

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



# KLASIFIKASI SENTIMEN TRANSFORMASI DAN REFORMASI SEPAK BOLA PADA TWITTER MENGGUNAKAN ALGORITMA *BERNOULLI NAÏVE BAYES*

## TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Teknik  
Pada Jurusan Teknik Informatika

Oleh

**DESTRI PUTRI YANI**

**NIM. 1950125030**



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2023

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**LEMBAR PERSETUJUAN****KLASIFIKASI SENTIMEN TRANSFORMASI DAN  
REFORMASI SEPAK BOLA INDONESIA PADA TWITTER  
MENGUNAKAN ALGORITMA BERNOULLI NAÏVE BAYES****TUGAS AKHIR**

Oleh

**DESTRI PUTRI YANI****NIM. 11950125030**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir  
di Pekanbaru, pada tanggal 8 Juni 2023

Pembimbing

**Siska Kurnia Gusti, S.T., M.Sc.****NIP. 198610092022032001**



## LEMBAR PENGESAHAN

### KLASIFIKASI SENTIMEN TRANSFORMASI DAN REFORMASI SEPAK BOLA INDONESIA PADA TWITTER MENGUNAKAN ALGORITMA BERNOULLI NAÏVE BAYES

Oleh

**DESTRI PUTRI YANI**

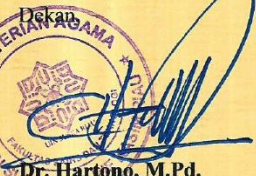
**NIM. 11950125030**

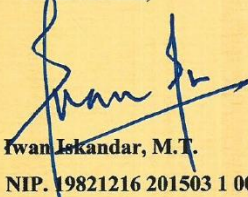
Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik  
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 8 Juni 2023

Mengesahkan,

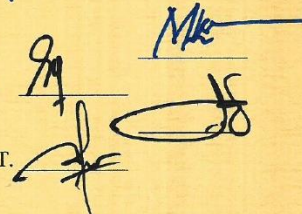
Ketua Jurusan,

  
Dekan  
KEMENTERIAN AGAMA  
UNIVERSITAS ISLAM NEGERI  
SULTAN SYARIF KASIM RIAU  
**Dr. Hartono, M.Pd.**  
NIP. 19640301 199203 1 003

  
**Iwan Iskandar, M.T.**  
NIP. 19821216 201503 1 003

#### DEWAN PENGUJI

Ketua : Muhammad Irsyad, S.T., M.T.  
Pembimbing : Siska Kurnia Gusti, S.T., M.Sc.  
Penguji I : Febi Yanto, S.Kom., M.Kom.  
Penguji II : Muhammad Affandes, S.T., M.T.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



## LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

© Hak Cipta Milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

UIN SUSKA RIAU



## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 13 Juni 2023

Yang membuat pernyataan,

**DESTRI PUTRI YANI**

**NIM 11950125030**

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**SURAT PERNYATAAN**

Saya yang bertanda tangan dibawah ini :

Nama : Destri Putri Yani  
NIM : 11950125030  
Tempat/Tgl.Lahir : Pekanbaru, 26 Desember 1999  
Fakultas : Sains dan Teknologi  
Prodi : Teknik Informatika  
Judul Skripsi : Klasifikasi Sentimen Transformasi dan Reformasi Sepak Bola Indonesia Pada Twitter Menggunakan Algoritma Bernoulli Naïve Bayes

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut diatas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu, Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai perundang-undangan.

Demikian Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 13 Juni 2023

Yang membuat pernyataan

  
  
AD189AKX474030443  
**DESTRI PUTRI YANI**  
NIM. 11950125030

## LEMBAR PERSEMBAHAN

Alhamdulillah Rabbil'alamin, Terima kasih kepada Allah SWT yang melimpahkan segala rahmatnya sehingga skripsi ini bisa terselesaikan dengan baik. Skripsi ini saya persembahkan untuk yang tersayang keluarga, Papa, Mama, Kakak, Abang dan Support System yang selama ini menemani dalam pembuatan skripsi ini serta teman-teman terdekat.

Semoga Tugas Akhir ini bisa bermanfaat bagi pembaca. Amiin ya rabb.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



# Klasifikasi Sentimen Transformasi dan Reformasi Sepak Bola Indonesia Pada Twitter Menggunakan Algoritma Bernoulli Naïve Bayes

Destri Putri Yani\*, Siska Kurnia Gusti, Febi Yanto, Muhammad Affandes

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: <sup>1</sup>\*11950125030@students.uin-suska.ac.id, <sup>2</sup>siskakurniagusti@uin-suska.ac.id, <sup>3</sup>febiyanto@uin-suska.ac.id, <sup>4</sup>affandes@uin-suska.ac.id

Email Penulis Korespondensi: 11950125030@students.uin-suska.ac.id

Submitted: 16/02/2023; Accepted: 16/03/2023; Published: 31/03/2023

**Abstrak**—Federation Internationale de Football Association (FIFA) melakukan Transformasi dan Reformasi terhadap Sepak Bola Indonesia dengan salah satunya Indonesia dipilih sebagai Tuan Rumah Piala Dunia U-20 pada tahun 2023. Transformasi dan Reformasi yang dilakukan menyebabkan masyarakat sering memberikan opini melalui media sosial Twitter. Opini yang diberikan oleh masyarakat bisa berupa positif atau negatif. Penelitian menggunakan Text Mining untuk mengklasifikasi sentimen pada 2 kategori dengan algoritma Bernoulli Naïve Bayes. Penelitian ini bertujuan untuk mengklasifikasi sentimen positif dan negatif serta mengetahui tingkat nilai akurasi dari hasil klasifikasi sentimen Transformasi dan Reformasi Sepak Bola Indonesia. Tahap penelitian yang dilakukan yaitu pengumpulan data, preprocessing text, pelabelan data, pembobotan TF-IDF, klasifikasi Bernoulli Naïve Bayes, dan evaluasi. Berdasarkan hasil penelitian dari 4907 data terdapat data duplikat dan hanya menggunakan 2125 data yang dibagi menjadi 90% data training dan 10% data testing, sehingga mendapatkan akurasi dengan nilai kategori tinggi sebesar 88%. Hasil klasifikasi menunjukkan bahwa banyak tweet yang merupakan sentimen positif.

**Kata Kunci:** Bernoulli Naïve Bayes; Klasifikasi; Reformasi; Sentimen; Sepak Bola; Transformasi; Twitter

**Abstract**—Federation Internationale de Football Association (FIFA) carried out Transformations and Reformations to Indonesian Football with one of them Indonesia was chosen as the Host of the U-20 World Cup in 2023. The transformations and reformations carried out cause people to often provide opinions through social media Twitter. Opinions given by the public can be positive or negative. The research uses Text Mining to classify sentiment in 2 categories with the Bernoulli Naïve Bayes algorithm. This research aims to classify positive and negative sentiments and determine the level of accuracy value of the sentiment classification results of Indonesian Football Transformation and Reformation. The research stages carried out are data collection, text preprocessing, data labeling, TF-IDF weighting, Bernoulli Naïve Bayes classification, and evaluation. Based on the research results from 4907 data there is duplicate data and only uses 2125 data which is divided into 90% training data and 10% testing data, so as to get accuracy with a high category value of 88%. The classification results show that many tweets are positive sentiments.

**Keywords:** Bernoulli Naïve Bayes; Classification; Reformation; Sentiment; Football; Transformation; Twitter

## 1. PENDAHULUAN

Sepak Bola merupakan olahraga yang banyak diminati di Indonesia sehingga menjadi negara yang memiliki penggemar terbesar didunia. [1]. Federasi Sepak Bola Internasional atau disebut Federation Internationale de Football Association (FIFA) melakukan Transformasi dan Reformasi terhadap Sepak Bola Indonesia dengan salah satunya Indonesia dipilih sebagai Tuan Rumah Piala Dunia U20 yang akan dilaksanakan pada tahun 2023. Transformasi dan Reformasi yang dilakukan oleh FIFA dapat memicu perdebatan oleh masyarakat pada media sosial Twitter terhadap perubahan tersebut. Pada tanggal 18-19 Oktober 2022 Sepak Bola menjadi trending topik pada media sosial Twitter yang disebabkan karena masyarakat banyak memberikan opini terhadap Transformasi dan Reformasi Sepak Bola Indonesia. Twitter dapat menghubungkan orang-orang di seluruh dunia dalam berbagai topik yang penggunaanya dapat dengan bebas mengeluarkan pendapat atau opini bahkan juga dapat mengekspresikan perasaan pengguna [2] [3].

Masyarakat memberikan opini terhadap Transformasi dan Reformasi Sepak Bola dalam bentuk opini negatif maupun opini positif, untuk mengetahui dan mengukur tingkat opini negatif dan opini positif dengan melakukan klasifikasi sentimen. Klasifikasi sentimen dilakukan berdasarkan proses pengolahan data dengan klasifikasi atau pengelompokan opini terhadap seseorang dengan penilaian pada objek tertentu [4]. Klasifikasi sentimen digunakan untuk mengidentifikasi teks yang dapat dikategorikan menjadi sentimen positif dan sentimen negatif [5]. Klasifikasi sentimen dilakukan dengan pendekatan Text Mining pada mengklasifikasi dokumen dalam menemukan berbagai bentuk atau model informasidari kumpulan data yang dikumpulkan dalam bentuk tidak teratur menjadi lebih teratur [6][7].

Algoritma Naive Bayes adalah perhitungan probabilitas dengan metode klasifikasi sederhana, tetapi memiliki akurasi tinggi meskipun menggunakan lebih sedikit data [8]. Algoritma Naïve Bayes termasuk kedalam metode Supervised Learning yang memiliki class atau label. Berdasarkan distribusi algoritma Naïve Bayes dibagi menjadi beberapa algortima yaitu Bernoulli Naive Bayes, Gaussian Naive Bayes, dan Multinomial Naive Bayes





1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[10]. Bernoulli Naïve Bayes digunakan dalam pengklasifikasi teks maupun dokumen [10]. Algoritma Bernoulli Naïve Bayes memiliki dua kategori ya dan tidak atau dalam biner 1 dan 0.

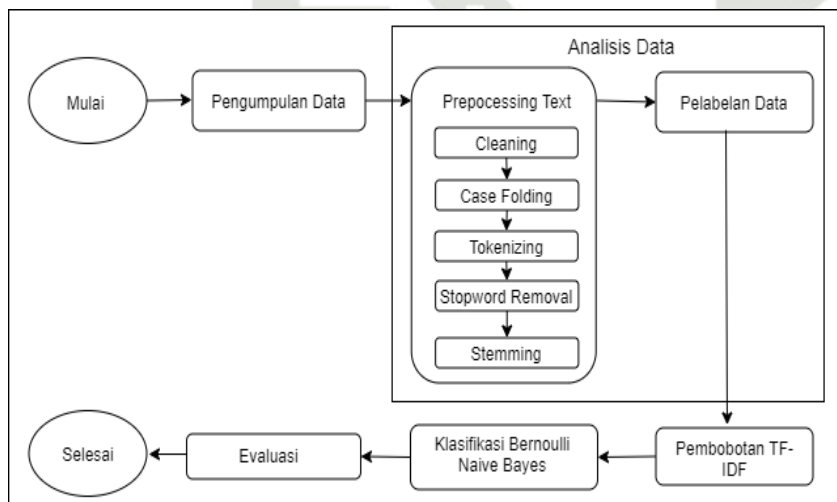
Penelitian-penelitian sebelumnya yang mengenai klasifikasi sentimen menggunakan Algoritma Bernoulli Naïve Bayes pada media sosial Twitter dengan tema vaksinasi Covid-19 [11]. Penelitian tersebut menggunakan algoritma Bernoulli Naïve Bayes dan algoritma Gaussian Naïve Bayes dengan 1200 data Tweet. Hasil dari penelitian ini menunjukkan bahwa algoritma Bernoulli Naïve Bayes lebih bagus dari pada algoritma Gaussian Naïve Bayes. Penelitian selanjutnya menggunakan algoritma Bernoulli Naïve Bayes dan Multinomial Naïve Bayes pada analisis sentimen pemindahan Ibu Kota Negara yang menggunakan Data dari komentar saluran Kompas TV sebanyak 1500. Hasil penelitian tersebut menunjukkan bahwa Bernoulli Naïve Bayes memiliki tingkat performa klasifikasi sebesar 93,45% dan Multinomial Naïve Bayes sebesar 90,19% yang dapat disimpulkan bahwa algoritma Bernoulli Naïve Bayes lebih bagus dari pada algoritma Multinomial Naïve Bayes [12]. Penelitian berikutnya mengenai “Analisis Sentimen Untuk Mengukur Kepercayaan Masyarakat Terhadap Pengadaan Vaksin Covid-19 Berbasis Bernoulli Naïve Bayes” [13]. Penelitian tersebut menggunakan data yang berjumlah 15.084 tweet yang menghasilkan akurasi cukup tinggi yaitu 80.58%.

Berdasarkan penelitian lainnya yang dilakukan pada “Deteksi Hoax Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes Dengan Ekstraksi Fitur Tf-Idf” [14]. Penelitian ini menggunakan bentuk prediksi yang dibuat dengan data berita sebanyak 8800, menunjukkan hasil nilai akurasi sebesar 98,5%. Implementasi penelitian ini mampu meningkatkan nilai akurasi sebesar 16,08%. Penelitian lainnya pada “Penerapan Bernoulli Naïve Bayes Untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Online Food Delivery di Indonesia” [15]. Penelitian ini menggunakan 1920 Tweet yang diambil pada 15 Maret – 30 Maret 2022 sehingga menghasilkan akurasi sebanyak 89%.

Berdasarkan uraian diatas dan beberapa penelitian sebelumnya bisa disimpulkan bahwa metode yang sangat efektif dalam menghasilkan akurasi terbaik dalam distribusi algoritma Naïve Bayes yaitu dengan menggunakan Bernoulli Naïve Bayes. Tujuan penelitian ini dilakukan untuk mengklasifikasi sentimen dan mengetahui tingkat nilai akurasi terhadap Transformasi dan Reformasi Sepak Bola Indonesia pada Twitter dengan menggunakan data yang diambil dari media sosial Twitter sebanyak 4907 data pada tanggal 18-19 Oktober 2022. Hasil penelitian ini dapat digunakan untuk melihat sentimen yang lebih dominan dari klasifikasi sentimen Transformasi dan Reformasi Sepak Bola Indonesia pada Twitter.

## 2. METODOLOGI PENELITIAN

Algoritma Naïve Bayes dalam penelitian ini menggunakan varian Bernoulli Naïve Bayes untuk memperoleh akurasi yang lebih unggul dalam klasifikasi sentimen terhadap Transformasi dan Reformasi Sepak Bola pada Twitter. Berikut adalah metode untuk mendapatkan tingkat kinerja yang tinggi.



Gambar 1. Alur Penelitian

### 2.1 Pengumpulan Data

Pada tahap pengumpulan data, penelitian ini menggunakan crawling data untuk memperoleh data pada media sosial Twitter. Crawling data dilakukan dengan aplikasi RapidMiner yang berguna sebagai alat dalam text mining [16]. Operator search Twitter digunakan dalam aplikasi RapidMiner dan mengatur parameter query sebagai kata kunci yang digunakan yaitu Transformasi, Reformasi, dan Sepak Bola serta language menjadi id untuk mengambil Tweet berbahasa Indonesia. Parameter select attribute digunakan untuk mengatur atribut yang akan digunakan dalam penelitian yaitu atribut text, untuk from-user, from-user-id, dan id. Operator write CSV digunakan untuk menyimpan hasil data yang telah diperoleh kedalam file dengan format CSV.



## 2.2 Preprocessing Text

Pada tahap selanjutnya adalah mempersiapkan text yang akan diproses untuk klasifikasi. Tahap preprocessing text atau pengolahan teks meliputi beberapa tahap sehingga teks yang sudah diubah dan dibersihkan siap untuk diproses di tahap klasifikasi. Berikut tahap preprocessing text yang digunakan oleh peneliti seperti yang tercantum dibawah

1. Di larang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Case Folding adalah mengganti seluruh huruf kapital yang terdapat dalam teks menjadi huruf kecil (lowercase) atau sebaliknya.

Cleaning adalah membersihkan teks dengan menghilangkan kata Retweet (RT) dan karakter seperti hastag, mention, url, symbol, dan angka sehingga menghasilkan teks Tweet yang asli.

Tokenizing adalah tahap ini disebut suku token yaitu terdiri dari memotong setiap kalimat menjadi kata.

Filtering atau stopword Removal: membuang kata yang tidak berarti atau tidak deskriptif seperti dari, di, dan, yang sehingga akan mengganggu saat proses klasifikasi.

Stemming adalah teknik mengganti kata menjadi bentuk dasarnya dengan membuang imbuhan-imbuhan pada kata dalam teks.

## 2.3 Pelabelan data

Pada tahap selanjutnya adalah pelabelan data dimana dilakukan proses pembuatan kelas atau label dengan 2 (dua) kategori yaitu positif dan negatif pada teks. Pemberian label positif dilakukan apabila teks tersebut memiliki sifat baik dengan mengandung kata-kata yang baik sedangkan pemberian label negatif dilakukan apabila teks tersebut memiliki sifat yang buruk dengan mengandung kata-kata seperti kata cacian dan kata yang kotor. Pelabelan data kemudian divalidasi oleh pihak ahli yaitu dosen Bahasa Indonesia dalam penentuan nilai aktual.

## 2.4 Pembobotan TF-IDF

TF-IDF adalah metode skema pembobotan kata yang paling populer dan efisien sehingga mempunyai nilai yang akurat (Term Weighting Scheme) [17]. Term Frequency (TF) merupakan jumlah dalam term yang muncul dari dokumen atau dataset. Inverse Document Frequency (IDF) merupakan jumlah pembobotan pada term yang telah didistribusikan pada semua dokumen atau dataset secara luas. [18]. Perhitungan metode TF-IDF dilakukan dengan menggunakan rumus.

$$TFIDF(t) = TF * \log \frac{N}{df} \tag{1}$$

Keterangan:

- (t) : term
- TF : jumlah yang muncul dari term t pada dokumen
- N : total dokumen
- df : jumlah dokumen yang ada dalam term t

## 2.5 Klasifikasi Bernoulli Naïve Bayes

Tahap selanjutnya setelah data siap diproses maka akan dilakukan klasifikasi. Pada tahap ini klasifikasi data menggunakan algoritma Naïve Bayes yang telah didistribusikan menjadi Bernoulli Naïve Bayes sebagai variable bernilai biner (Bernoulli, Boolean) [19]. Variable biner menunjukkan kata yang muncul dan tidak muncul serta mengabaikan frekuensi data dalam dokumen. Algoritma Bernoulli Naïve Bayes diberikan aturan keputusan sebagai persamaan berikut.

$$P(x_i|y) = P(i|y)^{x_i} + (1 - P(i|y))(1 - x_i) \tag{2}$$

## 2.6 Evaluasi

Tahap terakhir adalah melakukan evaluasi untuk menunjukkan hasil terbaik dari hasil uji. Pengujian digunakan untuk mengukur keakuratan hasil. Pengukuran yang tepat dari bentuk dengan confusion matrix. Pengukuran evaluasi dengan confusion matrix pada Tabel 1.

**Tabel 1.** Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Keterangan :

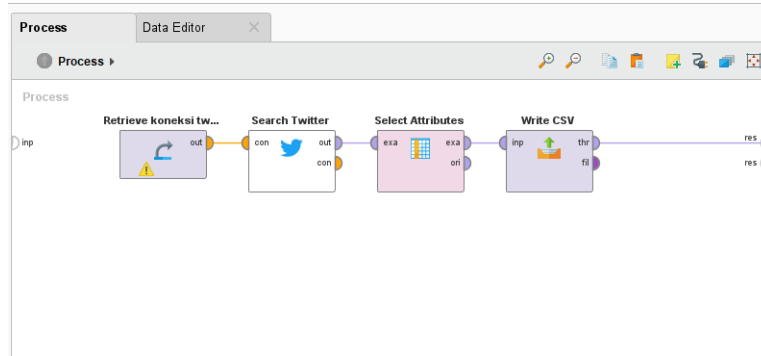
- TP (True Positive) = klasifikasi positif yang benar.
- TN (True Negative) = klasifikasi negatif yang benar.
- FP (False Positive) = klasifikasi negatif sebagai klasifikasi positif.
- FN (False Negative) = klasifikasi positif sebagai klasifikasi negatif.



### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengambilan Data

Dalam penelitian ini pengumpulan data menggunakan crawling data pada aplikasi RapidMiner dilakukan melalui koneksi Twitter API sedangkan operator search twitter digunakan sebagai pencarian katagori Transformasi, Reformasi, dan Sepak Bola. Selanjutnya, operator select attribute sebagai pemilihan atribut yang akan diambil dan operator write CSV sebagai penyimpanan hasil data dalam format CSV.



Gambar 2. Tahap Crawling Data

Hasil dari crawling data mendapatkan 4907 data Tweet. Data Tweet yang telah diperoleh ditunjukkan pada gambar 3.

	From-User	From-User-Id	Text	Id
0	PSSI	2363027508	Potret Presiden FIFA Gianni Infantino beserta ...	1580000000000000000
1	A. Ainur Rohman	132276059	Fitri lega saat suaminya, Hermawan, pulang. He...	1580000000000000000
2	Joko Widodo	366987179	Buah tangan dari Presiden FIFA Gianni Infantin...	1580000000000000000
3	Bang Aldy	392726733	RT @campusbois1923: Ada 18 klub Liga 1 dan ada...	1580000000000000000
4	Muhammad Rizky R	96552036	RT @ainurohman: Fitri lega saat suaminya, Herm...	1580000000000000000
...	...	...	...	...
4902	Harian Kompas	255866913	Jelang Kongres Luar Biasa PSSI, 16 Februari, h...	1614765495638429952
4903	CNBCinfo	853247416887201024	Tokoh muda Arief Rosyid Hasan berpendapat penc...	1614597079623670016
4904	iNews.id	944180887171742976	Dianggap bertalenta, Erick Thohir cocok pimpin...	1614457469568270080
4905	Bay.	244544980	melihat carut marutnya keadaan sepak bola di i...	1614172737827909888
4906	Woyoooo	1359423945770840064	RT @Gultombani: @MafiaWasit Ini juga harus kit...	1614102071372049920

Gambar 3. Dataset

#### 3.2 Preprocessing Text

Setelah hasil dari pengumpulan data maka selanjutnya data akan dilakukan preprocessing text untuk membersihkan text. Dalam tahap preprocessing text data dibagi menjadi beberapa langkah sebagai berikut.

##### a. Case Holding

Langkah ini mengonversikan seluruh huruf kapital yang ada pada teks menjadi huruf kecil (lowercase)

Tabel 2. Hasil Case Holding

Text	Case Holding
Potret Presiden FIFA Gianni Infantino beserta jajaran saat bermain sepak bola bersama Ketua Umum PSSI Mochamad Iriawan di Stadion Madya. #KitaGaruda https://t.co/oICdsF0sUr	potret presiden fifa Gianni Infantino beserta jajaran saat bermain sepak bola bersama ketua umum pssi mochamad iriawan di stadion madya. #kitagaruda https://t.co/oicdsf0sur

##### b. Cleaning

Setelah langkah case holding, cleaning dibutuhkan untuk membersihkan teks sehingga menghasilkan teks Tweet yang asli.

Tabel 3. Hasil Cleaning

Text	Cleaning
potret presiden fifa Gianni Infantino beserta jajaran saat bermain sepak bola bersama ketua umum pssi mochamad iriawan di stadion madya. #kitagaruda https://t.co/oicdsf0sur	potret presiden fifa Gianni Infantino beserta jajaran saat bermain sepak bola bersama ketua umum pssi mochamad iriawan di stadion madya. kitagaruda

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**Tokenizing**

Langkah ini bertujuan untuk memecah teks pada setiap kalimat menjadi sebuah token atau kata perkata.

**Tabel 4.** Hasil Tokenizing

Text	Tokenizing
potret presiden fifa Gianni Infantino beserta jajaran saat bermain sepak bola bersama ketua umum pssi mochamad iriawan di stadion madya. #kitagaruda https://t.co/oicdsf0sur	Potret, presiden, fifa, Gianni, Infantino, beserta, jajaran, saat, bermain, sepak, bola, bersama, ketua, umum, pssi, mochamad, iriawan, di, stadion, madya., kitagaruda,

**Filtering atau Stopword Removal**

Langkah filtering atau stopwords removal bertujuan untuk membuang kata yang tidak berarti dalam teks sehingga akan mengganggu saat proses klasifikasi. Ada beberapa stopwords sebagai berikut : merupakan, adalah, ada, adanya, akankah, anda, andalah, baik, baiklah, dan lain sebagainya.

**Tabel 5.** Hasil Stopword Removal

Tokenizing	Stopword Removal
Potret, presiden, fifa, Gianni, Infantino, beserta, jajaran, saat, bermain, sepak, bola, bersama, ketua, umum, pssi, mochamad, iriawan, di, stadion, madya., kitagaruda,	'potret', 'presiden', 'fifa', 'gianni', 'infantino', 'beserta', 'jajaran', ' ', 'bermain', 'sepak', 'bola', 'ketua', 'umum', 'pssi', 'mochamad', 'iriawan', 'stadion', 'madya', 'kitagaruda'

**Stemming**

Langkah ini bertujuan untuk mengembalikan kata kedalam bentuk dasar dengan menghapus imbuhan-imbuhan pada kata.

**Tabel 6.** Hasil Stemming

Stopword Removal	Stemming
'potret', 'presiden', 'fifa', 'gianni', 'infantino', 'beserta', 'jajaran', ' ', 'bermain', 'sepak', 'bola', 'ketua', 'umum', 'pssi', 'mochamad', 'iriawan', 'stadion', 'madya', 'kitagaruda'	'potret', 'presiden', 'fifa', 'gianni', 'infantino', 'serta', 'jajar', 'main', 'sepak', 'bola', ' ', 'ketua', 'umum', 'pssi', 'mochamad', 'iriawan', ' ', 'stadion', 'madya', 'kitagaruda',

Selanjutnya data akan dipilah untuk menghapus data Tweet yang terdapat duplikat maka dari itu data yang diperoleh ketika crawling data sebanyak 4907 data terdapat data duplikat sebanyak 2782 data sehingga data yang akan digunakan pada tahap selanjutnya yaitu 2125 data.

**3.3 Pelabelan Data**

Pada tahap ini, setiap teks Tweet diklasifikasikan sebagai dua kategori, yaitu positif dan negatif. Proses pelabelan dilakukan secara manual dengan melihat satu per satu Tweet yang berupa opini digunakan dalam klasifikasi. Nilai aktual dari data dapat divalidasi oleh tenaga ahli yaitu pihak ahli dari dosen Bahasa Indonesia sebagai bahan validasi dalam menentukan nilai sentimen. Beberapa hasil pelabelan ditunjukkan pada Tabel 7.

**Tabel 7.** Hasil Label Manual

Tweet	Sentimen
ada hal yang baik buat sepak bola indonesia etbola reformasi	positif
transformasi total sepak bola indonesia jokowi menerapkan teknologimitigasi	positif
kurang ajar salah lihat bola bukannya bola sepak yang diincar malah bola bilyar	negatif
setiap anak di dunia yg main sepak bola ingin menjadi pele	positif

**3.4 Pembobotan TF-IDF**

Dokumen teks yang telah melalui beberapa preprocess kemudian diteruskan pada metode pembobotan melalui penggunaan Term Frequency-Inverse Document Frequency (TF-IDF). Pembobotan kata diperlukan untuk mengganti term sebagai angka sehingga dapat dihitung dan diproses ketahap selanjutnya. Hasil pembobotan TF-IDF mengacu pada Tabel 8.

**Tabel 8.** Hasil TF-IDF

TERM	TF-IDF			
	d1	d2	d3	d4
baik	0,101	0	0	0
sepak	0	0	0	0
bola	0	0	0	0

Hak Cipta Dilindungi Undang-Undang  
 1. Diarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Sateh Hamid Jaryal, Kristyoni, Ridwan Syarif Kasim Ri



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

TERM	TF-IDF			
	d1	d2	d3	d4
indonesia	0,05	0,038	0	0
etbola	0,101	0	0	0
reformasi	0,101	0	0	0
transformasi	0	0,075	0	0
total	0	0,075	0	0
jokowi	0	0,075	0	0
terap	0	0,075	0	0
teknologimitigasi	0	0,075	0	0
kurang	0	0	0,086	0
ajar	0	0	0,086	0
salah	0	0	0,086	0
incar	0	0	0,086	0
bilyar	0	0	0,086	0
anak	0	0	0	0,101
dunia	0	0	0	0,101
main	0	0	0	0,101
pele	0	0	0	0,101

**3.5 Klasifikasi Bernoulli Naïve Bayes**

Proses klasifikasi menggunakan algoritma Naïve Bayes dengan Bernoulli Naïve Bayes. Data yang digunakan setelah melalui tahap preprocess sebanyak 2125 dengan positif 1799 dan negatif 314. Proses klasifikasi dilakukan dengan Jupyter Notebook dan menggunakan Bahasa pemograman Python. Data yang akan digunakan kemudian akan dibagi menjadi 90% data training dan 10% data testing. Proses klasifikasi mendapatkan akurasi sebesar 88% yang dinyatakan pada Gambar 4.

```
from sklearn.metrics import accuracy_score
score_naive = accuracy_score(predicted_naive, y_test)
print("Accuracy with Naive-bayes: ",score_naive)
print(predicted_naive)
Accuracy with Naive-bayes: 0.8820754716981132
```

**Gambar 4.** Hasil Akurasi Bernoulli Naïve Bayes

**3.6 Evaluasi**

Proses evaluasi berdasarkan hasil dari klasifikasi menggunakan Bernoulli Naïve bayes dapat dilakukan dengan baik. Confussion Matrix menyatakan beberapa laporan yang penting, seperti jumlah data uji manual dan jumlah data uji dari prediksi klasifikasi setiap kelasnya, presisi, dan recal dari setiap kelas targetnya. Confussion Matrix menunjukkan seberapa banyak nilai true positive dan true negative, false negative dan false positive dari data uji. Hasil Confussion Matrix ditunjukkan pada Tabel 9.

**Tabel 9.** Confussion Matrix

		Predicted	
		Negative	Positive
Actual	Positive	10	19
	Negative	6	177

Tahap selanjutnya hasil evaluasi ditunjukkan dalam nilai presisi, recal, dan akurasi pada hasil menggunakan python yang dinyatakan pada gambar 5.

```
from sklearn.metrics import classification_report
print(classification_report(y_test, predicted_naive, zero_division=0))
precision recall f1-score support
negatif 0.62 0.34 0.44 29
positif 0.90 0.97 0.93 183
accuracy 0.88 212
macro avg 0.76 0.66 0.69 212
weighted avg 0.87 0.88 0.87 212
```

**Gambar 5.** Hasil Evaluasi



Berdasarkan gambar diatas, hasil evaluasi mendapatkan nilai akurasi sebesar 88% yang dikategorikan tinggi dengan negatif presisi 62%, positif presisi 90%, negatif recal 34%, positif recal 97%, negatif f1-score 44%, dan positif f1-score 93%. Total rata-rata setiap kelas presisi sebesar 76%, recal sebesar 66%, dan f1-score sebesar 69%. Kemudian total rata-rata dari semua kelas presisi sebesar 87%, recal sebesar 88%, dan f1-score sebesar 87%.

#### 4. KESIMPULAN

Berdasarkan penelitian dan implementasi yang telah dilakukan oleh peneliti untuk mengklasifikasi sentimen Transformasi dan Reformasi Sepak Bola pada media sosial Twitter menggunakan algoritma Naïve Bayes dengan Bernoulli Naïve Bayes. Klasifikasi dapat dikelompokkan menjadi 2 kategori yaitu positif dan negatif. Pengujian klasifikasi Bernoulli Naïve Bayes dilakukan dengan 4907 dataset yang diperoleh dari media sosial Twitter, terdapat data duplikat sehingga menjadi 2125 data yang digunakan. Berdasarkan hasil pengujian dapat dengan menggunakan 2125 dataset yang telah dilakukan preprocessing menghasilkan akurasi sebesar 88% dengan nilai presisi negatif 62%, presisi positif 90%, recal negatif 34% dan recal positif 97%. Pengujian dilakukan dengan membagi data menjadi 90% data training dan 10% data testing. Hasil pengujian tersebut menunjukkan bahwa banyak tweet yang merupakan sentimen positif terkait Transformasi dan Reformasi sepak Bola. Hasil tersebut dapat disimpulkan bahwa implementasi yang dilakukan dengan Bernoulli Naïve Bayes mampu dilakukan dengan kategori tinggi, untuk penelitian selanjutnya pada klasifikasi sentimen Transformasi dan Reformasi sepak Bola pada media sosial Twitter dapat dilakukan dengan menambah jumlah data serta penambahan fitur-fitur pada preprocessing data. Kemudian pada proses klasifikasi dapat dilakukan dengan varian lain dari Naïve Bayes. Selain itu, klasifikasi sentimen pada Transformasi dan Reformasi dapat dilakukan menggunakan k-fold cross validation.

#### REFERENCES

- [1] S. Raja Manaek Pakpahan, "Analisis Sentimen Tentang Opini Performa Klub Sepak Bola Pada Dokumen Twitter Menggunakan Support Vector Machine Dengan Perbaikan Kata Tidak Baku," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 7, pp. 2548–964, 2019, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," *e-Proceeding of Engineering*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [3] L. D. Mahbubah and E. Zuliarso, "Analisa Sentimen twitter Pada PILPRES 2019 Menggunakan Algoritma Naive Bayes," *Proceeding SINTAK*, pp. 193–199, 2019.
- [4] M. D. Alizah, A. Nugroho, U. Radiyah, and W. Gata, "Sentimen Analisis Terkait 'Lockdown' pada Sosial Media Twitter," *CSRID (Computer Science Research and Its Development Journal)*, vol. 12, no. 3, p. 143, Mar. 2021, doi: 10.22303/csrid.12.3.2020.143-149.
- [5] F. Agus Maulana, I. Ernawati, and J. R. Fatmawati, "Analisa Sentimen Cyberbullying Di Jejaring Sosial Twitter Dengan Algoritma Naive Bayes," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, pp. 529–538, 2020.
- [6] L. Moh.Yasid & Junaedi, "Analisis Sentimen Maskapai Citilink Pada Twitter Dengan Metode Naïve Bayes KATA KUNCI," 2019. [Online]. Available: <http://www.Twitter.com>
- [7] I. Verawati and B. S. Audit, "Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, p. 1411, Jul. 2022, doi: 10.30865/mib.v6i3.4132.
- [8] M. R. Fanani, "Algoritma Naive Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa," *Jurnal DISPROTEK*, vol. 11, no. 1, pp. 13–22, 2020.
- [9] M. Saraswati and D. Riminarsih, "Analisis Sentimen Terhadap Pelayanan KRL Commuterline Berdasarkan Data Twitter Menggunakan Algoritma Bernoulli Naive Bayes," *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 3, pp. 225–238, 2020, doi: 10.35760/ik.2020.v25i3.3256.
- [10] H. Ashari, D. Arifianto, H. Azizah, and A. Faruq, "Perbandingan Kinerja Algoritma Multinomial Naive Bayes (MNB, Multivariate Bernoulli dan Rocchio Algorithim Dalam Klasifikasi Konten Berita Hoax Berbahasa Indonesia Pada Media Sosial," <http://repository.unmuhjember.ac.id>, 2020.
- [11] I. Puspita Dewi and K. Muslim Lhaksmana, "Prediksi Retweet Menggunakan Metode Bernoulli dan Gaussian Naive Bayes di Media Sosial Twitter Dengan Topik Vaksinasi Covid-19," *e-Proceeding of Engineering*, vol. 8, no. 5, 2021.
- [12] N. S. Wardani, A. Prahutama, and P. Kartikasari, "Analisis Sentimen Pemindahan Ibu Kota Negara Dengan Klasifikasi Naive Bayes Untuk Model Bernoulli dan Multinomial," *Jurnal Gaussian*, vol. 9, no. 3, pp. 237–246, 2020, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [13] H. Azizah, B. S. Rintyarna, and T. A. Cahyanto, "Sentimen Analisis Untuk Mengukur Kepercayaan Masyarakat Terhadap Pengadaan Vaksin Covid-19 Berbasis Bernoulli Naive Bayes," *BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer*, vol. 3, no. 1, pp. 23–29, Mar. 2022, doi: 10.37148/bios.v3i1.36.
- [14] A. Fodi Prayoga, A. Id Hadiana, and F. Rakhmat Umbara, "Deteksi Hoax pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naïve Bayes dengan Ekstraksi Fitur Tf-Idf," *Jurnal Health Sains*, vol. 2, no. 10, pp. 1808–1823, Oct. 2021, doi: 10.46799/jsa.v2i10.327.
- [15] D. F. A. Putri, Ir. Mohammad Masjkur, M.S., and I. Indahwati, "Penerapan Bernoulli Naïve Bayes untuk Analisis Sentimen Pengguna Twitter Terhadap Layanan Online Food Delivery di Indonesia," *Xplore: Journal of Statistics*, vol. 12, no. 1, pp. 50–62, Jan. 2023, doi: 10.29244/xplore.v12i1.1110.
- [16] R. R. Ramadhan and U. Saprudin, "Penerapan Rapidminer menggunakan metode K-Means untuk Pengelompokan Puskesmas pada Cakupan Imunisasi Dasar (Studi Kasus : Kota Bandung)," *Jurnal Teknologi Informatika dan Komputer*, vol. 8, no. 2, pp. 176–187, Sep. 2022, doi: 10.37012/jtik.v8i2.1238.

Hak Cipta Dilindungi Undang-Undang. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.  
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.  
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.  
 2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

[7] A. Muljono, D. Putri Artanti, A. Syukur, A. Prihandono, and D. I. Rosal Moses Setiadi, "Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," Konferensi Nasional Sistem Informasi , pp. 8–9, 2018. [Online]. Available: <http://twitter.com>

[8] W. Pratiwi Ali and Y. Sibaroni, "Analisis Sentimen Masyarakat Terhadap Kinerja Presiden Indonesia Dalam Aspek Ekonomi, Pembangunan Berdasarkan Opini dari Twitter," e-Proceeding of Engineering , vol. 6, pp. 8637–8649, 2019.

[K.] A. P. Hedva, W. Suharso, and Q. A'yun, "Pengklasifikasian Kanker Payudara Dan Kanker Paru-Paru Dengan Metode Gaussian Naïve Bayes, Multinomial Naïve Bayes, Dan Bernoulli Naïve Bayes," Jurnal Smart Teknologi, vol. 3, no. 1, pp. 2774–1702, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>

