

NAÏVE BAYES CLASSIFIER TERHADAP SENTIMENT ANALYSIS MENGENAI PSSI SETELAH TRAGEDI KANJURUHAN MALANG

Ahmad Faldhi Yudianto*¹, Statiswaty², Natalis Ransi³

^{1,2,3}Teknik Informatika, Teknik, Universitas Halu Oleo

e-mail: *faldhi.works@gmail.com , statiswaty@uho.ac.id, natalis.ransi@uho.ac.id

Abstrak

Informasi merupakan sesuatu yang penting dikarenakan dari informasi seseorang dapat mengetahui info yang ada seperti kondisi dalam masyarakat saat ini, dalam beberapa kumpulan informasi dapat menghasilkan informasi baru, hal ini disebut penggalian data atau biasa disebut *data mining*, penerapan data mining dapat membantu untuk menganalisa data yang diperoleh menjadi sebuah informasi yang sebelumnya tidak dapat terlihat oleh mata pembaca, salah satu variasi tersebut yaitu *text mining*. Salah satu bentuk implementasi *text mining* adalah analisis sentimen yang merupakan proses menganalisis potongan teks dengan menentukan nada emosional yang mereka bawa, apakah itu positif ataupun negatif. Penelitian ini berfokus pada analisis sentimen masyarakat pada platform Twitter tragedi Kanjuruhan yang berfokus kepada kinerja PSSI menggunakan metode *Naïve Bayes Classifier*. *Naïve Bayes Classifier* adalah metode klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan kelas. Berdasarkan hasil penelitian menggunakan 1500 dataset, dengan dataset 80:20 memiliki evaluasi yang cukup baik untuk efektivitas pada metode *Naïve Bayes Classifier* pada kasus analisis sentimen. Penggunaan metode tersebut memiliki hasil sentimen negatif yang banyak dari masyarakat terhadap PSSI daripada sentimen positif setelah tragedi Kanjuruhan.

Kata kunci— Analisis sentimen, *Naïve Bayes Classifier*, PSSI, *Preprocessing*, *Text mining*

Abstract

Information is something important because from information someone can find out existing information such as conditions in today's society, in some collections of information can produce new information, this is called data mining or commonly called data mining, the application of data mining can help to analyze the data obtained into an information that was previously invisible to the reader's eyes, one of these variations is text mining. One form of text mining implementation is sentiment analysis which is the process of analyzing pieces of text by determining the emotional tone they carry, whether it is positive or negative. This research focuses on analyzing public sentiment on the Twitter platform of the Kanjuruhan tragedy which focuses on PSSI's performance using the Naïve Bayes Classifier method. Naïve Bayes Classifier is a statistical classification method that can predict the probability of class membership. Based on the research results using 1500 datasets, with 80:20 it has a fairly good evaluation of the effectiveness of the Naïve Bayes Classifier method in the case of sentiment analysis. The use of this method has resulted in more negative sentiment from the public towards PSSI than positive sentiment after the Kanjuruhan tragedy.

Keywords— *Sentiment analysis, Naïve Bayes Classifier, PSSI, Preprocessing, Text mining*

1. PENDAHULUAN

Penerapan *data mining* dapat membantu untuk menganalisa data yang diperoleh dari transaksi pada sistem informasi sehingga dapat menggali pola-pola yang dapat dijadikan pengetahuan baru [1]. *Data mining* memiliki variasi salah satunya *text mining* yang berusaha menemukan pola yang menarik dari sekumpulan data tekstual yang berjumlah besar[2]. Data tekstual dapat berupa sebuah kalimat yang mengandung sentimen yang baik ataupun yang buruk untuk dilakukan penelitian *text mining*.

Salah satu isu yang paling banyak diperdebatkan di era saat ini adalah apa yang harus dilakukan terhadap ujaran "buruk" di internet, terutama di *platform* media sosial seperti Twitter. salah satu yang sering menjadi perbincangan di jagat media sosial yaitu mengenai sepak bola nasional sendiri, sehinggam masyarakat selalu tertuju kepada Persatuan Sepak Bola Indonesia (PSSI) yang mengatur megenai sepak bola nasional di Indonesia.

Berdiri sejak 19 April 1930 dan telah dipimpin oleh 16 orang berbeda. Dalam setiap era, banyak kontroversi dalam tubuh PSSI ini dalam setiap kepengurusan[3], salah satunya tragedi yang terjadi pada tanggal 1 Oktober 2022 yaitu tragedi Kanjuruhan yang dikabarkan menewaskan 129 suporter tuan rumah Arema Malang pasca laga *derby* Jawa Timur melawan Persebaya Surabaya, pada malam hari. Dan ratusan supporter lainnya mengalami luka-luka dan dilarikan ke rumah sakit terdekat [4], Komisi Nasional Hak Asasi Manusia kembali menegaskan bahwa penembakan gas air mata menjadi pemicu jatuhnya banyak korban, baik korban luka maupun meninggal dunia dalam peristiwa kericuhan di Stadion Kanjuruhan [5], penggunaan gas air mata pada di dalam stadion dilarang oleh *Fédération Internationale de Football Association (FIFA)* mengacu pada Pasal 19 FIFA tentang *Stadium Safety and Security Regulation*.

Berdasarkan latar belakang yang telah diuraikan sebelumnya, dalam penelitian ini penulis akan membangun “**Implementasi Naïve Bayes Classifier Terhadap Sentiment Analyst Masyarakat Pada Platform Twitter Mengenai PSSI Setelah Tragedi Kanjuruhan Malang**”. Metode *Naïve Bayes*

Classifier dipilih karena memiliki akurasi untuk menghasilkan data sentimen untuk di klasifikasikan dari literatur yang telah dibaca oleh penulis. Harapan adanya penelitian ini, penulis berharap dari penggalian data berupa sentimen terhadap masyarakat di media sosial pada kinerja PSSI pada saat ini.

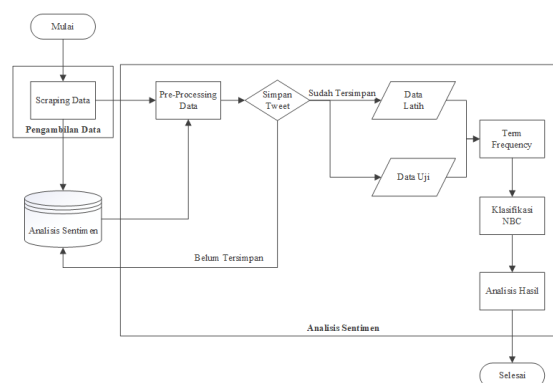
2. METODE PENELITIAN

Data yang digunakan untuk penelitian ini sebanyak 1500 data tweet yang diambil pada *platform* twitter tujuan dari penelitian ini untuk melihat hasil sentimen pada tweet tersebut menggunakan metode *Naïve Bayes Classifier*. *Dataset* yang digunakan akan dibagi menjadi 80:20 untuk data *training* dan *data testing*

Metode untuk pengumpulan data yang diperlukan dalam penyusunan tugas akhir ini, antara lain:

1. Kajian pustaka, dilakukan dengan cara mengumpulkan literatur atau sumber pustaka yang berkaitan dengan sistem yang dibuat serta dapat memperjelas teori-teori yang ada. Sumber literatur yang digunakan berupa buku, *e-book*, *paper*, dan jurnal.
2. *Web scraping*, teknik yang digunakan untuk mengambil beberapa data dalam jumlah banyak untuk keperluan analisis, dengan menggunakan cara mengambil data dari sebuah website, *dataset* yang telah dikumpulkan berupa ID twitter, *date and time*, *tweet* dan *username twitter*.

Proses untuk mendapatkan hasil sentimen dari sebuah data dengan menggunakan metode yang diinginkan harus melalui beberapa tahapan, tahapan tersebut dapat dilihat pada gambar 1.



Gambar 1 *Flowchart* Diagram Sistem

2.1 Text Mining

Text Mining merupakan penemuan pengetahuan di database dalam bentuk tekstual (*knowledge discovery in textual database* atau disingkat dengan KDT), bisa disebut juga penggalian atau pencarian data-data yang berbentuk teks yang di antaranya adalah ketertarikan terhadap pengetahuan yang baru dibuat, diartikan sebagai bagian dari proses penggalian atau pencarian data *text* yang sebelumnya tidak diketahui, sehingga dapat dimengerti, mempunyai potensi dan pola praktis atau pengetahuan dari koleksi data teks atau *corpus masif* (kumpulan teks yang menangkap penggunaan bahasa dalam bentuk tertulis atau lisan secara utuh dan padat) dan tidak terstruktur.

Text mining sebagai ilmu pengetahuan cabang dari *data mining*, dipercaya memiliki nilai komersial yang jauh lebih tinggi dibandingkan *data mining* itu sendiri, karena 80% pada setiap perusahaan terdapat dokumen informasi dalam bentuk teks. Namun, penggalian dan pencarian data teks lebih kompleks dengan pola text tidak terstruktur selalu menyulitkan. *Text mining* merupakan lingkup penelitian yang komprehensif, yang masuk hampir disetiap lini kehidupan kita. Pada makalah ini akan dijelaskan tentang sedikit sejarah *text mining* serta tujuan dan maksud penelitian. Dan akan dijelaskan juga beberapa pola atau model secara umum yang, serta melakukan pengelompokan masalah-masalah yang sering muncul pada *text mining*. Selain itu penulis juga berusaha membuat ringkasan dalam ulasan ini [6].

2.2 Analisis Sentimen

Analisis sentimen (*sentiment analysis*) adalah proses menganalisis potongan teks secara online dengan menentukan nada emosional yang mereka bawa, apakah itu positif, negatif atau netral. Sederhananya, analisis sentimen dapat membantu untuk menemukan sikap seseorang terhadap suatu topik [7]. Sentimen positif dapat diungkapkan melalui kata-kata seperti “baik”, “luar biasa”, “keren”, atau “mantap”. Sentimen negatif dapat diungkapkan melalui kata-kata seperti “buruk”, “memalukan”, atau “mengerikan”. Beberapa keuntungan ketika melakukan analisis sentimen, antara lain:

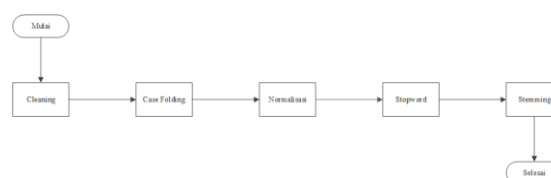
1. Dapat mengetahui pemahaman yang lebih baik tentang bagaimana perasaan seseorang terhadap objek.
2. Mendapat wawasan yang membantu untuk meningkatkan suatu hal.
3. Menjadikan sesuatu lebih responsif terhadap *feedback* seseorang.
4. Bereaksi lebih cepat terhadap sentimen negatif dan positif.
5. Dapat memantau reputasi sesuatu secara *real-time*.
6. Membuat sesuatu senang dengan selalu mengutamakan perasaan seseorang.

Informasi dari sentimen seseorang dapat berguna bagi pemilik bisnis organisasi, lembaga ataupun lainnya yang ingin memahami bagaimana perasaan seseorang tentang mereka.

2.3 Preprocessing

Preprocessing merupakan salah satu tahapan yang penting untuk data pada proses mining. Data yang digunakan dalam proses mining tidak selamanya dalam kondisi yang ideal untuk diproses. Terkadang pada data tersebut terdapat berbagai permasalahan yang dapat mengganggu hasil dari proses mining itu sendiri seperti diantaranya adalah *missing value*, data *redundant*, *outliers*, ataupun format data yang tidak sesuai dengan sistem. Oleh karenanya untuk mengatasi permasalahan tersebut dibutuhkan tahap *Preprocessing*.

Preprocessing merupakan salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data [8]. Tahapan *preprocessing* bisa dilihat pada gambar 2.



Gambar 2 Tahapan *Preprocessing*

Terdapat 5 tahapan *preprocessing* dalam penelitian ini, antara lain:

1. *Cleaning*, pembersihan yang dilakukan yaitu menghapus angka menghapus karakter selain huruf, menghapus username pengguna atau mentions (@), menghapus hashtag (#), menghapus URL atau link dari

setiap komentar serta simbol-simbol yang ada pada kalimat.

2. *Casse Folding*, merubah sebuah kata pada keseluruhan kalimat menjadi huruf kecil.
3. Normalisasi, diperlukan untuk melakukan transformasi pada sebuah kata atau mengubah kata asli menjadi format yang memungkinkan pemrosesan kata yang efisien.
4. *Stopword*, merupakan pembuangan kata-kata tidak penting. *Stopword* diproses pada sebuah kalimat jika mengandung kata-kata yang sering keluar dan di anggap tidak penting seperti waktu, penghubung, dan lain sebagainya.
5. *Stemming*, proses menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata (*root word*). Misalnya kata “mendengarkan”, “dengarkan” dan “didengarkan” akan ditransformasi menjadi kata “dengar”.

2.4 Term Frequency

Term frequency (TF) merupakan sebuah metode sederhana dari pembobotan kata. Term frequency memperhatikan kemunculan term di dalam dokumen sebuah dokumen. Pembobotan kata bertujuan untuk memberikan nilai atau bobot pada suatu *term* yang terdapat pada suatu dokumen[9].

2.5 Naïve Bayes Classifier

Naïve Bayes Classifier adalah metode klasifikasi statistik yang dapat memprediksi probabilitas keanggotaan kelas, seperti probabilitas bahwa sampel yang diberikan termasuk dalam kelas tertentu. Metode ini yang akan digunakan pada penelitian ini untuk klasifikasi data yang diambil dari Twitter dan data tersebut akan diklasifikasikan menjadi kelas positif dan kelas negatif. *Naïve Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes, yaitu memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa sebelumnya. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat dari masing-masing kondisi atau kejadian. Metode ini sangat cocok digunakan sebagai pengklasifikasian sentimen pada Tugas Akhir

ini dikarenakan memiliki beberapa kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi[10].

$$P(C_j) = \frac{N_j}{N} \quad (1)$$

Persamaan (1) untuk mencari nilai dari probabilitas prio, memiliki keterangan sebagai berikut:

N_j = jumlah data pada suatu kategori

N = jumlah total data

$$P(W_i|C_j) = \frac{P(C_j|W_i)P(C_j)}{P(W_i)} \quad (2)$$

Persamaan (2) merupakan notasi umum untuk mencari nilai dari probabilitas posterior, memiliki keterangan sebagai berikut:

$P(W_i|C_j)$ = peluang munculnya kategori j ketika kata i muncul

$P(C_j|W_i)$ = peluang kata i termasuk dalam kategori j

$P(W_i)$ = peluang munculnya kategori j probabilitas kejadian

$P(C_j)$ = probabilitas kemunculan kategori j

$$P(W_i|C_j) = \frac{n_i + 1}{n + |\text{kosakata}|} \quad (3)$$

Persamaan (3) merupakan bentuk dari persamaan (2) dari kasus analisis sentiment untuk menghitung setiap kosa kata dalam kategori negatif dan positif.

$$V_{MAP} = \text{argmax } P(C_j) \prod P(W_i|C_j) \quad (4)$$

Persamaan (4) merupakan perhitungan mencari nilai argumen maksima dari setiap kategori untuk menentukan klasifikasi sentimen.

2.6 Confusion Matrix

Evaluasi pada suatu model yang terbentuk akan diuji dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat *accuracy*, *precision*, *recall*, *specificity* dan *f1-score*. *Confusion Matrix* akan menggambarkan hasil dari kinerja sebuah model atau metode yang digunakan, mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar dan prediksi negatif yang salah. *Accuracy*, *precision*, *recall*, *specificity* dan *f1-score* akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif), untuk mencari nilai dari *Accuracy*, *precision*, *recall*, *specificity* dan *f1-score* harus mengetahui nilai *True Positive*,

True Negative, *False Positive* dan *False Negative*[11], sebagai berikut:

Tabel 1 *Confusion Matrix*

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

Keterangan:

1. TP = *True Positive*, yaitu jumlah data kelas positif yang terklasifikasi sebagai positif oleh sistem.
2. TN = *True Negative*, yaitu jumlah data kelas negatif yang terklasifikasi sebagai negatif oleh sistem.
3. FP = *False Positive*, yaitu jumlah data kelas negatif yang terklasifikasi sebagai positif oleh sistem.
4. FN = *False Negative*, yaitu jumlah data kelas positif yang terklasifikasi sebagai negatif oleh sistem.

3. HASIL DAN PEMBAHASAN

Hasil yang telah didapat setelah melakukan penelitian mengenai analisis sentimen terhadap PSSI setelah tragedi Kanjuruhan Malang yaitu dari 1500 *dataset* yang diambil dari tanggal 2 Oktober 2022 sampai 16 Oktober 2022. Salah satu sampel tweet “Tanggungjawab moral, seluruh pimpinan PSSI silahkan mundur!!!” dari *dataset* untuk dilakukannya proses *preprocessing* sebagai berikut:

Tabel 2 Tahap *Preprocessing*

Tahap <i>Preprocessing</i>	Hasil
<i>Cleaning</i>	Tanggungjawab moral seluruh pimpinan PSSI silahkan mundur
<i>Case Folding</i>	tanggungjawab moral seluruh pimpinan pssi silahkan mundur
Normalisasi	tanggung jawab moral seluruh pimpinan pssi silahkan mundur
<i>Stopword</i>	tanggung moral pimpinan silahkan mundur
<i>Stemming</i>	tanggung moral pimpin silah mundur

Dari hasil *preprocessing* tersebut maka akan dilakukan proses perhitungan untuk melakukan klasifikasi sentimen sebagai berikut:

Mencari nilai prior menggunakan persamaan (1).

$$P(\text{Label} = \text{positif}) = \frac{380}{1200} = 0.316666$$

$$P(\text{Label} = \text{negatif}) = \frac{820}{1200} = 0.683333$$

Kemudian dicari nilai probabilitas posterior pada setiap kata dari *dataset training* yang sudah dilakukan proses *preprocessing* pada tweet berlabel negatif dan positif. Salah satu perhitungan probabilitas posterior pada sebuah tweet menggunakan persamaan (4).

Kata Tanggung:

$$P(\text{tanggung} | \text{positif})$$

$$= \frac{11 + 1}{3448 + 10730} = \frac{12}{14178} = 0.000846$$

$$P(\text{tanggung} | \text{negatif})$$

$$= \frac{43 + 1}{7282 + 10730} = \frac{44}{18012} = 0.002442$$

Kata Moral:

$$P(\text{budaya} | \text{positif})$$

$$= \frac{3 + 1}{3448 + 10730} = \frac{4}{14178} = 0.000282$$

$$P(\text{budaya} | \text{negatif})$$

$$= \frac{3 + 1}{7282 + 10730} = \frac{4}{18012} = 0.000222$$

Kata Pimpin:

$$P(\text{kotor} | \text{positif})$$

$$= \frac{5 + 1}{3448 + 10730} = \frac{6}{14178} = 0.000423$$

$$P(\text{kotor} | \text{negatif})$$

$$= \frac{8 + 1}{7282 + 10730} = \frac{9}{18012} = 0.000499$$

Kata Silah:

$$P(\text{kusam} | \text{positif})$$

$$= \frac{3 + 1}{3448 + 10730} = \frac{4}{14178} = 0.000282$$

$$P(\text{kusam} | \text{negatif})$$

$$= \frac{1 + 1}{7282 + 10730} = \frac{2}{18012} = 0.000111$$

Kata Mundur:

$$P(\text{kusam} | \text{positif})$$

$$= \frac{33 + 1}{3448 + 10730} = \frac{34}{14178} = 0.002398$$

$P(\text{kusam}|\text{negatif})$

$$= \frac{150 + 1}{7282 + 10730} = \frac{151}{18012} = 0.008383$$

Setelah sudah ditentukan nilai probabilitas posterior setiap kata kemudian dijumlahkan dengan cara melakukan perkalian.

$P(\text{tweet}|\text{negatif})$

$$\begin{aligned} &= P(\text{tanggung} = 11|\text{positif}) \times P(\text{moral} \\ &\quad = 3|\text{positif}) \times P(\text{pimpin} \\ &\quad = 4|\text{positif}) \times P(\text{silah} \\ &\quad = 3|\text{positif}) \times P(\text{mundur} \\ &\quad = 11|\text{positif}) \end{aligned}$$

$$= 0.000846 \times 0.000282 \times 0.000423 \\ \times 0.000282 \times 0.002398$$

$P(\text{tweet}|\text{negatif})$

$$\begin{aligned} &= P(\text{tanggung} = 43|\text{negatif}) \times P(\text{moral} \\ &\quad = 3|\text{negatif}) \times P(\text{pimpin} \\ &\quad = 8|\text{negatif}) \times P(\text{silah} \\ &\quad = 1|\text{negatif}) \times P(\text{mundur} \\ &\quad = 150|\text{negatif}) \end{aligned}$$

$$= 0.002442 \times 0.000222 \times 0.000499 \\ \times 0.000111 \times 0.008383$$

$$= 2.51722e - 16$$

Setelah nilai tersebut sudah dihitung langkah terakhir mencari nilai argumen maksimal dari kategori positif dan negatif untuk melakukan klasifikasi sentimen menggunakan persamaan (4).

$V_{MAP}(\text{tweet}|\text{positif})$

$$= P(C_j|\text{positif}) \times P(\text{class} = \text{"positif"})$$

$$= 0.316666 \times 6.8243E - 17$$

$$= 2.16103e - 17$$

$V_{MAP}(\text{tweet}|\text{negatif})$

$$= P(C_j|\text{negatif}) \times P(\text{class} = \text{"negatif"})$$

$$= 0.683333 \times 2.51722e - 16$$

$$= 1.7201e - 16$$

Dari perhitungan argumen maksimal pada kelas positif dan kelas negatif di tweet tersebut, jadi dapat diketahui bahwa tweet

tersebut dalam menggunakan metode *Naïve Bayes Classifier* mendapatkan sentimen negatif dikarenakan nilai argumen maksimal kelas negatif lebih besar dari kelas positif. Klasifikasi tersebut juga berlaku pada semua *dataset*.

Dapat dilakukannya pengujian menggunakan *confusin matrix* untuk melihat efektivitas *Naïve Bayes Classifier* sebagai berikut:

Perhitungan pembagian *dataset* 80:20

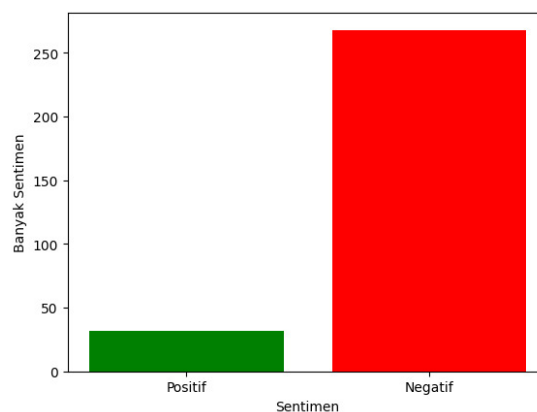
$$\begin{aligned} \text{Data Training} &= 1500 \times 80\% \\ &= 1200 \end{aligned}$$

$$\begin{aligned} \text{Data Testing} &= 1500 \times 20\% \\ &= 300 \end{aligned}$$

Dari banyaknya *dataset testing* diketahui TP sebanyak 21, TN sebanyak 193, FP sebanyak 11 dan FN sebanyak 75.

Tabel 3 Pengujian

Label Aktual	Label Prediksi		
		Positif	Negatif
	Positif	21	75
Negatif	11	193	



Gambar 3 Diagram Perbandingan Hasil Prediksi

Kemudian melakukan evaluasi *accuracy*, *precision*, *recall*, *specificity* dan *f1-score*. Sebagai berikut:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\ &= \frac{21 + 193}{21 + 193 + 11 + 75} \times 100\% \\ &= \frac{214}{300} \times 100\% \\ &= \frac{214}{300} \times 100\% \\ &= 71,33333\% \end{aligned}$$

Perhitungan nilai dari *accuracy* sebesar 0,7133333 atau 71,33333% dengan menggunakan metode *Naïve Bayes Classifier* dengan data *training* sebanyak 10% dari 1500 *dataset* yang ada.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{21}{21 + 11} \times 100\% \\ &= \frac{21}{32} \times 100\% \\ &= 65,625\% \end{aligned}$$

Perhitungan nilai dari *precision* sebesar 0,62625 atau 62,625% dengan menggunakan metode *Naïve Bayes Classifier* dengan data *training* sebanyak 20% dari 1500 *dataset* yang ada.

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{21}{21 + 75} \times 100\% \\ &= \frac{21}{96} \times 100\% \\ &= 21,875\% \end{aligned}$$

Perhitungan nilai dari *recall* sebesar 0,21875 atau 21,875% dengan menggunakan metode *Naïve Bayes Classifier* dengan data *training* sebanyak 20% dari 1500 *dataset* yang ada.

$$\begin{aligned} \text{Specificity} &= \frac{TN}{TN + FP} \times 100\% \\ &= \frac{193}{193 + 11} \times 100\% \\ &= \frac{193}{204} \times 100\% \\ &= 94,6078\% \end{aligned}$$

Perhitungan nilai dari *specificity* sebesar 0,946078 atau 94,6078% dengan menggunakan metode *Naïve Bayes Classifier* dengan data *training* sebanyak 20% dari 1500 *dataset* yang ada.

$$\begin{aligned} \text{F1 - score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \\ &= \frac{2 \times 0,65625 \times 0,21875}{0,65625 + 0,21875} \times 100\% \\ &= \frac{1,3125 \times 0,21875}{0,65625 + 0,21875} \times 100\% \end{aligned}$$

$$\begin{aligned} &= \frac{0,287109375}{0,875} \times 100\% \\ &= 32,8125\% \end{aligned}$$

Perhitungan nilai dari *f1-score* sebesar 0,328125 atau 32,8125% dengan menggunakan metode *Naïve Bayes Classifier* dengan data *training* sebanyak 20% dari 1500 *dataset* yang ada.

Tabel 4 Evaluasi

Evaluasi	Hasil
<i>Accuracy</i>	71,33333%
<i>Precision</i>	65,625%
<i>Recall</i>	21,875%
<i>Specificity</i>	94,6078%
<i>F1-score</i>	32,8125%

Jadi dari evaluasi yang didapatkan dari hasil klasifikasi menggunakan metode *Naïve Bayes Classifier* didapatkan efektivitas yang cukup baik dilihat dari evaluasi yang telah dilakukan.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah didapat maka metode *Naïve Bayes Classifier* dapat dikatakan cukup baik dalam melakukan klasifikasi dari penelitian analisis sentimen karena memiliki *accuracy* sebesar 71,3% dari *dataset* sebanyak 1500 dengan pembagian 80% untuk data *training* dan 20% data *testing*. Banyaknya klasifikasi menggunakan metode *Naïve Bayes Classifier* dari data *testing* yang digunakan menampilkan sentimen positif sebanyak 32 dan negatif sebanyak 268 hal ini menunjukkan banyak masyarakat yang berpendapat negatif terhadap PSSI.

5. SARAN

Saran dari penelitian untuk dapat meningkatkan efektivitas metode *Naïve Bayes Classifier* dengan cara menambahkan kata normalisasi dan *stopword* serta pengujian lebih lanjut.

DAFTAR PUSTAKA

- [1] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020.
- [2] R. Ghaniy and K. Sihotang, "Penerapan Metode Naive Bayes Classifier," *J. Ilm. Teknol. - Inf. dan Sains*, vol. 9, no. 1, pp. 63–72, 2019.
- [3] A. D. Sukma, Mayasari, and F. Hariyanto, "Analisis Wacana Kritis Pernyataan Mantan Ketua Umum Pssi Dan Gubernur Sumatera Utara 'Kalau Wartawan Nya Baik Nanti Timnas Nya Baik' Dalam Tayangan Youtube Kompas Tv," *JLT – J. Linguist. Terap.*, vol. 9, no. November, 2019.
- [4] R. Madjulekka, "Tragedi Kanjuruhan, Siapa Bertanggung Jawab? – Jurnal IDN," *Jurnal-IDN*, 2022. <https://www.jurnal-idn.com/tragedi-kanjuruhan-siapa-bertanggung-jawab/> (accessed Nov. 10, 2022).
- [5] D. Noorca, "Komnas HAM Tegaskan Gas Air Mata Pemicu Timbulnya Korban Tragedi Kanjuruhan - Suara Surabaya," *suarasurabaya*, 2022. [Online]. Available: <https://www.suarasurabaya.net/kelanakota/2022/komnas-ham-tegaskan-gas-air-mata-pemicu-timbulnya-korban-tragedi-kanjuruhan/> (Accessed Nov. 10, 2022).
- [6] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," *J. JUPITER*, vol. 13, no. 1, p. 66, 2021.
- [7] S. Sari, U. Khaira, P. Pradita, and T. S. Tri, "Analisis Sentimen Terhadap Komentar Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength," *IJIRSE (Indonesian J. Inform. Res. Softw. Eng.)*, vol. 1, no. 1, pp. 71–78, 2021.
- [8] M. A. Nur and N. Wardhani, "Optimasi Normalisasi Kata Pada Data Twitter Untuk Meningkatkan Akurasi Analisis Sentimen (Studi Kasus Respon Masyarakat Terhadap Layanan Teman Bus)," vol. 07, no. 04, pp. 237–243, 2022.
- [9] M. K. Anam, B. N. Pikir, and M. B. Firdaus, "Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 139–150, 2021.
- [10] M. Astiningrum, M. Haniah, and Y. rahmat yoga Pradana, "Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter," *Semin. Inform. Apl. Polinema*, p. 35— 39, 2020.
- [11] A. B. Yilmaz, Y. S. Taspinar, and M. Koklu, "Classification of Malicious Android Applications Using Naive Bayes and Support Vector Machine Algorithms," *IJISAE*, vol. 10, no. 2, pp. 269–274, 2022.