



# KLASIFIKASI SENTIMEN TRAGEDI KANJURUHAN PADA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES

Hak cipta milik 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.

## TUGAS AKHIR

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

Oleh

**IQBAL SALIM THALIB**

**NIM. 11950111696**



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2023

University of Sultan 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.

**LEMBAR PERSETUJUAN****KLASIFIKASI SENTIMEN TRAGEDI KANJURUHAN PADA  
TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES****TUGAS AKHIR**

Oleh

**IQBAL SALIM THALIB**

**NIM. 11950111696**

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 TRAGEDI KANJURUHAN PADA  
TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES**

Oleh

**IOBAL SALIM THALIB**

**NIM. 11950111696**

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,



**Dr. Hartono, M.Pd.**

**NIP. 19640301 199203 1 003**

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

**DEWAN PENGUJI**

Ketua : Surya Agustian, S.T., M.Kom.

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

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.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim R

### 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 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

**IQBAL SALIM THALIB**

**NIM 11950111696**

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.



**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 : Iqbal Salim Thalib  
NIM : 11950111696  
Tempat/Tgl.Lahir : Pekanbaru, 21 April 1999  
Fakultas : Sains dan Teknologi  
Prodi : Teknik Informatika  
Judul Skripsi : Klasifikasi Sentimen Tragedi Kanjuruhan Pada Twitter Menggunakan Algoritma Naive 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

  
E8275AIOX474030442  
**IQBAL SALIM THALIB**  
NIM. 11950111696

## LEMBAR PERSEMBAHAN

Alhamdulillah dan Terima kasih kepada Allah SWT yang melimpahkan segala rahmatnya sehingga skripsi ini bisa terselesaikan dengan baik. Skripsi ini saya persembahkan untuk yang tersayang Ayah, ibu, kakak, adik dan Support System yang selama ini menemani.

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 Tragedi Kanjuruhan Pada Twitter Menggunakan Algoritma Naïve Bayes

Iqbal Salim Thalib\*, 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>11950111696@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: 11950111696@students.uin-suska.ac.id

Submitted: 20/02/2023; Accepted: 21/03/2023; Published: 31/03/2023

**Abstrak**—Peristiwa Kanjuruhan Malang terjadi pada tanggal 1 Oktober dengan mengakibatkan 132 orang meninggal, 96 orang luka berat, dan 484 luka ringan. Penyebab kerusuhan itu terjadi karena dipicu provokasi antara suporter Arema Malang dan suporter Persebaya Surabaya yang menyebutkan kata-kata kasar dan tindakan provokatif lainnya yang menimbulkan kemarahan di kedua sisi. Analisis sentimen tragedi Kanjuruhan dengan metode Naive Bayes dilakukan melalui tweet yang diambil melalui Twitter untuk memahami persepsi masyarakat terhadap kejadian tersebut. Algoritma Naïve Bayes dilakukan untuk klasifikasi sentimen data tweet yang diterapkan dengan mengolah teks tweet dan mengklasifikasikannya menjadi positif, negative, dan netral. Dalam penelitian ini menggunakan data sebanyak 4843 data dan dilakukan dengan data tweet yang telah di crawling menghasilkan sebanyak 2.042 data. Penelitian ini bertujuan untuk mengklasifikasi sentimen serta mengetahui tingkat akurasi pada algoritma Multinomial Naïve Bayes dalam tragedi Kanjuruhan dengan menggunakan dataset yang berupa tweet dari media sosial twitter. Data tweet yang telah diproses dibagi menjadi dua jenis yaitu data latih mendapatkan sebanyak 90% dan data uji sebanyak 10%. Hasil klasifikasi ini mendapatkan accuracy Naïve Bayes sebesar 75% dengan precision sebesar 73%, recall sebesar 75%, dan nilai f1-score sebesar 74%. Hasil data tweet yang digunakan dipenelitian ini bisa ditanggapi bahwa algoritma Naïve Bayes mempunyai nilai akurasi yang cukup baik.

**Kata Kunci:** Klasifikasi; Multinomial Naïve Bayes; Sentimen; Tragedi Kanjuruhan; Twitter

**Abstract**—The Kanjuruhan Malang incident occurred on October 1 and resulted in 132 deaths, 96 serious injuries and 484 minor injuries. The cause of the riot occurred due to provocation between Arema Malang supporters and Persebaya Surabaya supporters who mentioned harsh words and other provocative actions that caused anger on both sides. Sentiment analysis of the Kanjuruhan tragedy using the Naïve Bayes method was conducted through tweets taken through Twitter to understand the public's perception of the incident. The Naïve Bayes algorithm is performed for the sentiment classification of tweet data which is applied by processing the tweet text and classifying it into positive, negative, and neutral. In this study using data as much as 4843 data and carried out with tweet data that has been crawled resulting in 2,042 data. This research aims to classify sentiment and determine the level of accuracy in the Multinomial Naïve Bayes algorithm in the Kanjuruhan tragedy using a dataset in the form of tweets from twitter social media. The processed tweet data is divided into two types, namely 90% training data and 10% test data. The results of this classification get a Naïve Bayes accuracy of 75% with a precision of 73%, recall of 75%, and f1-score value of 74%. The results of the tweet data used in this study can be concluded that the Naïve Bayes algorithm has a fairly good accuracy value.

**Keywords:** Classification; Multinomial Naïve Bayes; Sentiment; Kanjuruhan Tragedy; Twitter

## 1. PENDAHULUAN

Stadion Kanjuruhan merupakan stadion sepak bola milik Pemerintah Provinsi Malang. Terjadi kerusuhan pada tanggal 1 Oktober 2022 pasca pertandingan Arema Malang vs Persebaya Surabaya. Sekelompok suporter Aremania atau Arema FC masuk ke lapangan dan tidak terima dengan hasil pertandingan tersebut. Polisi menanggapi situasi tersebut dengan menembakkan gas air mata dilapangan dan dalam stadion. Gas air mata membuat suporter panik dan lari keluar sampai mereka diinjak-injak dengan mengakibatkan 132 orang meninggal, 96 orang luka berat, dan 484 luka ringan. Kerusuhan itu terjadi karena penyelenggara liga sepak bola Indonesia yang tidak memahami tugas dan tidak profesional serta tidak bertanggung jawab. Sampai saat ini tragedi Kanjuruhan masih belum terselesaikan dan banyak diperbincangkan oleh masyarakat Indonesia terutama di sosial media Twitter. [1].

Twitter dapat digunakan melalui handphone atau komputer sehingga memudahkan penggunaannya untuk menyampaikan pendapatnya secara bebas. Pada saat ini banyak menggunakan Twitter karena sosial media Twitter mudah digunakan di berbagai kalangan masyarakat, karena melalui Twitter kita dapat dengan mudah bertukar pendapat dan juga mendapatkan teman baru. Indonesia menjadi salah satu negara terbesar dalam menggunakan media Twitter untuk tweet setiap hari [2]. Data mining merupakan proses penelitian mendalam dan analisis informasi sejumlah data yang besar dalam mencapai model tersembunyi diantara kelompok dari sejumlah besar data [3]. Metode Data Mining dapat menerapkan metode matematika, artificial intelligence, dan Machine Learning [4] [5]. Metode Naïve Bayes pada twitter dilakukan untuk memahami opini masyarakat dalam tragedi Kanjuruhan Malang.

Algoritma Naïve Bayes memiliki teknik klasifikasi [6] yang menggunakan metode dalam pengambilan keputusan untuk memperoleh hasil yang cukup baik terhadap masalah klasifikasi. Algoritma Naïve Bayes dibagi





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.

menjadi beberapa algoritma yaitu Multinomial Naive Bayes dan Bernoulli Naive Bayes. Algoritma Multinomial Naive Bayes dipilih karena memiliki akurasi yang tinggi untuk perhitungan sederhana. Analisis sentiment tragedi Kanjuruhan Malang dilakukan melalui tweet yang diambil pada Twitter untuk memahami persepsi masyarakat terhadap kejadian tersebut menggunakan Algoritma Naive Bayes. Selain itu analisis sentimen menggunakan metode Naive Bayes telah dilakukan oleh beberapa penelitian terdahulu seperti opini pada komentar terhadap E-Trip [7], penggunaan sistem plat pada kendaraan nomor ganjil atau genap [8], sentimen terhadap opini film [9]–[12], dan sentimen terhadap data Tweet Lazada [13].

Berdasarkan penelitian sebelumnya mengenai “Analisis Sentimen Pembelajaran Daring di Twitter Selama Pandemi COVID-19 Implementasi Naive Bayes”. Tujuan penelitian ini untuk menganalisis opini publik tentang daring selama pandemi COVID-19 di Indonesia pada awal November 2020. Hasil dari penelitian ini menghasilkan 30% positif, 69% negatif, dan 1% netral beberapa tweet mengungkapkan kekecewaannya karena kata "stres" dan "malas" adalah kata yang umum digunakan dalam dialog selama ini. Berlandaskan hasil tersebut dapat disimpulkan pembelajaran online di Indonesia belum maksimal dilaksanakan selama masa pandemi [14].

Berdasarkan penelitian lainnya dengan Judul Analisis Opini Calon Gubernur Jawa Timur 2018 Memakai Naive Bayes Classifier. Penelitian ini menerapkan Algoritma Naive Bayes Classifier dan data Twitter melalui kata kunci Gus Ipul dan Khofifah. Penelitian ini bertujuan untuk pengembangan analisis sentimen calon Gubernur Jawa Timur 2018 dengan optimal dan maksimal pada Twitter. Hasil dari penelitian ini menunjukkan nilai dari dataset Khofifah akurasi 77% %, presisi 79,2%, recal 77%. Untuk hasil dataset GusIpul akurasi 76%, presisi 74,4%, recal 76%. [15].

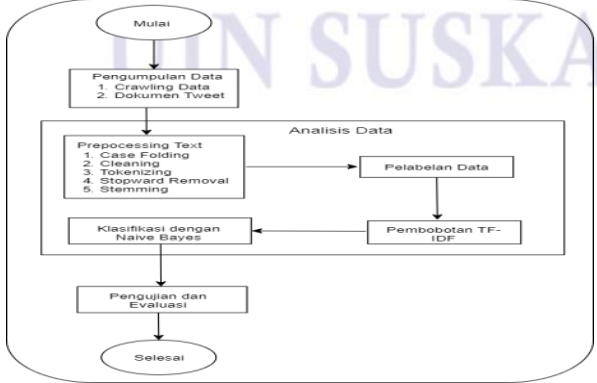
Penelitian yang sejenis juga dengan judul Analisis perasaan sarkasme dalam bahasa Indonesia. Penelitian ini menerapkan metode Naive Bayes dan Support vector Machine. Tujuan penelitian ini mendapatkan menganalisis pengaruh interjeksi dan unigram dalam mengklasifikasikan frasa sarkasme dalam bahasa Indonesia. Hasil dari penelitian ini pengembangan akurasi yang dicapai dengan menggunakan fitur interjeksi cara metode Classifier Naive Bayes mendekati dari 8%, dan pada metode SVM peningkatan lebih dari 13%. Pada penjabaran sarkasme ini Naive Bayes menaruh akurasi yg lebih baik dibandingkan SVM kernel polynomial [16].

Penelitian yang sejenis juga dengan judul “Sentimen Analisis Pelanggan Shopee pada Twitter menerapkan Algoritma Naive Bayes”. Penelitian ini memakai metode Naive Bayes dan data Shopee di Twitter. Maksud dari penelitian ini mengetahui jumlah komentar positif dan negatif di media sosial Twitter Shopee. Hasil dari penelitian ini jumlah komentar positif dan negatif setara yaitu 150:150, karena 300 data yang diperoleh dan tingkat keakurasiannya 97% [17]. Penelitian yang sejenis juga dengan judul Analisis opini berbasis text mining terhadap rating tempat wisata di Kota Tegal. Penjelasan ini menggunakan metode Naive Bayes dan Decision Tree. Penjelasan ini menyediakan informasi yang membawa keputusan bagi pendatang dan pengelola objek wisata untuk digunakan sebagai sumber informasi tentang objek wisata. Hasil penelitian ini mendapatkan metode Naive Bayes dengan tingkat accuracy 73,33%, dan Decision Tree 60.83% [18].

Pada permasalahan yang ada saat ini maka untuk mendapatkan suatu pengetahuan sentimen pada setiap opini perlu adanya Analisis Sentimen. Analisis sentiment salah satunya bidang penelitian Text Mining yang terkait dengan ruang lingkup yang lebih lebar [19]. Tujuan pembahasan ini buat mengetahui opini masyarakat yang telah diklasifikasi terkait tragedi Kanjuruhan menggunakan Teknik Text Mining. Text Mining mengarah pada Teknik yang digunakan untuk menganalisis data teks terstruktur dan semi-terstruktur [20]. Algoritma yang dipakai pada penelitian ini Multinomial Naive Bayes yang menciptakan metode untuk klasifikasi text Mining dalam analisis sentiment. Metode algoritma ini diaplikasikan untuk memperoleh sentimen dengan kelas positif, negatif, dan netral dengan data Twitter Adapun diambil sebesar 4843 data dari query pencarian “Kanjuruhan”.

## 2. METODOLOGI PENELITIAN

Algoritma Multinomial Naive Bayes metode yang dipakai pada penelitian ini karena memiliki ketelitian yang tinggi untuk perhitungan yang sederhana. Berikut adalah proses untuk mendapatkan tingkat kinerja yang tinggi.



Gambar 1. Alur penelitian



**2.1 Pengumpulan Data**

Data pada penelitian ini bermula dari tweet di situs media sosial Twitter. Pengumpulan data dilakukan untuk mendapatkan data dari media sosial Twitter yang berikutnya proses analisis sentimen dilakukan untuk mendapatkan informasi baru berupa sentimen yang muncul baik sentimen positif, negatif, dan netral. Data dikumpulkan dengan cara Crawling data dengan menggunakan RapidMiner.

**2.2 Preprocessing Text**

Tahapan Preprocessing Text dilakukan untuk mengubah teks dan membersihkan text agar siap untuk dikelola ke klasifikasi. Adapun tahapan yang dilakukan dalam preprocessing text yaitu Case Folding, Cleaning, Tokenizing, Filtering, dan Stemming.

Case Folding berguna untuk mengganti seluruh informasi dalam text dari huruf besar menjadi kecil atau sebaliknya. Tahapan kedua, yaitu Cleaning untuk membersihkan data dari karakter yang tidak dibutuhkan seperti tanda baca dan beberapa karakter lainnya. Selanjutnya, Tokenizing untuk tahap ini memotong setiap kalimat menjadi kata. Selanjutnya adalah Stopword Removal untuk membuang kata yang tidak dibutuhkan, Tahap terakhir, Lemming untuk pemotongan kata imbuhan menjadi kata dasar.

**2.3 Pelabelan Data**

Pada tahap sebelumnya preprocessing, akan diperoleh data yang telah bersih. Data tersebut kemudian diberikan label negative, positif, dan netral. Pada proses pelabelan data label dibuat menentukan apakah tweet tersebut termasuk dalam label positif yang berisi pujian. Pada label negatif mencakup sindiran atau kritik. Sedangkan label netral mencakup memberi informasi komentar yang biasa saja tanpa menggunakan tanggapan positif dan negatif. Proses pelabelan data dilakukan dengan manual memakai metode crowd sourcing.

**2.4 Pembobotan Tf-Idf**

Pada proses preprocessing yang sudah selesai, jadi langkah selanjutnya melakukan perhitungan menggunakan metode TF-IDF. Metode TF-IDF akan membantu hitung kata-kata yang muncul. Perhitungan TF-IDF dilakukan dengan cara menggunakan rumus.

$$TF(d,t) = f(d,t) \tag{1}$$

$$IDF(t) = \log\left(\frac{Nd}{df(t)}\right) \tag{2}$$

Keterangan :

$TF(d,t)$  : Term frequency

$f(d,t)$  : Terusan term t pada dokumen d

$IDF(t)$  : Inverse document frequency

$Nd$  : Total dokumen

$df(t)$  : Dokumen secara keseluruhan term t

Perhitungan TF-IDF dari suatu kata dilakukan dengan Persamaan

$$TF - IDF = TF(d,t) \cdot IDF(t) \tag{3}$$

**2.5 Klasifikasi Naïve Bayes**

Metode Naïve Bayes proses klasifikasi yang dipakai dalam klasifikasi sentimen pada text mining. Klasifikasi Naïve Bayes membentuk teknik machine learning yang populer bagi klasifikasi dan pemilihan teks penampilan yang bagus [21]. Metode ini terbukti akurat dan bila diterapkan pada kumpulan data besar. Langkah-langkah mengklasifikasi sentiment dengan metode Naïve Bayes yaitu menetapkan nilai term pada dokumen, menetapkan jumlah seluruh term dan yang terakhir menetapkan jumlah IDF pada term. Banyak metode Naïve Bayes digunakan dalam teknik klasifikasi khususnya di sosial media Twitter menggunakan berbagai metode seperti Gaussian Naive Bayes, Bernoulli Naive Bayes, dan Multinomial Naive Bayes.

**2.6 Evaluasi**

Pada tahap evaluasi, hasil klasifikasi diverifikasi dengan menghitung nilai keaslian sistem. Skala statistik yang dilakukan buat menghitung nilai efisiensi sistem disebut akurasi. Akurasi dihasilkan oleh faktor dari penjumlahan yang muncul dalam persamaan.

$$\text{akurasi} = \frac{TP+TN}{TP+FP+TF+TN} \tag{4}$$

Keterangan :

TP = True Positif, positif diakui secara benar

TN = True Negatif, negatif diakui secara benar

FP = False Positif, negatif terdeteksi positif

1. Diarangi 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 mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



1. Diarangi 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.

FN = False Negatif, positif terdeteksi negative

### 3. HASIL DAN PEMBAHASAN

#### Hasil Pengumpulan Data

Data yang diperoleh tersebut ditemukan melalui proses Crawling di media sosial Twitter. Jumlah data yang didapatkan menghasilkan 4843 tweet. Contoh data bisa diperhatikan pada Gambar 2.

	Created-At	From-User	From-User-Id	Text
0	2022-10-19 06:59:33 SE Asia Standard Time	tempodotco	18129942	Hari Ini Komnas HAM akan Periksa Dinut PT LIB ...
1	2022-10-19 06:59:33 SE Asia Standard Time	suhonosuardi	1246770000000000000	@ImminentDangers @awemany Bukannya yg menyuruh...
2	2022-10-19 06:59:32 SE Asia Standard Time	kontenisiam_com	1275930000000000000	Stadion Kanjuruhan Bakal Dibangun Ulang Sesuai...
3	2022-10-19 06:59:26 SE Asia Standard Time	mirhan_sandy	7074250000000000000	@republikaonline Proyek lagi rakus bener.. Km5...
4	2022-10-19 06:58:50 SE Asia Standard Time	hendrideen1	1472360000000000000	DUHI Begini Nasib Cahayu Gadis Cilik Korban Ka...
...	...	...	...	...
4838	2023-01-17 12:56:00	#UsutTuntasTragediKanjuruhan	1471104000000000000	RT @panditfootball: Sidang perdana Tragedi Kan...
4839	2023-01-17 12:56:00	As'ad Akbar	117579316	RT @dribble9: Bayangkan ada organisasi yang sa...
4840	2023-01-17 12:56:00	???? ???? ?	1106590000000000000	RT @OngisnadeCultur: Para penembak GAM di Trag...
4841	2023-01-17 12:56:00	vivianne ?	1344450000000000000	Sidang Tragedi Kanjuruhan: Tiga polisi didakwa...
4842	2023-01-17 12:56:00	Subik Yantoro	1572990000000000000	RT @GSyndicate27: dikiro nang kanjuruhan be.e ...

4843 rows x 4 columns

Gambar 2. Hasil Crawling Data

#### 3.2 Preprocessing Text

Pada tahap preprocessing, sebelum menggunakan data pada tweet, data bersih harus didapatkan dan diberi label. Langkah-langkah preprocessing sebagai berikut.

##### a. Case Holding

Prosedur mengubah semua data yang ada ditext dari huruf besar menjadi kecil atau sebaliknya. Dapat diperhatikan pada Tabel 1.

Tabel 1. Case Holding

Sebelum	Sesudah
Bertemu presiden FIFA, Jokowi akan bangun Kanjuruhan jadi stadion mewah Internasional	bertemu presiden fifa, jokowi akan bangun kanjuruhan jadi stadion mewah internasional

##### b. Cleaning

Selanjutnya cleaning yaitu untuk membersihkan data dari karakter yang tidak diperlukan seperti tanda baca dan beberapa karakter lainnya. Dapat diperhatikan pada Tabel 2.

Tabel 2. Cleaning

Sebelum	Sesudah
Bertemu presiden FIFA, Jokowi akan bangun Kanjuruhan jadi stadion mewah Internasional	bertemu presiden fifa jokowi akan bangun kanjuruhan jadi stadion mewah internasional

##### c. Tokenizing

Selanjutnya tokenizing yaitu tahap ini memotong setiap kalimat menjadi kata. Dapat diperhatikan pada Tabel 3.

Tabel 3. Tokenizing

Sebelum	Sesudah
Bertemu presiden FIFA, Jokowi akan bangun Kanjuruhan jadi stadion mewah Internasional	bertemu, presiden, fifa, jokowi, akan, bangun, kanjuruhan, jadi, stadion, mewah, internasional

##### d. Stopword Removal

Sesudah itu stopwords removal yaitu melepaskan kata yang tidak penting. Dapat diperhatikan pada Tabel 4.

Tabel 4. Stopword Removal

Sebelum	Sesudah
Bertemu presiden FIFA, Jokowi akan bangun Kanjuruhan jadi stadion mewah Internasional	presiden, fifa, jokowi, kanjuruhan, stadion, mewah, internasional



**Stemming**

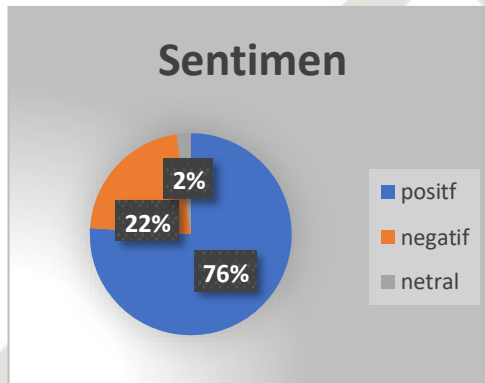
Selanjutnya yang terakhir stemming adalah pemotongan kata-kata imbuhan menjadi kata dasar. Dapat diperhatikan pada Tabel 5.

**Tabel 5.** Stemming

Sebelum	Sesudah
Bertemu presiden FIFA, Jokowi akan bangun Kanjuruhan jadi stadion mewah Internasional	presiden, fifa, jokowi, kanjuruhan, stadion, mewah, internasional

**Pelabelan Data**

Berdasarkan pelabelan data sebanyak 2.042 tweet diklasifikasikan menjadi 3 bagian, yaitu positif, negatif, dan netral. Pelabelan data sentimen positif menghasilkan sebanyak 1.550 tweet, pelabelan data sentimen negatif menghasilkan sebanyak 444, dan pelabelan data sentimen netral menghasilkan sebanyak 46 tweet dapat diperhatikan pada gambar 3.



**Gambar 3.** Presentase Sentimen Tragedi Kanjuruhan

**3.4 Pembobotan TF-IDF**

Selanjutnya tahapan pembobotan TF-IDF membantu untuk menghitung kata yang muncul. Nilai bobot dalam dokumen Term Frequency-Inversed Document Frequency (TF-IDF) diperoleh berlandaskan frekuensi munculnya kata-kata. Hasil pembobotan TF-IDF dapat diperhatikan pada Tabel 6.

**Tabel 6.** TF-IDF

TERM	TF-IDF			
	d1	d2	d3	d4
Pasien	0,151	0	0	0
Korban	0,151	0	0	0
Kanjuruhan	0	0	0	0
Sembuh	0,151	0	0	0
Brutal	0	0,2	0	0
Jokowi	0	0,1	0	0,043
Rekontruksi	0	0	0,12	0
Lapangan	0	0	0,12	0
Bola	0	0	0,12	0
Jatim	0	0	0,12	0
Presiden	0	0	0	0,086
Fifa	0	0	0	0,086
Stadion	0	0	0	0,086
Mewah	0	0	0	0,086
Internasional	0	0	0	0,086

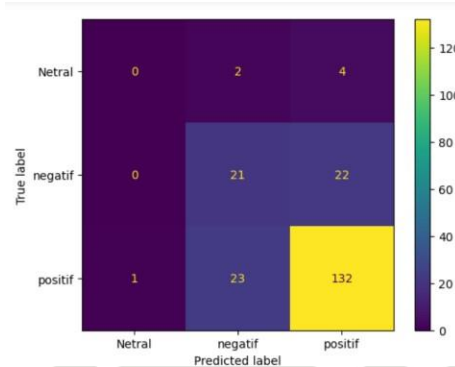
**3.5 Klasifikasi Multinomial Naïve Bayes**

Setelah bobot Term Frequency-Inversed Document Frequency dihitung, selanjutnya menerapkan algoritma Naïve Bayes melalui varian Multinomial Naïve Bayes. Klasifikasi tersebut dibagi menjadi 3 kelas, yaitu positif, negatif, dan netral dengan 90% data latih dan 10% data testing. Proses klasifikasi diterapkan dengan menggunakan Jupyter Notebook dan bahasa pemrograman Python.

1. Diarangi 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.

## 6 Evaluasi

Setelah proses klasifikasi dan analisis sentimen, Langkah berikutnya adalah pengujian. Dilakukan proses klasifikasi dan evaluasi menggunakan model klasifikasi teks Multinomial Naïve Bayes. Hasil evaluasi digunakan untuk mengevaluasi akurasi model naïve bayes dan hasil menggunakan Confusion Matrix dapat diperhatikan pada Gambar 4.



**Gambar 4.** Hasil Confusion Matrix

Berdasarkan gambar diatas mendapatkan hasil evaluasi sebagai uji accuracy Naïve Bayes sebesar 75% yang terbilang cukup baik dengan precision sebesar 73%, recall sebesar 75%, dan nilai f1-score sebesar 74%. Hasil evaluasi diperlihatkan pada gambar 5.

	precision	recall	f1-score	support
Netral	0.00	0.00	0.00	6
negatif	0.46	0.49	0.47	43
positif	0.84	0.85	0.84	156
accuracy			0.75	205
macro avg	0.43	0.44	0.44	205
weighted avg	0.73	0.75	0.74	205

**Gambar 5.** Hasil Evaluasi

## 4. KESIMPULAN

Berdasarkan hasil yang didapatkan pada proses penelitian diatas pada klasifikasi sentimen terhadap tragedi Kanjuruhan dengan algoritma Multinomial Naïve Bayes memakai data yang diperoleh dari media sosial Twitter. Algoritma Naïve Bayes dilakukan untuk klasifikasi sentimen data tweet yang diterapkan dengan mengolah teks tweet dan mengklasifikasikannya menjadi positif, negative, dan netral. Klasifikasi dilakukan dengan data tweet yang telah di crawling sebanyak 2.042. Data tweet yang telah diproses dibagi menjadi dua jenis yaitu data latih mendapatkan sebanyak 90% dan data uji sebanyak 10%. Hasil dari uji accuracy Naïve Bayes sebesar 75% dengan precision sebesar 73%, recall sebesar 75%, dan nilai f1-score sebesar 74%. Dari hasil dataset yang berupa tweet pada sosial media twitter yang diterapkan dalam penelitian ini, dari sini dapat disimpulkan bahwa Multinomial Naive Bayes mempunyai akurasi klasifikasi sentimen yang cukup baik pada tweet, namun untuk meningkatkan akurasi dalam klasifikasi sentimen dapat digunakan distribusi lain dari Naïve Bayes dan juga dengan menambahkan jumlah data, kemudian dapat menambahkan fitur-fitur lain seperti Lexion Based.

## REFERENCES

- [1] T. Prasetyo, H. Zakaria, and P. Wiliantoro, "Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier," OKTAL : Jurnal Ilmu Komputer dan Sains, vol. 1, no. 06, pp. 573-582, 2022, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [2] W. P. Angraini and M. S. Utami, "Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia," Faktoria Exacta, vol. 13, no. 4, p. 255, Feb. 2021, doi: 10.30998/faktoriaexacta.v13i4.7964.
- [3] S. Taleka, F. Alameka, and N. Wanti Wulan Sari, "Implementasi Data Mining Untuk Memprediksi Penjualan Dan Penempatan Stok Barang Pada Cv Pasti Jaya Houseware Menggunakan Algoritma Apriori," SEMINASTIKA, vol. 3, no. 1, pp. 115-123, Nov. 2021, doi: 10.47002/seminastika.v3i1.258.
- [4] S. Anastassia Amellia Kharis and A. Haqqi Anna Zili, "Learning Analytics dan Educational Data Mining pada Data Pendidikan," Jurnal Riset Pembelajaran Matematika Sekolah, vol. 6, 2022.
- [5] M. Rianto, R. Rusdiah, and H. Ichwan, "Penerapan Data Mining Dengan Metode Naïve Bayes Dan Learning Vector Quantization Credit Rating Dalam Memprediksi Kelayakan Pemberian Kredit Oleh PT. BPR Lebak Sejahtera," Jurnal Teknologi Informasi, vol. 1, 2022.
- [6] N. D. Ramadhani, "ANALISIS KINERJA ALGORITMA NAÏVE BAYES," 2019.



- Hak Cipta Dilindungi Undang-Undang**
1. Mibhandayani, E. Feriyanto, Syarham, and Kusriani, "Opinion Mining Pada Komentar Twitter E-KTP Menggunakan Naive Bayes Classifier," Seminar Nasional Teknologi Informasi dan Multimedia , 2018.
  2. N. Ruhana, "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada Twitter Dengan Metode Klasifikasi Naive Bayes," Jurnal IKRA-ITH Informatika, vol. 3, no. 1, 2019, [Online]. Available: [www.situs.com](http://www.situs.com)
  3. T. Ansyor Lorosae and B. Dwi Prakoso, "Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan Naive Bayes," UNIVERSITAS AMIKOM Yogyakarta, 2018.
  4. R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," JURNAL INOVTEK POLBENG -SERI INFORMATIKA, vol. 3, no. 1, 2018.
  5. Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," INSOLOGI: Jurnal Sains dan Teknologi, vol. 1, no. 4, pp. 411–423, Aug. 2022, doi: 10.55123/insologi.v1i4.770.
  6. F. Rozy, S. Rangkuti, M. A. Fauzi, Y. A. Sari, E. Dewi, and L. Sari, "Analisis Sentimen Opini Film Menggunakan Metode Naive Bayes dengan Ensemble Feature dan Seleksi Fitur Pearson Correlation Coefficient," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, no. 12, pp. 6354–6361, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
  7. D. Setian and I. Seprina, "Analisis Sentimen Masyarakat Terhadap Data Tweet Lazada Menggunakan Text Mining Dan Algoritma Naive Bayes Classifier," Bina Darma Conference on Computer Science, 2019.
  8. W. A. Prabowo and C. Wiguna, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 1, p. 149, Jan. 2021, doi: 10.30865/mib.v5i1.2604.
  9. G. A. Buntoro, "Analisis sentimen Calon Gubernur Jawa Timur 2018 Dengan Metode Naive Bayes Classifier," Journal Of Informatics Pelita Nusantara, vol. 4, no. 1, 2019.
  10. L. Septiani and Y. Sibaroni, "Sentiment Analysis Terhadap Tweet Bernada Sarkasme Berbahasa Indonesia," 2019. [Online]. Available: <https://twitter.com/hipwee/status/734249192273645568?y>
  11. A. H. Ruger, S. Analisis, P. Shopee, M. Suyanto, and M. P. Kurniawan, "Sentimen Analisis Pelanggan Shopee di Twitter menggunakan Algoritma Naive Bayes," JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY, vol. 1, no. 2, 2021.
  12. O. Somantri, "Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining," JEPIN (Jurnal Edukasi dan Penelitian Informatika) , vol. 5, no. 2, 2019, [Online]. Available: [www.google.com/maps](http://www.google.com/maps)
  13. D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," Jurnal TEKNO KOMPAK, vol. 15, no. 1, 2021.
  14. B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," e-Proceeding of Engineering, 2018.
  15. S. Sandino Berutu, "Text Mining dan Klasifikasi Sentimen Berbasis Naive Bayes Pada Opini Masyarakat terhadap Makanan Tradisional," Jurnal Sistem Komputer dan Informatika (JSON) Hal: 254–, vol. 262, no. 2, 2022, doi: 10.30865/json.v4i2.5138.

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.