

SENTIMEN ANALISIS PENGGUNA TWITTER TERHADAP SEA GAMES 2023 DENGAN METODE NAIVE BAYES

Wahyu Gumilang¹, Albert Riyandi²

¹Program Studi Sistem Informasi, Universitas Nusa Mandiri, Tangerang

E-mail: ¹1122003@nusamandiri.ac.id, ²Albert.abe@nusamandiri.ac.id

Abstract – The development of computers and telecommunications has made people often express their opinions through social media Twitter. As recently, after the 2023 SEA Games were held in Cambodia, various kinds of tweets appeared from various sports. From this opinion, it can be seen that the country's success in holding the Sea Games by conducting a sentiment analysis. So this research was conducted to find out how the public opinion regarding the holding of the 2023 SEA Games using the Naive Bayes Classifier algorithm. The data used was 1888 data obtained by crawling Twitter data using the keywords SEA Games and RapidMiner tools. The division of training data and testing data is carried out into two comparisons, namely 50:50 and 60:40. The best accuracy results are obtained with a data comparison of 50:50, which is 70.15%, while a data comparison of 60:40 has a smaller result, namely 70.07%.

Keywords : Sea Games, Sentiment Analysis, Naive Bayes, RapidMiner

I. PENDAHULUAN

1.1. Latar Belakang

Penggunaan internet dan media sosial saat ini sangat diminati oleh masyarakat di seluruh dunia. Seperti yang ditunjukkan oleh data We Are Social dan Hootsuite, per Januari 2023, akan ada 5,16 miliar pengguna internet dan 4,76 miliar pengguna media sosial di seluruh dunia. Menurut laporan tersebut, terdapat 167 juta pengguna aktif jejaring sosial di Indonesia pada Januari 2023. Teknologi yang dulu hanya bisa kita imajinasikan atau lihat saat menonton film-film sains fiksi, kini dapat dinikmati secara nyata[1].

Media sosial seperti twitter biasanya menjadi tempat untuk menuangkan opini masyarakat tentang berita yang sedang hangat dibicarakan. Adapun opini yang disampaikan dapat berupa opini positif, negative, ataupun netral. Bukan hanya mengenai peristiwa yang terjadi di Indonesia, tetapi juga terhadap peristiwa yang terjadi diluar negeri. Saat ini berbagai bidang tak lepas dari berbagai macam opini, baik dibidang politik, bisnis, keuangan sampai bidang olahraga.

Seperti setelah dilaksanakannya SEA Games 2023 di Kamboja baru-baru ini juga tak lepas dari berbagai macam cuitan disegala jenis cabang olahraga. Dari tweet yang banyak beredar, dapat diidentifikasi keberhasilan negara dalam penyelenggaraan SEA Games dengan melakukan sebuah opinion mining.

Opinion mining atau sentimen analisis adalah proses memahami, mengekstrak, dan mengolah data secara tekstual untuk memperoleh informasi sentimen yang ada dalam suatu kalimat opini. Sentimen analisis dilakukan untuk memperoleh pendapat atau opini didalam sebuah masalah maupun objek seseorang, apakah hasilnya cenderung berpandangan positif, negatif atau netral[2].

Sebuah penelitian mengenai analisis sentimen pada opini masyarakat tentang dampak virus corona di Twitter melakukan perbandingan dua algoritma yaitu Naive Bayes dan Support Vector Machine. Hasil yang diperoleh dari penelitian tersebut yaitu metode Naive Bayes mampu mengklasifikasi sentiment dengan nilai akurasi sebesar 81.07%, sedangkan SVM menghasilkan nilai yang lebih kecil yaitu 79.96%[3]. Dengan mengacu pada beberapa studi terdahulu, penelitian ini akan dilakukan menggunakan algoritma Naive Bayes.

Maka berdasarkan penjelasan di atas, penelitian ini diangkat dengan judul “**Sentimen Analisis Pengguna Twitter Terhadap SEA Games 2023 Dengan Metode Naive Bayes**”.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka rumusan masalah pada penelitian itu yaitu sebagai berikut:

1. Bagaimana mengetahui bahwa metode Naive Bayes mampu secara otomatis melakukan klasifikasi tweet pada Twitter dengan baik ?
2. Bagaimana informasi umum mengenai opini masyarakat tentang pelaksanaan SEA Games pada media sosial Twitter ?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

1. Menganalisis penggunaan metode Naive Bayes dalam mengklasifikasi tweet.
2. Mengetahui informasi umum mengenai opini masyarakat tentang pelaksanaan SEA Games pada media sosial Twitter.

II. TINJAUAN PUSTAKA

2.1 Teks Mining

Teks mining adalah salah satu bidang khusus dari *data mining*. *Teks mining* dapat didefinisikan sebagai suatu proses untuk menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi[4].

2.2 Sentimen Analisis

Sentimen analisis atau *opinion mining* mengacu pada bidang yang luas dari pengolahan bahasa alami, komputasi linguistik dan *text mining* yang bertujuan menganalisa pendapat, sentimen, evaluasi, sikap, penilaian dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Sentimen analisis juga bisa diartikan sebagai teknik atau metode untuk menentukan bagaimana suatu emosi diekspresikan melalui sebuah teks dan bagaimana emosi tersebut dapat diklasifikasikan sebagai positif atau negatif. Dengan demikian, sentimen analisis berfokus pada pengolahan opini yang mengandung polaritas, yaitu memiliki opini positif, negatif, atau netral[4].

2.3 Twitter

Twitter adalah sebuah situs web yang dimiliki dan dioperasikan oleh Twitter Inc., yang menawarkan jaringan sosial berupa mikroblog sehingga memungkinkan penggunaannya untuk mengirim dan membaca pesan *Tweets*[5].

Tweet dapat dilihat secara publik, tetapi pengirim dapat membatasi pesan ke daftar teman mereka. Pengguna dapat melihat tweet pengguna lain, yang dikenal sebagai pengikut atau follower. Semua pengguna dapat mengirim dan menerima tweet melalui situs web Twitter, aplikasi pihak ketiga yang kompatibel (ponsel). Pengguna dapat menulis pesan berdasarkan topik menggunakan tanda # (hashtag). Sementara itu, Anda bisa mention atau membalas postingan pengguna lain menggunakan simbol @. Twitter termasuk media sosial yang sangat populer dikalangan pengguna internet, hal ini dikarenakan kesederhanaan dan kemudahan dalam penggunaannya, serta pengguna dapat dengan bebas mengeluarkan pendapat atau opini mereka[6].

2.4 RapidMiner Studio

RapidMiner adalah salah satu alat data besar yang memiliki banyak keunggulan, beberapa di antaranya disediakan secara bebas sebagai perangkat lunak sumber terbuka, komunitas gratis sebagai

pengguna dan pengembang, dan nilai pengembangan dengan RapidMiner relatif rendah dibandingkan dengan solusi pengembangan data lainnya[7].

RapidMiner memproses data dengan mengekstrak pola dari kumpulan data dan kemudian menggabungkannya dengan database, kecerdasan buatan, dan metode statistik untuk mendapatkan informasi berkualitas tinggi dari teks yang diproses[8].

2.5 Naïve Bayes Classifier

Naive Bayes Classifier (NBC) merupakan metode klasifikasi berbasis probabilitas atau peluang, dimana dihitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. *Naive Bayes* menjadi salah satu metode klasifikasi yang memiliki keakuratan tinggi dalam mengklasifikasi juga memiliki waktu klasifikasi yang singkat sehingga dapat mempercepat proses sentimen analisis. Metode ini merupakan metode berdasarkan teorema dan probabilitas bayes dengan asumsi independensi yang kuat. Algoritma Naive Bayes sangat cocok untuk melakukan klasifikasi pada dataset yang bertipe nominal[9].

Klasifikasi ini merupakan klasifikasi yang bersifat *supervised learning* karena memiliki *supervisor* (manusia melakukan klasifikasi secara manual pada data yang digunakan dalam pelatihan) selaku pengajar dalam proses belajar atau learning. Ciri utama dari NBC ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian[10].

2.6 Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency - Inverse Document Frequency atau TF-IDF adalah suatu algoritma yang berfungsi untuk menghitung bobot kata yang terkadang digunakan pada suatu kumpulan kata. Secara sederhana, algoritma TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen. Algoritma ini menghitung sebuah nilai Term Frequency (TF) yaitu banyaknya kata yang muncul pada setiap dokumen, semakin banyak frekuensi kemunculan kata maka bobotnya semakin tinggi dan juga Inverse Document Frequency (IDF) yaitu perhitungan bagaimana kata terdistribusikan dalam dokumen, semakin sedikit frekuensi kemunculan kata maka bobotnya semakin tinggi[11]. Formula dari metodenya adalah sebagai berikut:

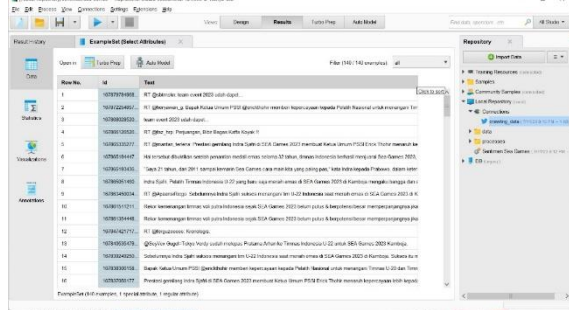
$$TFIDF(t) = TF * \log \frac{N}{df}$$

Huruf t merupakan term, TF merupakan jumlah dari term t yang terdapat dalam dokumen, N merupakan total dokumen, dan df merupakan jumlah dokumen yang terdapat term t[12].

III. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Proses crawling data dilakukan dengan kata kunci “SEA Games” dan “ASEAN Games”. Data diambil menggunakan tools Rapidminer Studio.



Gambar 1. Data hasil crawling.

Proses pengambilan data menggunakan RapidMiner menggunakan kata kunci tersebut menghasilkan data dengan jumlah 1888 data. Data tersebut kemudian disimpan dalam bentuk excel.

3.2 Pelabelan

Proses selanjutnya adalah pelabelan. Proses ini dilakukan secara manual yakni memberikan label positif, negatif dan netral pada setiap tweet. Pada proses ini hanya mengelompokkan data dengan cara memberi label kelas yang digunakan oleh penulis. Proses ini diperlukan sebagai data uji untuk proses selanjutnya yaitu klasifikasi.

3.3 Preprocessing

Proses selanjutnya adalah preprocessing. Proses ini akan dilakukan beberapa proses untuk mendapatkan data acuan yang siap untuk diproses. Preprocessing data dilakukan dengan beberapa tahapan, diantaranya *cleansing*, *case folding*, *tokenizing*, *filtering* dan *stemming*. Berikut ini contoh data yang belum melalui tahap preprocessing.

1	@albinovaliesh @AdhityaAchri squad sea games aja gk masuk baik sty maupun indra safri lebih milih arhan drpd firza di situasi persija yg buntu juga bisa dibantu dgn throw in arhan
2	RT @volleyballids: Performa tim -tim Asean di Challenge Cup jauh lebih baik dibandingkan pas Sea Games kemarin
3	Indonesia tentu gak boleh s...
4	Kenapa orang-orang pada ngeremehin thailand sih? Padahal mereka lumayan ngeri pas sea games kemarin walaupun ke comeback sama kamboja
5	@kim0nouch @aimrohman Hongkong deh spertnya kalo liat kemenangan sea games Indonesia mampu ngalahin thailand
6	Indra Sjafri, Pelatih Timnas Indonesia U-22 yang baru saja meraih emas di SEA Games 2022 di Kamboja mengaku bangga dan salut dengan komitmen untuk sepak bola Indonesia, termasuk yang dilakukan Prabowo #PrabowoUlutuSema #Neruskepakde #Prabowo Subianto
7	Ternyata dari perspektif lain kasian juga ke player yang dicemooh kayak si Jabran yang dibully sejak SEA Games sampe sekarang. Terus juga v'yn out, 'Acil out' dsb.
8	@dehadestu @verududie @tanvikantri Tapi slama nonton badminton beneran memang sbg tuan rumah ada apesnya n masalahnya sih, indonesia pernah apes atlet 1 hotel keracunan lampu tiba2 mati, malaysia apes karpel di pinggir court suka sobek sea games kamboja apes penginapan bocor hal2 yg mencoreng nama bgt
9	@D32NA @nOneNews Tapi ini kan melibatkan negara2 Asia Tenggara, pasti ada agenda yg lebih besar lagi. Apa mau pada minta transparansi bisa ikut SEA Games?
10	@INABadminton Di Sea Games dipanggil team tp ga pernah main krna posisi save mlulu nah skrg Alvi tunjukan kelasnya Nie anak emang punya 10 potensi besar So jgn smpe salah urus yaa PBSI kedepannya

Gambar 2. Data sebelum tahap preprocessing.

3.3.1 Cleansing

Tahap *cleansing* yaitu pembersihan data dari tanda baca (*punctuation*), *link*, *hashtag*, *mention*,

dan berbagai atribut yang tidak berpengaruh terhadap proses klasifikasi. Contoh data sesudah melalui tahap *cleansing* bisa dilihat pada gambar di bawah ini.

1	squad sea games aja gk masuk baik sty maupun indra safri lebih milih arhan drpd firza di situasi persija yg buntu juga bisa dibantu dgn throw in arhan
2	Performa tim -tim Asean di Challenge Cup jauh lebih baik dibandingkan pas Sea Games kemarin
3	Indonesia tentu gak boleh s...
4	Kenapa orang-orang pada ngeremehin thailand sih? Padahal mereka lumayan ngeri pas sea games kemarin walaupun ke comeback sama kamboja
5	Hongkong deh spertnya kalo liat kemenangan sea games Indonesia mampu ngalahin thailand setau gw jg indo 3 kali berturut
6	Indra Sjafri pelatih Timnas Indonesia U22 yang baru saja meraih emas di SEA Games Kamboja menyatakan bangga dan salutnya
7	Ternyata dari perspektif lain kasian juga ke player yang dicemooh kayak si Jabran yang dibully sejak SEA Games sampe sekarang
8	Terus juga v'yn out, 'Acil out' dsb
9	Tapi slama nonton badminton beneran memang sbg tuan rumah ada apesnya n masalahnya sih Indonesia pernah apes atlet 1 hotel keracunan lampu tiba2 mati, malaysia apes karpel di pinggir court suka sobek sea games kamboja apes penginapan bocor hal2 yg mencoreng nama bgt
10	Tapi ini kan melibatkan negara2 Asia Tenggara pasti ada agenda yg lebih besar lagi Apa mau pada minta transparansi bisa ikut SEA Games?
11	Di Sea Games dipanggil team tp ga pernah main krna posisi save mlulu nah skrg Alvi tunjukan kelasnya Nie anak emang punya 10 potensi besar So jgn smpe salah urus yaa PBSI kedepannya

Gambar 3. Data hasil cleansing

3.3.2 Case Folding

Tahap *case folding* merupakan tahap untuk mengubah seluruh huruf kapital jadi huruf kecil (*lowercase*). Hal ini dilakukan bertujuan untuk menghindari redudansi data yang hanya disebabkan oleh perbedaan penulisan jenis huruf. Contoh data sesudah melalui tahap *case folding* bisa dilihat pada gambar di bawah ini.

1	squad sea games aja gk masuk baik sty maupun indra safri lebih milih arhan drpd firza di situasi persija yg buntu juga bisa dibantu dgn throw in arhan
2	performa tim -tim asean di challenge cup jauh lebih baik dibandingkan pas sea games kemarin
3	indonesia tentu gak boleh s...
4	kenapa orang-orang pada ngeremehin thailand sih? padahal mereka lumayan ngeri pas sea games kemarin walaupun ke comeback sama kamboja
5	hongkong deh spertnya kalo liat kemenangan sea games indonesia mampu ngalahin thailand setau gw jg indo 3 kali berturut
6	indra sjafri pelatih Timnas Indonesia U22 yang baru saja meraih emas di sea games kamboja menyatakan bangga dan salutnya
7	ternyata dari perspektif lain kasian juga ke player yang dicemooh kayak si jabran yang dibully sejak sea games sampe sekarang
8	terus juga v'yn out, 'acil out' dsb
9	tapi slama nonton badminton beneran memang sbg tuan rumah ada apesnya n masalahnya sih indonesia pernah apes atlet 1 hotel keracunan lampu tiba2 mati, malaysia apes karpel di pinggir court suka sobek sea games kamboja apes penginapan bocor hal2 yg mencoreng nama bgt
10	tapi ini kan melibatkan negara2 asia tenggara pasti ada agenda yg lebih besar lagi apa mau pada minta transparansi bisa ikut sea games?
11	di sea games dipanggil team tp ga pernah main krna posisi save mlulu nah skrg alvi tunjukan kelasnya nie anak emang punya 10 potensi besar so jgn smpe salah urus yaa pbsi kedepannya

Gambar 4. Data hasil case folding.

3.3.3 Tokenizing

Data yang telah *cleansing* dan *case folding* kemudian akan dilanjutkan ke proses selanjutnya yaitu *tokenizing*. Tahap ini akan memecah *tweet* yang berupa kalimat menjadi kata untuk keperluan proses selanjutnya. Contoh data sesudah melalui tahap *tokenizing* bisa dilihat pada gambar di bawah ini.

1	['squad', 'sea', 'games', 'aja', 'gk', 'masuk', 'baik', 'sty', 'maupun', 'indra', 'safri', 'lebih', 'milih', 'arhan', 'drpd', 'firza', 'di', 'situasi', 'persija', 'yg', 'buntu', 'juga', 'bisa', 'dibantu', 'dgn', 'throw', 'in', 'arhan']
2	['performa', 'tim', 'tim', 'asean', 'di', 'challenge', 'cup', 'jauh', 'lebih', 'baik', 'dibandingkan', 'pas', 'sea', 'games', 'kemarin']
3	['indonesia', 'tentu', 'gak', 'boleh', 's...']
4	['asia', 'tenggara', 'indonesia', 'masih', 'no', '1', 'lah', 'ya', 'tapi', 'emang', 'lama', 'gak', 'ketemu', 'thailand', 'di', 'sea', 'games', 'semoga', 'saja']
5	['kenapa', 'orang-orang', 'pada', 'ngeremehin', 'thailand', 'sih', 'padahal', 'mereka', 'lumayan', 'neri', 'pas', 'sea', 'games', 'kemarin', 'walaupun', 'ke', 'comeback', 'sama', 'kamboja']
6	['hongkong', 'deh', 'spertnya', 'kalo', 'liat', 'kemenangan', 'sea', 'games', 'indonesia', 'mampu', 'ngalahin', 'thailand', 'setau', 'gw', 'jg', 'indo', '3', 'kali', 'berturut', 'jua', 'satu', 'gold', 'medal', 'sea', 'games', 'di', '2019', 'smp', 'kmaren', '2023']
7	['indra', 'sjafri', 'pelatih', 'timnas', 'indonesia', 'u22', 'yang', 'baru', 'saja', 'meraih', 'emas', 'di', 'sea', 'games', 'kamboja', 'menyatakan', 'bangga', 'dan', 'salutnya', 'untuk', 'prabowo', 'prabowountuksema', 'neruskepakde', 'prabowo', 'subianto']
8	['ternyata', 'dari', 'perspektif', 'lain', 'kasian', 'juga', 'ke', 'player', 'yang', 'dicemooh', 'kayak', 'si', 'jabran', 'yang', 'dibully', 'sejak', 'sea', 'games', 'sampe', 'sekarang', 'terus', 'juga', 'v'yn', 'out', 'acil', 'out', 'dsb']
9	['tapi', 'slama', 'nonton', 'badminton', 'beneran', 'memang', 'sbg', 'tuan', 'rumah', 'ada', 'apesnya', 'n', 'masalahnya', 'sih', 'indonesia', 'pernah', 'apes', 'atlet', '1', 'hotel', 'keracunan', 'lampu', 'tiba2', 'mati', 'malaysia', 'apes', 'karpel', 'di', 'pinggir', 'court', 'suka', 'sobek', 'sea', 'games', 'kamboja', 'apes', 'penginapan', 'bocor', 'hal2', 'yg', 'mencoreng', 'nama', 'bgt']
10	['tapi', 'ini', 'kan', 'melibatkan', 'negara2', 'asia', 'tenggara', 'pasti', 'ada', 'agenda', 'yg', 'lebih', 'besar', 'lagi', 'apa', 'mau', 'pada', 'minta', 'transparansi', 'bisa', 'ikut', 'sea', 'games?']
11	['di', 'sea', 'games', 'dipanggil', 'team', 'tp', 'ga', 'pernah', 'main', 'krna', 'posisi', 'save', 'mlulu', 'nah', 'skrg', 'alvi', 'tunjukan', 'kelasnya', 'nie', 'anak', 'emang', 'punya', 'potensi', 'besar', 'so', 'jgn', 'smpe', 'salah', 'urus', 'yaa', 'pbsi', 'ke depannya']

Gambar 5. Data hasil tokenizing.

3.3.4 Filtering

Proses ini berfungsi untuk menghilangkan kata yang tidak bermakna dan tidak berpengaruh terhadap hasil klasifikasi (*stopwords*). Contoh data sesudah melalui tahap *filtering* bisa dilihat pada gambar di bawah ini.

Table with 10 rows of stopwords and their corresponding Indonesian words, such as 'squad', 'sea', 'games', 'aja', 'gk', 'masuk', 'baik', 'sty', 'maupun', 'indra', 'safari', 'lebih', 'milih', 'arhan', 'drpd', 'firza', etc.

Gambar 6. Data hasil filtering.

3.3.5 Stemming

Stemming merupakan proses untuk mengubah kata berimbuhan menjadi kata dasar. Contoh data sesudah melalui tahap *stemming* bisa dilihat pada gambar di bawah ini.

Table with 10 rows of stemmed words, such as 'squad', 'sea', 'games', 'aja', 'gk', 'masuk', 'baik', 'sty', 'maupun', 'indra', 'safari', 'lebih', 'milih', 'arhan', 'drpd', 'firza', etc.

Gambar 7. Data hasil stemming.

3.4 Pembobotan

Tahap pembobotan yaitu tahap untuk melakukan pembobotan dengan menggunakan seleksi fitur. Hal ini dilakukan untuk mengubah data yang berupa nominal menjadi numerik. Adapun pada penelitian ini, seleksi fitur yang akan digunakan yaitu TF-IDF. Berikut ini merupakan hasil pembobotan TF-IDF menggunakan Rapidminer.

Table showing TF-IDF results for various words, with columns for word, TF, and IDF, and rows for words like 'squad', 'sea', 'games', 'aja', 'gk', 'masuk', 'baik', 'sty', 'maupun', 'indra', 'safari', 'lebih', 'milih', 'arhan', 'drpd', 'firza', etc.

Gambar 8. Data hasil TF-IDF.

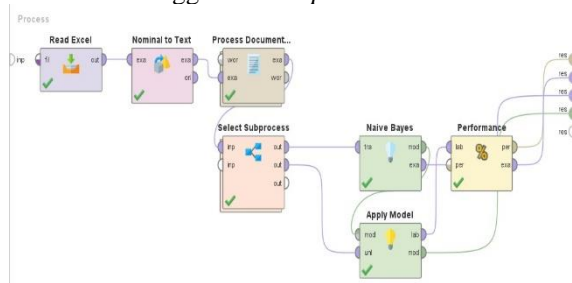
Contoh TF-IDF dalam satu kalimat:

Table with 3 columns: Kata, Kemunculan, Term Frequency Manual. Rows include words like 'timnas', 'indonesia', 'setia', 'samping', 'taeyong', 'debut', 'hingga', 'sekarang', 'nomor', 'kapten', 'games'.

Gambar 9. Contoh TF-IDF satu kalimat.

3.5 Klasifikasi

Tahap ini untuk memberikan pelatihan dan implementasi algoritma *Naive Bayes* pada data yang sudah di *preprocessing*. Berikut ini adalah proses klasifikasi menggunakan *Rapidminer*.



Gambar 10. Proses Klasifikasi Data Naive Bayes Classifier.

Setelah proses tersebut, dapat dilihat hasil pemodelan klasifikasi menggunakan *Naive Bayes Classifier* dan dilakukan pengujian, yang pertama dengan komposisi 50% data latih dan 50% data uji, selanjutnya dengan komposisi 60% data latih dan 40% data uji.

Pada pengujian pertama dilakukan dengan menggunakan 50% data latih dan 50% data uji. Algoritma *Naive Bayes* dilatih dengan menggunakan data latih sehingga menghasilkan suatu model, lalu selanjutnya dilakukan penerapan model tersebut kedalam data uji, secara keseluruhan tingkat akurasi yang dihasilkan sebesar 70.15% dengan 918 jumlah data uji. Berikut presentase lengkapnya pada tabel dibawah ini.

Tabel 1. Tabel hasil pengujian pertama 50:50

Confusion matrix table with columns: Benar Negatif, Benar Netral, Benar Positif, Class Precision. Rows: Prediksi Negatif, Prediksi Netral, Prediksi Positif, Class Recall.

Pada pengujian selanjutnya dilakukan dengan menggunakan 60% data latih dan 50% data uji. Algoritma *Naive Bayes* dilatih dengan menggunakan data latih sehingga menghasilkan suatu model, lalu selanjutnya dilakukan penerapan model tersebut kedalam data uji, secara keseluruhan tingkat akurasi yang dihasilkan sebesar 70.07% dengan 735 jumlah data uji. Berikut presentase lengkapnya pada tabel dibawah ini.

Tabel 2. Tabel hasil pengujian kedua 60:40

	Benar Negatif	Benar Netral	Benar Positif	Class Precision
Prediksi Negatif	35	67	2	33.65%
Prediksi Netral	57	466	46	81.90%
Prediksi Positif	9	39	14	22.58%
Class Recall	34.65%	81.47%	22.58%	

Maka dari dua pengujian yang dilakukan, diperoleh hasil akurasi terbaik untuk klasifikasi opini masyarakat mengenai SEA Games menggunakan metode *Naive Bayes* yaitu dengan pembagian data latih 50% dan data uji 50% yang menghasilkan akurasi 70.15%.

IV. PENUTUP

4.1 Kesimpulan

1. Pada pengujian pertama dengan komposisi 50% data latih dan 50% data uji menghasilkan rata-rata tingkat akurasi sebesar 70.15%, Dan pada pengujian selanjutnya dengan komposisi 60% data latih dan 40% data uji menghasilkan rata-rata tingkat akurasi sebesar 70.07%. Maka dapat dikatakan metode *Naive Bayes Classifier* dapat melakukan klasifikasi cukup baik.
2. Pada pengujian pertama didapatkan 126 data negatif dan 77 data positif maka didapatkan persentase 62% data negatif dan 38% data positif, lalu di pengujian selanjutnya didapatkan 101 data negatif dan 62 data positif maka didapatkan persentase 62% data negatif dan 38% data positif.

4.2 Saran

Dari kesimpulan yang sudah dipaparkan diatas, penulis memiliki saran yang sekiranya berguna untuk pihak terkait agar penelitian Sentimen Analisis Pengguna Twitter Terhadap SEA Games 2023 Dengan Metode *Naive Bayes* bisa berlanjut atau lebih bermanfaat, sebagai berikut.

1. Akurasi data dapat ditingkatkan lagi menjadi lebih dari 80% dengan catatan data yang diproses lebih berkualitas.
2. Banyaknya *Tweet* yang bernilai negatif menjadi pembelajaran untuk bangsa ini kelak ketika menjadi tuan rumah SEA Games selanjutnya.

DAFTAR REFERENSI

- T. Anwar, *Internet of Things (IoT): Panduan Pengenalan Dasar*. 2021. [Online]. Available: <http://www.commonroom.info>
- L. Ardiani, H. Sujaini, and T. Tursina, "Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.
- C. F. Hasri and D. Alita, "Penerapan Metode *Naive Bayes Classifier* Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- D. Ariyanti and K. Iswardani, "Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma *Naive Bayes*," *J. IKRA-ITH Inform.*, vol. 4, no. 3, pp. 125–132, 2020.
- A. Mukminin, "Analisis sentimen publik terhadap pelayanan tes," *Univ. Din.*, 2021.
- A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, "Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi," *J. Teknoinfo*, vol. 14, no. 2, p. 115, 2020, doi: 10.33365/jti.v14i2.679.
- E. D. Madyatmadja, S. I. Jordan, and J. F. Andry, "Big data analysis using rapidminer studio to predict suicide rate in several countries," *ICIC Express Lett. Part B Appl.*, vol. 12, no. 8, pp. 757–764, 2021, doi: 10.24507/icicelb.12.08.757.
- Ainurrohmah, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *Pros. Semin. Nas. Mat.*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- J. Suntoro, "Data Mining Algoritma dan Implementasi dengan Pemrograman PHP," *Elex Media Komputindo*, vol. 1. p. 196, 2019.
- S. Juniarsih, E. F. Ripanti, and E. E. Pratama, "Implementasi *Naive Bayes Classifier* pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam

Aspek Ekonomi,” *J. Sist. dan Teknol. Inf.*,
vol. 8, no. 3, p. 239, 2020, doi:
10.26418/justin.v8i3.39118.

10.32493/informatika.v5i3.6110.

- A. W. Andriyan Maulana Harry, “AUTOMATIC
SUMMARIZING DOKUMEN
REPOSITORY DENGAN TERM
FREQUENCY - INVERSE DOCUMENT
FREQUENCY METHOD,” *J. Tek. Inform.
Atmaluhur*, vol. 6, no. 1, p. 40, 2023.
- R. B. Trianto, A. Triyono, and D. M. P. Arum,
“Klasifikasi Rating Otomatis pada Dokumen
Teks Ulasan Produk Elektronik
Menggunakan Metode N-gram dan Naïve
Bayes,” *J. Inform. Univ. Pamulang*, vol. 5,
no. 3, p. 295, 2020, doi:

IDENTITAS PENULIS

Nama : Wahyu Gumilang
NIM : 1122008
TTL : Bekasi, 27 Mei 1993
Alamat Rumah : Binong Permai B32/22,
Tangerang
Telp. : 085693237045
Email : 11220083@nusamandiri.ac.id