

KLASIFIKASI DAN ANALISIS SENTIMEN PADA DATA TWITTER MENGUNAKAN ALGORITMA NAÏVE BAYES (STUDI KASUS: TIMNAS INDONESIA SENIOR, U-23, DAN U-19)

Reino Prajamukti¹, Jayanta², Mayanda Mega Santoni³
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Email: reinopraja@gmail.com¹, jayanta@upnvj.ac.id², megasantoni@upnvj.ac.id³
Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Jawa Barat 12450

Abstrak

Media sosial tidak hanya berfungsi sebagai media komunikasi yang efektif, namun juga dapat menjadi wadah untuk menampung opini masyarakat. Salah satu media sosial yang banyak dipakai untuk mewadahi opini-opini tersebut adalah Twitter. Pengguna Twitter di Indonesia sering mengutarakan pendapatnya pada semua bidang termasuk olahraga, khususnya Timnas Sepakbola Indonesia. Berbagai macam sentimen terhadap Timnas Sepakbola Indonesia ini dapat dilihat di Twitter. Berlatarkan kondisi tersebut, maka diperlukan penelitian tentang opini masyarakat tentang performa Timnas Sepakbola Indonesia. Salah satu cara yang dilakukan adalah dengan melakukan analisis sentimen terhadap Timnas Sepakbola Indonesia di jejaring sosial Twitter menggunakan metode klasifikasi dan algoritma *Naïve Bayes* untuk mengklasifikasikan tweet positif atau negatif yang masyarakat berikan tentang Timnas Sepakbola Indonesia. Hasil dari analisis sentimen menggunakan algoritma *Naïve Bayes* pada klasifikasi *tweet* tentang opini terhadap Timnas Sepakbola Indonesia berjalan dengan baik dengan nilai akurasi 83%, kemudian nilai *precision* nilai positif sebesar 86%, nilai *precision* nilai negatif sebesar 81%, nilai *recall* sebesar 78%, dan nilai *specificity* penelitian ini 87,5% dengan menggunakan metode *confusion matrix* berdasarkan data yang diambil pada Januari sampai Mei 2021.

Kata Kunci: *Twitter, Analisis Sentimen, Timnas Sepakbola Indonesia, Algoritma Naïve Bayes*

1 Pendahuluan

Timnas Sepakbola Indonesia pada jejaring sosial Twitter. Pada awal tahun 2021 sampai Mei 2021 Timnas Sepakbola Indonesia senior, usia 23 dan usia 19 tahun sedang banyak menjalani training camp di luar negeri, uji coba, hingga klasifikasi piala dunia 2022. Banyak dari pengguna media sosial Twitter yang membahas tentang Timnas Sepakbola Indonesia ini. Penggunaan algoritma *Naïve Bayes* digunakan karena dapat mengolah data yang berjumlah banyak dan memiliki akurasi ketepatan yang cukup tinggi. Dari penelitian yang dilakukan [1], hasil pengujian algoritma *Naïve Bayes* menggunakan 80% data *training* dan 20% data *testing* didapatkan nilai akurasi sebesar 84%, saat menggunakan 90% data *training* dan 10% data *testing* didapatkan nilai akurasi sebesar 87%. Penelitian tersebut membuktikan bahwa tingkat akurasi ketepatan yang tinggi algoritma *Naïve Bayes*.

Berlatarkan kondisi tersebut, maka penulis melakukan penelitian tentang sentimen pengguna media sosial Twitter tentang performa Timnas Indonesia. Salah satu cara yang dilakukan adalah dengan melakukan klasifikasi dan analisis sentimen terhadap *tweet* yang membahas performa Timnas Indonesia di media sosial Twitter. Penelitian ini menggunakan metode klasifikasi dan algoritma *Naïve Bayes* untuk mengklasifikasikan *tweet* positif dan negatif yang masyarakat berikan terhadap performa Timnas Indonesia.

2 Tinjauan Pustaka

2.1 Twitter

Twitter merupakan situs jejaring sosial yang dapat berinteraksi dengan pengguna lainnya dari perangkat komputer ataupun ponsel dimanapun dan kapanpun. Sebagai sebuah situs jejaring sosial, Twitter memberikan akses kepada penggunanya untuk dapat mengirimkan pesan teks, foto, dan video [2].

2.2 Analisis Sentimen

Tujuan dari analisis sentimen adalah untuk mendapatkan informasi dari sebuah platform, yang mana platform dalam penelitian ini adalah Twitter. Seringkali, informasi yang didapat untuk sebuah penelitian memiliki banyak

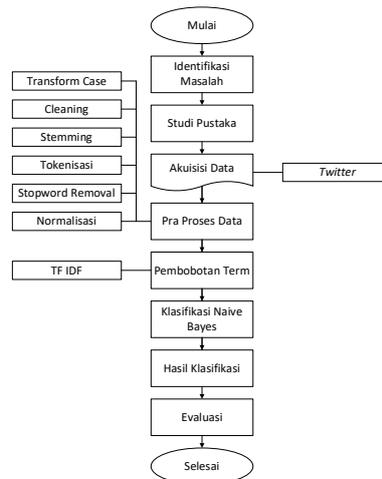
noise yang membuat informasi tersebut tidak terstruktur. Dengan menggunakan analisis sentimen, informasi yang sebelumnya tidak terstruktur dapat diolah menjadi informasi yang lebih terstruktur.

2.3 Algoritma Naïve Bayes

Naïve Bayes Classifier adalah suatu algoritma dalam metode klasifikasi sebagai pengetahuan pembelajaran terhadap sistem yang digunakan. Menurut [3] Klasifikasi *Naïve Bayes* merupakan proses klasifikasi yang menggunakan probabilitas bersyarat untuk menghitung probabilitas suatu peristiwa/objek ke dalam kategori yang tersedia/diberikan. Pada klasifikasi *Naïve Bayes* dibutuhkan dua himpunan utama, yaitu *training set* dan *testing set*. *Training set* digunakan sebagai data latih untuk mengidentifikasi objek yang akan diklasifikasi, sedangkan *testing set* akan digunakan sebagai pengukur akurasi dan ketepatan dari klasifikasi *Naïve Bayes*.

3 Metode Penelitian

Dalam melakukan penelitian ini, penulis menjabarkan alur penelitian yang dilakukan secara bertahap seperti yang tertera pada gambar di bawah ini.



Gambar 1. Metode Penelitian

1. Identifikasi Masalah
2. Studi Literatur
3. Akuisisi Data Dari Twitter
4. Pra Proses Data
5. Pembobotan Term TF-IDF
6. Klasifikasi Naïve Bayes
7. Hasil Klasifikasi
8. Evaluasi

4 Hasil dan Pembahasan

4.1 Data

Saat proses crawling data dari Twitter penulis dengan menggunakan beberapa kata kunci seperti 'timnas', 'timnas u 23', 'timnas u 19', 'bagus kahfi', 'sty', 'shin tae yong', 'mental timnas', 'fisik timnas', '#BanggaSepakBolaKita', dan 'uji coba timnas'. Crawling data dilakukan secara bertahap dari tanggal 10 Februari 2021 sampai dengan 31 Mei 2021. Jumlah data yang digunakan pada penelitian ini dari hasil *crawling* data Twitter yaitu sebanyak 232 tweet.

```

"RT @siromuhkowoc: kok timnas indonesia sering bgt ya uji coba sama klub indonesia sendiri. kek gapernah liat inggris vs arsenal, italia vs\u2026"
"Tim nasional Indonesia persiapan SEA Games 2021 meraih kemenangan kedua dalam laga uji coba setelah menaklukkan Bal\u2026"
https://t.co/SAbau6Prz"
"Timnas kalo mau uji coba mending lawan liverpool deh, pasti menang"
  
```

Gambar 2. Hasil Crawling Data

Setelah proses *crawling* data, maka proses berikutnya yaitu pelabelan data. Proses pelabelan data tweet dilakukan oleh dua pihak secara manual dengan melakukan penilaian suatu data tweet berupa kalimat yang bernilai positif atau negatif berdasarkan dari sudut pandang subjektif masing-masing pihak. Jika ada perbedaan penilaian data tweet antar kedua belah pihak maka dilakukan diskusi untuk menentukan hasil kesepakatan yang akan diberikan terhadap data tweet tersebut. Setelah melakukan pelabelan, maka dapat diketahui label dari 232 data tweet tersebut terdiri dari 116 tweet positif dan 116 tweet negatif. Untuk mengevaluasi tingkat kesepakatan hasil pelabelan data oleh 2 pihak, maka digunakan perhitungan *kappa value*.

Pihak 1	Pihak 2	Jumlah Tweet
Positif	Positif	103
Positif	Negatif	9
Negatif	Positif	4
Negatif	Negatif	116

Gambar 3. Hasil Pelabelan Data

$$P(A) = \frac{103 + 116}{232} = 0.9439$$

$$P(\text{Negatif}) = \frac{9 + 4 + 116 + 116}{464} = 0.5280$$

$$P(\text{Positif}) = \frac{103 + 103 + 9 + 4}{464} = 0.4719$$

$$P(E) = (0.528)^2 + (0.4719)^2 = 0.5014$$

$$Kappa\ Value = \frac{(0.9439 - 0.5014)}{(1 - 0.5014)} = 0.8874$$

Gambar 4. Perhitungan Dengan Kappa Value

Berdasarkan evaluasi hasil pelabelan data menggunakan kappa value, didapatkan hasil sebesar 0.8874 maka kekuatan kesepakatan dari pelabelan data memiliki nilai yang sangat baik.

4.2 Pra Proses Data

Pada tahap ini, penulis melakukan enam tahapan dalam pra proses data yaitu *transform cases*, *cleaning*, *stemming*, tokenisasi, *stopword removal*, dan normalisasi. Pada Tabel 6 merupakan data yang penulis gunakan untuk dijadikan sampel pada tahap pra proses data.

Sampel Data Sebelum Pra Proses
@PSSI goal berkelas, proses membangun serangan dari bawah, kecerdikan syahrhan dan penempatan posisi yang bagus https://t.co/s8cz1FyNMq
Baru bela timnas u19 doang udah ngerasa kyk superstar, bela real madrid dulu baru kamu bikin skandal https://t.co/EkzyvP0PtQ

Gambar 5. Sampel Data Pra Proses

Setelah Transform Cases
@pssi goal berkelas, proses membangun serangan dari bawah, kecerdikan syahrhan dan penempatan posisi yang bagus https://t.co/s8cz1fynmq
baru bela timnas u19 doang udah ngerasa kyk superstar, bela real madrid dulu baru kamu bikin skandal https://t.co/ekzyvp0ptq

Gambar 6. Hasil Transform Cases

Setelah Cleaning
goal berkelas proses membangun serangan dari bawah kecerdikan syahrhan dan penempatan posisi yang bagus
baru bela timnas u doang udah ngerasa kyk superstar bela real madrid dulu baru kamu bikin skandal

Gambar 7. Hasil Cleaning

Setelah Stemming
goal kelas proses bangun serang dari bawah cerdik syahrhan dan tempat posisi yang bagus
baru bela timnas u doang udah ngerasa kyk superstar bela real madrid dulu baru kamu bikin skandal

Gambar 8. Hasil Stemming

Setelah Tokenisasi
['goal', 'kelas', 'proses', 'bangun', 'serang', 'dari', 'bawah', 'cerdik', 'syahrhan', 'dan', 'tempat', 'posisi', 'yang', 'bagus']
['baru', 'bela', 'timnas', 'u', 'doang', 'udah', 'ngerasa', 'kyk', 'superstar', 'bela', 'real', 'madrid', 'dulu', 'baru', 'kamu', 'bikin', 'skandal']

Gambar 9. Hasil Tokenisasi

Setelah <i>Stopword Removal</i>	Setelah Normalisasi
['goal', 'kelas', 'proses', 'bangun', 'serang', 'cerdik', 'syahrian', 'posisi', 'bagus']	['goal', 'kelas', 'proses', 'bangun', 'serang', 'cerdik', 'syahrian', 'posisi', 'bagus']
['bela', 'timnas', 'ngerasa', 'superstar', 'bela', 'real', 'madrid', 'baru', 'bikin', 'skandal']	['bela', 'timnas', 'merasa', 'superstar', 'bela', 'real', 'madrid', 'baru', 'bikin', 'skandal']

Gambar 10. Hasil *Stopword Removal*

Gambar 11. Hasil Normalisasi

4.3 Pembobotan TF-IDF

Setelah melakukan pra proses pada tahap ini penulis akan menggunakan data latih sebagai bahan untuk menguji data uji. Sebelum melakukan klasifikasi terhadap data latih dan data uji, penulis melakukan pembobotan TF-IDF.

Data Tweet		Kelas
D1	['moga', 'sukses', 'karir', 'sepakbola', 'nasional', 'baik']	Positif
D2	['yakin', 'timnas', 'indonesia', 'baik']	Positif
D3	['asal', 'banget', 'timnas', 'bobrok']	Negatif
D4	['mental', 'fisik', 'skواد', 'garuda', 'lemah']	Negatif

Gambar 12. Sampel Data Klasifikasi

Data selanjutnya penulis melakukan proses pembobotan TF-IDF. Pembobotan TF-IDF merupakan sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata dalam dokumen atau dalam sekelompok kata. Berikut merupakan contoh perhitungan TF-IDF pada kata “moga”:

- Jumlah kemunculan kata “moga” pada D1 yaitu 1

Sehingga $tf_{ij} = 1$

- Total seluruh dokumen sebanyak 4 dokumen (D1, D2, D3, D4)

Sehingga $D = 4$

- Dari 4 dokumen yang ada yaitu sebanyak 1 dokumen yang memuat kata “moga”, yaitu D1.

Sehingga $df_i = 1$

Dari uraian di atas perhitungan TF-IDF dari kata “moga”, dapat dihitung dengan persamaan rumus (3.1). Maka bobot kata “moga” adalah sebagai berikut.

$$w_{ij} = tf_{ij} \times \log \frac{D}{df_i}$$

$$w = 1 \times \log \frac{4}{1}$$

$$w = 0,602$$

Term	Dokumen				DF	IDF	TF IDF			
	D1	D2	D3	D4			D1	D2	D3	D4
Moga	1	0	0	0	1	0,602	0,602	0	0	0
Sukses	1	0	0	0	1	0,602	0,602	0	0	0
Karir	1	0	0	0	1	0,602	0,602	0	0	0
Sepakbola	1	0	0	0	1	0,602	0,602	0	0	0
Nasional	1	0	0	0	1	0,602	0,602	0	0	0
Baik	1	1	0	0	2	0,301	0,301	0,301	0	0
Yakin	0	1	0	0	1	0,602	0	0,602	0	0
Timnas	0	1	1	0	2	0,301	0	0,301	0,301	0
Indonesia	0	1	0	0	1	0,602	0	0,602	0	0
Asal	0	0	1	0	1	0,602	0	0	0,602	0
Banget	0	0	1	0	1	0,602	0	0	0,602	0
bobrok	0	0	1	0	1	0,602	0	0	0,602	0
Mental	0	0	0	1	1	0,602	0	0	0	0,602
Fisik	0	0	0	1	1	0,602	0	0	0	0,602
Skuad	0	0	0	1	1	0,602	0	0	0	0,602
Garuda	0	0	0	1	1	0,602	0	0	0	0,602
lemah	0	0	0	1	1	0,602	0	0	0	0,602

Gambar 13. Hasil Pembobotan TF-IDF

4.4 Proses Klasifikasi

Probabilitas class (prior) digunakan untuk menghitung hasil dari probabilitas data uji untuk setiap kelas.

$$P(\text{Positif} / \text{Negatif}) = \frac{d|\text{Positif}/\text{Negatif}|}{|\text{Jumlah data}|}$$

$$P(\text{Positif}) = \frac{2}{4}, P(\text{Negatif}) = 0,5$$

Maka dapat diketahui bahwa probabilitas kelas positif yaitu 0.5

$$P(\text{Positif}) = \frac{2}{4}, P(\text{Negatif}) = 0.5$$

Maka dapat diketahui bahwa probabilitas kelas negatif yaitu 0.5

Keterangan:

$P(\text{Positif} / \text{Negatif})$: Peluang kemunculan data dari class positif dan negatif

$d|\text{Positif} / \text{Negatif}|$: Jumlah seluruh data dari kelas positif atau negatif

$|\text{Jumlah data}|$: Jumlah data latih.

4.4.1 Data Latih

Pada tahap ini, sebelumnya telah diperoleh data dari pembobotan TF-IDF di Tabel 14 berupa bobot dari setiap kata atau *tweet* yang selanjutnya akan dimasukkan ke dalam proses klasifikasi dengan algoritma *Naïve Bayes*.

$$P(W_i|c) = \frac{\text{Count}(W_i, c) + 1}{|c| + |V|}$$

Keterangan

$P(W_i|c)$: Probabilitas kata W_i pada kelas c

$Count(W_i, c)$: Jumlah kemunculan kata W_i pada kelas c .

$|c|$: Jumlah kata pada kelas c

$|V|$: Jumlah semua kata

Dari hasil Tabel 14 maka diketahui sebagai berikut:

Kelas : Positif = 10, Negatif = 9, $|V| = 17$

$$P(\text{"moga"}|\text{Positif}) = \frac{(\text{"moga"}|\text{Positif})+1}{(\text{Positif})+17} = \frac{0,602+1}{10+17} = 0.0593$$

$$P(\text{"moga"}|\text{Negatif}) = \frac{(\text{"moga"}|\text{Negatif})+1}{(\text{Negatif})+17} = \frac{0+1}{9+17} = 0.0384$$

Hasil Data Latih					
Term	TF-IDF	Positif	Negatif	$P(W_i Pos)$	$P(W_i Neg)$
Moga	0,602	1	0	0.0593	0.0384
Sukses	0,602	1	0	0.0593	0.0384
Karir	0,602	1	0	0.0593	0.0384
sepakbola	0,602	1	0	0.0593	0.0384
Nasional	0,602	1	0	0.0593	0.0384
Baik	0,301	1	0	0.0481	0.0384
Yakin	0,602	1	0	0.0593	0.0384
Timnas	0.301	1	1	0.0481	0.05
Indonesia	0,602	0	1	0.0593	0.0384
Asal	0,602	0	1	0.037	0.0616
Banget	0,602	0	1	0.037	0.0616
Bobrok	0,602	0	1	0.037	0.0616
Mental	0,602	0	1	0.037	0.0616
Fisik	0,602	0	1	0.037	0.0616
Skuad	0,602	0	1	0.037	0.0616
Garuda	0,602	0	1	0.037	0.0616
Lemah	0.602	0	1	0.037	0.0616

Gambar 14. Hasil Perhitungan Data Latih

4.4.2 Data Uji

Pada proses data uji, sebelumnya sudah didapatkan probabilitas setiap kata pada data latih maka selanjutnya dilakukan pengujian dengan sampel data uji. Berikut merupakan sampel data untuk melakukan pengujian dari data latih.

Data Uji
['menpora', 'nilai', 'laga', 'timnas', 'momentum', 'baik', 'sepakbola', 'nasional']

Gambar 15. Sampel Data Uji

Berikut merupakan perhitungan probabilitas setiap data pada kelas positif dan negatif.

$$P(\text{"baik"}|\text{Positif}) = \frac{(\text{"baik"}|\text{Positif})+1}{(\text{Positif})+17} = \frac{0.301+1}{10+17} = 0.0481$$

$$P(\text{"baik"}|\text{Negatif}) = \frac{(\text{"baik"}|\text{Negatif})+1}{(\text{Negatif})+17} = \frac{0+1}{9+17} = 0.0384$$

Hasil Data Uji					
Term	TF IDF	Positif	Negatif	$P(W_i Pos)$	$P(W_i Neg)$
Menpora	0	0	0	0.037	0.0384
Nilai	0	0	0	0.037	0.0384
Laga	0	0	0	0.037	0.0384
Timnas	0.301	1	1	0.0481	0.05
Momentum	0	0	0	0.037	0.0384
Baik	0.301	1	0	0.0481	0.0384
Sepakbola	0.602	1	0	0.0593	0.0384
Nasional	0.602	1	0	0.0593	0.0384

Gambar 16. Hasil Perhitungan Data Uji

Tahap selanjutnya yaitu perhitungan untuk menentukan *class* data uji tersebut.

$$c_{MAP} = \underset{c \in V}{\arg \max} P(c) \prod_i P(W_i|c)$$

Keterangan :

$P(c)$: Peluang kemunculan suatu dokumen yang memiliki kelas c .
 $P(W_i|c)$: Peluang kemunculan W_i pada kelas c

Berikut merupakan perhitungan probabilitas data uji pada kelas positif:

$$\begin{aligned} &P(\text{data uji}|\text{Positif})= \\ &P(\text{Positif}) \times P(\text{"Menpora"}|\text{Positif}) \times P(\text{"Nilai"}|\text{Positif}) \times P(\text{"Laga"}|\text{Positif}) \times P(\text{"Timnas"}|\text{Positif}) \times \\ &P(\text{"Momentum"}|\text{Positif}) \times P(\text{"Baik"}|\text{Positif}) \times P(\text{"Sepakbola"}|\text{Positif}) \times P(\text{"Nasional"}|\text{Positif}) \\ &= 0.5 \times 0.037 \times 0.037 \times 0.037 \times 0.0481 \times 0.037 \times 0.0481 \times 0.0593 \times 0.0593 \\ &= 0,0000000000762388681468682645 \end{aligned}$$

Berikut merupakan perhitungan probabilitas data uji pada kelas negatif:

$$\begin{aligned} &P(\text{data uji}|\text{Negatif})= \\ &P(\text{Negatif}) \times P(\text{"Menpora"}|\text{Negatif}) \times P(\text{"Nilai"}|\text{Negatif}) \times P(\text{"Laga"}|\text{Negatif}) \times P(\text{"Timnas"}|\text{Negatif}) \times \\ &P(\text{"Momentum"}|\text{Negatif}) \times P(\text{"Baik"}|\text{Negatif}) \times P(\text{"Sepakbola"}|\text{Negatif}) \times P(\text{"Nasional"}|\text{Negatif}) \\ &= 0.5 \times 0.0384 \times 0.0384 \times 0.0384 \times 0.05 \times 0.0384 \times 0.0384 \times 0.0384 \times 0.0384 \\ &= 0,0000000000307792887033102336 \end{aligned}$$

4.5 Hasil Klasifikasi

Hasil dari proses klasifikasi menggunakan algoritma Naïve Bayes secara keseluruhan dapat dilihat pada Tabel 18. Dalam proses klasifikasi tersebut penulis menggunakan pembobotan TF-IDF dan dari 232 data yang diperoleh, memuat 116 kelas positif dan 116 kelas negatif. Data tersebut dibagi menjadi data latih dan data uji dengan perbandingan 80:20 sehingga didapatkan jumlah data latih sebanyak 185 data dan data uji sebanyak 47 data.

	Label Positif	Label Negatif	Total Jumlah
Data Latih	93	92	185
Data Uji	23	24	47
Total Jumlah	116	116	232

Gambar 17. Pembagian Data

Selanjutnya data uji akan dijadikan model klasifikasi bagi data latih menggunakan algoritma Naïve Bayes. Setelah semua proses dilakukan maka didapatkan nilai sebagai berikut:

Data yang diprediksi positif dan ternyata positif (True Positive) : 18
 Data yang diprediksi negatif dan ternyata negatif (True Negative) : 21
 Data yang diprediksi positif dan ternyata negatif (False Positive) : 3
 Data yang diprediksi negatif dan ternyata positif (False Negative) : 5

4.6 Evaluasi

Dari hasil klasifikasi keseluruhan yang sudah didapat, maka pada proses ini dilakukan untuk menguji hasil klasifikasi dengan menggunakan metode confusion matrix dengan sejumlah data yang sudah diuji. Pada tahap evaluasi ini penulis mencari nilai akurasi, precision, recall, dan specificity. Berikut adalah Tabel dari nilai evaluasi yang sudah dibulatkan.

Variable Evaluasi		Nilai
Akurasi		0.83
Precision	Positif	0.86
	Negatif	0.81
Recall	Positif	0.78
Specificity	Negatif	0.875

Gambar 18. Variabel Evaluasi

Nilai akurasi pada sistem ini adalah 83%, nilai *precision* positif pada penelitian ini sebesar 86%, nilai *precision* negatif pada penelitian ini sebesar 81%, nilai *recall* penelitian ini sebesar 78%, dan nilai *specificity* penelitian ini sebesar 87,5% dengan perbandingan 80:20 data latih dan data uji yang diambil pada kurun waktu bulan Februari sampai Mei 2021 dengan nilai *true positive* sebanyak 18, *true negative* 21, *false positive* 5, dan *false negative*.

5 Kesimpulan

Berikut merupakan kesimpulan yang diperoleh penulis.

1. Penerapan pembobotan TF-IDF dan algoritma *Naive Bayes* dapat digunakan untuk klasifikasi sentimen terhadap Timnas Indonesia.
2. Algoritma *Naive Bayes* pada klasifikasi *tweet* tentang sentimen terhadap Timnas Indonesia berjalan dengan baik dengan nilai akurasi 83%, kemudian nilai *precision* nilai positif sebesar 86%, nilai *precision* nilai negatif sebesar 81%, nilai *recall* sebesar 78%, dan nilai *specificity* penelitian ini 87,5% dengan menggunakan metode *confusion matrix* berdasarkan data yang diambil pada Februari sampai Mei 2021.

6 Saran

Berdasarkan hasil penelitian yang telah dilaksanakan, penulis memiliki beberapa saran untuk penelitian-penelitian berikutnya, yaitu:

1. Menggunakan proses negation handling untuk mencari kata yang bersifat negasi, karena kalimat yang mengandung kata negasi dapat mengubah nilai positif atau negatif dari data. Contohnya “tidak jauh” menjadi “dekat”.
2. Meningkatkan jumlah keseluruhan data *tweet* yang diharapkan dapat menambah tingkat akurasi dari klasifikasi oleh system yang sudah dibuat.

Referensi

- [1] Astiningrum, Mungki, Mamluatul Hani, Yanuar Rahmat, & Yoga Pradana. 2020. “Analisis Sentimen Tentang Opini Terhadap Performa Timnas Sepak Bola Indonesia Pada Twitter.” : 35–39.
- [2] Wili. 2015. “Distributed Twitter Crawler Universitas Pendidikan Indonesia | Repository.Upi.Edu | Perpustakaan.Upi.Edu.” : 1–6.
- [3] Kosasih, Rifki, and Achmad Fahrurrozi. 2017. “Pengklasifikasian Bunga Dengan Menggunakan Metode Isomap Dan Naive Bayes Classifier Flowers Classification By Using Isomap Method and Naive Bayes Classifier.” *Jurnal Ilmiah Informatika dan Komputer* 22(3): 171–7.