

Analisis Sentimen Motogp Mandalika Pada Twitter Menggunakan Metode Naïve Bayes

Muhammad Agus Arianto^{1*}, Achmad Solichin²

^{1,2}Fakultas Teknologi Informasi, Teknik Informasi, Universitas Budi Luhur, Jakarta, Indonesia
Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260
Email: ^{1*}iraanagus@gmail.com, ²achmad.solichin@budiluhur.ac.id
(*: corresponding author)

Abstrak— Indonesia terpilih sebagai tuan rumah gelaran MotoGP seri kedua setelah MotoGP Qatar. Indonesia menggunakan Pertamina Mandalika International Street Circuit sebagai tempat berlangsungnya gelaran MotoGP, MotoGP Mandalika terbilang sukses, diperkirakan jumlah total penonton sebanyak 102.801 orang, MotoGP Mandalika sempat menjadi trending topic di Twitter yang tentu menuai reaksi yang beragam dari para netizen Indonesia. Untuk mengetahui tweet tersebut bernilai positif, negatif atau netral maka dilakukan analisis sentimen. Salah satu metode analisis sentimen yang dapat digunakan untuk menyelesaikan masalah tersebut adalah metode *Naïve Bayes*. Data yang terkumpul sebanyak 7195 Data yang diambil dalam kurun waktu 18 Maret 2022 sampai 18 April 2022. Data yang sudah didapat kemudian dilakukan remove duplikat, preprocessing, dan klasifikasi menggunakan program *Naïve Bayes Classifier* yang dibuat menggunakan Bahasa pemrograman PHP. Dari hasil remove duplikat dan preprocessing data yang tersisa sebanyak 1108 data, data dengan sentiment positif 396 data, sentiment netral 494 data dan sentiment negatif 218 data. Hasil dari pengujian akurasi menggunakan perbandingan 70:30 dan 80:20 mendapat nilai akurasi sebesar 57.65% dan 58.10%.

Kata Kunci: Analisis Sentimen, Naïve Bayes Classifier, MotoGP Mandalika

Abstract— Indonesia was chosen as the host for the second MotoGP series after the Qatar MotoGP. Indonesia uses the Pertamina Mandalika International Street Circuit as the venue for the MotoGP event, the Mandalika MotoGP is fairly successful, it is estimated that the total number of spectators is 102,801 people, the Mandalika MotoGP has become a trending topic on Twitter which of course has received mixed reactions from Indonesian netizens. To find out if the tweet is positive, negative or neutral, sentiment analysis is carried out. One method of sentiment analysis that can be used to solve this problem is the *Naïve Bayes* method. The data collected was 7195. Data were taken from March 18, 2022 to April 18, 2022. The data that had been obtained was then carried out by removing duplicates, preprocessing, and classifying using the *Naïve Bayes Classifier* program which was created using the PHP programming language. From the results of removing duplicates and preprocessing data, the remaining 1108 data, data with positive sentiment 396 data, neutral sentiment 494 data and negative sentiment 218 data. The results of the accuracy test using a comparison of 70:30 and 80:20 got an accuracy value of 57.65% and 58.10%, respectively.

Keywords: Sentiment Analysis, Naïve Bayes Classifier, Mandalika MotoGP

I. PENDAHULUAN

Saat ini teknologi telah berkembang pesat seperti sekarang ini, juga disertai dengan perkembangan jumlah data digital, yang semakin meningkat. Untuk menjadi informasi yang bermanfaat, data digital harus diolah. Salah satu platform penghasil data terbesar di internet saat ini adalah media sosial, pengguna media sosial semakin meningkat di masyarakat mempengaruhi perkembangan opini publik. Oleh karena itu, dapat digunakan untuk menganalisis opini publik, apakah positif, negatif, atau netral. *Twitter* adalah media sosial yang cukup populer di dunia yang memungkinkan pengguna untuk berbagi pesan yang disebut tweet di berbagai platform[1]. Adanya layanan ini membuat masyarakat lebih memilih untuk menyampaikan pendapatnya melalui twitter daripada mengirimkannya secara langsung. Namun penggunaannya memerlukan analisis di bidang *text mining* itu sendiri, sehingga informasi yang dihasilkan dapat membantu pengambilan keputusan yang lebih akurat dan efektif.

Penelitian sebelumnya yang dilakukan pada analisis sentimen, termasuk penggunaan *Naive Bayes* dalam klasifikasi analisis sentimen MotoGP[2], menunjukkan bahwa metode *Naive Bayes* mampu mencapai skor akurasi klasifikasi sentimen sebesar 82,67%. Metode Naive Bayes juga digunakan untuk melakukan analisis sentimen pada pandemic covid-19[3], Kinerja Dewan Perwakilan Rakyat (DPR)[4] dan wisata TMII[5], dengan nilai akurasi sebesar 80%.

Salah satu topik yang pernah menjadi perbincangan oleh masyarakat Indonesia yaitu gelaran MotoGP Mandalika 2022. Indonesia resmi masuk dalam kalender MotoGP 2021 dan telah ditampilkan di situs resmi ajang balap MotoGP tersebut. Di situs resmi MotoGP, Indonesia akan menjadi tuan rumah seri ke-15. MotoGP Mandalika sendiri akan berlangsung dari tanggal 18 – 22 Maret 2022. Gelaran MotoGP mandalika sendiri terbilang sukses karena jumlah penonton yang hampir mencapai 103 ribu orang, terbagi menjadi 3 yaitu latihan bebas, kualifikasi, dan balapan MotoGP, dimana pada hari pertama yaitu latihan bebas jumlah penonton mencapai 9.857 orang, kemudian pada hari ke dua yaitu kualifikasi penonton mencapai 30.021 orang, jumlah penonton terbanyak jatuh pada hari ke tiga balapan MotoGP jumlah penonton mencapai 62.923 orang.

Dari penjelasan diatas, peneliti akan melakukan analisis sentimen tentang gelaran MotoGP Mandalika melalui media sosial Twitter menggunakan metode *Naive Bayes Classifier*. Adapun hasil dari penelitian ini untuk melihat respon dari

masyarakat Indonesia di media sosial *Twitter* tentang gelaran MotoGP Mandalika.

II. METODE PENELITIAN

A. Analisis Sentimen pada *Twitter*

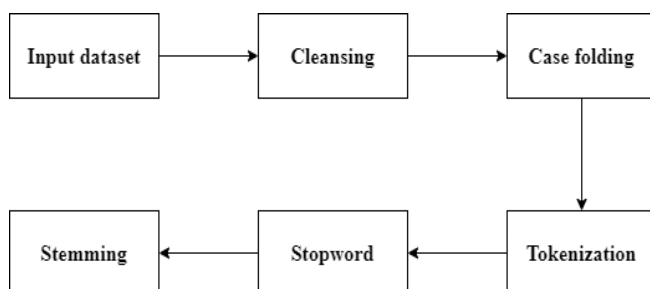
- 1) *Text mining* merupakan proses penambangan yang dilakukan oleh komputer untuk mendapatkan sesuatu yang baru, dan tidak diketahui sebelumnya, atau menemukan kembali informasi yang tersirat secara implisit. Temuan tersebut berasal dari informasi yang diekstrak secara otomatis dari sumber-sumber data teks yang berbeda[6].
- 2) Analisis sentimen merupakan sebuah metode yang digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini[7].
- 3) *Twitter* adalah media sosial yang cukup populer di dunia yang memungkinkan pengguna untuk berbagi pesan yang disebut *tweet* di berbagai platform. Jumlah *tweet* yang dikirim setiap hari mencapai lebih dari 350 juta per- hari. Ini digunakan oleh organisasi untuk memantau pergerakan pesaing berdasarkan pendapat pengguna mereka melalui *tweet* yang dikirim terintegrasi dengan merek mereka. Juga di sisi pengguna, interaksi dan konten melalui saluran-saluran resmi penyedia, mempengaruhi keputusan untuk membeli merek dan produk tertentu berdasarkan konten yang dihasilkan oleh pengguna lain[8].

B. Data Penelitian

Data yang digunakan penulis dalam penulisan ini merupakan data yang diambil dari media sosial *Twitter*. Data *tweet* yang digunakan dalam penelitian ini adalah data *tweet* dengan *keyword* “motogp mandalika” dan “sirkuit mandalika”. Dari *keyword* tersebut mendapatkan data yang belum diolah sebanyak 7195 *tweet*. Pengambilan data *tweet* ini dimulai dari tanggal 18 Maret 2022 sampai 18 April 2022. Pengumpulan data menggunakan Bahasa Indonesia saja.

C. Persiapan Data

Setelah proses pengambilan data pada tahap sebelumnya, maka data yang data yang sebelumnya sudah di dapat akan dilanjutkan ke tahap *preprocessing* seperti yang terlihat pada gambar 1.



Gambar 1. Text Preprocessing

- a. *Cleansing* adalah tahapan pembersihan data dari komponen-komponen yang tidak berhubungan dengan informasi yang terdapat pada dokumen, seperti karakter atau simbol, *emoticon* dan link URL.
- b. *Case folding* adalah proses pengubahan semua huruf dalam teks menjadi huruf kecil.
- c. *Tokenizing* adalah Suatu proses yang dilakukan untuk membagi kalimat menjadi bagian-bagian atau kata-kata. Kata yang dihasilkan dari proses ini disebut *tokenize*.
- d. *Stopword removal* adalah adalah proses menghilangkan kata-kata umum yang tidak memiliki arti. Penghilangan beberapa kata kerja, kata sifat, dan kata keterangan lainnya dapat ditambahkan ke daftar *stopword*, dalam proses *stopword* peneliti menggunakan kamus milik Kaggle.com
- e. *Stemming* adalah proses menemukan kata dasar dari setiap kata yang dihasilkan dari proses penyaringan sebelumnya. *Stemming* melakukan proses mengembalikan berbagai bentuk kata ke bentuk kata dasar dengan menghilangkan kalimat imbuhan, dalam proses *stemming* penulis melakukannya dengan cara manual dengan bantuan kamus besar Bahasa Indonesia.

D. Naïve Bayes Classifier (NBC)

Naïve Bayes adalah metode klasifikasi yang sangat sederhana dan efisien. Metode klasifikasi *Naïve Bayes* juga merupakan metode yang populer untuk klasifikasi teks dan memiliki kinerja yang baik, namun juga memiliki kelemahan yaitu sangat sensitif terhadap seleksi fitur. Terlalu banyak fitur tidak hanya meningkatkan waktu perhitungan, tetapi juga dapat mengurangi akurasi klasifikasi[9]. Klasifikasi Bayesian, berdasarkan teorema Bayes, mampu mengklasifikasikan dengan akurasi dan kecepatan tinggi ketika diterapkan pada database yang berisi data dalam jumlah besar [10]. Perhitungan Naïve Bayes secara manual dapat dilakukan dengan menggunakan rumus persamaan 1.

$$P(w_k|v_j) = \frac{nk + 1}{n + |\text{vocabulary}|} \quad (1)$$

P : adalah probability.

w_k : merupakan kata dalam semua dokumen yang memiliki label positif, negatif atau netral.

v_j : merupakan kata yang berada dalam kelas positif atau negatif, berbeda dengan **w_k** yang merupakan keseluruhan pada setiap label.

nk : berapa kali suatu kata muncul dalam kelas.

n : banyak kata dalam kelas.

vocabulary : total kata unik pada semua dokumen.

E. Evaluasi

Pada tahapan ini peneliti akan melakukan evaluasi metode klasifikasi dengan mengukur performa menggunakan *confusion matrix* terhadap algoritma *Naive Bayes* tersebut.

Confusion matrix adalah matrix 2x2 yang merepresentasikan hasil klasifikasi biner pada suatu dataset. Terdapat beberapa rumus umum yang dapat digunakan untuk menghitung performa klasifikasi pada tabel 1 dibawah ini terdapat contoh tabel *confusion matrix*.

Tabel 1. Confusion Matrix

	True Positive	True Negative
Positive	TP	FN
Negative	FP	TN

TP (True Positive) : yaitu data yang bernilai positif, yang terklasifikasi dengan benar oleh sistem.
 FP (False Positive) : yaitu data yang bernilai positif, tetapi terklasifikasi salah oleh sistem.
 FN (False Negative) : yaitu data yang bernilai negatif, tetapi terklasifikasi salah oleh sistem.
 TN (True Negative) : yaitu data yang bernilai negatif, yang terklasifikasi benar oleh sistem.

Akurasi adalah proporsi prediksi yang benar. Rumus untuk menghitung akurasi dapat dilihat pada persamaan 2 berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (2)$$

Precision adalah proporsi jumlah dokumen teks relevan yang diperiksa dengan semua dokumen teks yang dipilih oleh sistem. Rumus precision dapat dilihat pada persamaan 3 berikut:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

Recall adalah proporsi kasus positif yang di prediksi yang benar. Rumus recall dapat dilihat pada persamaan 4 berikut:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

F. Rancangan Pengujian

Pada tahap ini dilakukan pengujian terhadap aplikasi yang dibuat yaitu *Naive Bayes Classifier*, untuk mengukur keakuratan penghitungan suatu data yaitu sentimen positif, negatif dan netral, pengujian dilakukan dengan menggunakan sampel data dengan split data 70:30, dan 80:20

III. HASIL DAN PEMBAHASAN

A. Data Collection

Pada tabel 2 terdapat data hasil *crawling* sebanyak 1990 data dari *Twitter* yang akan digunakan.

Tabel 2. Jumlah Data Tweet Hasil Crawling

Keyword	Tanggal	Jumlah
MotoGP Mandalika	18 Maret – 18 April 2022	975 Data
Sirkuit Mandalika	18 Maret – 18 April 2022	1015 Data
	Total	1990 Data

Pada tabel 3 terdapat contoh dataset dari MotoGP Mandalika dan Sirkuit Mandalika.

Tabel 3. Contoh Isi Dataset

Text
RT @Helmi_Felis: Mandalika MotoGP gak sukses Formula E Jakarta

RT @republikaonline: Komika Ernest Prakasa menyindir perhelatan Jakarta E-Prix atau balapan Formula E yang berlangsung di Sirkuit Ancol
@aewin86 Saya jujur aja, karena ngga kesana mang, tapi dengan adanya motogp di mandalika makin mengangkat nama Indonesia di mata dunia #MandalikaMotoGP #PertaminaMotoGP #IndonesiaMotoGP
Dorna Batalkan MotoGP Finlandia, IMI Jajaki Mandalika Jadi Seri Pengganti - Lombok Post - https://t.co/XyNW7StiWQ

B. Pemberian Label Sentimen

Sebelum melakukan tahapan data *preprocessing*, data yang telah dilalui tahap *remove duplicate* harus diberikan label untuk menentukan sentimen dari setiap data. Label sentimen setiap data berupa positif, negatif dan netral. Berikut adalah contoh data yang sudah diberikan label sentiment. Pemberian label sentimen ini dilakukan manual dengan dibantu oleh pakar.

Pada tabel 4 berisi contoh tweet masyarakat terhadap gelatan MotoGP Mandalika yang memiliki sentimen positif, negative dan netral dapat dilihat pada tabel di bawah.

Tabel 4. Pemberian Label Sentimen

Tweet	Sentimen
alhamdulillah sirkuit mandalika indah	Positif
indonesia sirkuit mandalika bodoh bener	Negatif
perbedaan signifikan sirkuit formula e jakarta motogp mandalika	Netral

C. Tahapan Preprocessing

Pada tabel 5 terdapat langkah-langkah pembersihan data adalah *preprocessing*, yang terdiri dari *case folding*, *tokenization*, *stopword removal*, dan *lemmatization*. Dalam *stopword removal* peneliti menggunakan kamus *stopword* dan pada bagian *stemming* peneliti menggunakan kamus besar bahasa Indonesia sebagai panduannya. Berikut hasil *preprocessingnya*.

Tabel 5. Tahapan Preprocessing

Preprocessing	Sebelum	Sesudah
Cleansing	@amirkusumangara @ScherazadeMS Wah kalo mau bahas politik, mending bahas kenapa sirkuit Mandalika baru dipakai sekali (WSBK 2021) harus diaspal ulang di beberapa sektor sebelum event motogp ada penggunaan material aspal yang kurang baik kah?? Indikasi korupsi?	Wah kalo mau bahas politik mending bahas kenapa Sirkuit Mandalika baru dipakai sekali wsbk harus diaspal ulang di beberapa sektor sebelum event motogp ada penggunaan material aspal yang kurang baik kah indikasi korupsi

Case Folding	Wah kalo mau bahas politik mending bahas kenapa Sirkuit Mandalika baru dipakai sekali wsbk harus diaspal ulang di beberapa sektor sebelum event motogp ada penggunaan material aspal yang kurang baik kah indikasi korupsi.	wah kalo mau bahas politik mending bahas kenapa sirkuit mandalika baru dipakai sekali wsbk harus diaspal ulang di beberapa sektor sebelum event motogp ada penggunaan material aspal yang kurang baik kah indikasi korupsi.
Tokenize	wah kalo mau bahas politik mending bahas kenapa sirkuit mandalika baru dipakai sekali wsbk harus diaspal ulang di beberapa sektor sebelum event motogp ada penggunaan material aspal yang kurang baik kah indikasi korupsi	“wah”, “kalo”, “mau”, ”bahas”, “politik”, “mending”, “bahas”, “kenapa”, “sirkuit”, “mandalika”, “baru”, “dipakai”, “sekali”, “wsbk”, ”harus”, “diaspal”, “ulang”, “dibeberapa”, “sector”, “sebelum”, “event”, “motogp”, “ada”, “penggunaan”, “material”, “aspal”, “yang”, “kurang”, “baik”, “kah”, “indikasi”, “korupsi”.
Stopword	“wah”, “kalo”, “mau”, ”bahas”, “politik”, “mending”, “bahas”, “kenapa”, “sirkuit”, “mandalika”, “baru”, “dipakai”, “sekali”, “wsbk”, ”harus”, “diaspal”, “ulang”, “dibeberapa”, “sector”, “sebelum”, “event”, “motogp”, “ada”, “penggunaan”, “material”, “aspal”, “yang”, “kurang”, “baik”, “kah”, “indikasi”, “korupsi”.	“wah”, “kalo”, ”bahas”, “politik”, “mending”, “bahas”, “kenapa”, “sirkuit”, “mandalika”, “dipakai”, “wsbk”, “diaspal”, “ulang”, “sector”, “event”, “motogp “material”, “aspal”, “kah”, “indikasi”, “korupsi”.
Stemming	“wah”, “kalo”, ”bahas”, “politik”, “mending”, “bahas”, “kenapa”, “sirkuit”, “mandalika”, “dipakai”, “wsbk”, “diaspal”, “ulang”, “sector”, “event”, “motogp “material”, “aspal”, “kah”, “indikasi”, “korupsi”.	“wah”, “kalau”, ”bahas”, “politik”, “mending”, “bahas”, “kenapa”, “sirkuit”, “mandalika”, “pakai”, “wsbk”, “aspal”, “ulang”, “sektor”, “acara”, “motogp “material”, “aspal”, “kah”, “indikasi”, “korupsi”.

D. Term Frequency-Inverse Document Frequency (TF-IDF)

Merupakan proses untuk melakukan pembobotan kata, hal ini dilakukan dengan mencari nilai *Term Frequency* (TF), kemudian mencari nilai *Document Frequency* (DF), lalu mencari nilai *Invers Document Frequency* (IDF), dan menghitung bobot. *Term* adalah kata yang dapat digunakan untuk menentukan konteks suatu dokumen. Pada tabel 6 adalah contoh perhitungan TF-IDF.

Tabel 6. TF-IDF

Term	Tf					Df	IDF Log(n/df)	TF.IDF				
	D1	D2	D3	D4	D5			D1	D2	D3	D4	D5
allhamdulillah	1	0	0	0	0	1	0.699	0.175	0	0	0	0
sirkuit	1	0	1	0	1	3	0.222	0.0555	0	0.0555	0	0.0370
mandalika	1	1	1	0	1	4	0.097	0.0242	0.0194	0.0242	0	0.0162
indah	1	0	0	0	0	1	0.699	0.175	0	0	0	0
motogp	0	1	0	0	0	1	0.699	0	0.140	0	0	0
dampak	0	1	0	0	0	1	0.699	0	0.140	0	0	0
tumbuh	0	1	0	0	0	1	0.699	0	0.140	0	0	0
pariwisata	0	1	0	0	0	1	0.699	0	0.140	0	0	0
nikmat	0	0	1	0	0	1	0.699	0	0	0.175	0	0

mangkrak	0	0	1	0	0	1	0.699	0	0	0.175	0	0.398
anak	0	0	0	1	0	1	0.699	0	0	0	0	0
mahal	0	0	0	1	0	1	0.699	0	0	0	0.116	0
lahan	0	0	0	1	0	1	0.699	0	0	0	0.116	0
selesai	0	0	0	1	0	1	0.699	0	0	0	0.116	0
kalah	0	0	0	1	0	1	0.699	0	0	0	0.116	0
adil	0	0	0	1	0	1	0.699	0	0	0	0.116	0
pertamina	0	0	0	0	1	1	0.699	0	0	0	0	0.116
kayak	0	0	0	0	1	1	0.699	0	0	0	0	0.116
sepanjang	0	0	0	0	1	1	0.699	0	0	0	0	0.116
petronas	0	0	0	0	1	1	0.699	0	0	0	0	0.116

E. Pemodelan dengan Naïve Bayes

Naïve Bayes adalah metode klasifikasi yang sangat sederhana dan efisien. Metode klasifikasi Naïve Bayes juga merupakan metode yang populer untuk klasifikasi teks dan memiliki kinerja yang baik, namun juga memiliki kelemahan yaitu sangat sensitif terhadap seleksi fitur. Terlalu banyak fitur tidak hanya meningkatkan waktu perhitungan, tetapi juga dapat mengurangi akurasi klasifikasi. Naïve Bayes dapat dihitung dengan menggunakan persamaan 1, berikut merupakan contoh perhitungan Naïve Bayes, data sampel Naïve Bayes dapat terlihat pada tabel 7 dibawah ini:

Tabel 7. Contoh Sample tweet

Tweet	Label
alhamdulillah sirkuit mandalika indah	Positif
motogp mandalika dampak tumbuh pariwisata	Positif
motogp mandalika emang top	Positif
yup setuju event balap motogp mandalika	Positif
emang kamu pikir sirkuit mandalika hoax	Negatif
indonesia sirkuit mandalika bodoh bener	Negatif

Dalam Penelitian ini akan memakai 6 sampel tweet. Maka probabilitas dokumen positif adalah $P(\text{positif})=4/6 = 0.6$ dan probabilitas dokumen negatif adalah $P(\text{negatif})=2/6 = 0.3$ berdasarkan rumus yang tertulis pada (1) maka dapat dilakukan perhitungan probabilitas untuk menentukan kelas negatif atau positif sebagai berikut:

$$P(\text{alhamdulillah}|\text{positif}) = 1/19 + 1/30 = 2/49 = 0.04081$$

$$P(\text{sirkuit}|\text{positif}) = 1/19 + 1/30 = 2/49 = 0.04081$$

$$P(\text{mandalika}|\text{positif}) = 4/19 + 1/30 = 5/49 = 0.10204$$

$$P(\text{motogp}|\text{positif}) = 2/19 + 1/30 = 3/49 = 0.0612$$

$$P(\text{hoax}|\text{negatif}) = 1/11 + 1/30 = 2/41 = 0.0487$$

$$P(\text{bodoh}|\text{negatif}) = 1/11 + 1/30 = 2/41 = 0.0487$$

Contoh penentuan kelas dari sebuah tweet, dapat terlihat pada tabel 8 berikut ini:

Tabel 8. Menentukan Sentimen

Tweet	Sentimen
motogp mandalika petronas sponsor fine aja	-

Dapat dihitung nilai Naive Bayes untuk mendapatkan kelas positif dan negatif berdasarkan tweet diatas:

$$P(\text{motogp}|\text{positif}) \cdot p(\text{mandalika}|\text{positif}) \cdot p(\text{petronas}|\text{positif}) \cdot p(\text{sponsor}|\text{positif}) \cdot p(\text{fine}|\text{positif}) \cdot p(\text{aja}|\text{positif})$$

$$0.6 \cdot 0.06122 \cdot 0.10042 \cdot 0.04081 \cdot 0.04081 \cdot 0.04081 \cdot 0.04081 = 9.556 \times 10^{-6}$$

$$P(\text{Negatif}) \cdot P(\text{motogp}|\text{negatif}) \cdot p(\text{mandalika}|\text{negatif}) \cdot p(\text{petronas}|\text{negatif}) \cdot p(\text{sponsor}|\text{negatif}) \cdot p(\text{fine}|\text{negatif}) \cdot p(\text{aja}|\text{negatif})$$

$$0.3 \cdot 0.0487 \cdot 0.0487 \cdot 0.0487 \cdot 0.0487 \cdot 0.0487 \cdot 0.0487 = 3.669 \times 10^{-9}$$

Maka dapat disimpulkan bahwa nilai hasil perkalian menunjukkan nilai positif lebih tinggi yaitu 9.556×10^{-6} sementara nilai negatif lebih rendah yaitu 3.669×10^{-9} sehingga dapat disimpulkan kalimat tersebut mendapat sentimen positif.

F. Evaluasi

Tahap evaluasi model dilakukan dengan menggunakan Confusion Matrix dengan hasil sebagaimana terlihat di tabel 9.

Tabel 9. Confusion Matrix

	True Positive	True Negative
Positive	104	6
Negative	197	27

$$Recall = \frac{TP}{TP+FN} = \frac{104}{104+6} = 94.54\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{104}{104+197} = 34.51\%$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{104+27}{104+27+197+6} = 58.10\%$$

Berdasarkan perhitungan pada Tabel 9 dapat dilihat bahwa hasil dari pengolahan data sentimen dengan menggunakan *Naïve Bayes* menghasilkan *precision* sebesar 34.51% perhitungan *precision* menggunakan persamaan 3, *recall* sebesar 94.54% menggunakan persamaan 4, dan *accuracy* sebesar 58.10% menggunakan persamaan 2.

G. Hasil Analisis Sentimen

Tabel 10. Hasil Analisis Sentimen

Hasil Sentimen	Jumlah
Positif	396
Negatif	218
Netral	494

Pada tabel 10 terdapat total 1108 data yang telah masuk tahapan *preprocessing* maka didapatkan hasil sentimen positif sebanyak 396 *tweet*, lalu untuk sentimen negatif sebanyak 218 *tweet*, dan 494 sentimen netral, dapat disimpulkan bahwa hasil analisis sentimen masyarakat terhadap gelaran MotoGP Mandalika menggunakan 1108 data adalah netral.

IV. PENUTUP

Berdasarkan pengujian dan analisis yang telah dilakukan pada analisis sentimen masyarakat terhadap gelaran MotoGP Mandalika pada media sosial Twitter dengan metode *Naïve Bayes* dengan jumlah dataset yang digunakan sebanyak 1108 data dengan sentimen positif berjumlah 396 data, sentimen negatif berjumlah 218 data dan data netral berjumlah 494 data. Pengujian dilakukan dengan menggunakan metode *Naïve Bayes* dengan menggunakan perbandingan 80:20 mendapat nilai akurasi sebesar 58.10%. Semoga penelitian ini bisa disempurnakan oleh para peneliti di masa yang akan datang.

REFERENSI

- [1] S. Thaufik Rizaldi, A. al Khairi, "Text Mining Classification Opini Publik Terhadap Provider di Indonesia," 2021.
- [2] J. Saptia Kurnia, "Perhitungan Analisis Sentimen Berbasis Komparasi Algoritma Naive Bayes Dan K-Nearest Neighbour Berbasis Particle Swarm Optimization Pada Komentar Insiden Pembalap Motogp 2015," 2019. doi: 10.35968/jsi.v6i2.317.
- [3] A. K. Fauziyyah, "Analisis Sentimen Pandemi Covid19 Pada Streaming Twitter Dengan Text Mining Python," *Jurnal Ilmiah SINUS*, vol. 18, no. 2, p. 31, Jul. 2020, doi: 10.30646/sinus.v18i2.491.
- [4] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022.
- [5] R. Y. Hayuningtyas and R. Sari, "Analisis Sentimen Opini Publik Bahasa Indonesia Terhadap Wisata Tmii Menggunakan Naive Bayes Dan Pso," *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 37, 2019, [Online]. Available: <http://nusamandiri.ac.id/>
- [6] F. V. Sari and A. Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naive Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [7] J. Samuel *et al.*, "Feeling Positive about Reopening? New Normal Scenarios from COVID-19 US Reopen Sentiment Analytics," *IEEE Access*, vol. 8, pp. 142173–142190, 2020, doi: 10.1109/ACCESS.2020.3013933.
- [8] A. Z. Amrullah, A. Sofyan Anas, M. Adrian, and J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan

Seleksi Fitur Chi Square," *Jurnal*, vol. 2, no. 1, 2020, doi: 10.30812/bite.v2i1.804.

- [9] M. Christianto, J. Andjarwirawan, and A. Tjondrowiguno, "Aplikasi Analisa Sentimen Pada Komentar Berbahasa Indonesia Dalam Objek Video di Website YouTube Menggunakan Metode Naive Bayes Classifier," 2020.
- [10] L. Aji Andika and P. Amalia Nur Azizah, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," 2019.