingan status dari kriteria yang ada sehingga dapat dilihat berdasarkan kriteria tersebut berita yang akan diuji dapat tergolong hoax atau tidak.

3.5. Pelatihan Data

Pada tahap ini dilakukan pelatihan terhadap data yang sudah dikumpulkan. Pelatihan data dilakukan dengan membagi dua data tersebut menjadi dataset pelatihan dan dataset pengujian. Dataset pelatihan merupakan bagian dataset yang dilatih untuk membuat prediksi atau menjalankan fungsi dari sebuah algoritma machine learning. Dataset pelatihan inilah yang nantinya akan menjadi data yang dilatih untuk memprediksi sebuah data. Selanjutnya ada dataset pengujian yang merupakan bagian dataset yang kita tes untuk melihat keakuratannya dari pelatihan yang dilakukan, atau dengan kata lain melihat performanya. Pada penelitian kali ini kami mengambil 20% dari dataset yang ada sebagai data test atau pengujian.

3.6. Hardware dan software

1. Processor : 11th Gen Intel(R) Core(TM) i5-11300H @ 3.10GHz 3.11 GHz
2. Sistem Operasi : Windows 11
3. Memory : 16 GB
4. Storage : 256 GB
5. Software : Google Colab

4. HASIL

4.1. Pengumpulan

Data yang dikumpulkan merupakan kumpulan berita mengenai Ganjar Pranowo sebagai Calon Presiden 2024 Indonesia. Data tersebut didapat dari Media Sosial salah satunya adalah Facebook. Data yang telah dikumpulkan akan diklasifikasikan. Untuk setiap berita terdapat Judul Berita, Alamat Situs Berita, Foto Berita, dan Sumber Berita.

Judul Berita	Alamat Situs Berita	Cek Foto Berita	Sumber Berita	Jenis
1. PAN Perimbangan Dukung Garis Persepsi Sebagai Capres 2024	Provokatif - terverifikasi (nonmedia tempel)	Asli	Tidak Publik (Resmi)	fakta
2. PPP Belud PKN Secara Implisit Dukung Garis Persepsi Sebagai Capres 2024	Provokatif - terverifikasi (nonmedia tempel)	Asli	Tidak Publik (Resmi)	fakta
3. Kalah Tolak Dari Garis Persepsi Prabowo Subianto Memutuskan Mundur Sebagai Capres	Provokatif - terverifikasi (media tempel)	Palsu	Parang (Tidak Resmi)	hoax
4. Garis Persepsi Resmi Ketul dan PDIP	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
5. Garis Stok Sahat Publik Hanya Hari Libur, Saya Masih Gubernur	Provokatif - terverifikasi (nonmedia tempel)	Asli	Reporter (Resmi)	fakta
6. Garis Persepsi akan Isut Aksi May Day, KSPSI: Calon Presiden Lain Tidak Isut	Provokatif - terverifikasi (nonmedia tempel)	Asli	Parang (Resmi)	fakta
7. KPK TEMUKAN DANA 300 TRILIUN UNTUK BIAYA KAMPANYE GANJAR	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
8. LUKOWI PERINTAKKAN KAPOLRI UNTUK PENJAJARAN GANJAR	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
9. Video Garis dan Anak menjadi Tersangka Pencucian Uang 300 Triliun untuk Modal Kampanye Capres	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
10. Garis Capres, PDIP Surabaya Aktifkan Lagi Pasok-pesok di Perkampungan	Provokatif - terverifikasi (nonmedia tempel)	Asli	Reporter (Resmi)	fakta
11. Manas Buruh Tertakut Saat Ditensi Garis, Presiden	Provokatif - terverifikasi (nonmedia tempel)	Asli	Salah Publik (Resmi)	fakta
12. Bendera Delapan Anies-Garis untuk Pilpres 2024	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
13. Baris SMC: Edebiatua PDIP Nuk Ura Usasab Garis Persepsi Capres	Provokatif - terverifikasi (nonmedia tempel)	Asli	Youtube (Resmi)	fakta
14. Satu Pulu Petik Bersewa Masore Petik Setelah PDIP Resmi Hadirkan Garis Persepsi Capres 2024	Provokatif - terverifikasi (nonmedia tempel)	Asli	Salah Publik (Resmi)	fakta
15. GANJAR GANDENG MAJUD MD MAU PILPRES 2024	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax
16. Identi Dohar Asin Bersewa Menopra, Renti KIB Dan KIR Deklarasikan Prabowo Garisgar	Provokatif - terverifikasi (nonmedia tempel)	Palsu	Parang (Tidak Resmi)	hoax

4.2. Klasifikasi Manual

Pada tahap ini akan dilakukan pengujian secara manual menggunakan Naive Bayes Classifier. Algoritma menggunakan teorema bayes dan mengasumsikan secara atribut independen pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberi nilai input. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu.

4.2.1. Data Training

Tabel 4. Data Training

Judul Berita	Alamat Situs	Cek Foto	Sumber Berita	Jenis Berita
Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Terverifikasi	Palsu	Tidak Resmi	Hoax
Provokatif	Terverifikasi	Palsu	Tidak Resmi	Hoax
Non Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Non Provokatif	Terverifikasi	Asli	Resmi	Fakta
Non Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Non Provokatif	Terverifikasi	Asli	Resmi	Fakta
Non Provokatif	Terverifikasi	Asli	Resmi	Fakta
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	Hoax

4.2.2. Kriteria dan Probabilita

$$P(C_i)$$

$$P(\text{Jenis Berita} = \text{"Hoax"}) = \frac{8}{16} = 0,5$$

$$P(\text{Jenis Berita} = \text{"Fakta"}) = \frac{8}{16} = 0,5$$

a. Probabilitas Kriteria Judul Berita

$$P(\text{Provokatif} | \text{Hoax}) = \frac{8}{8} = 1$$

$$P(\text{Non Provokatif} | \text{Hoax}) = \frac{0}{8} = 0$$

$$P(\text{Provokatif} | \text{Fakta}) = \frac{3}{8} = 0,375$$

$$P(\text{Non Provokatif} | \text{Fakta}) = \frac{5}{8} = 0,675$$

b. Probabilitas Kriteria Alamat Situs Berita

$$P(\text{Terverifikasi} | \text{Hoax}) = \frac{2}{8} = 0,25$$

$$P(\text{Tidak Terverifikasi} | \text{Hoax}) = \frac{6}{8} = 0,75$$

$$P(\text{Terverifikasi} | \text{Fakta}) = \frac{8}{8} = 1$$

$$P(\text{Tidak Terverifikasi} | \text{Fakta}) = \frac{0}{8} = 0$$

c. Probabilitas Kriteria Sumber Foto

$$P(\text{Asli} | \text{Hoax}) = \frac{0}{8} = 0$$

$$P(\text{Palsu} | \text{Hoax}) = \frac{8}{8} = 1$$

$$P(\text{Asli} | \text{Fakta}) = \frac{8}{8} = 1$$

$$P(\text{Palsu} | \text{Fakta}) = \frac{0}{8} = 0$$

d. Probabilitas Kriteria Sumber Berita

$$P(\text{Resmi} | \text{Hoax}) = \frac{0}{8} = 0$$

$$P(\text{Tidak Resmi} | \text{Hoax}) = \frac{8}{8} = 1$$

$$P(\text{Resmi} | \text{Fakta}) = \frac{8}{8} = 1$$

$$P(\text{Tidak Resmi} | \text{Fakta}) = \frac{0}{8} = 0$$

Tabel 2. Contoh Penulisan Tabel Kedua

Judul Kolom	Kolom A (t)	Kolom B (t)
Baris pertama	1	2
Baris kedua	3	4
Baris selanjutnya	5	6

4.2.3. Tabel Probabilitas

Tabel probabilitas dari setiap kriteria adalah sebagai berikut:

a. Tabel Probabilitas Kriteria Judul Berita

Judul Berita	Jumlah Kriteria Jenis Berita		Probabilistik	
	Hoax	Fakta	Hoax	Fakta
Provokatif	8	3	1	0.375

Judul Berita	Jumlah Kriteria Jenis Berita		Probabilistik	
	Hoax	Fakta	Hoax	Fakta
Non provokatif	0	5	0	0.375
Jumlah	8	8	1	1

b. Tabel Probabilitas Kriteria Alamat Situs Berita

Alamat Situs Berita	Jumlah Kriteria Jenis Berita		Probabilistik	
	Hoax	Fakta	Hoax	Fakta
Terverifikasi	2	8	0.25	1
Tidak Terverifikasi	6	0	0.75	0
Jumlah	8	8	1	1

c. Tabel Probabilitas Kriteria Cek Foto Berita

Cek Foto	Jumlah Kriteria Jenis Berita		Probabilistik	
	Hoax	Fakta	Hoax	Fakta
Asli	0	8	0	1
Palsu	8	0	1	0
Jumlah	8	8	1	1

d. Tabel Probabilitas Kriteria Sumber Berita

Sumber Berita	Jumlah Kriteria Jenis Berita		Probabilistik	
	Hoax	Fakta	Hoax	Fakta
Resmi	2	8	0.25	1

Tidak Resmi	6	0	0.75	0
Jumlah	8	8	1	1

4.2.4. Data Testing

Judul Berita	Alamat Situs	Cek Foto	Sumber Berita
Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Terverifikasi	Palsu	Tidak Resmi
Provokatif	Terverifikasi	Palsu	Tidak Resmi
Non Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi
Non Provokatif	Terverifikasi	Asli	Resmi
Non Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi
Non Provokatif	Terverifikasi	Asli	Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi
Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi

Menguji Data 1

Judul Berita	Alamat Situs	Cek Foto	Sumber Berita
Provokatif	Terverifikasi	Asli	Resmi

$$P(\text{Berita}|\text{Hoax}) = P(\text{Provokatif}|\text{Hoax})P(\text{Terverifikasi}|\text{Hoax})P(\text{Asli}|\text{Hoax})P(\text{Resmi}|\text{Hoax})P(\text{Hoax}) = 1 \times 0,25 \times 0 \times 0,25 = 0$$

$$P(\text{Berita}|\text{Fakta}) = P(\text{Provokatif}|\text{Fakta})P(\text{Terverifikasi}|\text{Fakta})P(\text{Asli}|\text{Fakta})P(\text{Resmi}|\text{Fakta})P(\text{Fakta}) = 0,375 \times 1 \times 1 \times 1 = 0,375$$

Normalisasi

$$P(\text{Berita}|\text{Hoax})' = \frac{P(\text{Berita}|\text{Hoax})}{P(\text{Berita}|\text{Hoax}) + P(\text{Berita}|\text{Fakta})} = \frac{0}{0 + 0,375} = 0$$

$$P(\text{Berita}|\text{Fakta})' = \frac{P(\text{Berita}|\text{Fakta})}{P(\text{Berita}|\text{Hoax}) + P(\text{Berita}|\text{Fakta})} = \frac{0,375}{0 + 0,375} = 1$$

maka

$$P(\text{Berita}|\text{Fakta})' > P(\text{Berita}|\text{Hoax})'$$

Sehingga dengan kondisi berita yang diberikan, prediksi klasifikasinya adalah **'Fakta'**

4.3. Klasifikasi Menggunakan Python

1. Import library yang diperlukan seperti gambar dibawah ini

```
[ ] import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer

from jcopml.pipeline import num_pipe, cat_pipe
```

2. Import data ke dalam python dan simpan dalam DataFrame. Lalu drop kolom label.

```
df = pd.read_csv("datadamin.csv")
df
```

	Judul Berita	Alamat Situs Berita	Cek Foto Berita	Sumber	Label
0	Provokatif	Terverifikasi	Asli	Resmi	fakta
1	Provokatif	Terverifikasi	Palsu	Tidak Resmi	hoax
2	Provokatif	Terverifikasi	Palsu	Tidak Resmi	hoax
3	Non Provokatif	Terverifikasi	Asli	Resmi	fakta
4	Provokatif	Terverifikasi	Asli	Resmi	fakta
5	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax
6	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax
7	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax
8	Non Provokatif	Terverifikasi	Asli	Resmi	fakta
9	Non Provokatif	Terverifikasi	Asli	Resmi	fakta
10	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax
11	Non Provokatif	Terverifikasi	Asli	Resmi	fakta
12	Non Provokatif	Terverifikasi	Asli	Resmi	fakta
13	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax
14	Provokatif	Tidak Terverifikasi	Palsu	Tidak Resmi	hoax

```
[ ] X = df.drop(columns="Label")
y = df.Label
```

3. Bagi data menjadi data train dan data test. Lalu cetak dimensi atau bentuk dari masing-masing variabel

```
[7] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
((12, 4), (3, 4), (12,), (3,))
```

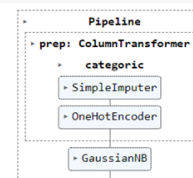
4. Transformatikan pada kolom-kolom kategorikal dari dataset menggunakan 'ColumnTransformer' menggunakan pipeline untuk melatih dan menguji model.

```
[11] preprocessor = ColumnTransformer([
    ('categorical', cat_pipe(encoder='onehot'), ['Judul Berita', 'Alamat Situs Berita',
    'Cek Foto Berita', 'Sumber']),
])

[9] from sklearn.naive_bayes import GaussianNB
pipeline = Pipeline([
    ('prep', preprocessor),
    ('algo', GaussianNB())
])
```

5. Memanggil 'fit()' dan 'score()' pada pipeline untuk melatih model. Mengukur akurasi pada data train dan test. Lalu tampilkan tampilkan dalam confusion matrix.

```
[12] pipeline.fit(X_train, y_train)
```



```
[13] pipeline.score(X_train, y_train)
```

1.0

```
[14] pipeline.score(X_test, y_test)
```

1.0

```
[15] from jcopml.plot import plot_confusion_matrix
plot_confusion_matrix(X_train, y_train, X_test, y_test, pipeline)
```



6. Lakukan pengujian dengan menginput data 1 dan 3. Maka dapat kita lihat prediksi klasifikasi data 1 adalah fakta dan data 3 hoax

```
[18] X_pred = pd.read_csv("testing.csv")
X_pred
```

	Judul Berita	Alamat Situs Berita	Cek Foto Berita	Sumber
0	Provokatif	Terverifikasi	Asli	Resmi
1	Provokatif	Terverifikasi	Palsu	Tidak Resmi

```
[19] pipeline.predict(X_pred)
```

```
array(['fakta', 'hoax'], dtype=<U>)
```

```
[20] X_pred["Label"] = pipeline.predict(X_pred)
X_pred
```

	Judul Berita	Alamat Situs Berita	Cek Foto Berita	Sumber	Label
0	Provokatif	Terverifikasi	Asli	Resmi	fakta
1	Provokatif	Terverifikasi	Palsu	Tidak Resmi	hoax

5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan, dapat disimpulkan bahwa algoritma Naive Bayes Classifier mampu mengklasifikasi berita sebagai Hoax atau Fakta dengan akurasi yang memuaskan. Pada data training, algoritma ini mampu mencapai akurasi yang cukup tinggi, demikian juga pada data testing pada data split ratio. Hal ini menunjukkan bahwa Naive Bayes Classifier dapat digunakan sebagai salah satu metode untuk mendeteksi berita Hoax atau Fakta yang tersebar di sosial media. Dengan begitu, dapat membantu masyarakat dalam menghindari penyebaran informasi yang tidak benar dan memastikan kebenaran informasi yang mereka terima. Meskipun demikian, diperlukan pengembangan lebih lanjut terhadap algoritma ini agar dapat meminimalisir kesalahan klasifikasi dan meningkatkan akurasi secara keseluruhan.

DAFTAR PUSTAKA

- [1] Ardhana, Fikri. 2021. *Sentiment Analysis Tweet Vaksinasi COVID-19 Menggunakan Metode Naive Bayes Classifier*. Skripsi. Tidak Diterbitkan. Fakultas Teknologi Informasi. Universitas Andalas: Padang.
- [2] Diskominfo. 2022. *Pengertian Hoax dan Cara Menangkalnya*. Diakses pada 30 Mei 2023, dari [Pengertian Hoax dan Cara Menangkalnya | Diskominfo Badung \(badungkab.go.id\)](https://diskominfo.badung.badungkab.go.id).
- [3] Gravina, Hesti. 2020. *Pemanfaatan Image Mining Untuk Klasifikasi Kanker Paru-Paru Menggunakan Jaringan Syaraf Tiruan (Studi Kasus: Semen Padang Hospital Dan RSUP M Djamil Padang)*. Skripsi. Tidak Diterbitkan. Fakultas Teknologi Informasi. Universitas Andalas: Padang.
- [4] Hendriyanto, M, D., B, N, Sari. 2022. *Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Judul Berita Hoax*. Jurnal Ilmiah Informatika 10 (2).
- [5] Herlambang, MB. 2018. *Training dan test set*. Diakses pada 30 Mei 2023, dari [Training dan test set - Artificial intelligence and data science \(megabagus.id\)](https://megabagus.id).
- [6] Ibrahim, M., E, Bu'ulolo., I, Lubis. 2020. *Penerapan Algoritma Naive Bayes Classifier Untuk Mendeteksi Tingkat Kredibilitas Hoax News/ Fake News Pada Sosial Media Di Indonesia Berbasis Android (Studi Kasus: Kantor Tribun Medan)*. Jurnal RESOLUSI: Rekayasa Teknik Informatika dan Informasi 1 (1): 9-17.
- [7] Pangarso, Ditto Afik. 2021. *Klasifikasi Genre Film Berdasarkan Wikipedia Movie Plots Menggunakan Naive Bayes Classifier*. Skripsi. Tidak Diterbitkan. Fakultas Teknologi Informasi. Universitas Andalas: Padang.
- [8] Populix. 2023. *Apa itu Data Mining? Pengertian, Metode, Tahapan, Contoh*. Diakses pada 30 Mei 2023, dari [Apa itu Data Mining? Pengertian, Metode, Tahapan, Contoh \(populix.co\)](https://populix.co).
- [9] Rehearsal. 2021. *Prediksi lulus with Naive Bayes.ipynb*
- [10] Setiawan, Rony. 2021. *Apa itu Data Mining dan Bagaimana Metodenya? Dicoding*. Diakses pada 30 Mei 2023, dari [Apa itu Data Mining dan Bagaimana Metodenya? - Dicoding Blog](https://www.dicoding.com).
- [11] STIKOM, SURABAYA. 2019. *BAB 2 TINJAUAN PUSTAKA*. Diakses pada 30 Mei 2023, dari [BAB II.pdf \(dinamika.ac.id\)](https://dinamika.ac.id).
- [12] Team, Jagoan Hosting. 2023. *Data Mining: Pengertian, Fungsi, Metode & Penerapannya*. Diakses pada 30 Mei 2023, dari [Data Mining: Pengertian, Fungsi, Metode & Penerapannya \(jagoanhosting.com\)](https://jagoanhosting.com).
- [13] Tineges, Rian. 2022. *Mengenal Naive Bayes Sebagai Salah Satu Algoritma Data Science*. Diakses pada 30 Mei 2023, dari [Mengenal Naive Bayes Sebagai Salah Satu Algoritma Data Scien... \(dqlab.id\)](https://dqlab.id).
- [14] Zubayr, Abu. 2016. *Algoritma Naive Bayes Classifier Menggunakan Java*. Diakses pada 30 Mei 2023, dari [Algoritma Naive Bayes Classifier Menggunakan Java | Programming Tutorial \(celiksoftware.blogspot.com\)](https://celiksoftware.blogspot.com).