

**SKRIPSI**

**IMPLEMENTASI ALGORITMA BERNOULLI NAÏVE BAYES  
DALAM KLASIFIKASI *CYBERBULLYING* PADA MEDIA  
SOSIAL TWITTER**

**Disusun dan diajukan oleh:**

**NURUL UTAMI RUSLI  
D121181004**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
GOWA  
2024**

**LEMBAR PENGESAHAN SKRIPSI**

**IMPLEMENTASI ALGORITMA BERNOULLI NAÏVE BAYES  
DALAM KLASIFIKASI *CYBERBULLYING* PADA MEDIA SOSIAL  
TWITTER**

Disusun dan diajukan oleh

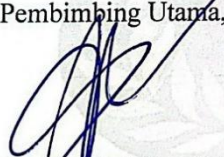
**Nurul Utami Rusli  
D121181004**


Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin Pada tanggal 5 April 2024 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,

  
Dr. Amil Ahmad Ilham, S.T., M.IT.  
NIP 19731010 199802 1 001

  
Anugrayani Bustamin, S.T., M.T.  
NIP 19901201 201807 4 001



Ketua Program Studi,  
Prof. Dr. Ir. Indrabayu, S.T., M.Bus.Sys., IPM, ASEAN. Eng  
NIP. 19750716 200212 1 004

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini ;  
Nama : Nurul Utami Rusli  
NIM : D121181004  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

{Implementasi Algoritma Bernoulli Naïve Bayes dalam Klasifikasi *Cyberbullying*  
pada Media Sosial Twitter}

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 5 April 2024

Yang Menyatakan



Nurul Utami Rusli

## ABSTRAK

**NURUL UTAMI RUSLI.** *Implementasi Algoritma Bernoulli Naïve Bayes dalam Klasifikasi Cyberbullying pada Media Sosial Twitter* (dibimbing oleh Amil Ahmad Ilham dan Anugrayani Bustamin)

Seiring peningkatan pengguna media sosial, fenomena *cyberbullying* juga semakin meningkat yang menimbulkan dampak negatif seperti gangguan mental bahkan bunuh diri. Pemerintah Indonesia bersama POLRI telah meresmikan beroperasinya polisi virtual untuk memantau dan memperingatkan pengguna yang melanggar UU ITE melalui unggahan teks atau gambar. Dari sejumlah akun yang mendapat peringatan, tercatat bahwa akun Twitter menyumbang sebanyak 79 akun yang mendapat teguran. Analisis *text mining* dengan metode klasifikasi teks, menggunakan algoritma Bernoulli Naïve Bayes, diusulkan sebagai solusi untuk meminimalisir *cyberbullying*.

Tujuan dari penelitian ini adalah untuk mengimplementasikan algoritma Bernoulli Naïve Bayes dalam melakukan klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter, serta untuk mengevaluasi algoritma tersebut.

Penelitian ini mengimplementasikan algoritma Bernoulli Naïve Bayes dan menggunakan *dataset* dari Twitter dengan menggunakan kata kunci "puan". Setelah tahap pelabelan manual, langkah berikutnya adalah melakukan tahap *preprocessing text* dengan tujuan untuk menghilangkan *noise* pada *dataset*. *Dataset* yang digunakan sebanyak 3.541 cuitan untuk kelas bukan *cyberbullying* dan 3.227 cuitan untuk kelas *cyberbullying*. Proses berikutnya yaitu melakukan ekstraksi fitur dengan metode *bag of words*. Selanjutnya, data dibagi menjadi 90% sebagai data latih dan 10% sebagai data uji.

Evaluasi model menggunakan *confusion matrix* menghasilkan tingkat *accuracy* sebesar 0.91, *recall* sebesar 0.96, *precision* sebesar 0.86, dan *f-measure* sebesar 0.91. Meskipun demikian, model belum mampu mengklasifikasikan makna kata secara pragmatik.

Kata Kunci: *Cyberbullying*, Twitter, Bernoulli Naïve Bayes, Klasifikasi Teks, *Bag of Words*.

## ABSTRACT

**NURUL UTAMI RUSLI.** *Implementation of the Bernoulli Naïve Bayes Algorithm in Cyberbullying Classification on Twitter Social Media* (supervised by Amil Ahmad Ilham and Anugrayani Bustamin)

As social media users increase, the phenomenon of cyberbullying is also increasing, which has negative impacts such as mental disorders and even suicide. The Indonesian government together with the POLRI have launched the operation of a virtual police force to monitor and warn users who violate the ITE Law by uploading text or images. Of the number of accounts that received warnings, it was recorded that Twitter accounts accounted for 79 accounts that received warnings. Text mining analysis with text classification methods, using the Bernoulli Naïve Bayes algorithm, is proposed as a solution to minimize cyberbullying.

The aim of this research is to implement the Bernoulli Naïve Bayes algorithm in classifying cyberbullying and non-cyberbullying tweets on Twitter social media, as well as to evaluate the algorithm.

This research implements the Bernoulli Naïve Bayes algorithm and uses a dataset from Twitter using the keyword "puan". After the manual labeling stage, the next step is to carry out the text preprocessing stage with the aim of removing noise in the dataset. The dataset used was 3.541 tweets for the non-cyberbullying class and 3.227 tweets for the cyberbullying class. The next process is to extract features using the bag of words method. Next, the data is divided into 90% as training data and 10% as test data.

Model evaluation using the confusion matrix produces an accuracy rate of 0.91, recall of 0.96, precision of 0.86, and f-measure of 0.91. However, the model is not yet able to classify the meaning of words pragmatically.

Keywords: Cyberbullying, Twitter, Bernoulli Naïve Bayes, Text Classification, Bag of Words.

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK.....	iii
ABSTRACT.....	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR.....	vii
DAFTAR TABEL.....	viii
DAFTAR SINGKATAN DAN ARTI SIMBOL.....	ix
DAFTAR LAMPIRAN.....	x
KATA PENGANTAR.....	xi
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Ruang Lingkup.....	3
BAB 2 TINJAUAN PUSTAKA.....	4
2.1 <i>Cyberbullying</i> .....	4
2.2 <i>Crawling</i> .....	6
2.3 <i>Text Mining</i> .....	6
2.4 <i>Twitter</i> .....	7
2.5 <i>Bag of Words</i> .....	8
2.6 <i>CountVectorizer</i> .....	9
2.7 <i>Preprocessing Text</i> .....	9
2.8 <i>Naïve Bayes</i> .....	11
2.9 <i>Bernoulli Naïve Bayes</i> .....	12
2.10 <i>Confusion Matrix</i> .....	14
2.11 <i>Klasifikasi Teks</i> .....	16
2.12 <i>Penelitian Terkait</i> .....	16
BAB 3 METODE PENELITIAN.....	19
3.1 Tahapan Penelitian.....	19
3.2 Waktu dan Lokasi Penelitian.....	20
3.3 Instrumen Penelitian.....	20
3.4 Rancangan Sistem.....	21
3.5 Pengumpulan Data.....	21
3.6 Pelabelan.....	22
3.7 Tahap <i>Preprocessing Text</i> .....	24
3.8 Ekstraksi Fitur.....	29
3.9 <i>Klasifikasi Bernoulli Naïve Bayes</i> .....	30
3.10 <i>Pengklasifikasian Data Uji menggunakan Bernoulli Naïve Bayes</i> .....	33
BAB 4 HASIL DAN PEMBAHASAN.....	39
4.1 Implementasi Algoritma <i>Bernoulli Naïve Bayes</i> .....	39
4.2 Evaluasi Kinerja Algoritma <i>Klasifikasi Bernoulli Naïve Bayes</i> .....	42
BAB 5 KESIMPULAN DAN SARAN.....	51
5.1 Kesimpulan.....	51

5.2 Saran.....	51
DAFTAR PUSTAKA .....	53
LAMPIRAN.....	55

## DAFTAR GAMBAR

Gambar 1. Naive bayes .....	11
Gambar 2. Bernoulli naive bayes .....	14
Gambar 3. Tahapan penelitian .....	19
Gambar 4. Rancangan sistem.....	21
Gambar 5. Grafik perbandingan data .....	28
Gambar 6. <i>Bag of words sample data</i> .....	29



**DAFTAR TABEL**

Tabel 1. <i>Confusion matrix</i> .....	15
Tabel 2. Twitter API .....	21
Tabel 3. Hasil <i>crawling</i> data .....	22
Tabel 4. Pelabelan .....	24
Tabel 5. <i>Cleaning</i> .....	25
Tabel 6. <i>Case folding</i> .....	25
Tabel 7. <i>Tokenization</i> .....	26
Tabel 8. <i>Stopword removal</i> .....	27
Tabel 9. <i>Stemming</i> .....	28
Tabel 10. Hasil <i>preprocessing text</i> .....	29
Tabel 11. <i>Sample</i> data .....	30
Tabel 12. Pembagian data latih dan data uji .....	40
Tabel 13. Ganjar dalam kelas <i>cyberbullying</i> .....	41
Tabel 14. Rakyat dalam kelas bukan <i>cyberbullying</i> .....	42
Tabel 15. <i>Confusion matrix</i> .....	43
Tabel 16. <i>False negative</i> .....	45
Tabel 17. <i>False positive</i> .....	46
Tabel 18. Aspek <i>false negative</i> .....	47
Tabel 19. <i>True positive</i> .....	49

## DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
SARA	Suku, Agama, Ras, dan Antargolongan
GIF	<i>Graphics interchange format</i>
URL	<i>Uniform Resource Locator</i>
ASCII	<i>American Standart Code for Information Interchange</i>
NLP	<i>Natural language processing</i>
NLTK	<i>Natural Language Toolkit</i>
API	<i>Application Programming Interface</i>
CSV	<i>Comma Separated Value</i>
RE	<i>Regular Expression</i>
POLRI	Kepolisian Negara Republik Indonesia
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
BoW	<i>Bag of Words</i>
Doc2Vec	<i>Document to Vector</i>
Word2Vec	<i>Word to Vector</i>
BPSO	<i>Binary Particle Swarm Optimization</i>
UU ITE	Undang-undang Informasi dan Transaksi Elektronik
SVM	<i>Support Vector Machines</i>

## DAFTAR LAMPIRAN

Lampiran 1. <i>Crawling data</i> .....	56
Lampiran 2. <i>Preprocessing data</i> .....	57
Lampiran 3. Bernoulli naive bayes .....	59
Lampiran 4. <i>False negative</i> .....	61
Lampiran 5. <i>True positive</i> .....	62
Lampiran 6. Berita acara seminar hasil.....	63
Lampiran 7. Berita acara ujian skripsi .....	67
Lampiran 8. Lembar perbaikan skripsi .....	71

## KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT karena berkat Rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi berjudul “Implementasi Algoritma Bernoulli Naïve Bayes dalam Klasifikasi *Cyberbullying* pada Media Sosial Twitter”. Dalam proses penyelesaian skripsi ini, penulis menyadari sepenuhnya bahwa penulis memperoleh banyak bantuan, bimbingan, serta dukungan dari berbagai pihak baik di masa perkuliahan maupun masa penyusunan skripsi ini. Oleh karena itu pada kesempatan ini penulis ingin berterima kasih kepada:

1. Kedua orang tua penulis, Bapak Rusli dan Ibu Ramasiah. R. yang selalu mendoakan, memfasilitasi dan memberikan dukungan kepada penulis dalam menyelesaikan perkuliahan.
  2. Bapak Dr. Amil Ahmad Ilham, S.T., M.IT. selaku dosen pembimbing utama yang telah menyempatkan waktu, tenaga, dan pikiran untuk mengarahkan penulis selama masa perkuliahan hingga penyusunan skripsi.
  3. Ibu Anugrayani Bustamin, S.T., M.T. selaku dosen pembimbing pendamping yang telah menyempatkan waktu, tenaga, dan pikiran untuk mengarahkan penulis selama masa perkuliahan hingga penyusunan skripsi.
  4. Bapak Prof. Dr. Ir. Indrabayu, S.T., M.Bus.Sys., IPM, ASEAN. Eng. selaku ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.
  5. Seluruh dosen dan staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas didikan dan bantuan kepada penulis selama masa perkuliahan.
  6. Synchronous18 sebagai teman seperjuangan atas segala bantuan dan momen bersama yang turut mewarnai masa perkuliahan.
  7. Semua pihak yang telah membantu penulis dari masa perkuliahan hingga dalam menyelesaikan skripsi yang tidak dapat disebutkan satu persatu.
- Akhir kata, penulis mohon maaf atas kekeliruan yang terdapat dalam skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis dan pembaca khususnya.

Gowa, 22 Desember 2023

Yang Menyatakan

Nurul Utami Rusli

# BAB 1 PENDAHULUAN

## 1.1 Latar Belakang

Pengguna media sosial di Indonesia pada Januari 2023 mencapai 167 juta pengguna atau setara dengan 60,4% dari total populasi Indonesia (Simon Kemp, 2023). Salah satu manfaat media sosial adalah sarana berdiskusi dan berekspresi. Namun, terkadang opini yang diungkapkan tidak diikuti dengan penerapan prinsip etika dalam berkomunikasi sehingga mengarah pada tindakan *cyberbullying* (Fortuna et al., 2022).

Sekitar 49% pengguna media sosial pernah menjadi sasaran *bullying*. Dampak buruk dari *cyberbullying* seperti gangguan mental bahkan bunuh diri. Jumlah korban *cyberbullying* yang mencoba melakukan bunuh diri hampir dua kali lebih banyak dibandingkan yang bukan korban *cyberbullying*. Kebebasan berekspresi di media sosial menjadikan seseorang lebih terbuka untuk mengungkapkan pendapatnya tanpa bertentangan pada norma sosial, seperti pada interaksi secara langsung. Terdapat lima faktor yang dapat menyebabkan terjadinya *cyberbullying*, yaitu: (a) dukungan sosial dari teman sebaya, (b) harga diri rendah, (c) dukungan sosial dari orang tua, (d) kesopanan dalam menggunakan media sosial, dan (e) Masalah dalam menggunakan internet. (Ristie Nashaya Faidatu' Nissa & Muhammad Ilmi Hatta, 2022).

Pemerintah Indonesia bekerja sama dengan POLRI tertanggal 24 Februari 2021 telah meresmikan beroperasinya polisi virtual yang bertugas untuk memantau dan memperingatkan pengguna yang mengunggah postingan baik berupa teks maupun gambar yang dapat melanggar pidana pada UU ITE. Direktorat *cyber* Bareskrim POLRI telah mencatat bahwa sebanyak 125 akun media sosial sudah mendapat teguran oleh polisi virtual. Dari akun-akun yang mendapat teguran tersebut, akun Twitter paling banyak menyumbang 79 akun. Angka ini tercatat pada periode 23 Februari – 11 Maret 2021 (Perwira et al., 2021).

Media sosial menjadi wadah terjadinya *cyberbullying*. *Cyberbullying* terjadi dengan cepat dan dapat terjadi pada siapa saja yang menggunakan media sosial (Filzah et al., 2022). *Cyberbullying* pada media sosial, seperti Twitter dilakukan

dengan cara menulis cuitan yang bersifat merugikan orang lain. Tindakan ini dapat berupa penggunaan kata-kata yang menyinggung, kasar, bahkan berpotensi SARA.

Berdasarkan uraian di atas, diperlukan suatu cara untuk meminimalisir *cyberbullying* di media sosial. Di masa lalu, fitur linguistik tradisional digunakan untuk mengidentifikasi *cyberbullying* yang membutuhkan banyak waktu dan tenaga. Namun saat ini, berbagai metode telah dikembangkan melalui pembelajaran mesin yang memungkinkan sistem untuk mengakses dan belajar dari data dengan cepat, sehingga dapat menyelesaikan masalah dengan lebih efisien (Filzah et al., 2022). Hal ini dapat dilakukan dengan menerapkan analisis *text mining* menggunakan metode klasifikasi teks. Klasifikasi tersebut bertujuan untuk membedakan antara komentar yang merupakan *cyberbullying* dan bukan *cyberbullying*. Penelitian ini bertujuan untuk mendeteksi *cyberbullying* di media sosial Twitter dengan menggunakan algoritma klasifikasi Bernoulli Naïve Bayes. Diharapkan penelitian ini dapat membantu instansi terkait dalam meminimalisir terjadinya *cyberbullying*.

Pada penelitian ini, *search key* yang digunakan dalam proses *crawling* data adalah "Puan". Pada tanggal 25 September 2022, Puan Maharani mengunggah foto bersama Muhaimin Iskandar yang sedang menyantap hidangan di warung pecel lele. Unggahan tersebut memicu beragam komentar dari pengguna Twitter, termasuk beberapa komentar yang bersifat *bully*. Oleh karena itu, nama "Puan" menjadi *trending topics* di Twitter.

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang maka dirumuskan permasalahan sebagai berikut:

1. Bagaimana implementasi algoritma Bernoulli Naïve Bayes dalam melakukan klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter?
2. Bagaimana akurasi algoritma Bernoulli Naïve Bayes dalam melakukan klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter?

### 1.3 Tujuan Penelitian

Dari permasalahan di atas, penelitian ini dilakukan dengan tujuan:

1. Untuk mengimplementasikan algoritma Bernoulli Naïve Bayes dalam melakukan klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter.
2. Untuk mengetahui akurasi algoritma Bernoulli Naïve Bayes dalam melakukan klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter.

### 1.4 Manfaat Penelitian

Manfaat yang diperoleh dalam penelitian ini adalah diharapkan sistem ini dapat membantu instansi terkait dalam mengidentifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter.

### 1.5 Ruang Lingkup

Agar penelitian lebih spesifik dan terarah, ruang lingkup yang diterapkan pada penelitian ini adalah:

1. *Dataset* yang digunakan berbahasa Indonesia.
2. Pengumpulan *dataset* menggunakan *Application Programming Interface* (API) Twitter.
3. *Dataset* diperoleh dengan cara *crawling*.
4. *Search key* yang digunakan dalam proses *crawling* data yaitu Puan.
5. *Dataset* yang dikumpulkan sebanyak 12.075 data cuitan.
6. Sistem ini dibangun menggunakan bahasa pemrograman Python.

## BAB 2 TINJAUAN PUSTAKA

### 2.1 *Cyberbullying*

*Cyberbullying* merupakan perilaku yang merugikan orang lain dengan mengirim atau mengunggah konten berbahaya serta melakukan agresi sosial menggunakan internet atau teknologi digital lainnya (Willard, 2007). Willard menjelaskan aspek-aspek *cyberbullying* antara lain:

1. *Flaming*

*Flaming* adalah serangan yang lebih pribadi terhadap seseorang, biasanya dilakukan di lingkungan sosial, seperti grup media sosial atau forum obrolan. *Flaming* biasanya ditandai dengan penggunaan bahasa yang tidak senonoh (*The 10 Types of Cyberbullying*, 2019). *Flaming* berasal dari kata *flame*, Istilah *flame* ini pun merujuk pada kata-kata yang berapi-api (Willard, 2007). Contoh kalimat *flaming* adalah "Kamu bodoh sekali! Pendapatmu sampah!". Kalimat ini merupakan *flaming* karena kalimat ini menggunakan bahasa yang tidak senonoh.

2. *Harassment*

*Harassment* adalah perilaku berulang kali mengirim pesan jahat dan kasar kepada orang lain. *Harassment* merupakan hasil dari tindakan *flaming* dalam jangka panjang. *Harassment* dilakukan dengan saling berbalas pesan atau bisa disebut perang teks (Alvian, 2020). Contoh kalimat *harassment* adalah "Kamu memang tidak berguna, tidak ada yang suka denganmu, dan kamu seharusnya tidak pernah dilahirkan".

3. *Denigration*

*Denigration* adalah tindakan menyebarkan gosip atau fitnah secara *online* tentang seseorang untuk merusak reputasi, nama baik, dan persahabatan orang tersebut. Pelesetan nama seseorang atau penggunaan nama palsu untuk mengejek atau mencemarkan nama baik seseorang dapat dikategorikan sebagai bentuk *denigration* (Jalal et al., n.d.). Contoh kalimat *denigration* adalah "Dia bahagia dengan uang hasil korupsi suaminya!". Kalimat ini



merupakan *denigration* karena kalimat ini dapat merusak reputasi dan nama baik dari individu yang disebutkan.

#### 4. *Impersonation*

*Impersonation* adalah tindakan berpura-pura menjadi orang lain dengan menyebarkan informasi tentang orang yang ditiru sehingga membahayakan orang tersebut (Willard, 2007). Contoh kalimat *impersonation* adalah "Halo, ini saya (nama orang lain). Saya dari bank (nama bank). Kami mendeteksi aktivitas mencurigakan di akun Anda. Silahkan hubungi kami segera". Kalimat ini adalah kalimat *impersonation* karena menyamar sebagai orang lain untuk mendapatkan keuntungan dan dapat merugikan orang lain.

#### 5. *Outing*

*Outing* adalah tindakan menyebarkan rahasia orang lain secara *online*. Rahasia yang disebarkan dapat berupa foto-foto pribadi orang lain dengan maksud mengumbar keburukan atau privasi orang tersebut. Perbedaan *outing* dengan *denigration* adalah terletak pada jenis objek medianya. *Outing* lebih menggunakan pada foto-foto dan video pribadi, sedangkan *denigration* lebih pada pendeskripsian melalui tulisan. Akan tetapi, tujuannya adalah sama-sama menjatuhkan harga diri seseorang. Contoh perilaku *outing* adalah kasus penyebaran video skandal Ariel dan Luna Maya (Wahyuningtyas et al., 2019).

#### 6. *Trickery*

*Trickery* adalah membujuk seseorang dengan tipu daya agar mendapatkan rahasia atau foto pribadi orang tersebut yang bertujuan untuk disebarkan secara *online* (Willard, 2007). Contoh kalimat *trickery* adalah "Kamu boleh mengirimkan foto-foto pribadi mu, aku tidak akan menyebarkannya". Kalimat ini merupakan kalimat *trickery* karena membujuk seseorang dengan tipu daya untuk mendapatkan rahasia orang tersebut.

#### 7. *Exclusion*

*Exclusion* adalah tindakan tidak menganggap seseorang pada grup *online* dengan sengaja, seperti secara sengaja mengeluarkan seseorang dari grup *online* (Willard, 2007). Contoh kalimat *exclusion* adalah "Kamu tidak diperbolehkan untuk masuk grup *chat* kami karena kamu dari ras X". Kalimat

ini merupakan *exclusion* karena mengecualikan seseorang dari suatu kelompok atau kegiatan secara *online*.

#### 8. *Cyberstalking*

*Cyberstalking* adalah perilaku berulang mengirimkan ancaman membahayakan atau pesan-pesan yang mengintimidasi melalui media elektronik yang dapat menyebabkan rasa takut pada korban (Willard, 2007). Contoh kalimat *cyberstalking* adalah "Aku akan mengikuti kamu kemanapun kamu pergi". Kalimat tersebut dikategorikan sebagai *cyberstalking* karena kalimat tersebut dapat menyebabkan rasa takut pada korban.

## 2.2 *Crawling*

*Crawling* merupakan kegiatan pengambilan informasi atau data dari *World Wide Web* menggunakan perangkat lunak komputer, dan hasilnya dapat disimpan dalam *database*. Data yang diperoleh dari proses *crawling* dapat berupa teks atau gambar. Dan juga *crawling* dapat dilakukan pada tulisan di media sosial seperti *tweet* atau *cuitan*. Proses *crawling*, secara khusus merujuk pada pengambilan data dalam bentuk teks melalui API. Twitter memberikan izin kepada penggunanya untuk mengakses data melalui *endpoint* yang disediakan, data yang bisa diakses berupa *tweet, users, spaces, direct message, list, trends, media, places* (Badan Sistem Informasi UII).

## 2.3 *Text Mining*

*Text mining* didefinisikan sebagai proses untuk mendapatkan informasi tersembunyi dari data teks dalam jumlah yang besar. *Text mining* merupakan bagian dari *information retrieval*, yang merupakan ilmu yang mempelajari metode untuk mencari informasi yang relevan. Secara umum, algoritma yang digunakan dalam *text mining* memiliki kinerja yang serupa dengan algoritma dalam *data mining*. Perbedaan utama terletak pada tipe data yang digunakan. *Data mining* bekerja pada data terstruktur (*structured*), sedangkan *text mining* bekerja pada data tidak terstruktur (*unstructured*). Secara umum terdapat tiga bagian utama dari *text mining* yaitu *text preprocessing, feature selection, dan text analytic* (Nugraha et al., n.d.).

## 2.4 Twitter

Twitter adalah layanan jejaring sosial atau dikenal mikroblog daring yang pertama kali digagas oleh Jack Dorsey pada tahun 2006. Pada awal tahun 2023, jumlah pengguna Twitter di Indonesia mencapai 24 juta pengguna, yang setara dengan 8,7% dari total populasi (Simon Kemp, 2023). Seiring berkembangnya, aplikasi ini telah menghadirkan fitur tambahan, antara lain:

1. *Tweet*

*Tweet* atau cuitan merupakan fitur utama yang digunakan oleh pengguna untuk membagikan tulisan, foto, GIF, atau video secara publik di Twitter. Maksimum panjang tulisan dalam sebuah cuitan adalah 280 karakter.

2. *Follow, Followers, dan Unfollow*

Pengguna Twitter menggunakan istilah *follow* untuk mengikuti pengguna lain. Sementara itu, istilah *followers* merujuk pada pengguna yang mengikuti pengguna tersebut. Jika pengguna memutuskan untuk tidak mengikuti lagi pengguna tersebut, disebut dengan istilah *unfollow*.

3. *Hashtag* atau Tagar

*Hashtag* atau tagar adalah kata yang diawali dengan tanda tagar ('#'). Dalam aplikasi ini, simbol tagar menjadi penting karena digunakan untuk memulai topik pembicaraan tertentu yang sedang berlangsung. Fitur ini bermanfaat dalam mengidentifikasi topik khusus yang sedang menjadi perbincangan hangat.

4. *Username* atau Nama Pengguna

Nama pengguna atau *username* adalah fitur yang digunakan untuk mengidentifikasi suatu akun dan membedakan satu akun dengan akun lainnya. Pada media sosial Twitter, nama akun biasanya diawali dengan *at* ('@').

5. *Retweet*

*Retweet* adalah fitur yang memungkinkan pengguna untuk membagikan kembali unggahan dari pengguna lain atau dari unggahan mereka sendiri.

6. *Trending Topics*

*Trending topics* adalah topik-topik yang sedang populer dan banyak dibicarakan oleh pengguna lain dalam aplikasi tersebut, berdasarkan tagar,

kata, atau frasa tertentu. Topik-topik populer ini ditampilkan secara berurutan berdasarkan jumlah pembicaraan yang terkait dengan topik tersebut.

#### 7. URL yang Dipersingkat

URL yang dipersingkat adalah ketika pengguna membagikan suatu unggahan, Twitter secara otomatis mempersingkat tautan URL tersebut. Hal ini memungkinkan pengguna untuk tidak perlu khawatir mengenai panjang URL saat membuat unggahan yang berisi tautan.

#### 8. Tiga Mode Tampilan

Tiga mode tampilan memungkinkan pengguna Twitter untuk mengubah tampilan dengan tiga mode pilihan warna yang tersedia, yaitu terang, temaram, dan lampu mati. Selain itu, pengguna juga dapat mengubah gaya, ukuran, dan warna *font*.

#### 9. *Swipe*

*Swipe* memungkinkan pengguna untuk menggeser atau *swipe* pada lini masa untuk beralih ke tampilan-tampilan lainnya, seperti cuitan yang direkomendasikan dan diikuti, *trending topics*, dan lain sebagainya.

Twitter menyediakan akses data kepada perusahaan, pengembang dan pengguna API. Salah satu API yang disediakan oleh twitter adalah *search API*, yang memungkinkan perusahaan, pengembang, atau pengguna untuk mendapatkan akses ke cuitan yang telah diunggah ke media sosial Twitter (*About Twitter's APIs*, 2023).

## 2.5 *Bag of Words*

Algoritma *machine learning* tidak dapat bekerja secara langsung dengan teks mentah. Jadi sebelum menerapkan algoritma terlebih dahulu dilakukan ekstraksi fitur dari data teks. Metode yang umum digunakan dalam pemrosesan bahasa alami adalah *bag of words*. *Bag of words* dapat diilustrasikan sebagai tas yang berisi kumpulan kata-kata unik dari seluruh dokumen, di mana posisi kata-kata tidak diperhatikan. Jumlah total kolom dalam *matrix* yang dihasilkan sama dengan jumlah kata unik yang terdapat di seluruh dokumen (George, 2022).

Ada dua jenis hitungan dalam pendekatan *bag of words*, yaitu biner dan absolut. Dalam hitungan biner, setiap kata dalam vektor fitur diberi nilai satu jika

kata tersebut muncul dalam dokumen dan nilai nol jika kata tersebut tidak muncul dalam dokumen. Dengan kata lain, hitungan ini hanya memperhitungkan keberadaan atau ketidakhadiran kata dalam dokumen. Sementara itu, dalam hitungan absolut, nilai yang terkait dengan setiap kata adalah seberapa sering kata tersebut muncul di setiap dokumen. Dalam hal ini, metode *bag of words* menghitung frekuensi kata-kata tersebut dalam dokumen (Raschka, 2014).

Pilihan antara biner dan absolut tergantung pada model probabilitas yang digunakan untuk pengklasifikasian. Kedua hitungan dapat digunakan dalam konteks klasifikasi teks, tetapi pilihan tergantung pada kebutuhan dan tujuan spesifik dari tugas pemrosesan bahasa alami yang sedang dijalankan (Raschka, 2014).

## 2.6 *CountVectorizer*

*CountVectorizer* merupakan metode implementasi dari konsep *Bag of words* dalam pemrosesan teks. *CountVectorizer* digunakan untuk mengubah teks menjadi vektor berdasarkan kemunculan setiap kata dari keseluruhan dokumen. Vektor tersebut dibangun berdasarkan dimensi ukuran dari keseluruhan kata unik, yang jumlahnya bertambah setiap kali sebuah kata ditambahkan, sehingga dimensinya juga bertambah. Dengan demikian, terbentuklah istilah dokumen matriks (Turki & Roy, 2022).

Dalam penggunaannya, *CountVectorizer* menghasilkan matriks di mana setiap kata unik dalam dokumen yang direpresentasikan sebagai kolom dalam tabel, sementara setiap dokumen direpresentasikan sebagai baris dalam tabel tersebut. Kata-kata dalam kolom disusun berdasarkan abjad, dan nilai dari setiap sel dalam matriks menunjukkan kemunculan kata dalam dokumen. Dalam konteks klasifikasi dokumen, *count vectorizer* pada dasarnya adalah model representasi biner yang ditandai dengan parameter '*binary = true*' (Turki & Roy, 2022).

## 2.7 *Preprocessing Text*

*Preprocessing text* adalah proses pengolahan data yang belum terstruktur menjadi terstruktur sesuai dengan kebutuhan untuk proses *mining* lebih lanjut. Pada tahap ini, data dibersihkan dengan menghapus *noise* dan karakter *non-ASCII*,

sehingga data menjadi siap olah (Perkins et al., 2016). Berikut adalah langkah-langkah dalam *preprocessing text*.

1. *Cleaning*

*Cleaning* adalah proses yang dilakukan untuk membersihkan teks seperti penghapusan URL, karakter *non-ASCII*, angka, simbol, tanda baca, *hashtag*, *username*, dan *retweet* untuk mengurangi *noise* saat proses klasifikasi (Perkins et al., 2016).

2. *Case Folding*

*Case folding* adalah tahap menyeragamkan semua bentuk huruf menjadi bentuk yang sama, misalnya berupa huruf kapital atau huruf kecil (Efrizoni et al., 2022).

3. *Tokenization*

*Tokenization* adalah proses pemisahan atau pemecahan *string* menjadi token atau kata-kata dengan memisahkannya menggunakan spasi. Token adalah unit terkecil yang dapat dipahami dan diproses oleh mesin. Oleh karena itu, *string* teks tidak dapat diproses, tanpa melalui proses *tokenization* (Perkins et al., 2016).

4. *Stopword Removal*

*Stopword removal* adalah menghapus kata-kata yang umum muncul dalam dokumen. Biasanya, kata-kata ini termasuk dalam kategori kata berhenti. Kata-kata tersebut tidak memiliki arti yang signifikan dalam beberapa tugas NLP seperti klasifikasi teks. *Library* NLTK menyediakan daftar *stopword* untuk berbagai bahasa (Perkins et al., 2016).

5. *Stemming*

*Stemming* adalah tahap untuk mengubah kata-kata menjadi bentuk dasar tanpa mengubah maknanya. *Stemming* dilakukan dengan menghilangkan imbuhan, awalan, sisipan, akhiran dan kombinasi awalan dan akhiran pada sebuah kata tanpa mengubah makna kata. Tujuan dari *stemming* adalah untuk mengurangi variasi kata dalam dokumen. Pada penelitian ini, tahap *stemming* menggunakan bantuan *library* Stemmer Porter (Perkins et al., 2016).

## 2.8 Naïve Bayes

Salah satu contoh dari algoritma *supervised learning* adalah Naïve Bayes yang merupakan algoritma berdasarkan probabilitas (Manning et al., 2008). Naive Bayes adalah algoritma klasifikasi berdasarkan teorema Bayes yang biasa digunakan dalam klasifikasi teks. Dikatakan *naïve* karena algoritma ini mengasumsikan semua fitur tidak bergantung pada fitur-fitur lainnya. Dengan kata lain algoritma ini mengasumsikan bahwa ada atau tidaknya *term* tertentu dari kelas tidak berhubungan dengan ada atau tidaknya *term* yang lain. Algoritma Naïve Bayes menggunakan keseluruhan probabilitas, yaitu probabilitas dokumen terhadap kelas yang disebut *prior*. Kemudian teks akan terkategori berdasarkan probabilitas maksimum yaitu *posterior* (George, 2022). Variasi dari algoritma Naive Bayes yang digunakan dalam klasifikasi teks yaitu Multinomial Naive Bayes. Implementasi dari algoritma Naïve Bayes digambarkan pada Gambar 1.

```

TRAINMULTINOMIALNB(C, D)
1  V ← EXTRACTVOCABULARY(D)
2  N ← COUNTDOCS(D)
3  for each c ∈ C
4  do Nc ← COUNTDOCSINCLASS(D, c)
5     prior[c] ← Nc / N
6     textc ← CONCATENATETEXTOFALLDOCSINCLASS(D, c)
7     for each t ∈ V
8     do Tct ← COUNTTOKENSOFTERM(textc, t)
9     for each t ∈ V
10    do condprob[t][c] ←  $\frac{T_{ct}+1}{\sum_{t'}(T_{ct'}+1)}$ 
11  return V, prior, condprob

APPLYMULTINOMIALNB(C, V, prior, condprob, d)
1  W ← EXTRACTTOKENSFROMDOC(V, d)
2  for each c ∈ C
3  do score[c] ← log prior[c]
4     for each t ∈ W
5     do score[c] += log condprob[t][c]
6  return arg maxc∈C score[c]

```

Gambar 1. Naive bayes

Sumber: Manning et al., (2008)

Algoritma Naïve Bayes menghitung *conditional probability* untuk setiap *term* berdasarkan frekuensi kemunculan dalam setiap kelas untuk keseluruhan dokumen seperti pada tahap Train baris ke-10 di Gambar 1. Selanjutnya, pada tahap *apply* pada baris ke-5 di Gambar 1, menghitung skor dengan menambahkan log *conditional probability* dari setiap *term* dalam dokumen. Pendekatan ini

menghitung skor berdasarkan *conditional probability* yang telah dihitung selama pelatihan, mengasumsikan bahwa setiap *term* dalam data latih memberikan kontribusi terhadap kelasnya. Sedangkan, setiap *term* yang tidak terdapat dalam data latih, tidak mempengaruhi hasil dari klasifikasi (Manning et al., 2008).

## 2.9 Bernoulli Naïve Bayes

Bernoulli Naïve Bayes merupakan salah satu variasi dari algoritma Naïve Bayes. Bernoulli Naïve Bayes direpresentasikan dengan nilai biner untuk setiap *term* dalam dokumen, yaitu nol yang menunjukkan tidak adanya *term* dalam dokumen dan nilai satu yang menunjukkan adanya *term* dalam dokumen. Algoritma ini hanya memperhatikan kemunculan kata pada dokumen dan mengabaikan frekuensi kemunculan kata (Manning et al., 2008). Representasi dokumen pada Bernoulli Naïve Bayes ditunjukkan pada Formula 1.

$$d = \langle e_1, \dots, e_i, \dots, e_M \rangle; e_i \in \{0,1\} \quad (1)$$

di mana,

- $d$  = kumpulan nilai biner yang menunjukkan untuk setiap *term* itu muncul atau tidak dalam dokumen,
- $M$  = banyaknya kata.

Probabilitas dokumen  $d$  terhadap kelas  $c$  digambarkan pada Formula 2.

$$P(c|d) \propto P(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} 1 - P(t_i|c) \quad (2)$$

di mana,

- $P(t_i|c)$  = *conditional probability* dari *term*  $t_i$  dalam dokumen kelas  $c$ . *Conditional probability* adalah ukuran seberapa banyak kontribusi bahwa *term*  $t_i$  di dalam dokumen kelas  $c$  yang benar,

$$P(c) = \text{prior probability dari dokumen yang terjadi di kelas } c.$$

Tujuan dalam klasifikasi teks adalah untuk mengidentifikasi kelas yang paling sesuai dengan dokumen. Kelas yang paling sesuai dengan dokumen disebut



sebagai kelas yang memiliki kemungkinan (*likely*) tertinggi atau disebut juga sebagai kelas *Maximum A Posteriori* (MAP) dengan  $c$  adalah himpunan {positif, negatif}.

$$c_{map} = \operatorname{argmax}_{c \in \mathcal{C}} \hat{P}(c) \prod_{\substack{1 \leq i \leq M \\ e_i=1}} P(t_i|c) \prod_{\substack{1 \leq i \leq M \\ e_i=0}} 1 - P(t_i|c) \quad (3)$$

Untuk menghitung  $\hat{P}(c)$  digunakan Formula 4.

$$\hat{P}(c) = \frac{N_c}{N} \quad (4)$$

di mana,

- $N_c$  = jumlah dokumen pada kelas  $c$ ,
- $N$  = jumlah total dari dokumen.

Untuk menghitung nilai  $\hat{P}(t_i|c)$  Bernoulli Naïve Bayes menerapkan *add-one* atau *laplace smoothing* untuk menghindari probabilitas nol. Probabilitas nol diakibatkan karena adanya kata pada data uji yang tidak terdapat pada data latih. Penerapan *laplace smoothing* pada baris ke-8 pada Gambar 2 pada implementasi data *train* dianalogikan dengan Formula 5.

$$\hat{P}(t_i|c) = \frac{N_{ct} + 1}{N_c + 2} \quad (5)$$

di mana,

- $N_{ct}$  = jumlah total kemunculan dari *term*  $t$  dari kelas  $c$  pada data latih, Misalnya, jika sebuah *term* muncul dua kali dalam sebuah dokumen, dalam posisi  $t_1$  dan  $t_2$ , jadi  $\hat{P}(t_{i1}|c) = \hat{P}(t_{i2}|c)$ ,
- $N_c$  = jumlah dokumen pada kelas  $c$ ,
- 1 = Nilai dari *laplace smoothing*,
- 2 = karena *constant* untuk algoritma Bernoulli Naïve Bayes adalah 2, karena ada dua kasus yang perlu dipertimbangkan untuk setiap *term*, yaitu kejadian dan tidak terjadinya.

Menuliskan  $\hat{P}$  untuk  $P$  pada formula di atas karena belum diketahui nilai sebenarnya dari parameter  $P(c)$  dan  $P(t_i|c)$ . Perhitungan  $\hat{P}(t|c)$  untuk kelas  $c$  yang mengandung *term* diibaratkan pada Gambar 2 *train* Bernoulli Naïve Bayes baris

ke-8. Dalam mengklasifikasikan data uji, Bernoulli Naïve Bayes menggunakan representasi biner dan mengabaikan jumlah kejadian. Oleh karena itu, Bernoulli Naïve Bayes biasanya membuat kesalahan ketika mengklasifikasikan dokumen yang panjang. Misalnya, algoritma Bernoulli Naïve Bayes mengklasifikasikan suatu cuitan ke dalam kelas positif karena satu kemunculan *term* tersebut di kelas positif. Dalam Bernoulli Naïve Bayes, probabilitas tidak terjadinya kejadian diperhitungkan saat menghitung  $P(c|d)$  seperti pada implementasi *apply* Bernoulli Naïve Bayes baris ke-7. Hal ini karena model Bernoulli Naïve Bayes memodelkan tidak adanya istilah secara eksplit (Manning et al., 2008).

Implementasi dari algoritma Bernoulli Naïve Bayes (Manning et al., 2008) dipaparkan dalam Gambar 2.

```

TRAINBERNOULLINB(C, D)
1  V ← EXTRACTVOCABULARY(D)
2  N ← COUNTDOCS(D)
3  for each c ∈ C
4  do Nc ← COUNTDOCSINCLASS(D, c)
5     prior[c] ← Nc/N
6     for each t ∈ V
7     do Nct ← COUNTDOCSINCLASSCONTAININGTERM(D, c, t)
8        condprob[t][c] ← (Nct + 1)/(Nc + 2)
9  return V, prior, condprob

APPLYBERNOULLINB(C, V, prior, condprob, d)
1  Vd ← EXTRACTTERMSFROMDOC(V, d)
2  for each c ∈ C
3  do score[c] ← log prior[c]
4     for each t ∈ V
5     do if t ∈ Vd
6        then score[c] += log condprob[t][c]
7        else score[c] += log(1 - condprob[t][c])
8  return arg maxc∈C score[c]

```

Gambar 2. Bernoulli naive bayes

Sumber: Manning et al., (2008)

## 2.10 Confusion Matrix

*Confusion matrix* adalah metode yang membandingkan akurasi klasifikasi data menjadi kategori positif dan negatif. *Confusion matrix* memvisualisasikan jumlah data uji yang diklasifikasikan dengan benar dan yang diklasifikasikan dengan salah berdasarkan kategori aktual dan prediksi (George, 2022). Tabel *Confusion matrix* ditunjukkan pada Tabel 1.

Tabel 1. *Confusion matrix*

Kelas Aktual	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Berikut merupakan penjelasan dari masing-masing nilai pada *confusion matrix*:

1. TP (*True Positive*) adalah jumlah kelas positif yang diklasifikasikan sebagai kelas positif.
2. FP (*False Positive*) adalah jumlah kelas negatif yang diklasifikasikan sebagai kelas positif.
3. FN (*False Negative*) adalah jumlah kelas positif yang diklasifikasikan sebagai kelas negatif.
4. TN (*True Negative*) adalah jumlah kelas negatif yang diklasifikasikan sebagai kelas negatif.

Berdasarkan pada nilai-nilai yang merupakan representasi kinerja klasifikasi tersebut dapat dihitung:

#### 1. *Accuracy*

*Accuracy* adalah rasio antara jumlah prediksi yang benar dengan keseluruhan jumlah data. Untuk menghitung *accuracy*, digunakan Formula 6.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

#### 2. *Precision*

*Precision* adalah rasio antara jumlah prediksi positif yang benar dengan total prediksi positif. Untuk menghitung *Precision*, digunakan Formula 7.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

#### 3. *Recall*

*Recall* adalah rasio antara jumlah prediksi benar positif dibandingkan dengan jumlah keseluruhan data yang benar positif. *Recall* juga dikenal sebagai *Sensitivity*. Untuk menghitung *recall*, digunakan Formula 8.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

#### 4. *F-measure*

*F-measure* adalah rata-rata harmonik antara *precision* dan *recall*. Nilai maksimum dari ukuran ini adalah 1. Untuk menghitung *F-measure*, digunakan Formula 9.

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (9)$$

## 2.11 Klasifikasi Teks

Dalam pembelajaran mesin, klasifikasi teks dipelajari secara otomatis dari data pelatihan. Kelas juga bisa disebut sebagai topik, kategori atau label. Klasifikasi teks juga disebut kategorisasi teks, klasifikasi topik, atau pencarian topik. Dalam klasifikasi teks, memerlukan sejumlah data latih untuk setiap kelas. Dengan menggunakan data latih, pembelajaran mesin mampu membangun pengetahuan berdasarkan data tersebut. Proses pelabelan secara manual tetap dilakukan untuk data latih, di mana pelabelan mengacu pada proses memberi anotasi kelas pada setiap data. Pembelajaran seperti ini disebut pembelajaran yang diawasi atau *supervised learning*. Disebut *supervised learning* karena diibaratkan manusia yang mendefinisikan kelas pada dokumen pelatihan yang berfungsi sebagai guru yang mengarahkan proses pembelajaran (Manning et al., 2008).

## 2.12 Penelitian Terkait

Penelitian dengan judul "*Comparison of Feature Extraction in Multilabel Text Classification Using Machine Learning Algorithm*". Penelitian ini membandingkan kinerja ekstraksi fitur seperti BoW, TF-IDF, Doc2Vec dan Word2Vec dalam melakukan klasifikasi teks dengan menerapkan algoritma *machine learning* seperti Naive Bayes, Support Vector Machines, Decision Tree, K-Nearest Neighbors, Random Forest, Logistic Regression. *Dataset* yang digunakan terdiri dari 1000 *sample* yang diambil dari sumber *tribunnews.com*, dengan pembagian data sebanyak 50:50, 70:30, 80:20 dan 90:10. Hasil dari penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi tertinggi menggunakan ekstraksi fitur TF-IDF sebesar 87% dan BoW sebesar 83%. Sementara itu, ekstraksi fitur Doc2Vec mencapai akurasi tertinggi pada algoritma

SVM sekitar 81%. Sedangkan ekstraksi fitur Word2Vec dengan algoritma *machine learning* menunjukkan akurasi model di bawah 50% (Efrizoni et al., 2022).

Penelitian berjudul "*Machine Learning Approach on Cyberstalking Detection in Social Media Using Naive Bayes and Decision Tree*" bertujuan untuk mengklasifikasikan perilaku *cyberstalking* dan *cyberbullying*. Dalam penelitian ini, algoritma yang digunakan yaitu Multinomial Naïve Bayes dan Decision Tree, dengan ekstraksi fitur menggunakan BoW. Hasil penelitian menunjukkan Naïve Bayes mampu mengidentifikasi *dataset* dengan tingkat akurasi sebesar 0.957953, sementara Decision Tree mampu mengklasifikasikan *dataset* dengan akurasi 0.943812. Dalam hal *precision*, Naïve Bayes mencapai 0,819484, sedangkan dan Decision Tree memiliki nilai 0,597518. *Recall* untuk Naïve Bayes dan Decision Tree berturut-turut adalah 0,511628 dan 0,602862. Selanjutnya, *f1-score* dari Naïve Bayes mencapai 0.629956 sedangkan Decision Tree memiliki 0.600178 (Filzah et al., 2022).

Penelitian ketiga berjudul "*Penerapan Binary Particle Swarm Optimization Support Vector Machine untuk Klasifikasi Komentar Cyberbullying di Instagram*" bertujuan untuk membandingkan kinerja model klasifikasi antara kondisi tanpa seleksi fitur dan dengan seleksi fitur. Data yang digunakan berasal dari hasil *scraping* komentar pada unggahan Instagram beberapa akun Instagram yang pernah mengalami *cyberbullying* seperti @rahmawatikekeyiputricantikka23, @cimoyluv, @viavallen, @riaricis1795, dan @aurelie.hermansyah. *Dataset* terdiri dari 7.866 komentar yang diambil dari 9 unggahan. Pemodelan klasifikasi dilakukan dalam dua kondisi, yaitu tanpa seleksi fitur dan dengan seleksi fitur menggunakan BPSO. Algoritma yang digunakan adalah SVM, dengan ekstraksi fitur menggunakan metode TF-IDF. Hasil penelitian menunjukkan bahwa pemodelan klasifikasi *cyberbullying* dengan SVM menghasilkan performa klasifikasi yang cukup akurat, melebihi 72% untuk semua parameter klasifikasi pada setiap nilai C. Penerapan seleksi fitur BPSO mampu meningkatkan performa klasifikasi dengan peningkatan nilai akurasi dan spesifisitas. Namun, dalam penelitian ini, model tanpa seleksi fitur pada nilai C = 0,1 terpilih dengan nilai akurasi sebesar 82,6%, sensitivitas 83,7%, dan spesifisitas 82,5% (Fortuna et al., 2022).

Penelitian berjudul "Penerapan Bernoulli Naïve Bayes Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan *Online Food Delivery* Di Indonesia" menggunakan data dari Twitter sebanyak 1.920 cuitan dengan kata kunci "*online food delivery*", "*shopeefood*", "*gofood*", dan "*grabfood*". Pada Penelitian ini, dilakukan seleksi fitur menggunakan *information gain* dan membangun model menggunakan algoritma Bernoulli Naive Bayes. Hasil penelitian menunjukkan bahwa akurasi yang didapatkan sebesar 89%, 87%, 86%, dan 85% untuk seluruh data serta data khusus *GoFood*, *ShopeeFood*, dan *GrabFood*. Evaluasi pada *confusion matrix* seluruh data, yaitu 1.179 cuitan, menunjukkan tingkat ketepatan hasil klasifikasi yang benar positif 98% dan benar negatif 70%. Hasil ini menandakan bahwa model telah cukup baik dalam melakukan prediksi *tweet* dengan tingkat kesalahan klasifikasi yang rendah (Fisyahri, 2023).

Berdasarkan penelitian-penelitian sebelumnya mengenai *text mining* dalam klasifikasi teks, yang menunjukkan bahwa implementasi algoritma Naive Bayes dengan metode ekstraksi fitur *bag of words* menghasilkan nilai akurasi yang baik. Kemudian penelitian lain juga menyimpulkan bahwa implementasi algoritma Bernoulli Naïve Bayes menghasilkan akurasi yang baik dalam melakukan klasifikasi cuitan dengan tingkat kesalahan yang rendah. Oleh karena itu, penelitian ini mengimplementasikan algoritma Bernoulli Naïve Bayes dalam klasifikasi cuitan *cyberbullying* dan bukan *cyberbullying* pada media sosial Twitter dengan menggunakan metode ekstraksi fitur *bag of words*.