



UNIVERSITAS  
**MERCU BUANA**

PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST  
NEIGHBOR* UNTUK KLASIFIKASI BERITA *HOAX* KESEHATAN DI MEDIA  
SOSIAL TWITTER

*TUGAS AKHIR*

Dheo Alpian Nanda Krisna  
41518010028

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022**



PERBANDINGAN ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR* UNTUK KLASIFIKASI BERITA *HOAX* KESEHATAN DI MEDIA SOSIAL TWITTER

*Tugas Akhir*

Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer

Oleh:

Dheo Alpian Nanda Krisna  
41518010028

UNIVERSITAS  
MERCU BUANA  
PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS MERCU BUANA  
JAKARTA  
2022

---

## LEMBAR PERNYATAAN ORISINALITAS

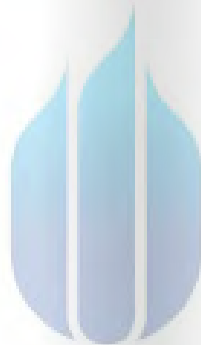
Yang bertanda tangan dibawah ini:

NIM : 41518010028

Nama : Dheo Alpian Nanda Krisna

Judul Tugas Akhir : Perbandingan Algoritma Naive Bayes dan K Nearest Neighbor Untuk Klasifikasi Berita Hoax Di Media Sosial Twitter

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 04 Juli 2022



Dheo Alpian Nanda Krisna

UNIVERSITAS  
MERCU BUANA

## SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Dheo Alpian Nanda Krisna  
NIM : 41518010028  
Judul Tugas Akhir : Perbandingan Algoritma Naive Bayes dan K Nearest Neighbor Untuk Klasifikasi Berita Hoax Di Media Sosial Twitter

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 04 Juli 2022

UNIVERSITAS  
MERCU BUANA



Dheo Alpian Nanda Krisna

### SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa : Dheo Alpian Nanda Krisna  
NIM : 41518010028  
Judul Tugas Akhir : Perbandingan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Berita Hoax Di Media Sosial Twitter

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis	Status
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi	Diajukan ✓
		Jurnal Nasional Terakreditasi	
		Jurnal International Tidak Bereputasi	Diterima
		Jurnal International Bereputasi	
Disubmit/dipublikasikan di :	Nama Jurnal	: JTIK (Jurnal Teknik Informatika Kaputama)	
	ISSN	: 2548 - 9704	
	Link Jurnal	: <a href="https://jurnal.kaputama.ac.id/index.php/JTIK/index">https://jurnal.kaputama.ac.id/index.php/JTIK/index</a>	
	Link File Jurnal Jika Sudah di Publish	:	

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 04 Juli 2022




Dheo Alpian Nanda Krisna

## LEMBAR PERSETUJUAN PENGUJI

NIM : 41518010028  
Nama : Dheo Alpian Nanda Krisna  
Judul Tugas Akhir : Perbandingan *Algoritma Naive Bayes dan K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 01 Agustus 2022



(Ir. Emil R. Kaburuan, Ph.D., IPM)

UNIVERSITAS  
MERCU BUANA

**LEMBAR PERSETUJUAN PENGUJI**

NIM : 41518010018  
Nama : Dheo Alpian Nanda Krisna  
Judul Tugas Akhir : Perbandingan Algoritma *Naive Bayes* dan *K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 01 Agustus 2022



(Vina Ayumi, S.Kom., M.Kom)



UNIVERSITAS  
MERCU BUANA

**LEMBAR PERSETUJUAN PENGUJI**

NIM : 41518010028  
Nama : Dheo Alpian Nanda Krisna  
Judul Tugas Akhir : Perbandingan Algoritma *Naive Bayes* dan *K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 01 Agustus 2022

  
(Eugenius Kau Suni, ST, MT)

UNIVERSITAS  
MERCU BUANA



## LEMBAR PENGESAHAN

NIM : 41518010028  
Nama : Dheo Alpian Nanda Krisna  
Judul Tugas Akhir : Perbandingan Algoritma *Naive Bayes* dan *K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 09 Agustus 2022

Menyetujui,



Umniy Salamah, ST.,  
MMSI  
Dosen Pembimbing

Mengetahui,

UNIVERSITAS  
MERCU BUANA



(Wawan Gunawan, S.Kom, MT)  
Tugas Akhir Teknik Informatika



(Ir. Emil R. Kaburuan, Ph.D., IPM.) Koord.  
Ka. Prodi Teknik Informatika

## KATA PENGANTAR

Puji serta syukur tak lupa penulis haturkan atas Kehadirat Allah S.W.T, karena berkat rahmat dan hidayah-Nya, penulis mampu untuk menyelesaikan skripsi yang berjudul Perbandingan Algoritma *Naive Bayes* dan *K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter “.Sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) pada Program Sarjana Fakultas Ilmu Komputer Jurusan Informatika Universitas Mercu Buana.Tentunya di dalam penyusunan skripsi ini banyak pihak yang membantu, baik membantu dalam moral maupun dari material. Karena itu, penulis tidak lupa untuk menyampaikan terima kasih kepada :

1. Orang Tua yang selalu memberikan dukungan, baik dukungan berupa doa bahkan sampai dukungan material sehingga penulis mampu menyelesaikan tugas akhir ini tepat pada waktunya.
2. Ibu Umniy Salamah, ST., MMSI selaku dosen pembimbing saya yang telah membantu dan memberikan arahan kepada saya dalam menyelesaikan tugasakhir ini.
3. Bapak Emil R. Kaburuan Ph.D selaku Kepala Program Studi Informatika yang telah memberikan arahan dan bimbingannya selama perkuliahan.
4. Bapak Wawan Gunawan, S.Kom., M.T. selaku Koordinator Tugas Akhir Prodi Informatika yang telah memberikan arahan selama periode tugas akhir.
5. Bapak Dr. Harwikarya, MT selaku dosen pembimbing akademik saya yang telah membimbing selama perkuliahan hingga sampai pada tugas
6. Teman-teman Toonite , sahabat dan kerabat yang selalu memberikan support dan waktu untuk selalu mengingatkan dalam menyelesaikan tugas akhir ini.

Tentunya, penulis menyadari bahwasanya masih banyak kekurangan di penyusunan Jurnal Tugas Akhir ini.Akhir kata, penulis berharap jika ada kritik dan juga saran yang sifatnya membangun tentu akan penulis terima dengan senang hati.

Jakarta, 04 Juli 2022

Penulis

## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL .....	i
LEMBAR PERNYATAAN ORISINALITAS .....	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR...iii	
SURAT PERNYATAAN LUARAN TUGAS AKHIR.....	iv
LEMBAR PERSETUJUAN PENGUJI .....	v
LEMBAR PENGESAHAN.....	viii
ABSTRAK.....	ix
ABSTRACT .....	x
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xii
NASKAH JURNAL .....	1
KERTAS KERJA .....	12
BAB 1. LITERATUR REVIEW.....	13
BAB 2. SOURCE CODE.....	25
BAB 3. DATASET .....	32
BAB 4. TAHAPAN EKSPERIMEN.....	41
BAB 5. HASIL SEMUA EKSPERIMEN.....	47
DAFTAR PUSTAKA.....	51
LAMPIRAN DOKUMEN HAKI .....	54
LAMPIRAN KORESPONDENSI.....	56

## NASKAH JURNAL

Jurnal Teknik Informatika Kaputama (JTIK) Vol.6, No. 2, Juli 2022

P-ISSN: 2548-9704

E-ISSN: 2686-0880

### PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI BERITA HOAX KESEHATAN DI MEDIA SOSIAL TWITTER

Dheo Alpian Nanda Krisna<sup>1)</sup>, Umniy Salamah<sup>2)</sup>

<sup>1,2</sup>Universitas Mercu Buana

Jl. Raya, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Jakarta

E-mail: dheoalpian7@gmail.com<sup>1)</sup>, umniy.salamah@mercubuana.ac.id<sup>2)</sup>,

#### ABSTRACT

Along with the rapid development of technology, making the spread of information and news unstoppable. Only by opening social media on the smart devices that we use, we can get all kinds of news. Of course, not all the news we get is factual information, but fake news or hoax news. This study focuses on fake news that is widely spread in social media, especially on Twitter social media. Later this research can be useful for classifying hoax news on Twitter social media. The presence of fake news or hoaxes has a very extraordinary impact that can have the potential to make people's perceptions or views become misguided where they think that fake news or hoaxes are the truth. Not only that, because it is easier for fake news or hoaxes to spread on social media, making it difficult for us to determine which news is true and which is fake news or hoax. Examples of such negative impacts are what we must avoid. In this study, two algorithms were used, namely the K-Nearest Neighbor (KNN) and Naïve Bayes (NB) algorithms. The two algorithms are used to classify hoax news on Twitter social media. The results obtained are the K-Nearest Neighbor (KNN) algorithm provides a better accuracy value than the algorithm and Naïve Bayes (NB). With an accuracy value of 0.68 for the K-Nearest Neighbor KNN algorithm and 0.66 for the algorithm and Naïve Bayes (NB).

**Keywords:** Hoax, Twitter, K-Nearest Neighbor, Naïve Bayes, Classification

#### ABSTRAK

Seiring dengan perkembangan teknologi yang sangat pesat, membuat persebaran informasi maupun berita menjadi tidak terbendung. Hanya dengan membuka sosial media yang ada di dalam perangkat pintar yang kita gunakan, maka segala macam berita bisa kita peroleh. Tentunya, tidak semua berita yang di dapat merupakan informasi yang bersifat fakta, melainkan berita bohong atau berita hoax. Penelitian ini berfokus pada berita bohong yang marak tersebar di dalam media sosial khususnya di media sosial Twitter. Nantinya penelitian ini bisa bermanfaat untuk mengklasifikasi berita hoax yang ada di media sosial Twitter. Dengan hadir nya berita bohong atau hoax memiliki dampak yang sangat luar biasa yang bisa berpotensi untuk membuat persepsi ataupun pandangan masyarakat menjadi sesat dimana menganggap berita bohong atau hoax itu merupakan sebuah kebenaran. Bukan hanya itu, karena semakin mudahnya berita bohong atau hoax menyebar di sosial media, membuat kita kesulitan untuk

836

menentukan mana berita yang benar dan mana yang berita bohong atau hoax. Contoh dampak negatif seperti itulah yang harus kita hindari. Di dalam penelitian ini menggunakan dua algoritma, yaitu algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes (NB). Kedua algoritma tersebut digunakan untuk mengklasifikasi berita hoax yang ada di media sosial Twitter. Hasil yang diperoleh adalah algoritma K-Nearest Neighbor (KNN) memberikan nilai akurasi yang lebih baik dibanding dengan algoritma dan Naïve Bayes (NB). Dengan nilai akurasi sebesar 0.68 untuk algoritma K-Nearest Neighbor KNN dan 0.66 untuk algoritma dan Naïve Bayes (B).

**Kata kunci: Hoax, Twitter, K-Nearest Neighbor, Naïve Bayes, Klasifikasi**

## I. PENDAHULUAN

Seiring dengan semakin majunya teknologi dan ditambah dengan situasi pandemic covid – 19, maka penggunaan sosial media makin meningkat secara drastis. Tentunya masing – masing masyarakat punya gaya tersendiri dalam memanfaatkan media sosial. Sebagai contoh, ada sebagian masyarakat yang menggunakan media sosial untuk menyebarkan sebuah berita [1]. Namun, berdasarkan fakta yang ada tidak semua orang menggunakan sosial media dengan bertanggung jawab dalam menyebarkan sebuah berita. Dampak dari berita hoax itu sendiri bisa merubah persepsi orang menjadi sesat [2]. Hal ini dilihat dari maraknya berita hoax yang tersebar di salah satu media sosial yaitu Twitter. Twitter itu sendiri merupakan sebuah media sosial dengan salah satu fitur yang memungkinkan kita untuk bercengkrama dengan para sesama pengguna bahkan dengan artis sekalipun [3]. Hadirnya berita hoax itu sendiri hanya akan menimbulkan sebuah permasalahan. Permasalahan lainnya adalah dengan beredar berita hoax di media sosial Twitter menyebabkan beberapa pihak tertentu bisa mengalami kerugian [4]. Beberapa penelitian terdahulu yang memiliki hubungan dengan klasifikasi hoax diantaranya penelitian [5] yang dilakukan oleh Berlian mengenai

klasifikasi berita dengan menggunakan algoritma K-Nearest Neighbor memberikan hasil bahwa algoritma ini mampu untuk mengklasifikasi berita hoax namun dengan hasil yang tidak begitu baik, selanjutnya penelitian yang dilakukan oleh R.Santhi yang menunjukkan bahwa metode KNN memberikan nilai akurasi sebesar 75,89%

untuk mendeteksi hoax terkait covid – 19 [6]. Sedangkan penelitian yang dilakukan oleh Nova dkk yang menggunakan algoritma Naïve Bayes untuk mendeteksi berita palsu memberikan hasil akurasi sebesar 81% [7]. Kemudian, penelitian yang dilakukan oleh Eri Zuliarso mengenai deteksi hoax menggunakan KNN dan mendapatkan hasil yang baik dimana didalam penelitiannya memanfaatkan 74 data hoax memberikan nilai akurasi sebesar 83,6% [8]. Dan penelitian yang dilakukan oleh Faisal Rahutomo dkk dengan topik mendeteksi berita hoax didalam bahasa Indonesia memberikan hasil akurasi sebesar 78,6% dengan tiga kali pengujian [9]. Berdasarkan hal itu untuk mengatasi permasalahan yang ada, penulis nantinya akan mengimplementasikan algoritma K-Nearest Neighbor dan Naïve bayes untuk mengklasifikasi berita hoax dengan harapan akan membantu bagi para pembaca untuk bisa memilah mana berita yang benar mana berita yang hoax. Dimana kelebihan didalam penelitian ini adalah dengan menggunakan dua metode yang berbeda

ditambah dengan penggunaan dataset dengan jumlah yang banyak.

## 2. METODOLOGI

Metodologi yang digunakan didalam penelitian ini diawali dengan mengidentifikasi masalah, selanjutnya akan dikumpulkan beberapa literatur review yang digunakan untuk menjadi referensi didalam penelitian ini. Setelah itu akan dilakukan pengumpulan data yang dimana data itu akan di preprocessing sebelum diuji dengan menggunakan Algoritma K-Nearest Neighbor dan Naïve Bayes. Dibawah ini merupakan alur dari metode penelitian :

Gambar 1 Tahapan Penelitian



### 1. Identifikasi Masalah

Tahapan awal yang peneliti lakukan adalah dengan mengidentifikasi masalah yang ada dan mencari solusi untuk mengatasi permasalahan tersebut. Permasalahan yang ada didalam penelitian yang penulis lakukan adalah maraknya tersebar berita hoax di media sosial, yaitu media sosial Twitter. Dilihat dari dampak yang ada tentunya keberadaan berita hoax hanya memberikan dampak buruk saja. Maka dari itu dibutuhkan sebuah klasifikasi untuk membagi berita yang ada di media sosial Twitter dengan memanfaatkan algoritma K-nearest Neighbor dan Naïve Bayes.

### 2. Pengumpulan Literatur Review

Tentunya disaat pembuatan penelitian ini diperlukannya sebuah literatur yang menjadi pedoman untuk bagaimana penelitian ini akan dibuat. Literatur review yang digunakan didalam penelitian ini adalah dengan membaca setiap jurnal yang memiliki kaitan ataupun hubungan dengan topik yang peneliti pilih.

### 3. Pengumpulan Data Berita Hoax

Pengumpulan data berita hoax disini menggunakan teknik crawling. Teknik crawling itu sendiri merupakan sebuah teknik yang memang diperuntukan untuk mengambil data yang ada didalam web [10]. Untuk teknik crawling yang peneliti gunakan adalah dengan memanfaatkan bahasa pemrograman python. Berita yang dimaksud adalah berita yang berasal dari *tweet* para pengguna media sosial Twitter.

### 4. Preprocessing

Preprocessing merupakan tahapan awal untuk mengolah data yang telah dikumpulkan. Preprocessing itu sendiri merupakan sebuah tahapan dalam mengolah data yang diperuntukan untuk mengolah data supaya sesuai dengan kebutuhan. Preprocessing juga bisa dimanfaatkan untuk mengubah data yang tidak terstruktur menjadi terstruktur [11]. Beberapa hal seperti menghapus, membersihkan, sekaligus menghapus karakter maupun angka yang ada didalam data merupakan beberapa hal yang akan dilakukan didalam preprocessing ini [12]. Pada penelitian ini, beberapa tahapan preprocessing yang dilakukan adalah data cleansing, stopword removal, tokenizing, case folding, stemming dan labelling.

### 5. Penerapan Algoritma K-Nearest Neighbor dan Naïve Bayes

Setelah data selesai di preprocessing tahapan selanjutnya adalah diimplementasikannya algoritma yang akan digunakan yaitu algoritma K-Nearest Neighbor dan Naïve Bayes. Kedua algoritma tersebut nantinya akan dimanfaatkan untuk mengklasifikasi berita hoax.

#### 6. Pengujian Algoritma K-Nearest Neighbor dan Naïve Bayes

Selanjutnya pada tahapan ini akan diuji bagaimana performa dari kedua algoritma yang digunakan untuk mengklasifikasi berita hoax yang berasal dari media sosial Twitter. Selanjutnya akan dibandingkan kedua algoritma tersebut untuk dilihat algoritma mana yang memberikan hasil terbaik untuk mengklasifikasi.

##### 2.1. Klasifikasi

Jika dilihat berdasarkan pengertiannya, klasifikasi merupakan sebuah proses yang diperuntukan untuk menjelaskan dan juga membedakan kelas – kelas data. Sedangkan jika dilihat berdasarkan Kamus Besar Bahasa Indonesia pengertian dari klasifikasi ialah penyusunan bersistem dalam kelompok atau golongan menurut kaidah atau standar yang ditetapkan.

##### 2.2. K-Nearest Neighbor

Algoritma K-Nearest Neighbor merupakan sebuah algoritma yang masih masuk ke dalam metode Machine Learning. Algoritma ini merupakan salah satu algoritma yang tidak terlalu sulit untuk diimplementasikan. Algoritma ini bisa dimanfaatkan untuk menghindari outlier maupun noise. Untuk dataset yang digunakan pun bisa dengan jumlah yang besar maupun sedikit. Untuk cara kerjanya, algoritma ini akan menghitung jarak antara data uji dengan data lain yang ada di dataset. Selanjutnya, K data dengan jarak terdekat akan dipilih. Kelas yang memiliki jumlah data terbanyak dalam daftar K data tersebut kemudian akan dipilih sebagai kelas dari data uji [13]. Didalam algoritma ini terdapat persamaan sebagai berikut :

Keterangan :

- $d(x_i, x_j)$  : jarak antara data  $x_i$  dan  $x_j$
- $n$  : jumlah fitur yang dimiliki data
- $ar(x_i)$  : nilai fitur  $r$  dari data  $x_i$
- $ar(x_j)$  :  $h$  nilai fitur  $r$  dari data  $x_j$

##### 2.3. Naïve Bayes

UNIVERSITAS  
MERCU BUANA

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

Naïve Bayes merupakan sebuah algoritma yang dikemukakan oleh Thomas Bayes. Algoritma Naïve Bayes merupakan sebuah pengklasifikasian probabilistik yang tidak kompleks untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Teori keputusan bayes merupakan pendekatan statistic yang fundamental dalam pengenalan pola. Apabila kita ingin mengimplementasikan dalam hal klasifikasi. Naïve Bayes didasari pada asumsi sederhana bahwa nilai atribut secara kondisional saling tidak terkait jika diberikan nilai output, nantinya probabilitas akan mengamati secara Bersama yang hasilnya adalah produk dari probabilitas individu [14]. Metode bayes itu sendiri adalah sebuah metode yang diperuntukan untuk inferensi induksi untuk permasalahan klasifikasi [15]. Dalam algoritma ini terdapat persamaan seperti dibawah ini :

Keterangan :

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (*posteriori probabilitas*)
- P(H) : Probabilitas hipotesis H (*prior probabilitas*)
- P(X|H) : Probabilitas X berdasarkan kondisi pada

hipotesis H

- P(X) : Probabilitas X

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Dataset

Dataset yang digunakan adalah data yang berasal dari media sosial Twitter dengan menggunakan teknik crawling. Untuk dataset yang digunakan berjumlah 1000 data yang berasal dari *tweet* dari para pengguna media sosial Twitter. Dibawah ini merupakan tabel dari dataset yang digunakan :

Tabel 1 Dataset

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (ar(x_i) - ar(x_j))^2}$$



Waktu	Username	Teks
20/05/2022 00:56	andika setyawan aray	selamat hari kebangkitan nasional tahun momentum.....
20/05/2022 00:55	nadma malaysia	situasi terkini pusat kuarantin amp rawatan covid berisiko rendah.....
20/05/2022 00:55	lobak merah	update kes harian covid mei sumber kementerian.....
20/05/2022 00:55	cak uding	zaman soeharto juga ga ada covid luar biasa emang zaman soeharto.....
20/05/2022 00:54	prokopim pangandaran	selamat memperingati hari kebangkitan nasional .....
...	...	...
19/05/2022 15:08	strawberrymilk	covid punya pasal first time jumpa mertua.....
19/05/2022 15:07	jawa pos	kai tidak wajibkan penumpang.....
19/05/2022 15:05	sarawak news	rakyat indonesia sudah terima bansos ekonomi rakyat.....
19/05/2022 15:05	muhammad faisal	last year masih lagi steady km pace bulan haritu stop sebab positif covid.....
19/05/2022 15:05	rukundo mmed	kita wajib mempertahankan gaya hidup disiplin taat protokol kesehatan.....
19/05/2022 15:08	nananeranananera	subsidi biaya rawat pasien covid dihentikan.....

### 3.2. Preprocessing

Didalam tahapan ini dilakukan beberapa tahapan seperti dibawah berikut :

#### 1. Data Cleansing

Pada penelitian ini peneliti menggunakan data dengan total sebanyak 1000 record. Pada data cleansing ini peneliti menghapus setiap character yang terdapat didalam tweet yang ada.

#### 2. Stopword Removal

Dalam proses ini dilakukan melakukan stopwords removal dengan tujuan untuk melakukan penghapusan kata – kata yang tidak bermakna. Sebagai contohnya adalah *tweet* “@silverspoonnn Jaga kesehatan yaaa, jangan lupa mam vitamin sama minum air putih yang banyak jangan lupa istirahat take care yaδY’œ” akan diubah menjadi “silverspoonnn Jaga kesehatan yaaa jangan lupa mam vitamin sama minum air putih banyak jangan lupa istirahat take care ya”.

#### 3. Tokenizing

Proses tokenizing merupakan sebuah proses untuk menjadikan menghapus setiap angka yang terdapat didalam data. Sebagai contoh “ @nandarwa15 @strawjoyie diskon 10% (up to 50k), minbel 200k kak \*Hanya berlaku untuk pembayaran Tagihan Telkom, Penyedia Angsuran, PDAM dan BPJS Kesehatan “ menjadi “nandarwa strawjoyie diskon up to k minbel k kak Hanya berlaku pembayaran Tagihan Telkom Penyedia Angsuran PDAM BPJS kesehatan”.

#### 4. Case Folding

Tahapan case folding

diperuntukan untuk menjadikan setiap huruf yang ada menjadi huruf kecil. Sebagai contoh menjadi "nandarwa strawjoyie diskon up to k minbel k kak Hanya berlaku pembayaran Tagihan Telkom Penyedia Angsuran PDAM BPJS kesehatan", menjadi " " nandarwa strawjoyie diskon up to k minbel k kak hanya berlaku pembayaran tagihan telkom penyedia angsuran pdam bpjs kesehatan ".

#### 5. Stemming

Proses ini merupakan proses yang digunakan untuk menghapus setiap infleksi kata ke bentuk kata dasarnya. Beberapa contoh kata yang sudah di stemming :

- kehadiran : hadir
- memastikan : pasti
- rangkaian : rangkai

#### 6. Labelling

Penulis juga melakukan proses labelling terhadap dataset yang telah berhasil dikumpulkan dengan teknik crawling data menggunakan API. Dimana proses labelling itu sendiri merupakan proses untuk menentukan tiap kelas dari *tweet* yang telah diunggah. Untuk labelling itu sendiri disimbolkan menjadi 3 bagian yaitu :

- 0 = Netral
- 1 = Hoax
- 2 = Fakta

### 3.3. Hasil Pengujian Algoritma K-Nearest Neighbor dan Naïve Bayes

#### 1. Algoritma K-Nearest Neighbor

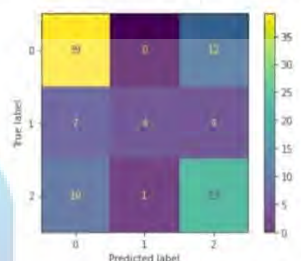
Hasil Accuracy, Precision and Recall menggunakan split validation :

Tabel 2 Hasil Accuracy, Precision and

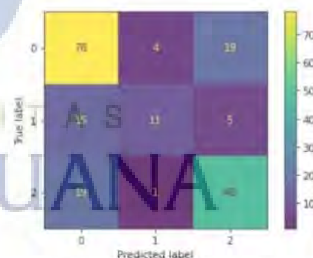
Eksperiment	Accuracy	Precision	Recall
1	0.66	0.68	0.66
2	0.69	0.68	0.69
3	0.70	0.69	0.70

Recall

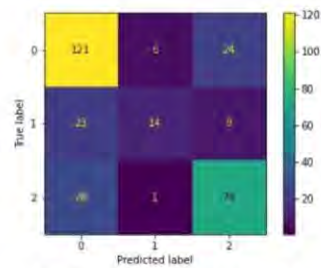
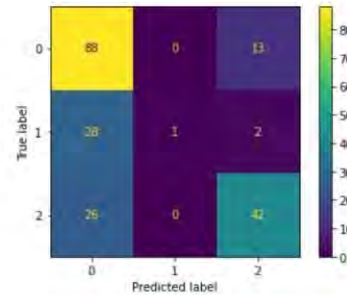
Confusion Matrix Ke-3 Percobaan :



Gambar 2 Matrix Percobaan 1



Gambar 3 Matrix Percobaan 2

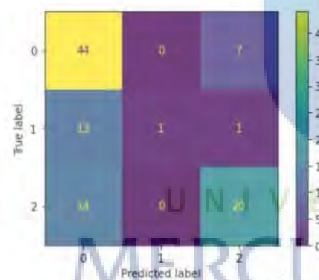
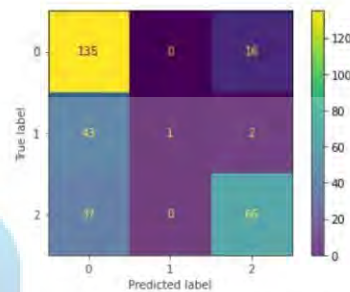
Gambar 4 *Matrix* Perobaan 3Gambar 6 *Matrix* Perobaan 2

## 2. Algoritma Naïve Bayes

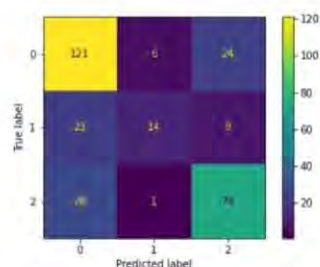
Hasil Accuracy, Precision and Recall menggunakan split validation :

Eksperimen	Accuracy	Precision	Recall
1	0.65	0.71	0.65
2	0.66	0.72	0.66
3	0.67	0.74	0.67

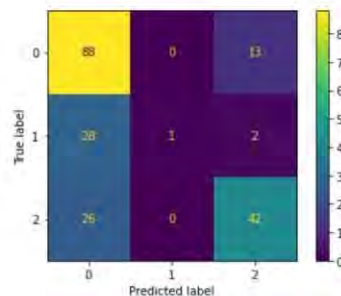
Confusion Matrix Ke-3 Percobaan :

Gambar 5 *Matrix* Perobaan 1Gambar 7 *Matrix* Perobaan 3

Berdasarkan dari semua percobaan yang telah dilakukan, algoritma K Nearest Neighbor (KNN) memberikan nilai accuracy sebesar 0.68 sedangkan untuk algoritma Naive Bayes (NB) memberikan nilai accuracy sebesar 0.66. Namun untuk nilai precision yang diperoleh saat menggunakan algoritma K Nearest Neighbor (KNN) hanya sebesar 0.68, sedikit lebih rendah jika dibandingkan dengan algoritma Naive Bayes (NB) yang memperoleh nilai precision sebesar 0.72. Dan hasil yang berbeda terjadi di bagian Recall dimana algoritma K Nearest Neighbor (KNN) unggul sedikit dibandingkan dengan algoritma Naive



Gambar 4 Matrix Percobaan 3



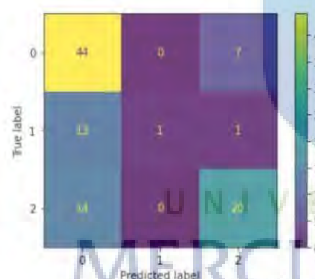
Gambar 6 Matrix Percobaan 2

2. Algoritma Naïve Bayes

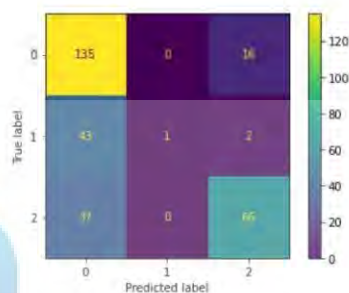
Hasil Accuracy, Precision and Recall menggunakan split validation :

Eksperimen	Accuracy	Precision	Recall
1	0.65	0.71	0.65
2	0.66	0.72	0.66
3	0.67	0.74	0.67

Confusion Matrix Ke-3 Percobaan :



Gambar 5 Matrix Percobaan 1



Gambar 7 Matrix Percobaan 3

Berdasarkan dari semua percobaan yang telah dilakukan, algoritma K Nearest Neighbor (KNN) memberikan nilai accuracy sebesar 0.68 sedangkan untuk algoritma Naive Bayes (NB) memberikan nilai accuracy sebesar 0.66. Namun untuk nilai precision yang diperoleh saat menggunakan algoritma K Nearest Neighbor (KNN) hanya sebesar 0.68, sedikit lebih rendah jika dibandingkan dengan algoritma Naive Bayes (NB) yang memperoleh nilai precision sebesar 0.72. Dan hasil yang berbeda terjadi di bagian Recall dimana algoritma K Nearest Neighbor (KNN) unggul sedikit dibandingkan dengan algoritma Naive

Bayes (NB). Dengan hasil Recall sebesar 0.68 untuk algoritma K Nearest Neighbor KNN dan 0.66 untuk Naive Bayes (NB).

#### 4. KESIMPULAN

Maka kesimpulan yang didapat adalah algoritma K Nearest Neighbor (KNN) memberikan hasil yang lebih baik jika dibandingkan dengan algoritma Naive Bayes (NB) untuk menilai seberapa tepat kedua metode itu untuk mengklasifikasi berita hoax di media sosial Twitter.

#### 5. SARAN

Peneliti menyadari bahwa pembuatan jurnal ini masih jauh dari kata sempurna, maka dari itu peneliti menyarankan bahwa perlu dikembangkannya lebih dimana memanfaatkan metode lain sehingga bisa dijadikan pembandingan, supaya hasil yang diperoleh bisa lebih baik lagi.

#### DAFTAR PUSTAKA

- [1] R. Wati, "Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 159–164, 2020, doi: 10.33480/jitk.v5i2.1034.
- [2] A. Afriza and J. Adisantoso, "Metode Klasifikasi Rocchio untuk Analisis Hoax Rocchio Classification Method for Hoax Analysis," *J. Ilmu Komput. Agri-Informatika*, vol. 5, pp. 1–10, 2018, [Online]. Available: <http://journal.ipb.ac.id/index.php/jika>
- [3] S. Azeharie and O. Kusuma, "Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis Di Jakarta," pp. 83–98, 2014. T. Limbong and dkk, *Sistem Pendukung Keputusan : Metode dan Implementasi*. Medan: Yayasan Kita Menulis, 2020.
- [4] H. Muhabatin, C. Prabowo, I. Ali, C. L. Rohmat, and D. R. Amalia, "Klasifikasi Berita Hoax Menggunakan Algoritma Naive Bayes Berbasis PSO," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 2, pp. 156–165, 2021, [Online]. Available: <http://ejournal-binainsani.ac.id/index.php/ITBI/article/view/1531>.
- [5] B. K. Palma, D. T. Murdiansyah, and W. Astuti, "Klasifikasi Teks Artikel Berita Hoaks Covid-19 dengan Menggunakan Algoritma K-Nearest Neighbor," *eProceedings ...*, vol. 8, no. 5, pp. 10637–10649, 2021, [Online].
- [6] E. Utami, A. F. Iskandar, W. Hidayat, A. B. Prasetyo, and A. D. Hartanto, "Covid-19 Hoax Detection Using KNN in Jaccard Space," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 15, no. 3, p. 255, 2021, doi: 10.22146/ijccs.67392.
- [7] N. Agustina and M. Hermawati, "Implementasi Algoritma Naive Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media," vol. 14, no. 4, pp. 1979–276, 2021, doi: 10.30998/faktorexacta.v14i4.11259.
- [8] E. Zuliarso, M. T. Anwar, K. Hadiono, and I. Chasanah, "Detecting Hoaxes in Indonesian News Using TF/TDM and K Nearest Neighbor," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 835, no. 1, 2020, doi: 10.1088/1757-

- 899X/835/1/012036.
- [9] I. F. of I. T. D. of I. Institut Teknologi 10 Nopember (Surabaya, Institute of Electrical and Electronics Engineers. Indonesia Section, and Institute of Electrical and Electronics Engineers, "ICTS 2017 : proceedings of the 11th International Conference on Information & Communication Technology and Systems (ICTS) 2017 : Department of Informatics, Faculty of Information and Communication Technology, Institut Teknologi Sepuluh Nopember (ITS), October 31st, 2017," pp. 73–78, 2017.
- [10] P. Y. Saputra, "Implementasi Teknik Crawling untuk Pengumpulan Data dari Media Sosial Twitter," *Din. Dotcom*, vol. 8, pp. 160–168, 2017.
- [11] Saruni Dwiasnati and Yudo Devianto, "Classification of forest fire areas using machine learning algorithm," *World J. Adv. Eng. Technol. Sci.*, vol. 3, no. 1, pp. 008–015, 2021, doi: 10.30574/wjaets.2021.3.1.0048.
- [12] F. A. Muttaqin and A. M. Bachtiar, "Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak 'Dodo Kids Browser,'" *J. Ilm. Komput. dan Inform.*, pp. 1–8, 2016.
- [13] C. Dhaneswara, Y. Azhar, and N. Hayatin, "Deteksi Berita Hoax Pada Dokumen Berbahasa Indonesia Menggunakan Metode Modified K - Nearest Neighbor," pp. 165–170, 2020.
- [14] H. D. Wijaya and S. Dwiasnati, "Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat," *J. Inform.*, vol. 7, no. 1, pp. 1–7, 2020, doi: 10.31311/ji.v7i1.6203.
- [15] Saruni Dwiasnati and Yudo Devianto, "Classification of forest fire areas using machine learning algorithm," *World J. Adv. Eng. Technol. Sci.*, vol. 3, no. 1, pp. 008–015, 2021, doi: 10.30574/wjaets.2021.3.1.0048.

## KERTAS KERJA

### Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul “Perbandingan Algoritma *Naive Bayes* dan *K Nearest Neighbor* Untuk Klasifikasi Berita *Hoax* Di Media Sosial Twitter”. Kertas kerja berisi semua material hasil penelitian Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan : literature review, dataset yang digunakan, source code dan hasil eksperimen secara keseluruhan.

