

Analisis Sentimen Opini Publik Terhadap Pekan Olahraga Nasional Pada Instagram Menggunakan Metode Naïve Bayes Classifier

Priza Pandunata, Caesarina Kurnia Ananta, Yanuar Nurdiansyah

Universitas Jember

priza@unej.ac.id, caesarinak@gmail.com, yanuar_pssi@unej.ac.id

ABSTRACT

National Sports Week (Pekan Olahraga Nasional) held in October 2021 in Papua has brought many pros and cons from the public. This topic allows public for give criticism, suggestions, and opinions about the National Sport Week 2021. Instagram in one of social media that popular place for deliver public opinion. The process of sentiment analysis can find and solve problems based on public opinion on social media such as Instagram. The classification method used in this research is Naïve Bayes Classifier. The dataset can be obtained from data crawling process using the google chrome extension: IGCommentExport. The data the labelled as positive, neutral, or negative. The labelling process result showed 965 negative data, 256 neutral data, and 770 positive data. Then pre-processing is carried out on the data that has been labeled before, also word weighting process using TF-IDF. After that modelling is carried out using Naïve Bayes Classifier and the last process is evaluation-testing. The high accuracy of the result from fourth experiment which compare 90% data training with 10% data testing produce 75% accuracy. While the result of sentiment test show that negative sentiment more than positive sentiment and neutral sentiment.

Keyword: *Sentiment analysis, National Sports Week, TF-IDF algorithm, naive bayes classifier method.*

1. Pendahuluan

Pekan Olahraga Nasional (PON) merupakan event olahraga terbesar di Indonesia yang diadakan setiap empat tahun sekali. Pada tahun 1984, PON pertama kali di gelar di Solo, Jawa Tengah. Tujuan Pekan Olahraga Nasional I adalah menunjukkan kepada dunia, khususnya London bahwa Indonesia telah menjadi negara yang berdaulat sepenuhnya [1]. Pada Oktober 2021, Pekan Olahraga Nasional ke 20 yang diselenggarakan di Provinsi Papua. Pelaksanaan PON XX di Provinsi Papua terbilang sukses. Menurut Menteri Pemuda dan Olahraga Zainudin Amali menyebutkan bahwa adanya dampak PON XX terhadap perekonomian papua yaitu usaha kecil dan menengah (UKM) diuntungkan hingga empat kali lipat dari biasanya dan PON XX sukses menciptakan 3 rekor internasional dan 90 rekor nasional [2] dan [3]. Menurut Presiden Joko Widodo, PON pertama kali di Papua menunjukkan pertumbuhan Papua dengan adanya kesiapan infrastruktur untuk menyelenggarakan acara olahraga di tingkat nasional sekaligus bangga memiliki Stadion Lukas Enembe yang merupakan stadion terbaik di Asia Pasifik [4]. Dibalik kesuksesan PON XX terdapat beberapa masalah pada saat penyelenggaraan seperti terjadinya konflik saat pertandingan futsal antara Jawa Barat dengan Jawa Timur, pembangunan venue akuatik di Kabupaten Jayapura sempat dipalang warga sekitar karena mereka menuntut pembayaran hak ulayat, dan konflik pada saat pemilihan duta (brand ambassador), menurut arie kriting adanya perempuan Papua sebagai duta PON akan mencegah cultural appropriation [5], [6], dan [7].

Dengan adanya hal tersebut, masyarakat Indonesia mengungkapkan pendapatnya melalui media sosial, dan popularitas media sosial tidak dapat disangkal karena sebagai titik fokus penyebaran opini dan informasi. Namun seiring berjalannya waktu, masyarakat telah memanfaatkan media sosial untuk menyebarkan opini dan informasi [8]. Meskipun demikian, penggunaan media sosial cenderung mengganggu lingkungan sekitarnya dan kebebasan berekspresi dilakukan tanpa memperhatikan moralitas dan peraturan yang berlaku di masyarakat [9]. Instagram merupakan salah satu platform media sosial yang banyak digunakan masyarakat selama masa pandemi COVID-19 [10]. Instagram adalah aplikasi berbagi foto yang memungkinkan pengguna mengunggah foto secara langsung dari ponsel mereka [11]. Instagram juga memiliki fitur komentar postingan yang menjadi kelebihan instagram. Karena komentar instagram dibatasi hingga 2.200 karakter, sedangkan twitter hanya 280 karakter [12]. Sebanyak 36.4% pengguna instagram di Indonesia berusia antara 18 sampai dengan 24 tahun [13]. Angka-angka tersebut menunjukkan bahwa instagram memiliki peran yang penting di Indonesia, sehingga menyebabkan meningkatnya jumlah instansi pemerintahan yang tertarik memanfaatkan instagram untuk berkomunikasi dengan publik [14]. Hal ini karena daya tarik media sosial yang meluas, penggunaan yang relatif sederhana, kemampuan untuk menjangkau khalayak luas, biaya rendah [15].

Instagram juga dimanfaatkan untuk mencari informasi, bersosialisasi dan menyampaikan pesan atau komentar berdasarkan pandangan mereka [16].

Lantaran semakin banyaknya pengguna media sosial yang memanfaatkan instagram untuk mengungkapkan pendapatnya, diperlukan tindakan berupa analisis sentimen untuk memecahkan masalah berdasarkan opini publik saat ini. Analisis sentimen dapat digunakan sebagai alat untuk menilai pandangan publik terhadap peristiwa tertentu, sehingga dapat digunakan sebagai bahan untuk mengevaluasi peristiwa di masa mendatang [17]. Analisis sentimen memeriksa polaritas teks dalam komentar untuk menentukan bahwa pendapat dalam komentar tersebut positif, netral, atau negatif [18]. Metode klasifikasi yang digunakan yaitu Naïve Bayes Classifier. Metode Naïve Bayes Classifier merupakan algoritma yang populer untuk menentukan nilai probabilitas terbaik untuk mengkategorikan data uji dalam suatu kelompok kelas [19]. Metode klasifikasi menggunakan metode probabilistik dan statistik untuk memperkirakan probabilitas berdasarkan peristiwa sebelumnya (Teorema Bayes), dengan asumsi yang kuat (naif) antara setiap keadaan atau peristiwa. Penelitian ini berfokus pada sentimen komentar penyelenggaraan Pekan Olahraga Nasional 2021 pada instagram yang mengandung maksud positif, netral atau negatif. Uji evaluasi juga dilaksanakan untuk menguji model klasifikasi yang telah dihasilkan. Penelitian ini diharapkan dapat memberikan kontribusi bagi masyarakat untuk mengetahui komentar tentang Pekan Olahraga Nasional 2021 yang bersifat netral, negatif atau positif, serta membantu pemerintah sebagai bahan evaluasi dan untuk menentukan langkah-langkah atau kebijakan Pekan Olahraga Nasional selanjutnya.

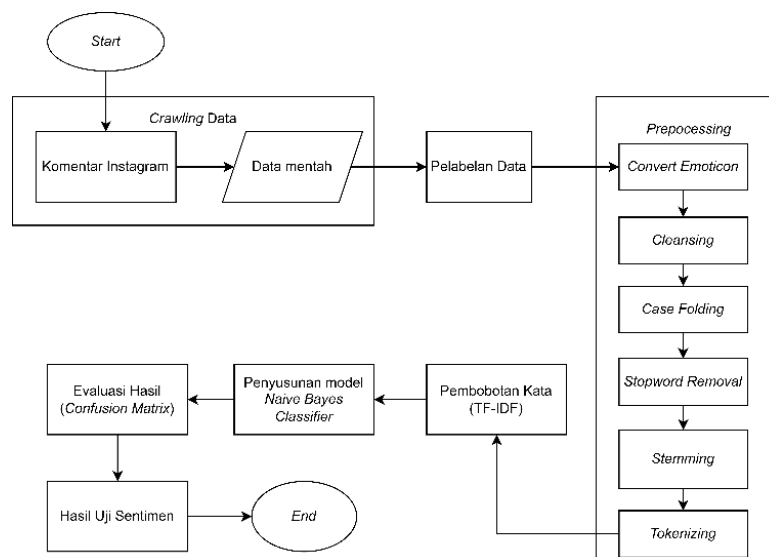
2. Metode Penelitian

2.1. Metode Pengumpulan Data

Metode pengumpulan data pada penelitian ini dilakukan dengan mengambil data komentar dari instagram menggunakan IGCommentExport yang merupakan *extensions google chrome*. Pengumpulan data dikumpulkan dari postingan instagram yang dibuat oleh @ponxx2020papua antara tanggal 24 Juni hingga 17 Oktober 2021. Jumlah data yang diambil sebanyak 2.000 data komentar.

2.2. Tahapan Penelitian

Tahapan penelitian merupakan urutan langkah-langkah penelitian. Urutan tersebut dilakukan mulai dari pengumpulan data, pengolahan data, hingga hasil akhir penelitian. Alur tahapan penelitian dapat dilihat pada gambar 1:



Gambar 1. Tahapan Penelitian

Dari flowchart gambar 1 menunjukkan bahwa penelitian ini diawali dengan proses *crawling* data di instagram menggunakan IGCommentExport yang merupakan *extensions google chrome*. Dataset tersebut selanjutnya diberi label positif, negatif atau netral oleh ahli bahasa. Selanjutnya data diolah pada tahap *text preprocessing* yang terdiri dari *convert emoticon*, *cleansing*, *case folding*, *stopword removal*, *stemming*, dan *tokenizing*. Fungsi *text preprocessing* untuk mengurangi *noise* pada dataset. Tahap selanjutnya adalah pembobotan kata dengan memanfaatkan algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF). Perhitungan TF-IDF dapat dilihat pada persamaan (1):

$$W_{t,d} = TF_{t,d} \times Idf_t \quad (1)$$

- Keterangan :
- TF_{t,d} : nilai dari *Term Frequency*
 - Idf_t : nilai dari *Inverse Document Frequency*.

Dimana nilai *IDF* diperoleh dari persamaan (2):

$$Idf_t = \log_{10} \left(\frac{N}{df_{(t)}} \right) \quad (2)$$

- Keterangan :
- N : jumlah total semua dokumen
 - df_(t) : jumlah total kemunculan frekuensi dalam term.

Tahap selanjutnya adalah penyusunan model dengan metode *Naïve Bayes Classifier*. Persamaan dari *teorema bayes* dirumuskan menggunakan persamaan (3) sebagai berikut:

$$P(C_j|X_i) = \frac{P(X_i|C_j) P(C_j)}{P(X_i)} \quad (3)$$

- Keterangan :
- i* : indeks kata yang berawal dari 1 sampai dengan kata ke-k
 - j* : indeks kategori yang berawal dari 1 sampai dengan kategori ke-n
 - $P(C_j|X_i)$: posterior, adalah kemunculan peluang pada kategori *j* tertentu ketika terdapat kemunculan kata *i*
 - $P(C_j)$: prior, adalah peluang kemunculan dokumen pada kategori *j*
 - $P(X_i|C_j)$: conditional, adalah peluang sebuah kata *i* masuk ke dalam kategori *j*
 - $P(X_i)$: peluang kemunculan sebuah kata

Rumus perhitungan *prior probability* dirumuskan menggunakan persamaan (4) sebagai berikut:

$$P(C_j) = \frac{NC_j}{N} \quad \dots\dots\dots (4)$$

- Keterangan :
- $P(C_j)$: *prior probability*
 - NC_j : jumlah dokumen latihan yang masuk ke dalam kategori *cj*
 - N* : jumlah seluruh dokumen latihan yang digunakan

Rumus perhitungan *conditional probability* dirumuskan menggunakan persamaan (5) sebagai berikut:

$$P(X_i|C_j) = \frac{T_{(X_i|C_j)} + 1}{\sum_{t' \in VT(XC_j)} t' + B'} \quad (5)$$

- Keterangan :
- $P(X_i|C_j)$: conditional probability
 - $T_{(X_i|C_j)}$: frekuensi kemunculan suatu kata dalam kategori *cj* pada dokumen latihan
 - $\sum_{t' \in VT(XC_j)} t'$: jumlah total kata dalam suatu kategori *cj*
 - B'* : total kosakata (kata unik) pada keseluruhan kategori dalam dokumen latihan

Rumus perhitungan *posterior probability* dirumuskan menggunakan persamaan (6) sebagai berikut:

$$P(C_j|X_i) = \arg \max_{c \in C} P(C_j) \prod_{1 \leq i \leq n_d} P(X_i|C_j) \quad \dots\dots\dots (6)$$

- Keterangan :
- arg max* : fungsi untuk mencari nilai posterior probability terbesar pada suatu kelas
 - $P(X_i|C_j)$: conditional probability, adalah peluang sebuah kata *i* dalam suatu kategori *j*

$P(C_j)$: prior probability, adalah peluang kemunculan dokumen pada kategori j

Dilanjutkan uji evaluasi model dengan *confusion matrix*. Terakhir yaitu dilakukan uji sentiment pada dataset yang berbeda.

3. Hasil dan Pembahasan

3.1. Hasil Dataset

Dataset untuk penelitian ini berasal dari proses *crawling* data di *instagram*. Data komentar *instagram* dapat diambil menggunakan *extension Google Chrome IGCommentExport*. Akun *google chrome* yang telah terpasang *extension* tersebut dapat melakukan *crawling* dengan menginputkan link postingan *instagram* yang hendak di *crawling*. Dataset yang diperoleh diberi label secara manual oleh Sarjana Pendidikan Bahasa dan Sastra Indonesia. Berdasarkan hasil pelabelan didapatkan 965 data negatif, 265 data netral dan 770 data positif. Sehingga jumlah keseluruhan dataset pada penelitian ini sebanyak 2000 data. Dataset tersebut berisi data *training* dan data *testing*. Pada tahap klasifikasi, data *training* digunakan untuk membuat bahan pembelajaran sehingga hasil klasifikasi dapat diterapkan pada data *testing*. Dataset untuk penelitian ini dibagi menjadi data latih dan data uji dengan perbandingan sebagai berikut: 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Ahli bahasa telah melabeli contoh data pada tabel 1 sebagai berikut:

Tabel 1. Contoh Pelabelan Dataset

Komentar	Sentimen
tolong ya panitia, evaluasi buat kedepannya. wasitnya berlisensi lah. kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya	Negatif
@disiniodhe pon papua kan boleh dihadiri penonton dengan syarat vaksin 2x dan kapasitas 25%	Netral
Keren banget Open ceremony Papua , ❤ bangga !	Positif

3.2. Hasil Pre-Processing

Pada tahap *preprocessing* dataset diawali dengan tahap *convert emoticon* yaitu tahap mengubah menjadi *string* yang bersesuaian. Hasil proses *convert emoticon* dapat dilihat pada tabel 2:

Tabel 2. Hasil Convert Emoticon

Komentar	Hasil Convert Emoticon
tolong ya panitia, evaluasi buat kedepannya. wasitnya berlisensi lah. kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya	tolong ya panitia, evaluasi buat kedepannya. wasitnya berlisensi lah. kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya
@disiniodhe pon papua kan boleh dihadiri penonton dengan syarat vaksin 2x dan kapasitas 25%	@disiniodhe pon papua kan boleh dihadiri penonton dengan syarat vaksin 2x dan kapasitas 25%
Keren banget Open ceremony Papua , ❤ bangga !	Keren banget Open ceremony Papua ,red heart bangga !

Setelah selesai proses *convert emoticon* selanjutnya merupakan tahap *cleansing*. Tahap ini menghapus angka, *username*, tanda baca, *hashtag*, dan link URL dari komentar karena tidak memiliki pengaruh dalam proses. Hasil proses *cleansing* dapat dilihat pada tabel 3:

Tabel 3. Hasil Cleansing

Komentar	Hasil Cleansing
tolong ya panitia, evaluasi buat kedepannya. wasitnya berlisensi lah. kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya	tolong ya panitia evaluasi buat kedepannya wasitnya berlisensi lah kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya
@disiniodhe pon papua kan boleh dihadiri penonton dengan syarat vaksin 2x dan kapasitas 25%	pon papua kan boleh dihadiri penonton dengan syarat vaksin x dan kapasitas
Keren banget Open ceremony Papua ,red heart bangga !	Keren banget Open ceremony Papua red heart bangga

Proses mengonversi data dari huruf besar ke huruf kecil (*lowercase*) dikenal sebagai *case folding*. Hasil proses *case folding* dapat dilihat pada tabel 4:

Tabel 4. Hasil *Case Folding*

Komentar	Hasil <i>Case Folding</i>
tolong ya panitia evaluasi buat kedepannya wasitnya berlisensi lah kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya	tolong ya panitia evaluasi buat kedepannya wasitnya berlisensi lah kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya
pon papua kan boleh dihadiri penonton dengan syarat vaksin x dan kapasitas Keren banget Open ceremony Papua red heart bangga	pon papua kan boleh dihadiri penonton dengan syarat vaksin x dan kapasitas keren banget open ceremony papua red heart bangga

Proses *stopword removal* kemudian digunakan untuk menghapus kata dari klasifikasi yang tidak memiliki arti yang penting. Contoh list kata *stopword removal* adalah 'haha', 'kah', 'wah', 'wkwk', dan seterusnya. Hasil proses *stopword removal* dapat dilihat pada tabel 5:

Tabel 5. Hasil *Stopword Removal*

Komentar	Hasil <i>Stopword Removal</i>
tolong ya panitia evaluasi buat kedepannya wasitnya berlisensi lah kayak apa ya bukan gondok sama pemainnya malah jadi gondok ke wasitnya	tolong panitia evaluasi buat kedepannya wasitnya berlisensi kayak apa bukan gondok sama pemainnya malah jadi gondok wasitnya
pon papua kan boleh dihadiri penonton dengan syarat vaksin x dan kapasitas	pon papua boleh dihadiri penonton dengan syarat vaksin kapasitas
keren banget open ceremony papua red heart bangga	keren banget open ceremony papua red heart bangga

Stemming adalah penghapusan imbuhan, seperti awalan dan akhiran untuk pengambilan kata dasar atau *root*. Pada proses ini dibantu oleh *library sastrawi*. Hasil proses *stemming* dapat dilihat pada tabel 6:

Tabel 6. Hasil *Stemming*

Komentar	Hasil <i>Stemming</i>
tolong panitia evaluasi buat kedepannya wasitnya berlisensi kayak apa bukan gondok sama pemainnya malah jadi gondok wasitnya	tolong panitia evaluasi buat depan wasit lisensi kayak apa bukan gondok sama main malah jadi gondok wasit
pon papua boleh dihadiri penonton dengan syarat vaksin kapasitas	pon papua boleh hadir tonton dengan syarat vaksin kapasitas
keren banget open ceremony papua red heart bangga	keren banget open ceremony papua red heart bangga

Proses *tokenizing* dilakukan untuk membagi kalimat menjadi bagian-bagian dari setiap kata. Hasil proses *tokenizing* dapat dilihat pada tabel 7:

Tabel 7. Hasil *Tokenizing*

Komentar	Hasil <i>Tokenizing</i>
tolong panitia evaluasi buat depan wasit lisensi kayak apa bukan gondok sama main malah jadi gondok wasit	['tolong', 'panitia', 'evaluasi', 'buat', 'depan', 'wasit', 'lisensi', 'kayak', 'apa', 'bukan', 'gondok', 'sama', 'main', 'malah', 'jadi', 'gondok', 'wasit']
pon papua boleh hadir tonton dengan syarat vaksin kapasitas	['pon', 'papua', 'boleh', 'hadir', 'tonton', 'dengan', 'syarat', 'vaksin', 'kapasitas']
keren banget open ceremony papua red heart bangga	['keren', 'banget', 'open', 'ceremony', 'papua', 'red', 'heart', 'bangga']

3.3. Hasil Pembobotan Kata

Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dimanfaatkan untuk pembobotan

kata dalam penelitian ini dengan menggunakan fungsi *feature_extraction* pada *library scikit learn*. Sebagai contoh perhitungan proses *TF-IDF* dalam penelitian ini menggunakan sampel data *testing* dan *training*. Contoh data *testing* ditunjukkan pada tabel 8 dan data *training* pada tabel 9:

Tabel 8. Contoh Data *Testing*

Dokumen	Kelas
Saya bukan orang papua tapi saya dukung penuh	Positif

Tabel 9. Contoh Data *Training*

Dokumen	Kelas
tolong panitia evaluasi buat depan wasit lisensi kayak apa bukan gondok sama main malah jadi gondok wasit	Negatif
pon papua boleh hadir tonton dengan syarat vaksin kapasitas	Netral
keren banget open ceremony papua red heart bangga	Positif

Pada tahap ini, *Term Frequency (TF)* ditentukan dengan menghitung frekuensi banyaknya kemunculan suatu kata dalam suatu dokumen. Tabel 10 sampai 12 menunjukkan perhitungan *Term Frequency (TF)*:

Tabel 10. Contoh Hasil Perhitungan Term Frequency (TF) Kelas Negatif

Kata	Nilai <i>TF</i>
tolong	1
panitia	1
evaluasi	1
buat	1
depan	1
wasit	2
lisensi	1
kayak	1
apa	1
bukan	1
gondok	2
sama	1
main	1
malah	1
jadi	1

Tabel 11. Contoh Hasil Perhitungan Term Frequency (TF) Kelas Netral

Kata	Nilai <i>TF</i>
pon	1
papua	1
boleh	1
hadir	1
tonton	1
dengan	1
syarat	1
vaksin	1
kapasitas	1

Tabel 12. Contoh Hasil Perhitungan Term Frequency (TF) Kelas Positif

Kata	Nilai <i>TF</i>
keren	1
banget	1
open	1
ceremony	1
papua	1
red	1
heart	1

bangga	1
--------	---

Nilai *document frequency (DF)* digunakan untuk menghitung *inverse document frequency (IDF)*. Nilai *IDF* ditunjukkan pada tabel 13:

Tabel 13. Contoh Hasil Perhitungan *Inverse Document Frequency (IDF)*

Kata	N	Nilai <i>DF</i>	$Idf_t = \log\left(\frac{N}{df(t)}\right)$
apa	3	1	0.4771212547
banget	3	1	0.4771212547
bangga	3	1	0.4771212547
boleh	3	1	0.4771212547
buat	3	1	0.4771212547
bukan	3	1	0.4771212547
ceremony	3	1	0.4771212547
dengan	3	1	0.4771212547
depan	3	1	0.4771212547
evaluasi	3	1	0.4771212547
gondok	3	2	0.1760912591
hadir	3	1	0.4771212547
heart	3	1	0.4771212547
jadi	3	1	0.4771212547
kapasitas	3	1	0.4771212547
kayak	3	1	0.4771212547
keren	3	1	0.4771212547
lisensi	3	1	0.4771212547
main	3	1	0.4771212547
malah	3	1	0.4771212547
open	3	1	0.4771212547
panitia	3	1	0.4771212547
papua	3	2	0.1760912591
pon	3	1	0.4771212547
red	3	1	0.4771212547
sama	3	1	0.4771212547
syarat	3	1	0.4771212547
tolong	3	1	0.4771212547
tonton	3	1	0.4771212547
vaksin	3	1	0.4771212547
wasit	3	2	0.1760912591
Total			13.8876689093

Bobot *TF-IDF* dihitung dengan mengalikan hasil *Term Frequency (TF)* dengan *inverse document frequency (IDF)*. Tabel 14 menunjukkan hasil *TF-IDF* kelas negatif, tabel 15 menunjukkan hasil *TF-IDF* kelas netral dan tabel 16 menunjukkan hasil *TF-IDF* kelas positif:

Tabel 14. Hasil *TF-IDF* Kelas Negatif

Kata	Nilai <i>TF</i>	Nilai <i>IDF</i>	$W_{t,d} = Tf_{t,d} \times Idf_t$
tolong	1	0.4771212547	0,4771212547
panitia	1	0.4771212547	0,4771212547
evaluasi	1	0.4771212547	0,4771212547
buat	1	0.4771212547	0,4771212547
depan	1	0.4771212547	0,4771212547
wasit	2	0.1760912591	0,3521825181
lisensi	1	0.4771212547	0,4771212547
kayak	1	0.4771212547	0,4771212547
apa	1	0.4771212547	0,4771212547
bukan	1	0.4771212547	0,4771212547
gondok	2	0.1760912591	0,3521825181

sama	1	0.4771212547	0,4771212547
main	1	0.4771212547	0,4771212547
malah	1	0.4771212547	0,4771212547
jadi	1	0.4771212547	0,4771212547
Total			6.9069413476

Tabel 15. Hasil *TF-IDF* Kelas Netral

Kata	Nilai <i>TF</i>	Nilai <i>IDF</i>	$W_{t,d} = T f_{t,d} \times I d f_t$
pon	1	0.4771212547	0.4771212547
papua	1	0.1760912591	0.1760912591
boleh	1	0.4771212547	0.4771212547
hadir	1	0.4771212547	0.4771212547
tonton	1	0.4771212547	0.4771212547
dengan	1	0.4771212547	0.4771212547
syarat	1	0.4771212547	0.4771212547
vaksin	1	0.4771212547	0.4771212547
kapasitas	1	0.4771212547	0.4771212547
Total			3.9930612968

Tabel 16. Hasil *TF-IDF* Kelas Positif

Kata	Nilai <i>TF</i>	Nilai <i>IDF</i>	$W_{t,d} = T f_{t,d} \times I d f_t$
kalau	1	0.4771212547	0.4771212547
belum	1	0.4771212547	0.4771212547
ada	1	0.4771212547	0.4771212547
ktp	1	0.4771212547	0.4771212547
bisa	1	0.4771212547	0.1760912591
ganti	1	0.4771212547	0.4771212547
apa	1	0.4771212547	0.4771212547
kak	1	0.4771212547	0.4771212547
Total			3.5159400421

3.4. Hasil Penyusunan Model *Naïve Bayes Classifier*

Perhitungan *Naïve Bayes Classifier* menggunakan data *testing* sebagai perhitungan dan dimulai dengan membagi jumlah dokumen dalam suatu kelas tertentu dengan jumlah total dokumen untuk mendapatkan nilai *prior probability*. Tabel 17 adalah contoh perhitungan nilai *prior probability*:

Tabel 17. Hasil Nilai *Prior Probability*

Kelas	$P(V_j) = \frac{D_j}{D}$
Negatif	0.33
Netral	0.33
Positif	0.33

Setelah didapatkan nilai *prior probability* untuk masing-masing kelas proses selanjutnya yaitu menghitung masing-masing kata dengan perhitungan *conditional probability*. Tabel 18 menunjukkan hasil *conditional probability* kelas negatif, positif, dan netral.

Tabel 18. Hasil *Conditional Probability* Kelas Negatif, Positif, dan Netral

	Negatif	Positif	Netral
<i>P(saya)</i>	0.0480893841	0.0574593467	0.0559261277
<i>P(bukan)</i>	0.0710338514	0.0574593467	0.0559261277
<i>P(orang)</i>	0.0480893841	0.0574593467	0.0559261277
<i>P(papua)</i>	0.0480893841	0.0675774354	0.0657742299
<i>P(tapi)</i>	0.0480893841	0.0574593467	0.0559261277

$P(saya)$	0.0480893841	0.0574593467	0.0559261277
$P(dukung)$	0.0480893841	0.0574593467	0.0559261277
$P(penuh)$	0.0480893841	0.0574593467	0.0559261277

Nilai *posterior probability* untuk setiap kelas kemudian ditentukan dengan menggunakan perhitungan pada tabel 19:

Tabel 19. Hasil Perhitungan *Posterior Probability* Negatif, Positif, dan Netral

$P(negatif/saya \text{ bukan orang papua tapi saya dukung penuh})$	$P(positif/saya \text{ bukan orang papua tapi saya dukung penuh})$	$P(netral/saya \text{ bukan orang papua tapi saya dukung penuh})$
$= 0.33 \times 0.0480893841 \times 0.0710338514 \times 0.0480893841 \times 0.0480893841 \times 0.0480893841 \times 0.0480893841 \times 0.0480893841 \times 0.0480893841 \times 0.0480893841$	$= 0.33 \times 0.0574593467 \times 0.0574593467 \times 0.0574593467 \times 0,0675774354 \times 0.0574593467 \times 0.0574593467 \times 0.0574593467 \times 0.0574593467 \times 0.0574593467$	$= 0.33 \times 0.0559261277 \times 0.0559261277 \times 0.0559261277 \times 0,0657742299 \times 0.0559261277 \times 0.0559261277 \times 0.0559261277 \times 0.0559261277 \times 0.0559261277$
$= 0.0000000000140827871$	$= 0.0000000000465806169$	$= 0.0000000000375178346$

Nilai probabilitas tertinggi yakni pada kelas positif sebesar 0.0000000000465806169 maka dapat disimpulkan bahwa data *testing* termasuk dalam kelas positif.

3.5. Hasil Uji Evaluasi

Proses uji evaluasi menggunakan *confusion matrix* yang berguna untuk menentukan keakuratan model yang sebenarnya. Proses perhitungan *confusion matrix* meliputi akurasi, presisi, *recall* dan *f-measure*. Pada penelitian ini, penulis melakukan empat percobaan dengan membandingkan data *training* dengan data *testing* dalam proporsi berikut: 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Hasil uji evaluasi dapat dilihat pada tabel 20:

Tabel 20. Hasil Uji Evaluasi

Keterangan	Percobaan			
	1	2	3	4
Data Training	60%	70%	80%	90%
Data Testing	40%	30%	20%	10%
Akurasi	70%	70%	74%	75%
Presisi	82%	82%	84%	84%
Recall	53%	53%	57%	59%
F-Measure	50%	51%	56%	58%

3.6. Hasil Uji Sentimen

Uji sentimen dilakukan agar dapat mengetahui sentimen dari komentar yang ditelaah diuji mengandung sentimen positif, netral atau negatif. Data komentar yang digunakan dalam tahap uji sentimen berbeda dari dataset. Proses uji sentimen dilakukan dalam empat percobaan dengan membandingkan data *training* dan data *testing* sebesar 60%:40%, 70%:30%, 80%:20%, dan 90%:10%. Contoh hasil percobaan uji sentimen dapat dilihat pada tabel 21:

Tabel 21. Contoh Hasil Uji Sentimen Perbandingan 60%:40%

Komentar	Hasil Uji Sentimen
@jovimaspaitella pernah tuh PON 2016 bahkan DUTA nya bukan dari kelahiran jabar tapi gak dipermasalahkan ini cuma jadi IKON kok rame sejagad raya ya?	Negatif
Betul tuh min, kearifan lokal jgn sampe di lupakan dan wajib di lestarikan.	Netral
@arie_kriting sangat sangat setuju jika yang di pilih adalah putra/i dari asli papua	Positif

4. Kesimpulan

Berdasarkan hasil pengujian didapatkan hasil akurasi yang berbeda pada keempat percobaan. Nilai

akurasi tertinggi didapatkan pada percobaan keempat yakni sebesar 75% dengan data *training* 90% dan *testing*%. 3. Disimpulkan bahwa 2000 data opini publik pada instagram terhadap kegiatan Pekan Olahraga Nasional 2021 pada rentang waktu 24 Juni – 17 Oktober 2021 menunjukkan hasil negatif yang lebih tinggi dibandingkan data positif atau netral. Hasil uji sentimen menunjukkan sentimen negatif ketika komentar yang diuji pada tahap uji sentimen mengandung konotasi negatif, penolakan, dan ketidaksetujuan. Sedangkan komentar yang diuji pada tahap uji sentimen tidak berisi pernyataan negatif maupun positif dan kalimat tersebut merupakan kalimat pernyataan yang tidak memihak maka hasil uji sentimen menunjukkan netral. Dan hasil uji sentimen akan menunjukkan hasil sentimen positif apabila komentar tersebut bermakna positif, setuju dan mendukung pernyataan tersebut.

Daftar Pustaka

- [1] L. Ponne, “Berlomba Di Daerah Bergolak: Penyelenggaraan Pekan Olahraga Nasional (Pon) Iv Di Kota Makassar,” *Lensa Budaya J. Ilm. Ilmu-Ilmu Budaya*, vol. 12, no. 1, pp. 70–81, 2017, doi: <http://dx.doi.org/10.34050/jlb.v12i1.3114>.
- [2] E. Dyah, “PON XX Papua Tingkatkan Pendapatan bagi Pelaku UKM hingga 4x Lipat,” *news.detik.com*, 2021. <https://news.detik.com/berita/d-5756185/pon-xx-papua-tingkatkan-pendapatan-bagi-pelaku-ukm-hingga-4x-lipat> (accessed Feb. 25, 2022).
- [3] CNN Indonesia, *Menpora: Ketakutan Gangguan Keamanan dan Lonjakan Covid-19 di PON Papua Tidak Terjadi*. Indonesia: www.youtube.com, 2021.
- [4] Humas, “Presiden Jokowi: Penyelenggaraan PON XX Gambarkan Kemajuan Papua,” *setkab.go.id*, 2021. <https://setkab.go.id/presiden-jokowi-penyelenggaraan-pon-xx-gambarkan-kemajuan-papua/> (accessed Feb. 25, 2022).
- [5] I. F. Rialdi, “Viral Permainan ‘Bar-bar’ Jatim vs Jabar, Kepemimpinan Wasit Dipertanyakan,” *bolatimes.com*, 2021. <https://www.bolatimes.com/arena/2021/10/01/191431/viral-permainan-bar-bar-jatim-vs-jabar-kepemimpinan-wasit-dipertanyakan> (accessed Feb. 25, 2022).
- [6] CNN Indonesia, “Kapolda Papua Janji Bayar Sengketa Lahan Adat Usai PON XX,” *cnnindonesia.com*, 2021. <https://www.cnnindonesia.com/nasional/20210831093436-12-687670/kapolda-papua-janji-bayar-sengketa-lahan-adat-usai-pon-xx> (accessed Feb. 25, 2022).
- [7] F. Janati, “Nagita Slavina Jadi Duta PON XX Papua, Arie Kriting: Seharusnya Sosok Perempuan Papua,” *Kompas.com*, 2021. <https://www.kompas.com/hype/read/2021/06/02/163654966/nagita-slavina-jadi-duta-pon-xx-papua-arie-kriting-seharusnya-sosok?page=all> (accessed Feb. 25, 2022).
- [8] K. L. Widyasari, F. S. Nuranna, and A. L. Sulasmi, *Memaksimalkan Penggunaan Media Sosial Dalam Lembaga Pemerintahan*, 1st ed. Jakarta: Direktorat Jenderal Informasi dan Komunikasi Publik, Kementerian Komunikasi dan Informatika, 2018.
- [9] H. S. Arifin, W. Widyowati, and T. Hernawaty, “Freedom of Expression Di Media Sosial Bagi Remaja Secara Kreatif Dan Bertanggung Jawab: Bagi Siswa Sma Al-Ma’Soem Rancaekek Dan Sma Muhammadiyah Pangandaran,” *J. Pengabd. Kpd. Masy.*, vol. 1, no. 5, pp. 332–337, 2017.
- [10] H. Junawan and N. Laugu, “Eksistensi Media Sosial , Youtube , Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia,” *Baitul Ulum J. Ilmu Perpust. dan Inf.*, vol. 4, no. 1, pp. 41–57, 2020.
- [11] M. Iqbal and M. A. Ghofur, “Efektivitas Unggahan Foto di Akun Instagram @indtravelers.co terhadap Minat Berlibur Komunitas Backpacker Malang,” *JISIP J. Ilmu Sos. dan Ilmu Polit.*, vol. 8, no. 1, pp. 52–64, 2019, doi: <https://doi.org/10.33366/jisip.v8i1.1532>.
- [12] A. Muhaddisi, B. N. Prastowo, D. Utami, and K. Putri, “Sentiment Analysis With Sarcasm Detection On Politician ’ s Instagram,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 4, pp. 349–358, 2021, doi: <https://doi.org/10.22146/ijccs.66375>.
- [13] K. Ashilah, “Instagram Sebagai Media Pemasaran Digital Wasabi Sushi Japanese Food Jember,” *J. Ekon. dan Bisnis Islam*, vol. 1, pp. 149–160, 2021.
- [14] L. Margaretha and D. M. Sunarya, “INSTAGRAM SEBAGAI MEDIA SOSIALISASI 9 PROGRAM UNGGULAN PEMERINTAH (Studi Kasus pada Bagian Hubungan Masyarakat Kementerian Koperasi dan Usaha Kecil dan Menengah Republik Indonesia pada Instagram @kemenkopUKM),” *Communication*, vol. 8, no. 2, pp. 88–105, 2017, doi: <http://dx.doi.org/10.36080/comm.v8i2.637>.
- [15] M. Graham and E. Avery, “Government Public Relations and Social Media: An Analysis of the Perceptions and Trends of Social Media Use at the Local Government Level,” *Public Relat. J.*, vol. 7, no. 4, pp. 1–21, 2013, [Online]. Available: <http://dev.prsa.org/Intelligence/PRJournal/Documents/2013GrahamAvery.pdf>.
- [16] I. G. Titah Pratyaksa and N. L. Wiwik Ela Putri, “PERANAN NEW MEDIA DALAM TRANSFORMASI FUNGSI KOMUNIKATOR DAN FUNGSI KONSTRUKSI (DRAMATURGY AKUN INSTAGRAM WALIKOTA DENPASAR RAI MANTRA TAHUN 2019),” *Ganaya J. Ilmu Sos. dan Hum.*, vol. 2, no. 2–3, pp. 92–103, 2020.

- [17] Y. Nurdiansyah, S. Bukhori, and R. Hidayat, "Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method," *J. Phys.*, 2018, doi: 10.1088/1742-6596/1008/1/012011.
- [18] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, 2008, doi: <https://doi.org/10.1561/1500000001>.
- [19] A. Lasarudin and Purwanto, "Klasifikasi Pengaduan Masyarakat Menggunakan Naive Bayes Berbasis Seleksi Atribut Information Gain," *J. Teknol. Inf.*, vol. 14, no. 2, pp. 59–72, 2018, [Online]. Available: <http://research.pps.dinus.ac.id/index.php/Cyberku/article/view/65>.