

# ANALISIS SENTIMEN TENTANG OPINI TERHADAP PERFORMA TIMNAS SEPAK BOLA INDONESIA PADA TWITTER

Mungki Astiningrum<sup>1</sup>, Mamluatul Hani'ah<sup>2</sup>, Yanuar Rahmat Yoga Pradana<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang  
<sup>1</sup>mungki.astiningrum@polinema.ac.id, <sup>2</sup>mamluatulhaniah@polinema.ac.id, <sup>3</sup>yanuar.ryp@gmail.com

**Abstrak**—Media sosial *Twitter* merupakan salah satu media yang digunakan untuk memberikan opini melalui *tweet*. *Tweet* tersebut dapat memuat opini dan komentar yang berkaitan dengan bidang ekonomi, sosial, hiburan, pendidikan, olahraga, dan lain- lain. Salah satu olahraga yang digemari di Indonesia adalah sepakbola khususnya Timnas Indonesia. Pengguna *Twitter* akan memberikan komentar dan opini tentang performa Timnas Indonesia melalui media sosial, salah satunya *Twitter*. Dalam opini dan komentar *supporter* terdapat penyampaian komentar yang beragam apabila Timnas mendapatkan hasil yang baik maka akan ada opini atau komentar bahagia maupun pujian dan jika Timnas mendapatkan hasil yang buruk maka akan ada opini dan komentar kritikan bahkan cacian. Dalam penelitian ini, dilakukan analisis sentimen terhadap *tweet* dengan mengklasifikasikan kedalam kategori positif atau negatif menggunakan algoritma *Naive Bayes*. *Dataset* didapatkan dari *Twitter API* dan *tweet* mengenai timnas sepak bola Indonesia dalam periode tertentu. Setelah mendapatkan *dataset* dari *Twitter* selanjutnya yang dilakukan adalah *preprocessing* dimana pada proses ini dilakukan *cleaning*, *casefolding*, normalisasi, *stopword* dan *stemming* lalu menghitung *term frequency* dokumen dan masuk ke tahap proses *training* menghasilkan model klasifikasi untuk proses *testing* sehingga dapat melakukan klasifikasi. Pada pengujian penelitian ini dilakukan tiga kali pengujian dan setiap pengujian presentase data *training* nya berbeda karena jumlah data *training* juga berpengaruh terhadap peningkatan nilai akurasi. Dari tiga kali pengujian didapatkan akurasi tertinggi yaitu 87%, *precision* positif tertinggi yaitu 93%, *precision* negatif tertinggi yaitu 85%, *recall* positif tertinggi yaitu 75%, dan *recall* negatif tertinggi yaitu 98% pada presentase 90% data *training* dan 10% data *testing*.

**Kata kunci**—Analisis Sentimen, Opini, Sepak Bola, *Naive Bayes*

## I. PENDAHULUAN

Media sosial *Twitter* adalah salah satu media komunikasi yang diminati oleh masyarakat di dunia. Hal ini dapat dilihat dari peningkatan pengguna *Twitter* yang tercatat di seluruh dunia. *Twitter* memiliki jumlah pengguna aktif sebesar 313 juta per bulan pada tahun 2016 dan sebagian besar pengguna mengakses *Twitter* melalui perangkat *mobile*, yaitu sebesar

82 persen. Jumlah pengguna yang banyak juga menimbulkan *tweet* yang banyak dari pengguna. Pengguna akan memberikan kabar terbaru atau komentar tentang hal yang sedang menjadi topik utama di dunia. Hal yang sedang menjadi topik utama dan banyak dikomentari oleh pengguna akan menimbulkan *trending* topik di *Twitter* [4].

Pengguna *Twitter* yang banyak akan menimbulkan peningkatan *tweet* yang di *posting*. Setiap *tweet* pada *Twitter* memiliki topik yang berbeda. *Tweet* tersebut dapat memuat opini dan komentar yang berkaitan dengan bidang ekonomi, sosial, hiburan, pendidikan, olahraga, dan lain- lain. Salah satu olahraga yang digemari di Indonesia adalah sepakbola khususnya Timnas Indonesia. Pengguna *Twitter* akan memberikan komentar dan opini tentang performa Timnas Indonesia melalui media sosial, salah satunya *Twitter*. Dalam opini dan komentar *supporter* terdapat penyampaian komentar yang beragam apabila Timnas mendapatkan hasil yang baik maka akan ada opini atau komentar bahagia maupun pujian dan jika Timnas mendapatkan hasil yang buruk maka akan ada opini dan komentar kritikan bahkan cacian. *Tweet* dari pengguna juga dapat menjadi evaluasi bagi pengelolaan Timnas Indonesia kedepannya agar performa sesuai yang diharapkan oleh *supporter*. Karena dukungan dari *supporter* sangat berpengaruh kepada Timnas mulai dari dukungan semangat, moril maupun pemasukan Timnas itu sendiri. Namun, pengguna akan mengalami kesulitan apabila melihat *tweet* secara langsung tanpa ada label *tweet* tersebut bernilai positif atau negatif. Sehingga diperlukan klasifikasi untuk memberikan kemudahan pada pengguna dalam melihat *tweet* yang bernilai positif atau negatif.

Penelitian tentang analisis sentiment pada dokumen *Twitter* telah banyak dilakukan salah satunya penelitian Analisis Sentimen Pada *Twitter* Mengenai Pasca Bencana Menggunakan Metode *Naive Bayes* Dengan Fitur *N-Gram* didapatkan hasil akurasi tertinggi pada unigram yaitu sebesar 93.33% dan bigram sebesar 86.67% [3]. Pada penelitian lain mengenai Analisis Sentimen *Twitter* Menggunakan Metode *Naive Bayes* Studi Kasus SAMSAT Kota Malang didapatkan hasil akurasi tertinggi pada setiap kategori positif, netral dan negatif masing- masing sebesar 82%, 92%, 80% dengan jumlah data latih masing- masing 200 *tweet* [2]. Akan tetapi

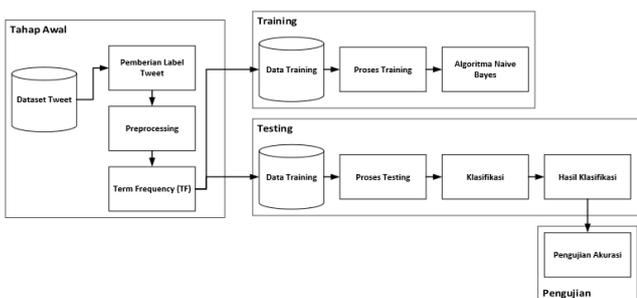
penelitian analisis sentimen tentang Timnas sepak bola Indonesia belum pernah dilakukan sehingga penulis mengusulkan penelitian ini.

Berdasarkan penjelasan diatas, maka pada penelitian ini dilakukan analisis sentimen Twitter untuk mengklasifikasikan *tweet* opini dan komentar pada performa Timnas Indonesia. Data tersebut akan diolah menggunakan algoritma *Naive Bayes* dan menghasilkan *tweet* yang telah diklasifikasikan menjadi positif atau negatif. Algoritma *Naive Bayes* merupakan salah satu algoritma untuk klasifikasi yang mempunyai akurasi yang tinggi. Pengujian akan dilakukan berdasarkan hasil klasifikasi. Pengujian bertujuan untuk mengetahui tingkat akurasi dari klasifikasi. Hasil klasifikasi akan memberikan kemudahan bagi pihak terkait untuk melihat dalam melihat *tweet* yang bernilai positif atau negatif.

## II. METODE PENELITIAN

### A. Analisis Sistem

Sistem yang akan dibuat dapat mengklasifikasikan *tweet* ke dalam positif dan negatif. Sistem tersebut dapat menghasilkan hasil klasifikasi sesuai dengan data yang diperoleh. Data yang digunakan diperoleh dari hasil *crawling* pada *Twitter*. Sistem memiliki empat sub proses sehingga dapat menghasilkan satu proses klasifikasi yang mengimplementasikan algoritma *Naive Bayes*. Berikut akan dijelaskan mengenai gambaran sistem yang akan dibuat dan tertera pada Gambar 1.



Gambar 1. Gambaran Sistem.

### B. Pengumpulan Data

Metode pengambilan data digunakan untuk mengumpulkan data pendukung yang dibutuhkan dalam pembuatan aplikasi. Untuk *dataset* yang digunakan diambil melalui *Twitter* API dengan proses *crawling*. *Dataset* yang diambil berupa file sql. Total *dataset* yang digunakan pada aplikasi analisis sentimen opini terhadap performa timnas sepakbola indonesia pada *twitter* sebanyak 530 *tweet* menggunakan 5 hashtag yaitu #TimnasDay, #TimnasJuara, #TimnasIndonesia, #InfoTimnas dan #KitaGaruda yang masing- masing *hashtag* periode pengumpulan pada tanggal 17 Februari 2020 dan telah di klasifikasikan oleh 3 responden yaitu jurnalis sepakbola. *Dataset* yang digunakan merupakan *tweet* dari pertandingan Indonesia vs Malaysia dengan skor 2-3 untuk Malaysia pada event kualifikasi piala dunia di Stadion Utama Gelora Bung Karno, Jakarta pada tanggal 5 September 2019 sebanyak 424 *tweet* dan pertandingan Indonesia vs Vietnam dengan skor 0-3 untuk Vietnam pada *event final* Seagames 2019 cabang olahraga

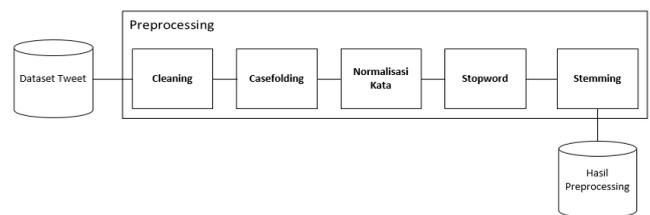
sepakbola di Rizal Memorial Stadium, Filipina pada tanggal 10 Desember 2019 sebanyak 106 *tweet*.

### C. Pengolahan Data

Data yang didapat dari *Twitter* API merupakan data yang langsung diambil dari *tweet* pengguna di *Twitter*. Data tersebut perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses *sentiment analysis*. *Tweet* akan mengalami penyeleksian dan perbaikan kata- kata sehingga *tweet* menjadi lebih bagus untuk proses klasifikasi. Proses ini dapat disebut dengan *preprocessing*. Setelah melalui *preprocessing*, data yang berupa teks akan diubah ke dalam bentuk angka melalui perhitungan TF (*Term Frequency*). Nilai TF ini yang akan menjadi masukan untuk algoritma *Naive Bayes*. Algoritma *Naive Bayes* akan menghasilkan klasifikasi untuk data yang dimasukkan.

#### 1) Preprocessing

Pada gambar 2 merupakan tahapan *preprocessing* dan selanjutnya akan dijelaskan cara kerja masing- masing *preprocessing*:



Gambar 2. Tahapan *Preprocessing*.

- *Cleaning*

Tahap *cleaning* akan dilakukan pembersihan *tweet* dari simbol karakter, URL *link*, *username*, serta *emoticon*. Dengan proses *cleaning* dapat mengurangi *noise* dari *tweet* yang didapatkan dalam proses *crawling* sehingga akan memudahkan dalam proses sentimen dan meningkatkan kinerja sistem.

- *Casefolding*

Tahap *casefolding* mengubah semua huruf besar atau kapital dalam *Twitter* menjadi huruf kecil. Hanya huruf 'a' sampai 'z' yang diterima. Dengan huruf yang sama akan lebih mudah diproses dalam proses sentimen yang akan mempengaruhi logika di dalam sistem.

- Normalisasi Kata

Tahap normalisasi kata mengubah bentuk suatu kata yang tidak baku dalam suatu kalimat pada *tweet* menjadi suatu kata yang baku dalam suatu kalimat yakni penyingkatan kata, penggunaan bahasa gaul atau *slang*, terjadinya kesalahan eja pada suatu kalimat, atau penggunaan bahasa yang masih belum sesuai dengan kamus.

- *Stopword*

Tahap *stopword* menghilangkan kata- kata yang tidak bermakna dalam *tweet* yang di *crawling*. Proses *stopword* kata pada *dataset* disesuaikan dengan koleksi *stopword list* yang dimiliki. Pada penelitian ini menggunakan *stopword list* dari Tala [1].

- *Stemming*

Tahap *stemming* menghilangkan semua imbuhan, dan menghasilkan kata dasar yang sesuai dengan struktur morfologi bahasa Indonesia. Pada penelitian ini menggunakan *stemming* Sastrawi yang berbasis algoritma Nazief dan Adriani [5]. Dalam penelitiannya Asian, dkk [7] melakukan beberapa pengembangan algoritma Nazief & Adriani sebagai berikut:

- Menggunakan kamus kata yang lebih lengkap
- Menambahkan aturan-aturan untuk kata-kata majemuk perulangan.
- Menambahkan aturan awalan dan akhiran, serta aturan lainnya, yaitu:
  - Menambahkan partikel (*inflection suffix*) “-pun”.
  - Penambahan aturan pemenggalan awalan.
  - Perubahan aturan pemenggalan untuk tipe awalan “me”.
- Perubahan urutan proses *stemming*, yaitu:
  - Kata dengan awalan “be-” dan akhiran “-lah”, hilangkan awalan terlebih dahulu kemudian akhiran.
  - Kata dengan awalan “be-” dan akhiran “-an”, hilangkan awalan terlebih dahulu kemudian akhiran.
  - Kata dengan awalan “me-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
  - Kata dengan awalan “di-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
  - Kata dengan awalan “pe-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran.
  - Kata dengan awalan “ter-” dan akhiran “-i”, hilangkan awalan terlebih dahulu kemudian akhiran. Dengan proses *stemming* dapat menghasilkan kata-kata yang tepat dalam melakukan proses sentimen. Dan hasil dari *preprocessing* akan di simpan dalam *database*.

Tabel 1 merupakan contoh hasil *preprocessing* yang menjadi masukan untuk *term frequency*.

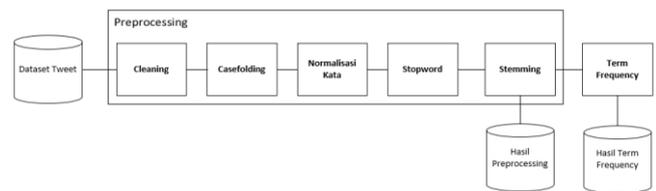
TABEL 1. HASIL PREPROCESSING.

No.	Preprocessing	Hasil
1.	<i>Tweet Awal</i>	Awas maen2 dibelakang lagi., penyakit Indonesia udh unggul pasti bertahan #TimnasDay
2.	<i>Cleaning</i>	Awas maen dibelakang lagi penyakit Indonesia udh unggul pasti bertahan
3.	<i>Casefolding</i>	awas maen dibelakang lagi penyakit indonesia udh unggul pasti bertahan
4.	Normalisasi	awas main dibelakang lagi penyakit indonesia sudah

		unggul pasti bertahan
5.	<i>Stopword</i>	awas main dibelakang penyakit indonesia unggul bertahan
6.	<i>Stemming</i>	awas main belakang sakit indonesia unggul tahan

### 2) Term Frequency

Perhitungan *TF* dilakukan dengan menghitung jumlah atau frekuensi kemunculan setiap kata dalam sebuah dokumen yang melalui proses *preprocessing*. Hasil dari perhitungan *TF* disimpan dalam basis data sistem untuk digunakan untuk proses klasifikasi menggunakan algoritma *Naive Bayes*. Pada gambar 3 terdapat penjelasan mengenai tahapan *term frequency*.



Gambar 3. Tahapan *Term Frequency*

### 3) Algoritma *Naive Bayes*

Klasifikasi menggunakan *Naive Bayes* dibagi menjadi 2 proses, yaitu proses *training* dan *testing*. Proses *training* digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan dalam klasifikasi dengan data *testing*. Proses *testing* yaitu membandingkan kata-kata pada dokumen uji dengan nilai probabilitas kata pada setiap kelas dokumen yang telah tersimpan dalam data *training*. Berikut adalah algoritma klasifikasi sentimen menggunakan *Naive Bayes*:

- Proses *Training*
  - Pada persamaan (3.1) menghitung peluang kemunculan dokumen pada tiap kategori positif atau negatif  $P(C_j)$ .
 
$$P(C_j) = \frac{|category_j|}{|corpus|} \quad (3.1)$$
  - Pada persamaan (3.2) menghitung peluang sebuah kata masuk ke dalam tiap kategori positif atau negatif  $P(W_i|C_j)$ .
 
$$P(W_i|C_j) = \frac{1+n_i}{n+|kosakata|} \quad (3.2)$$
- Proses *Testing*
  - Pada persamaan (3.3) menghitung untuk setiap kategori positif atau negatif lalu menentukan kategori dengan nilai tertinggi ( $V_{MAP}$ ).
 
$$V_{MAP} = \operatorname{argmax} P(C_j) \prod P(W_i|C_j) \quad (3.3)$$

## III. PENGUJIAN DAN ANALISA

Pada pengujian akurasi sistem dilakukan pengujian terhadap hasil klasifikasi pada proses *testing*. Untuk mengetahui keakuratan algoritma dalam proses klasifikasi maka akan dibandingkan hasil klasifikasi dengan hasil klasifikasi pelabelan manual oleh responden. Perbandingan tersebut akan dihitung tingkat keakurasiannya menggunakan *accuracy*, *precision* dan *recall*. Semakin tinggi nilai *accuracy*, *precision* dan *recall* pada sebuah algoritma yang digunakan maka menunjukkan bahwa algoritma tersebut berjalan dengan baik dan cocok untuk proses klasifikasi.

Pengujian dilakukan dengan memilih data *training* dan data *testing* secara acak dari 530 data yang akan digunakan.

Persamaan (3.4) merupakan rumus untuk pengujian akurasi:

$$akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (3.4)$$

Persamaan (3.5) merupakan rumus untuk pengujian *precision*:

$$precision = \frac{TP}{TP+FP} \quad (3.5)$$

Persamaan (3.6) merupakan rumus untuk pengujian *recall*:

$$recall = \frac{TP}{TP+FN} \quad (3.6)$$

Keterangan:

TP = *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.

TN = *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.

FP = *False Positive*, yaitu jumlah data positif yang terklasifikasi salah oleh sistem.

FN = *False Negative*, yaitu jumlah data negatif yang terklasifikasi salah oleh sistem.

Tabel 2 merupakan pengujian data *training* sebanyak 70% dan 30% untuk data *testing*. Pemilihan data *training* dan data *testing* dipilih secara acak dari 530 *dataset*.

TABEL 2. PENGUJIAN DENGAN KOMPOSISI 70% DAN 30%.

Data Training 70% dan Data Testing 30%				
Accuracy	Precision +	Precision -	Recall +	Recall -
78%	72%	80%	59%	88%

Tabel 3 merupakan pengujian data *training* sebanyak 80% dan 20% untuk data *testing*. Pemilihan data *training* dan data *testing* dipilih secara acak dari 530 *dataset*.

TABEL 3. PENGUJIAN DENGAN KOMPOSISI 80% DAN 20%.

Data Training 80% dan Data Testing 20%				
Accuracy	Precision +	Precision -	Recall +	Recall -
84%	85%	84%	75%	91%

Tabel 4 merupakan pengujian data *training* sebanyak 90% dan 10% untuk data *testing*. Pemilihan data *training* dan data *testing* dipilih secara acak dari 530 *dataset*.

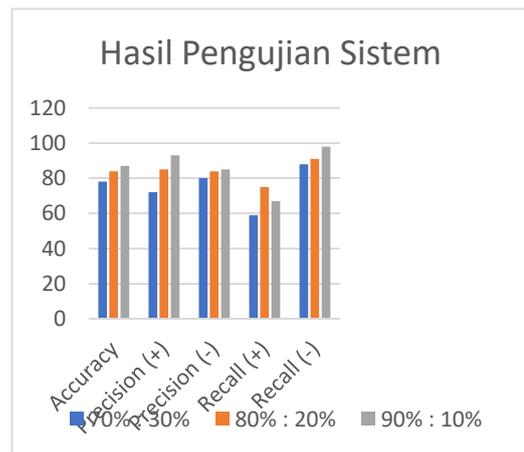
TABEL 4. PENGUJIAN DENGAN KOMPOSISI 90% DAN 10%.

Data Training 90% dan Data Testing 10%				
Accuracy	Precision +	Precision -	Recall +	Recall -
87%	93%	85%	67%	98%

Dari hasil pengujian diatas nilai algoritma *Naive Bayes* berturut-turut didapatkan sebesar 78%, 84% dan 87%. Nilai akurasi terendah adalah 78% dan tertinggi adalah 87%. Setiap terjadi penambahan komposisi data *training*, maka nilai dari *accuracy*, *precision* dan *recall* juga mengalami peningkatan. Hal ini dikarenakan algoritma *Naive Bayes*

merupakan algoritma yang sangat bergantung pada data *training*, kemungkinan akurasi dapat ditingkatkan lagi dengan menambahkan data *training* yang lebih banyak lagi. Seperti hasil dari penelitian ini pada Gambar 4 setiap terjadi penambahan data *training* maka akurasinya cenderung mengalami peningkatan.

Pada pertandingan Indonesia vs Malaysia, Indonesia mengalami kekalahan dengan skor 2-3 dan hasil klasifikasi sentimen positif sebanyak 185 *tweet* dan sentimen negatif sebanyak 239 *tweet*. *Tweet* pada pertandingan Indonesia vs Malaysia didapatkan *tweet* pada sebelum pertandingan, saat pertandingan dan selesai pertandingan di hari yang sama. *Tweet* positif pada pertandingan Indonesia vs Malaysia terjadi saat sebelum pertandingan, saat indonesia mencetak gol dan unggul pada pertandingan. *Tweet* negatif pada pertandingan Indonesia vs Malaysia terjadi saat malaysia



Gambar 4. Hasil Pengujian Sistem.

berhasil unggul mengembalikan keadaan, pemain bermain kurang maksimal dan saat hasil akhir pertandingan. Sehingga hasil analisis sentimen pada pertandingan Indonesia vs Malaysia cenderung mengarah ke negatif dikarenakan *tweet* negatif lebih besar daripada *tweet* positif.

Pada pertandingan Indonesia vs Vietnam, Indonesia mengalami kekalahan di *final* Seagames dengan skor 0-3 untuk Vietnam dan hasil klasifikasi sentimen positif sebanyak 28 *tweet* dan sentimen negatif sebanyak 78 *tweet*. *Tweet* pertandingan Indonesia vs Vietnam didapatkan *tweet* pada sebelum pertandingan, saat pertandingan dan selesai pertandingan di hari yang sama. *Tweet* positif pada pertandingan Indonesia vs Vietnam terjadi saat sebelum pertandingan karena pada saat itu penampilan Timnas sangat baik, saat pertandingan dan ada juga yang mengapresiasi pada akhir pertandingan karena penampilan Indonesia baik. *Tweet* negatif pada pertandingan Indonesia vs Vietnam terjadi saat Vietnam mencetak gol, unggul skor terhadap Indonesia dan hasil akhir pertandingan. Sehingga hasil analisis sentimen pada pertandingan Indonesia vs Vietnam cenderung mengarah ke negatif dikarenakan *tweet* negatif lebih besar daripada *tweet* positif.

Dari hasil klasifikasi dapat dilakukan pencarian opini tentang suatu target opini. Menurut Liu [6] target opini data berupa produk, individual, organisasi atau event. Pada penelitian ini dibatasi pada pemain, pelatih, dan federasi diharapkan dengan adanya target opini dapat digunakan sebagai salah satu masukan atau saran pendukung terhadap

timnas sepak bola Indonesia. Pengguna dapat memasukkan target opini pada sistem, kemudian sistem akan menelusur untuk mencari *document frequency* pada kelas positif dan negatif pada data testing sebanyak 53 dokumen didapatkan hasil sebagai berikut.

1. Target opini pemain hasil yang didapatkan adalah terdapat tweet negatif dengan target kata pemain sebanyak 20 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 1 dokumen. Dari hasil ini dapat dikatakan bahwa pada target pemain opini masyarakat cenderung negatif.
2. Target opini pelatih hasil yang didapatkan adalah terdapat tweet negatif dengan target kata pelatih sebanyak 6 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 1 dokumen. Dari hasil ini dapat dikatakan bahwa pada target pelatih opini masyarakat cenderung negatif.
3. Target opini federasi hasil yang didapatkan adalah terdapat tweet negatif dengan target kata federasi sebanyak 2 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 1 dokumen. Dari hasil ini dapat dikatakan bahwa pada target federasi opini masyarakat cenderung negatif.

Sedangkan pada keseluruhan data sebanyak 530 dokumen, didapatkan hasil sebagai berikut.

1. Target opini pemain hasil yang didapatkan adalah terdapat tweet negatif dengan target kata pemain sebanyak 119 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 50 dokumen. Dari hasil ini dapat dikatakan bahwa pada target pemain opini masyarakat cenderung negatif.
2. Target opini pelatih hasil yang didapatkan adalah terdapat tweet negatif dengan target kata pelatih sebanyak 25 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 5 dokumen. Dari hasil ini dapat dikatakan bahwa pada target pelatih opini masyarakat cenderung negatif.
3. Target opini federasi hasil yang didapatkan adalah terdapat tweet negatif dengan target kata federasi sebanyak 3 dokumen, sedangkan untuk tweet positif didapatkan sebanyak 2 dokumen. Dari hasil ini dapat dikatakan bahwa pada target federasi opini masyarakat cenderung negatif.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut:

1. Dari hasil penelitian dapat dikatakan bahwa pada opini masyarakat pada pertandingan Indonesia vs Malaysia tanggal 5 September 2019 pada event kualifikasi piala dunia 2022 dan pertandingan Indonesia vs Vietnam tanggal 10 Desember 2019 pada event final Seagames 2019 cenderung negatif. Pada target pemain didapatkan 119 dokumen negatif dan 50 dokumen positif pada keseluruhan dokumen, target pelatih didapatkan 25 dokumen negatif dan 5 dokumen positif, target federasi

didapatkan 3 dokumen positif dan 2 dokumen negatif.

2. Algoritma *Naive Bayes* dapat digunakan untuk mengklasifikasikan tweet kedalam positif atau negatif terutama *tweet* mengenai timnas sepak bola Indonesia. Dari tiga pengujian didapatkan hasil nilai algoritma *Naive Bayes* pada komposisi data *training* 371 dan data *testing* 159 sebesar 78%, komposisi data *training* 424 dan data *testing* 106 sebesar 84% dan komposisi data *training* 477 dan data *testing* 53 sebesar 87%. Nilai akurasi terendah adalah 78% dan tertinggi adalah 87%. Setiap terjadi penambahan komposisi data *training*, maka nilai dari *accuracy*, *precision* dan *recall* juga mengalami peningkatan. Hal ini dikarenakan algoritma *Naive Bayes* merupakan algoritma yang sangat bergantung pada data *training*, kemungkinan akurasi dapat ditingkatkan lagi dengan menambahkan data *training* yang lebih banyak lagi. Seperti hasil dari penelitian ini setiap terjadi penambahan data *training* maka akurasinya cenderung mengalami peningkatan.

#### DAFTAR PUSTAKA

- [1] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation Universeit Van Amsterdam*, 2003.
- [2] I. F. Rozi, E. N. Hamdana and M. B. I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naive Bayes Classifier SAMSAT Kota Malang," *Jurnal Informatika Polinema*, 2019.
- [3] I. F. Rozi, A. T. Firdausi and K. Islamiyah, "Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naive Bayes Dengan Fitur N-Gram," *Jurnal Informatika Polinema*, 2019.
- [4] F. Rahutomo, P. Y. Saputra and M. A. Fidyawan, "Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine," *Jurnal Informatika Polinema*, 2017.
- [5] B. Nazief and M. Adriani, "Stemming Indonesian: A confix-stripping approach'," *ACM Transactions on Asian Language Information Processing*, 2007.
- [6] B. Liu, "Sentiment Analysis: A Multi-Faceted Problem," *IEEE Intelligent Systems*, 2010.
- [7] J. Asian, H. E. Williams and S. M. M. Tahaghohi, "Stemming Indonesian," in *In Conferences in Research and Practice in Information Technology Series*, 2005.