



Analisis sentimen pandangan masyarakat terhadap piala dunia u-17 menggunakan Teknik *teks mining*

Aziz Musthafa¹, Dihin Muriyatmoko², R Muh Yusril Harmawan^{3*}

¹Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Darussalam Gontor
sisfotek@iaii.org (yusril@mhs.unida.gontor.ac.id)

Abstract

One of the most loved sports by people around the world is football. Indonesia is one of the countries with the most football fans in the world. Indonesia is one of the countries with the largest number of football fans in the world, with 77% of the Indonesian population interested in football. Based on research analysis, Indonesia was selected to host the 2023 U-17 World Cup. The decision was made after the International Football Federation (FIFA) granted hosting rights. Specifically to the President of the Indonesian Football Federation (PSSI). This research aims to classify public opinion related to the event from twitter social media into 3 class categories, namely neutral, positive and negative. In this research, the methods used are Naïve Bayes algorithm and Support Vector Machine (SVM) algorithm. The classification results show that the naïve bayes method has an accuracy result of 0.73 while for the Support Vector Machine method the accuracy value obtained is 0.84 which shows that Support Vector Machine has better accuracy than Naive Bayes. Based on the model classification, positive sentiment has the highest percentage of other classes with a percentage of 35%, followed by negative sentiment with a percentage of 31% and neutral sentiment is the minority class with a percentage of 33%. From the percentage obtained, it can be concluded that the public has a positive view of the organisation of the U-17 world cup in Indonesia. It is hoped that in the future this research can be improved and implemented better with additional algorithm methods or with a larger amount of data.

Keywords: World Cup U-17, Naïve Bayes, Support Vector Machine (SVM)

Abstrak

Salah satu olahraga yang paling digandrungi oleh masyarakat di seluruh dunia adalah sepak bola. Indonesia merupakan salah satu negara dengan penggemar sepak bola terbanyak di dunia. Indonesia merupakan salah satu negara dengan jumlah penggemar sepak bola terbesar di dunia, dengan 77% penduduk Indonesia meminati sepak bola. Berdasarkan analisa riset, Indonesia terpilih menjadi tuan rumah Piala Dunia U-17 2023. Keputusan itu diambil setelah Federasi Sepak Bola Internasional (FIFA) memberikan hak tuan rumah. Secara khusus kepada Presiden Federasi Sepak Bola Seluruh Indonesia (PSSI). Penelitian ini bertujuan untuk mengklasifikasi opini masyarakat terkait penyelenggaraan tersebut dari media sosial twitter ke dalam 3 kategori kelas, yaitu netral, positif dan negatif. Pada penelitian ini metode yang digunakan adalah algoritma Naïve Bayes dan algoritma Support Vector Machine (SVM). Hasil klasifikasi menunjukkan bahwa metode naïve bayes memiliki hasil akurasi sebesar 0.73 sedangkan untuk metode Support Vector Machine nilai akurasi yang didapatkan sebesar 0,84 yang menunjukkan bahwa Support Vector Machine memiliki akurasi yang lebih baik dari Naive Bayes. Berdasarkan dari pengklasifikasian model, sentimen positif memiliki presentase yang paling tinggi dari kelas lainnya dengan presentase 35%, kemudian disusul sentimen negatif dengan presentase 31% dan sentimen netral merupakan kelas minoritas dengan presentase 33%. Dari presentase yang didapat, dapat disimpulkan bahwa masyarakat memiliki pandangan yang positif terhadap penyelenggaraan piala dunia U-17 di indonesia. Diharapkan untuk kedepannya penelitian ini dapat ditingkatkan dan di implementasikan dengan lebih baik dengan tambahan metode algoritma ataupun dengan jumlah data yang lebih besar.

Kata kunci: Piala dunia U-17, Naïve bayes, Support Vector Machine (SVM).

1. Pendahuluan

Piala Dunia FIFA U17 adalah turnamen sepak bola internasional yang diselenggarakan oleh Federasi Sepak Bola Internasional (FIFA). Turnamen ini merupakan

ajang bergengsi yang menampilkan talenta-talenta muda dari berbagai negara di dunia. Berbeda dengan Piala Dunia senior, Piala Dunia U-17 diperuntukkan bagi pemain muda di bawah 17 tahun. Piala Dunia U-17 pertama kali diadakan pada tahun 1985 di Tiongkok.

Sejak saat itu, turnamen ini diadakan setiap dua tahun sekali. Beberapa pemain terkenal yang pernah berpartisipasi dalam Piala Dunia U-17 kemudian melanjutkan karier mereka di level senior dan mencapai kesuksesan yang besar.

Berdasarkan analisa riset, Indonesia terpilih menjadi tuan rumah Piala Dunia U-17 2023. Keputusan itu diambil setelah Federasi Sepak Bola Internasional (FIFA) memberikan hak tuan rumah. Secara khusus kepada Presiden Federasi Sepak Bola Seluruh Indonesia (PSSI) yang akan di selenggarakan pada tanggal 7 November sampai 2 Desember 2023.

Keputusan ini justru menimbulkan reaksi positif pada masyarakat Indonesia khususnya pada sosial media terlepas dari itu ini menjadi ajang Indonesia untuk meningkatkan kualitas sepakbola namun tidak sedikit juga reaksi masyarakat yang negative dan netral atas terselenggara nya piala dunia U-17 ini di Indonesia. Namun Indonesia sendiri merupakan salah satu negara dengan penggemar sepak bola terbanyak di dunia. Indonesia sendiri merupakan salah satu negara dengan jumlah penggemar sepak bola terbesar di dunia, dengan 77% penduduk Indonesia meminati sepak bola. Turnamen ini menjadi salah satu topik hangat di Twitter hingga memunculkan berbagai cuitan dari netizen sehingga menimbulkan munculnya pro dan kontra antara masyarakat dengan pihak tertentu.[1]

Media sosial sudah bukan lagi menjadi hal yang asing bagi sebagian besar masyarakat di dunia khususnya di Indonesia, Twitter menjadi salah satu platform yang hadir sebagai wadah bagi penggunaanya untuk mengungkapkan perasaan, pengalaman, dan pendapat yang mereka ungkapkan saat mengetik, dimana hal tersebut diketahui. bahwa Twitter merupakan platform terpopuler di dunia diantara media sosial lainnya.[2] Penentuan polaritas positif atau negatif suatu opini dapat dilakukan secara manual, namun seiring dengan bertambahnya jumlah sumber opini tentunya waktu dan tenaga yang diperlukan untuk melakukan klasifikasi juga semakin banyak, polaritas opini tersebut akan semakin banyak digunakan.

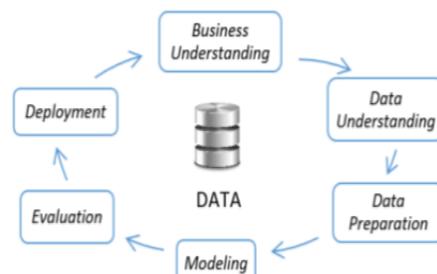
Analisis sentimen adalah cara mengumpulkan opini masyarakat dengan menggunakan jejaring sosial yang berisi layanan publik dan topik yang sedang tren. Analisis sentimen mengekstraksi opini publik terhadap topik, produk, atau layanan tertentu yang mengandung teks tidak terstruktur. Tujuan dari analisis sentimen adalah untuk mengetahui apakah pendapat seseorang dalam arti positif atau negatif, yang kemudian dapat digunakan untuk mengambil keputusan. Dalam Dalam penelitian analisis sentimen, terdapat beberapa metode yang telah terbukti efektif, yaitu *Support Vector Machines* (SVM) dan *Naive Bayes*. SVM adalah metode pembelajaran mesin yang mampu melakukan klasifikasi pada data dengan membangun model yang memisahkan antara kelas-kelas yang berbeda dengan batas keputusan yang optimal. Sedangkan *Naive Bayes* adalah metode

yang berdasarkan teorema *Bayes* dengan asumsi independensi antara setiap pasangan fitur.

Salah satu penelitian tentang analisis sentiment adalah analisis sentiment mengenai pembatalan piala dunia u 20 menggunakan metode naïve bayes. Penelitian ini bertujuan untuk mengkalsifikasi opini positif, negatif dan netral. Klasifikasi sentiment ini dengan metode Lexicon Based untuk menentukan sentimen positif, negatif dan netral, klasifikasi dengan metode Multinomial Naïve Bayes untuk perhitungan akurasi dengan Confusion Matrix. Berdasarkan sistem yang dibangun, didapatkan akurasi sebesar 85%.[1] Perbedaan penelitian saat ini dapat dilihat dari penelitian tersebut hanya menggunakan 1 metode yaitu metode naïve bayes dalam analisis sentiment dan objek penelitian yaitu pembatalan piala dunia u 20 di Indonesia. Sedangkan penelitian saat ini menggunakan 2 metode yaitu naïve bayes dan support vector machine dan berfokus pada analisis sentiment masyarakat terhadap penyelenggaraan piala dunia u 17. Diharapkan penelitian ini dapat memberi wawasan terhadap perkembangan sepak bola di Indonesia yang menyangkut pemain, Lembaga sepak bola, dan penyelenggaraan kejuaraan. Dengan mengumpulkan data teks yang berkaitan tentang piala dunia u-17 2023, penelitian ini menerapkan metode support vector machine dan naïve bayes untuk mengkalsifikasikan sentiment masyarakat.

2. Metode Penelitian

Tahap pada penelitian ini menggunakan model CrossIndustry Standard Process for Data Mining (CRISPDM). CRISP-DM adalah metode yang menyediakan standar baku pada data mining dan dapat diterapkan pada strategi pemecahan masalah umum.[3] Tahapan dalam CRISP-DM terlihat pada Gambar 1.



Gambar 1. CRISP-DM

2.1. Bussines Understanding

Tahap ini merupakan tahap pemahaman mengenai objek penelitian yang dilakukan. Penelitian dilakukan dengan mencari informasi melalui media sosial Twitter. Kemudian mencari komentar mengenai sentiment masyarakat terhadap piala dunia u-17 yang di selenggarakan di indonesia memungkinkan pengembang menggunakan API (Application Programming Interface) untuk mengambil data dari media sosialnya.

2.2. Data Understanding

Pada tahap ini proses pengambilan data melalui media social twitter. Dataset penelitian merupakan teks komentar dari twitter yang berbahasa Indonesia yang menggunakan kata kunci “piala dunia u 17 indonesia” denhan total data yang di dapatkan dari bulan November hingga Desember 2023 sebanyak 2044 komentar dari berbagai postingan dengan cara crawling dataset menggunakan bantuan dari google collab dan API (Application Programming Interface) dari twitter. Dengan jumlah data 2044. Dan data ini memiliki atribut: created_at, id_str, full_text, sentiment, quote_count, reply_count, retweet_count, favorite_count, lang, user_id_str, conversation_id_str, username, tweet_url. Dan memiliki kelas: positif, negatif, dan netral. setelah itu dilakukan penghapusan duplikat dan penghapusan data yang tidak penting, jumlah data yang awalnya berjumlah 4050 data menjadi 2044 data setelah dibersihkan.

2.3. Data Preparation

Pada tahap ini akan melalui persiapan dan pembersihan data yang mencakup positif, negatif dan netral untuk pengolahan data. Dalam penelitian ini pengolahan data menggunakan bahasa pemrograman Python. Berikut tahapan Data preparation, diantaranya:

Labeling, pada tahap ini data yang tweet yang telah terkumpul aka di beri label dengan tiga kategori yaitu positif, netral dan negatif. Data yang sudah diberi label harus menjalani proses validasi terlebih dahulu oleh ahli, yang mana data yang sudah diberikan label akan divalidasi, dimana pada penelitian ini validasi data dilakukan oleh salah satu Dosen Ilmu Komunikasi di Universitas Darussalam Gontor. Data yang sudah di beri label dan telah di validasi dapat di lihat pada Tabel 1.

Tabel 1. Data yang sudah diberi label

No	Date	Sentiment	Username	Kategori
1	Mon Nov 13 13:0 0:42 +000 0 2023	Emang gaboleh pelatih kek misal sekarang u-17 itu pelatih luar gitu, apa emang PSSI yang kepedean pakek LUKALPERIDE?? Soalnya yaa nonton timnas kek gini kalah bagus sama SSB yang kaga kebanyakan backpass longball hilang bola.	Eko27	Negatif
2		jam 03.45 WIB. Semoga Timnas Indonesia bisa memberikan penampilan terbaik dan meraih hasil positif. Mari dukung tim kita dengan penuh semangat! 👍 I D	dewianka	Positif

No	Date	Sentiment	Username	Kategori
3		Persiapkan Pengamanan Piala Dunia U-17, Polda Jatim Gelar TFG #PialaDuniaU17 #PengamananPialaD uniaU17	republikjja	Netral

Dari ketiga kategori yang paling mendominasi adalah sentimen netral dengan jumlah data sebanyak 1167, kemudian disusul sentimen positif sebanyak 695 data kemudia sentimen negatif dengan jumlah 182 data. Untuk hasil dari perbandingan ketiga kategori dapat dilihat pada Tabel 2.

Tabel 2. Data yang sudah diberi label

No	Kelas	Jumlah Data
1	Netral	1167
2	Positif	695
3	Negatif	182
	Total	2044

Preprocessing data adalah proses yang dilakukan pada data mentah. Tujuan dari preprocessing data adalah untuk membersihkan, mengubah format, dan menyesuaikan data agar sesuai dengan kebutuhan analisis atau pemodelan.[4] Adapun tahapan dari *preprocessing* ialah *case folding*, *Tokenizing*, *Filltering* dan *stemming*, untuk penjelasan masing-masing tahapan *preprocessing* adalah sebagai berikut:

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil, atau biasa di kenal sebagai lowercasting hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

Tokenizing adalah memisahkan setiap teks menjadi bagian yang lebih kecil, bagian-bagian tersebut umumnya disebut sebagai token. Tujuan dari *tokenizing* adalah memecah teks menjadi unit-unit yang lebih mudah diproses dalam analisis. Comtoh dari tokenizing adalah “Saya adalah mahasiswa teknik” akan di pecah menjadi (“saya”, “adalah”, “mahasiswa”, “teknik”).

Filltering adalah tahap mengambil katakata penting dari hasil term. Bisa menggunakan algoritma stoplist (membuang kata yang kurang penting) atau wordlist (menyimpan kata penting). Ditahapan ini tanda baca, symbol, emoji atau karakter khusus dan sejenisnya dihilangkan.

Stemming adalah dimana kata-kata yang diproses akan diubah menjadi kata dasar pada tahapan *stemming*. Contohnya seperti kata “Menendang” akan diubah menjadi “Tendang”, “Bertanding” akan diubah menjadi “Tanding” begitupun kata-kata lainnya, akan diubah menjadi kata dasar.

Pembobotan, TF-IDF (Term Inverse Frekuensi) adalah metode algoritmik yang memberikan bobot pada teks. TF adalah berapa kali kata tersebut muncul dalam

dokumen, sedangkan IDF adalah kebalikan dari dokumen yang memuat kata tersebut. TF dan IDF dikalikan untuk mendapatkan bobot kata.[5] Rumus pembobotan menggunakan TF-IDF adalah:

$$TF\text{-}IDF = TF * IDF$$

Prosedur pertama adalah term frekuensi (TF), yang menghitung berapa kali sebuah kata muncul dalam sebuah dokumen. Secara umum nilai TF akan dibagi dengan jumlah kata dalam dokumen karena panjang masing-masing dokumen berbeda. Berikut rumus TF:

$$TF = (\text{Kemunculan kata dalam document}) / (\text{Jumlah total kata dalam dokumen})$$

Prosedur kedua adalah inverse document frekuensi (IDF), yaitu nilai untuk mengukur pentingnya suatu kata, pada kata IDF yang sering muncul $br > 14$ dianggap sebagai kata yang kurang penting. Oleh karena itu, semakin kecil nilai IDF maka kata tersebut kurang penting dan sebaliknya. Rumus IDF adalah:

$$IDF = \log ((\text{Jumlah dokumen dalam korpus}) / (\text{Jumlah dokumen yang mengandung kata}))$$

2.4. Modeling

Pada tahap ini data dibagi menjadi *data training* dan *data testing*, *data training* merupakan data yang digunakan untuk mempelajari pola dan karakteristik data, sedangkan *data testing* merupakan data yang digunakan untuk menguji model yang telah mempelajari pola dan karakteristik data. Metode yang di gunakan dalam penelitian ini adalah metode Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM).

Algoritma Naive Bayes Classifier merupakan algoritma metode klasifikasi sederhana yang diturunkan dari teorema Bayes. Metode Naive Bayes merupakan salah satu metode klasifikasi text mining yang digunakan dalam analisis sentiment.[6] Adapun Persamaan *Naive Bayes* sesuai dengan Persamaan 1 berikut:

$$P(x|y) = \frac{p(y|x)p(x)}{p(y)} \quad (1)$$

y adalah Data dengan kelas yang belum di ketahui, x adalah Hipotesis data y merupakan suatu klas spesifik, dan $P(x|y)$ adalah robalitas hipotesis x berdasarkan kondisi y (posteriori probability), $P(x)$ adalah Probabilitas hipotesis x (prior probability), $P(y|x)$ adalah Probabilitas y berdasarkan kondisi pada hipotesis x, $P(y)$ adalah Probalitas dari y.

Support Vector Machine (SVM) menawarkan beberapa keuntungan dalam analisis sentiment, terutama karena kinerjanya yang baik pada data berdimensi tinggi dan kemampuannya untuk menemukan Solusi optimal bahkan Ketika data tidak sepenuhnya dapat dipisahkan secara linier[7]. Persamaan 2 adalah hyperplane untuk SVM linier diberikan sebagai:

$$f(x) = w^T x_i + b = 0 \quad (2)$$

W adalah Vektor bobot, X adalah Vektor fitur masukan, dan B adalah Istilah bias.

Vektor w diorientasikan tegak lurus terhadap fungsi pemisah yang disebut hyperplane. Dalam konteks ini, x mewakili titik data yang merupakan support vektor, titik yang berperan penting dalam menentukan lokasi hyperplane. Data x_i dibedakan menjadi sampel kelas -1 (negatif) dan sampel kelas +1 (positif), yang mencerminkan sifat klasifikasi biner SVM. Tujuan utamanya adalah memisahkan kedua kelas secara optimal. Persamaan 3 dan 4 hyperplane dapat dilihat pada rumus berikut.

$$w \cdot x_i + b \leq -1 \quad (3)$$

$$w \cdot x_i + b \leq +1 \quad (4)$$

Prinsip dasar yang mendasari metode ini adalah pembentukan hyperplane yang optimal dalam memisahkan dua kelas data dengan jarak sebesar mungkin terhadap sejumlah titik data dari masing-masing kelas. Pemisah tersebut dapat dirumuskan dengan Persamaan 5 sebagai berikut:

$$f(x) = w^T x + b \quad (5)$$

2.5. Evaluation

Pada tahanan ini setelah sistem berhasil berjalan Peneliti akan mengevaluasi metode klasifikasi dengan mengukur performa berupa akurasi presisi dan recall terhadap algoritma Model yang di gunakan menggunakan *Confusion Matrix* dan *cross validation*.

Confusion Matrix merupakan metode yang digunakan untuk mengevaluasi model algoritma klasifikasi yang digunakan, sehingga pengembang dapat mengetahui sejauh mana sistem klasifikasi dia berjalan dengan baik. Komponen utama Confusion Matrix adalah[8].

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <small>Type I Error</small>
	0 (Negative)	FN (False Negative) <small>Type II Error</small>	TN (True Negative)

Gambar 2. Confusion Matrix

- True Positive(TP) : Merupakan data Positif yang di prediksi benar.
- True Negative(TN) : Merupakan data Positif yang di prediksi benar.
- False Positive(FP) : Merupakan data Negatif namun diprediksi sebagai data Positif.
- False Negative(FN) : Merupakan data Positif namun diprediksi sebagai data Negative.

(2) *Cross Validation* adalah Teknik yang umum digunakan dalam *Machine Learning* dan *Teks Mining* untuk

mengevaluasi kinerja model ini melibatkan pembagian dataset kedalam himpunan bagian, pelatihan dan pengujian berulang-ulang himpunan bagian sehingga setiap titik data diambil sampelnya dan dilatih. Salah satu metode *Cross Validation* yang umum adalah 10-fold *Cross Validation*, Dimana data dibagi menjadi 10 bagian yang sama besar. Proses ini diulang dengan setiap bagian yang berbeda sebagai data uji. Tujuan dari *Cross Validation* untuk mengukur kualitas model yang digunakan, dengan metode *k-fold Cross Validation*. Teknik ini membantu dalam memastikan bahwa model tidak overfitting dan memiliki kemampuan generalisasi yang baik.

3. Hasil dan Pembahasan

Tahapan ini berisi pembahasan dan hasil dari proses implementasi yang telah dilakukan. Penelitian ini memiliki tujuan untuk memberi klasifikasi sentimen masyarakat pada media sosial twitter mengenai pelaksanaan piala dunia U-17 2023 di Indonesia menggunakan metode algoritma *Naïve Bayes* dan *SVM*.

3.1. Hasil Preprocessing

Data yang sudah terkumpul dan diberi label akan dibersihkan dengan tahapan-tahapan preprocessing diantaranya *case folding*, *tokenizing*, *filltering* dan *stemming*. Dengan demikian, *preprocessing* memiliki kontribusi yang besar dalam memastikan bahwa data yang digunakan sudah siap untuk dimasukkan kedalam model untuk analisis sentimen. Hasil dari preprocessing dapat di lihat pada Tabel 3.

Tabel 3. Hasil Preprocessing

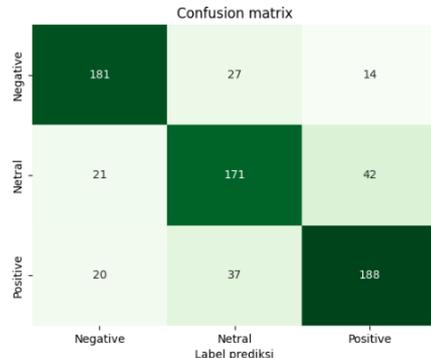
No	Preprocssing	Data Awal	Hasil
1	Case Folding	Semangat Timnas Indonesia U-17 kian membara jelang menghadapi Piala Dunia U-17 2023. #pialadunia	semangat timnas indonesia u-17 kian membara jelang menghadapi piala dunia u-17 2023 #pialadunia
2	Tokenizing	Semangat Timnas Indonesia U-17 kian membara jelang menghadapi Piala Dunia U-17 2023. #pialadunia	semangat', 'timnas', 'indonesia', 'u', 'kian', 'membara', 'jelang', 'menghadapi', 'piala', 'dunia', 'u', 'pialadunia'.
3	Filltering	Semangat Timnas Indonesia U-17 kian membara jelang menghadapi Piala Dunia U-17 2023. #pialadunia	semangat timnas indonesia u kian membara jelang menghadapi piala dunia u pialadunia
4	Stemming	Semangat Timnas Indonesia U-17 kian membara jelang menghadapi Piala Dunia U-17 2023. #pialadunia	emangat timnas indonesia u kian bara jelang hadap piala dunia u pialadunia

3.2. Hasil TF-IDF

Setelah proses *preprocessing* dan pembobotan menggunakan TF-IDF pemodelan dengan algoritma Data yang sudah melalui tahapan preprocessing selanjutnya diberi bobot pada tahapan ini, pada penelitian kali ini Teknik pembobotan yang digunakan adalah TF-IDF. Teknik ini merupakan teknik yang umum dipakai pada analisis teks khususnya analisis sentimen. Dari hasil TF-IDF ini terdapat 6065 dataset yang dihasilkan.

3.3. Evaluasi Model Naïve Bayes

Model pertama yang digunakan ialah metode *Naïve Bayes*. *Naïve Bayes* Naïve Bayes Classifier merupakan model probabilistik yang digunakan untuk proses klasifikasi berdasarkan teorema Bayes. Naïve Bayes Classifier dikenal sederhana yang berakar pada teorema bayes dan memiliki asumsi ketidaktergantungan (independent)[9]. Tahap evaluasi dilakukan menggunakan *Confusion Matrix*. Hasil *Confusion Matrix* dari metode *Naïve Bayes* dapat dilihat pada



Gambar 3. Confusion Matrix Naïve Bayes

Dari hasil confusion matrix diatas dapat disimpulkan bahwa hasil dari *true negative* berjumlah 181 dan *false negative* berjumlah 41. Jumlah dari *true netral* sebesar 171 sedangkan *false netral* berjumlah 63. *True positive* berjumlah 188 dan jumlah *false positive* ada 57. Hal ini dapat disimpulkan bahwa kinerja yang dihasilkan oleh model yang dipakai memiliki kinerja yang baik dalam pengklasifikasian data ke dalam tiga kategori. Jumlah yang tinggi dari *true negative*, *true positive* dan *true netral* menunjukkan bahwa model dapat memberikan klasifikasi ke sebagian besar data dengan benar. Dari hasil yang didapat, pemodelan menggunakan *Naïve Bayes* memiliki akurasi sebesar 0,77 dengan waktu *testing* selama 0,58s dengan rincian yang disajikan pada Tabel 4.

Tabel 4. Rincian Nilai Akurasi

Class	Precision	Recall	F1-Score
Negative	0,82	0,82	0,82
Netral	0,73	0,73	0,73
Positive	0,77	0,77	0,77
AVG	0,77	0,77	0,77

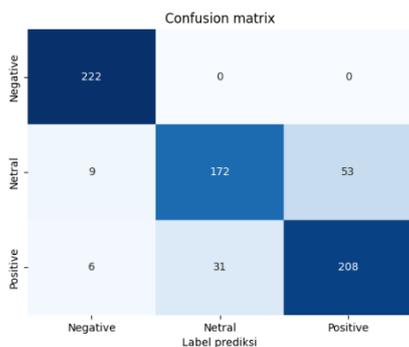
Selanjutnya proses validasi di penelitian ini menggunakan *k-fold cross validation*. Nilai akurasi yang didapatkan sebelum proses validasi dengan *k-fold cross validation* adalah 0.77 sedangkan hasil yang didapatkan setelah melalui tahap validasi dengan menggunakan *k-fold cross validation* adalah rata-rata 0.73 dan nilai tertingginya adalah 0.76. Digunakan 10 *k-fold* untuk validasi model *Naïve Bayes* dengan hasil yang dapat dilihat pada Tabel 5.

Tabel 5. Rincian Nilai

K-Fold	Hasil
1	0.74285714
2	0.73571429
3	0.71785714
4	0.72857143
5	0.725
6	0.73214286
7	0.73214286
8	0.73928571
9	0.75
10	0.76785714
AVG	0.7371428571428572

3.4 Evaluasi Model Support Vector Machine

Gunakan Metode selanjutnya yang digunakan adalah metode *SVM*. *SVM* mengharuskan sebuah teks diubah menjadi vektor sebelum digunakan untuk klasifikasi. Konsep dasar dari model ini adalah mencari *hyperplane* terbaik yang nantinya akan digunakan untuk melakukan proses klasifikasi pada data yang dipakai.[10] Sama seperti metode *Naïve Bayes*, *SVM* juga melalui proses pembobotan dengan menggunakan *TF-IDF* dan setelahnya masuk ke tahap klasifikasi dan diproses. Selanjutnya tahap yang dilakukan adalah tahap evaluasi dengan menggunakan *Confusion Matrix*. Untuk hasil *confusion matrix* dapat di lihat pada Gambar 4.



Gambar 4. Confusion Matrix SVM

hasil nilai yang tinggi menandakan bahwa model berhasil melakukan klasifikasi dengan benar ke sebagian besar data. Kali ini *SVM* mendapat nilai akurasi sebesar 0.85 dengan waktu testing 0,32s dengan detail yang disajikan pada Tabel 6.

Tabel 6. Rincian Nilai Akurasi

Class	Precision	Recall	F1-Score
Negative	0,94	1,00	0,97
Netral	0,85	0,74	0,79
Positive	0,80	0,85	0,82
AVG	0,86	0,86	0,86

Proses validasi dilakukan dengan menggunakan *k-fold cross validation*. Hasil nilai akurasi yang didapatkan sebelum tahap validasi dilakukan adalah 0.859, dan setelah tahap validasi dengan menggunakan *k-fold cross validation*, rata-rata nilai akurasi yang di dapatkan nilainya sama dengan nilai yang didapat sebelum tahap validasi yaitu sebesar 0.84 dengan nilai tertinggi sebesar 0.857. Digunakan 10 *k-fold* pada tahapan validasi model yang dibuat. Hasil yang didapat dari tahapan ini dapat dilihat pada Tabel 7.

Tabel 7. Rincian Nilai Akurasi

K-Fold	Hasil
1	0.82142857
2	0.83928571
3	0.83571429
4	0.81785714
5	0.84642857
6	0.84642857
7	0.85357143
8	0.85357143
9	0.83928571
10	0.85714286
AVG	0.8410714285714285

3.5 Hasil Perbandingan

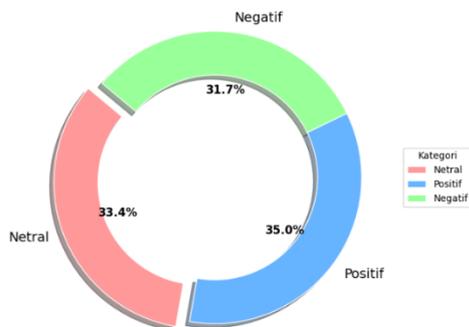
Hasil klasifikasi dari kedua model yang dipakai memiliki nilai akurasi yang berbeda, dimana untuk metode *Naïve Bayes* nilai akurasi yang didapat setelah proses *K-fold cross validation* sebesar 0.73, sedangkan untuk model *SVM* setelah tahapan *K-fold cross validation*, nilai akurasi yang didapat sebesar 0.84. Tabel perbandingan antara kedua model disajikan pada Tabel 8.

Tabel 8. Hasil Perbandingan

Metode	Hasil
Naïve Bayes	0,73
Support Vector Machine	0,84

Berdasarkan nilai akurasi yang didapat dari kedua model, setelah melewati *caross validation* *Naïve Bayes* menghasilkan akurasi sebesar 0,73 sedangkan *Support Vector Machine* menghasilkan nilai akurasi sebesar 0,84. Menunjukkan bahwa penggunaan metode *Support Vector Machine* lebih baik dibandingkan *Naïve Bayes* untuk proses klasifikasi pada analisis sentimen piala dunia u-17 pada penelitian ini. Berdasarkan hasil klasifikasi model, sentimen positif mendominasi dengan presentase sebesar 35%, untuk sentimen negatif memiliki presentase sebesar 31% dan pada sentimen netral presentase yang dimiliki sebesar 33%. Dari hasil presentase tersebut menunjukkan bahwa banyak data yang berisi pujian, dukungan dan kepuasan masyarakat terhadap penyelenggaraan piala dunia U-ss17 di indonesia dan di susul dengan data netral menunjukkan sikap informatif biasanya berisi berita, namun tidak sedikit data negatif yang menunjukkan kritik emosi yang kuat dan ketidak puasan masyarakat terhadap penyelenggarana piala dunia U-17 hal ini bisa mengindikasikan adanya pandangan buruk, kritik

dengan menjelekkkan suatu instansi, pemain, pelatih, penyelenggaraan piala dunia u-17, dan respon-respon buruk lainnya terhadap gelaran piala dunia u-17 2023 di Indonesia. Presentase dalam bentuk diagram dapat di lihat pada Gambar 5.



Gambar 5. Presentase Perbandingan Sentimen

4. Kesimpulan

Dari proses penelitian yang dilakukan melalui pengujian 2 model yang berbeda pada analisis sentimen mengenai gelaran piala dunia U-17, metode *naïve bayes* menghasilkan nilai akurasi 0,77 sedangkan nilai akurasi yang didapat dari metode Support Vector Machine adalah sebesar 0,85. Setelah melewati tahapan validasi menggunakan k-fold cross validation *naive bayes* memberikan rata-rata nilai akurasi sebesar 0,73 sedangkan Support Vector Machine memberikan rata-rata nilai akurasi sebesar 0,84. Berdasarkan nilai akurasi yang didapat, metode *SVM* memiliki nilai akurasi yang lebih tinggi dari *Naïve Bayes*, dapat disimpulkan bahwa metode *SVM* lebih baik daripada *Naïve Bayes* pada pengklasifikasian sentimen terkait penelitian ini. Dari hasil klasifikasi model *SVM*, sentimen positif memiliki presentase akurasi sebesar 35% disusul oleh negatif dengan presentase 31% kemudian sentimen netral dengan presentase sebesar 33%. Dengan jumlah sentimen positif yang tinggi dengan presentase 35% mengindikasikan dimana terdapat banyak data tweet yang berisi (pujian dan dukungan) terhadap penyelenggaraan in yang menunjukkan bahwa masyarakat memiliki pandangan yang baik terhadap gelaran piala dunia U-17 2023 di Indonesia. Presentase 31% pada sentimen negatif menunjukkan bahwa data tweet yang mengandung sentimen negatif juga cukup signifikan meskipun tidak sebanyak positif. Hal ini bisa mengindikasikan adanya pandangan buruk, kritik dengan menjelekkkan suatu instansi, pemain, pelatih, penyelenggaraan piala dunia U-17, dan respon-respon buruk lainnya gelaran piala dunia u-17 2023 di indonesia. Sedangkan sentiment netral sebesar 33%

menunjukkan sentiment informatif biasanya hanya berisi berita yang berkaitan tentang piala dunia U-17 yang diselenggarakan di Indonesia.

Daftar Rujukan

- [1] R. Sulastiyono, A. Setiawan, and S. Nugroho, "Sentimen Analisis Pembatalan Indonesia Menjadi Tuan Rumah Piala Dunia U-20 Menggunakan Metode Naïve Bayes," vol. 4, no. 4, pp. 1387–1394, 2023, doi: 10.47065/josh.v4i4.3737.
- [2] H. Ramanizar, A. Fajri, R. Binsar Sinaga, H. Mubarak, A. D. Pangestu, and D. S. Prasvita, "Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Naïve Bayesian Classification dan Support Vector Machine," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. September, pp. 166–175, 2021.
- [3] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [4] H. Rahmah, E. Putra, and U. Budiyanto, "KOMENTAR YOUTUBE INDONESIA TUAN RUMAH PIALA DUNIA U-20 IMPLEMENTATION OF KNN ALGORITHM FOR SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON THE INDONESIAN HOST OF FIFA U-20 WORLD CUP," vol. 2, no. September, pp. 369–378, 2023.
- [5] R. R. Sani, Y. A. Pratiwi, S. Winarno, E. D. Udayanti, and F. Alzami, "Analisis Perbandingan Algoritma Naive Bayes Classifier dan Support Vector Machine untuk Klasifikasi Berita Hoax pada Berita Online Indonesia," *J. Masy. Inform.*, vol. 13, no. 2, pp. 85–98, 2022, doi: 10.14710/jmasif.13.2.47983.
- [6] F. Sains, "The 2 st Seminar Nasional dan Prosiding Scitech 2023 SOSIAL MENGGUNAKAN METODE NAIVE BAYES) SENTIMENT ANALYSIS OF PUBLIC VIEWS ON QATAR GOVERNMENT POLICIES IN THE 2022 WORLD CUP ON SOCIAL MEDIA USING," pp. 559–565, 2023.
- [7] B. W. Sari and F. F. Haranto, "Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 171–176, 2019, doi: 10.33480/pilar.v15i2.699.
- [8] D. Rahma Putri, B. Arif Dermawan, I. Purnamasari, U. H. Singaperbangsa Karawang Jl Ronggo Waluyo, and T. Timur, "Implementasi Modified Enhanced Confix Stripping Stemmer pada Klasifikasi Fake News Covid-19," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 589–600, 2021.
- [9] C. H. Yutika, A. Adiwijaya, and S. Al Faraby, "Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 422, 2021, doi: 10.30865/mib.v5i2.2845.
- [10] P. R. Saputra Andri, Subing Mulia, "Perbandingan Metode Naïve Bayes Classifier Dan Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Mengenai Piala Dunia Fifa 2022," *Teknomatika*, vol. 13, no. 01, pp. 22–31, 2023, [Online]. Available: <http://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/616>