

ANALISIS SENTIMEN PADA MEDIA SOSIAL TWITTER(X) TERHADAP PEMAIN DIASPORA DI TIM NASIONAL SEPAK BOLA INDONESIA STUDI KASUS PESEPAKBOLA MEES HILGERS DAN ELIANO REINJDERS MENGGUNAKAN METODE NAÏVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)

¹Ade Bagus Ferdiyawan, ²Nurasiah, ³Imam Fauzy, ⁴Winanti, ⁵Sucipto Basuki

¹Mahasiswa Prodi Sistem Informasi, Universitas Insan Pembangunan Indonesia

^{2, 3, 4, 5}Dosen Tetap Prodi Sistem Informasi, Universitas Insan Pembangunan Indonesia

¹ferdytayu@gmail.com, ²nurash_ip@yahoo.com, ³imamipb42@gmail.com

⁴win_anti@yahoo.co.id, ⁵ciptainsan@yahoo.com

ABSTRAK

Pemain diaspora semakin memegang peranan penting dalam skuad tim nasional sepak bola Indonesia. Penelitian ini bertujuan untuk menganalisis sentimen publik di twitter (X) terkait dua pemain diaspora, Mees Hilgers dan Elliano Reindjs, dengan memanfaatkan algoritma *Naïve Bayes* dan *Support Vector Machine* serta membaginya menjadi kelas positif, negatif dan netral. Hasil penelitian ini menunjukkan bahwa sentimen masyarakat terhadap pemain diaspora di tim nasional sepak bola Indonesia cenderung positif dengan nilai presentase sebesar 71,2% atau 640 tweet positif, kemudian 25,5% atau 229 tweet negatif dan 3,3% atau 30 tweet netral. Penelitian ini menggunakan metode SEMMA (*Sample, Explore, Modify, Model, Assess*) sebagai acuan metodologi penelitian. Pada metode *Support Vector Machine* berasio 90:10 dengan parameter C: 100, gamma: 0,1 dan kernel: rbf memiliki hasil *accuracy* lebih baik dibandingkan dengan metode *Naïve Bayes* pada rasio yang sama, dengan presentase perbandingan 86% : 74%. Melalui kajian ini, peneliti ingin memahami persepsi publik terhadap pemain diaspora setra kontribusinya dalam perkembangan sepak bola Indonesia. Diharapkan, hasil penelitian ini dapat memberikan wawasan yang lebih mendalam tentang dinamika opini publik di media sosial terkait isu-isu sepak bola nasional.

Kata Kunci : Analisis, Sentimen, Diaspora, Naturalisasi, Tim Nasional, Sepak Bola, Indonesia, *Naïve Bayes*, *Support Vector Machine*.

PENDAHULUAN

Sepak bola merupakan salah satu olahraga yang paling popular hampir diseluruh dunia, termasuk di Indonesia. Bahkan di Indonesia, sepak bola mempunyai basis penggemar terbesar diantara olahraga lainnya. Segala perkembangan dunia sepak bola, khususnya yang terkait dengan tim nasional (Timnas) Indonesia selalu ramai diperbincangkan di berbagai media, baik tradisional maupun digital. Salah satu isu yang menarik perhatian publik akhir-akhir ini adalah penggunaan atlet diaspora, atau atlet keturunan Indonesia yang lahir dan besar di luar negeri. Disatu sisi para pemain diaspora ini kerap dianggap memiliki kualitas yang bisa meningkatkan performa timnas Indonesia, dikarenakan mereka bermain dan dibesarkan di daerah yang kualitas serta lingkungan sepak bolanya lebih baik daripada di Indonesia. Namun disisi lain dianggap dapat menghambat perkembangan talenta-talenta asli

Indonesia dan dapat pula menimbulkan ketergantungan jangka panjang sehingga kualitas pembibitan dan liga tidak akan berjalan dengan maksimal karena terus bergantung kepada pemain keturunan.

Dalam beberapa dekade terakhir perkembangan teknologi informasi dan komunikasi berkembang sangatlah pesat. Hal ini tentu membawa banyak perubahan dalam berbagai aspek kehidupan, termasuk cara orang-orang berkomunikasi dan mencari informasi. Sejak adanya *handphone*, ponsel pintar atau tablet. Semua orang dapat mengakses informasi dari mana saja dan kapan saja. Salah satu keuntungan dari perkembangan teknologi yang semakin canggih adalah terciptanya berbagai media sosial yang memiliki berbagai jenis dan ragam kegunaan. Jumlah pengguna media sosial di seluruh dunia telah mencapai 63,8% dari total

populasi global ini menunjukkan bahwa lebih dari setengah penduduk dunia memanfaatkan berbagai bentuk media sosial [1].

Salah satu *platform* media sosial yang banyak penggunanya adalah twitter(X). Twitter merupakan salah satu media sosial yang sangat populer di Indonesia sebagai *platform* untuk menyuarakan aspirasi melalui *tweet* atau cuitan [2]. Penulis memilih twitter(X) sebagai tempat penelitian karena twitter(X) memiliki jumlah cakupan data yang sangat besar serta datanya mudah diakses, selain itu *platform* ini memungkinkan orang-orang untuk bertukar komentar atau *tweet* secara langsung atau *real-time* ditambah dengan adanya fitur *hashtag* dan trending topik memudahkan dalam penyeleksian data yang akan digunakan dalam penelitian secara lebih spesifik.

Sentimen analisis atau *opinion mining* dapat diartikan sebagai bidang ilmu yang meneliti bagaimana klasifikasi sentimen. Topik ini berpusat pada kegiatan pengelompokan sentimen berlandaskan teks opini terhadap pembahasan masalah yang menarik [3]. Sentimen biasanya akan dikategorikan berdasarkan polaritasnya, apakah positif atau negatif dengan menggunakan teknik *Natural Language Processing* (NLP) [4]. Dalam analisis sentimen memiliki berbagai metode klasifikasi yang dapat digunakan. Pada penelitian ini penulis menggunakan metode *naïve bayes* dan *support vector machine* karena keduanya merupakan algoritma yang biasa digunakan untuk klasifikasi dan prediksi alasan lain adalah karena keduanya merupakan salah dua metode yang paling popular yang digunakan untuk analisis sentimen di media sosial. Studi kasus yang dipakai untuk bahan penelitian ini adalah bergabungnya 2 (dua) pemain timnas Indonesia, yaitu Mees Hilgers dan Eliano Reinjders.

Tujuan Penelitian:

- a. Guna mengetahui sentimen publik di media sosial twitter(X) terhadap proses naturalisasi Mees Hilgers dan Eliano Reinjders sebagai pemain diaspora di timnas Indonesia.
- b. Guna mengetahui perbandingan performa metode klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM) dalam

menganalisis sentimen publik terkait pemain diaspora di twitter(X).

- c. Guna mengetahui kata yang paling banyak muncul dalam *tweet* terkait naturalisasi pesepakbola Mees Hilgers dan Eliano Reinjders.

Analisis Sentimen

Analisis sentimen termasuk ke dalam salah satu bidang dari *Natural Language Processing* (NLP) dan merupakan suatu proses yang digunakan untuk membantu mengidentifikasi isi dari dataset yang berupa opini atau pandangan (sentimen) berbentuk teks terhadap suatu isu atau kejadian bersifat positif, negatif atau netral [5]. Dalam analisis sentimen terdapat proses penambangan data yang disebut *text mining*. *Text mining* adalah ekstraksi informasi dari data sumber yang belum terstruktur yang mengacu pada teknik penambangan data untuk menganalisis dan memproses data [6].

Media Sosial

Media sosial adalah seperangkat alat komunikasi dan kolaborasi baru yang memungkinkan terjadinya berbagai jenis interaksi yang sebelumnya tidak tersedia bagi orang awam [7].

Twitter(X)

Merupakan salah satu media sosial yang paling populer didunia dan di Indonesia. salah satu layanan yang disediakan oleh twitter(X) kepada penggunaanya adalah pembuatan pesan status (disebut “*tweet*”) yang dapat dibaca oleh pengguna twitter(X) lainnya dan biasanya berisi ungkapan pendapat pengguna dalam berbagai topik dengan batasan sebanyak 140 karakter, sehingga twitter(X) menjadi salah satu situs yang menyediakan kumpulan data opini dari masyarakat di seluruh dunia [8].

Metode SEMMA

SEMMA adalah singkatan dari (*Sample, Explore, Modify, Model, and Access*) merupakan teknik penggalian data yang dikembangkan oleh SAS Institut [9]. Metode ini digunakan untuk metodologi penambangan data untuk memecahkan berbagai masalah bisnis, termasuk identifikasi penipuan, retensi, pergantian pelanggan, pemasaran basis data, loyalitas pelanggan dan lain sebagainya [10].

Naïve Bayes

Merupakan suatu metode klasifikasi yang berasal dari teorema bayes yang menggunakan probabilitas dan statistik untuk memprediksi peluang berdasarkan pengalaman masa sebelumnya, dengan asumsi yang sangat sederhana tentang ketergantungan antara kondisi atau kejadian [11].

Support Vector Machine

Support Vector Machine atau SVM adalah salah satu algoritma dari *Machine Learning* yang berasal dari teori pembelajaran statistik yang biasa digunakan untuk tugas klasifikasi dan regresi [12].

Lexicon Based

lexicon-based merupakan fitur pembobotan kata dengan menggunakan kamus sentimen yang berisi kata-kata opini dan membandingkannya dengan data untuk mengidentifikasi nilai suatu kata [13].

Term Frequency-Invers Document Frequency

Merupakan metode pembobotan kata dengan menentukan frekuensi kemunculan kata dalam sebuah dokumen dengan cara memperhatikan frekuensi atau seringnya kata tersebut muncul dalam sebuah dokumen [14].

Confusion Matrix

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi proses kinerja dari model klasifikasi, matrik ini dimungkinkan untuk menentukan kualitas performa model klasifikasi[15].

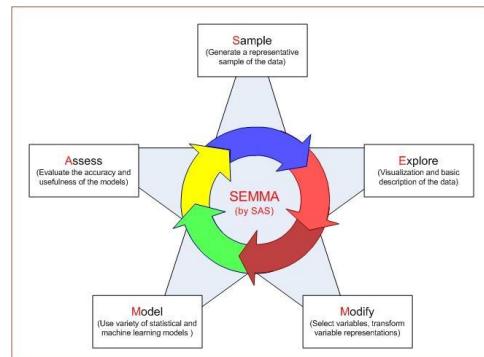
K-fold Cross Validation

Merupakan salah satu teknik pengujian *cross-validation* dimana data akan dibagi menjadi data training dan data testing yang akan diacak dan dibagi ke dalam k himpunan yang berbeda [16].

METODE

Data yang digunakan dalam penelitian ini adalah data *tweet* dari berbagai postingan yang berkaitan dengan timnas sepak bola indonesia, Mees Hilgers dan Eliano Reijnders. Terhitung dari tanggal 1 Juni 2024 hingga 31 Oktober 2024. Metodologi yang diterapkan dalam penelitian ini merujuk pada SEMMA *Data*

Mining Process. Penelitian ini menggabungkan pendekatan kualitatif dan kuantitatif dengan menerapkan metode SEMMA adapun alurnya terdapat pada Gambar 1.



Gambar 1. SEMMA data mining process (Hendarta 2014)

Sample

Tahap ini dilakukan dengan mencari teori-teori yang berkaitan dengan penelitian dari jurnal, buku atau website yang relevan. Data diambil dari twitter(X) dengan bantuan *tweet-harvest* dan *node.js* yang ada di *google colaboratory*. Data yang didapat kemudian disimpan dengan format csv dan dikumpulkan dalam satu folder.

Explore

Setelah terkumpul semua dataset dengan nama dokumen adalah *keyword* dari masing-masing pencarian, kemudian dijadikan kedalam satu dokumen. Setelah itu dataset dieksplorasi untuk menghilangkan kolom-kolom hasil *crawling* data yang tidak diperlukan dan memilih *tweet* yang relevan dengan topik yang dibahas. Menyisakan kolom teks yang berisi kumpulan *tweet* dan akan digunakan dalam tahap berikutnya.

Modify

Pada tahap ini, data *tweet* yang belum diolah akan melalui proses *pre-processing* data, seperti *cleaning*, *case folding*, *tokenizing*, *stopwords*, normalisasi dan *stemming* agar data yang dihasilkan menjadi lebih terstruktur. Kemudian dilanjutkan dengan melakukan visualisasi kata-kata yang sering muncul dengan *wordcloud*.

Model

Pelabelan dan perhitungan sentimen dilakukan setiap barisnya, menggunakan metode *lexicon based* dan *term frequency-invers document frequency* (TF-IDF). Perhitungan dan pelabelan skor sentimen adalah sebagai berikut:

$$\text{Skor} = (\sum \text{kata positif}) + (\sum \text{kata negatif})$$

Perhitungan skor tf-idf adalah sebagai berikut:

$$\text{Skor} = (\sum \text{kata setiap baris})$$

Perhitungan kombinasi dilakukan dengan membagi bobot atau rasio *lexicon* : tf-idf = 60%:40% dengan rumus sebagai berikut:

$$\text{Skor gabungan} = (\text{bobot lexicon} * \text{skor lexicon}) + (\text{bobot tfidf} * \text{skor tfidf})$$

Pemilihan besaran nilai sentimen akhir adalah positif $> 0,7$ untuk sentimen negatif adalah $< 0,5$ dan selain keduanya bernilai netral. Langkah berikutnya adalah *splitting* data menjadi rasio 90:10, 80:20, dan 70:30. Kemudian dilakukan perhitungan dengan metode *naïve bayes* dan *support vector machine* (SVM) untuk mengetahui *accuracy* pelabelan yang telah dilakukan.

Assess

Pada tahap ini, akan dilakukan evaluasi terhadap pemodelan yang telah dibuat. Evaluasi ini bertujuan untuk membandingkan hasil yang diperoleh dari model prediksi dengan data uji yang telah dilengkapi dengan label sentimen [19]. Hasil evaluasi akan dihitung menggunakan metrik-metrik seperti *precision*, *recall*, *f1-score*, dan *accuracy* menggunakan metode *10-fold cross validation* dan *confusion matrix* untuk mendapatkan hasil yang optimal dari masing masing model klasifikasi.

Pengujian Akhir

Tahap terakhir adalah pengujian dengan memasukkan komentar baru yang belum pernah diprediksi sebelumnya.

HASIL DAN PEMBAHASAN

Sample

Pencarian data menggunakan *tools google colab* dengan memanfaatkan *library node.js* dan *tweet harvest* dengan cara mengambil *twitter auth token*.

conversation_id_str	created_at	favorite_count	full_text
1851767462360158272	Wed Oct 30 23:25:41 +0000 2024	0	Tijani Reijnders Bangga Punya Darah Indonesia...
1851762147082277003	Wed Oct 30 23:04:34 +0000 2024	115	seneng banget meliat verdonk dan eliano yang s...
1851426235618115602	Wed Oct 30 22:16:39 +0000 2024	0	@FaktaSepakbola Congrats bng verdonk tetap sem...

Gambar 2. Data hasil *crawling*

Diperoleh data *tweet* sebanyak 1254 data pada rentang waktu 1 Juni 2024 sampai 31 Oktober 2024 didalamnya terdapat 15 kolom tiap dokumennya, kemudian data disimpan dengan format csv.

Explore

Data yang diperoleh kemudian dirapikan dan dilakukan pemilihan *tweet* yang relevan dengan penelitian, menggabungkan dokumen-dokumen hasil *crawling* data dan membuang kolom yang tidak digunakan, hanya menyisakan kolom *full_text* saja dengan jumlah *tweet* yang siap diolah sebanyak 917 *tweet* dan disimpan dengan nama master.csv.

full_text

0	Nge-fans sama Mees sebagai athlete aja tolong ...
1	Kan ada Surya Insomnia mirip2 lah sm Mees Hilgers
2	Gue sedih sm beberapa orang di TT yg tinggal d...
3	Bek Timnas Indonesia Mees Hilgers diyakini bis...
4	Omongan Pelatih Lazio soal Klub Mees Hilgers M...

Gambar 3. Seleksi data

Modify

Setelah data diseleksi kemudian dilakukan tahap *preprocessing* meliputi tahap *cleaning*, *case folding*, *tokenizing*, *stop words*, normalisasi dan *stemming* menghasilkan data bersih sebanyak 899 *tweet* dan disimpan dengan nama hasil.csv.

	stemming
0	['idola', 'mees', 'atlet', 'tolong', 'jangan', ...]
1	['surya', 'kurang tidur', 'sama', 'mees', 'hil...']
2	['sedih', 'sama', 'orang', 'tiktok', 'tinggal'...]
3	['main belakang', 'tim nasional', 'indonesia', ...]
4	['omong', 'latih', 'lazio', 'klub', 'mees', 'h...']
...	...
894	['ayo', 'mees', 'kamar', 'sama', 'es', 'sendy'...]
895	['aktif', 'sekali', 'mees']
896	['sambut', 'baru', 'tar', 'tim nasional', 'me...']
897	['mees', 'tidak', 'ihat', 'karena', 'tidak', ...]
898	['mees', 'sibuk', 'buka', 'latih', 'semanqat']

Gambar 4. Hasil *preprocessing*

Kemudian divisualisasikan dengan *word cloud* dan *plot bar*.



Gambar 5. Wordcloud hasil preprocessing

Gambar diatas merepresentasikan kata-kata yang sering muncul dengan ukuran kata yang besar berbanding terbalik dengan kata yang jarang muncul berukuran kecil.



Gambar 6. Kata yang sering muncul

Pada gambar diatas menunjukkan kata yang sering muncul adalah kata mees dengan jumlah 442 kata, indonesia 341 kata dan tim 329 kata.

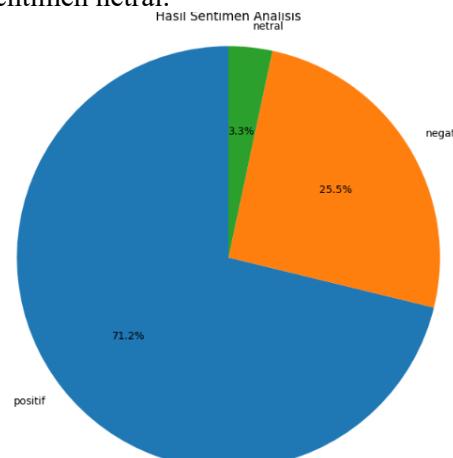
Model

Data yang sudah bersih kemudian dilakukan pelabelan dan perhitungan skor sentimen dengan metode *lexicon based* dan TF-IDF.

	stemming	skor_sentimen	skor_tfidf	skor_gabungan	sentimen_akhir
ingan spal bocor rana...		-8.0	3.203274	-3.518690	negatif
idur sama mees hilgers		3.0	2.256197	2.702479	positif
nggal belanda sampai...		-12.0	3.793367	-5.682653	negatif
il indonesia mees hilg...		25.0	4.437861	16.775145	positif
ees hilgers bukti sing...		-3.0	3.560951	-0.375620	negatif

Gambar 7. Hasil pelabelan dan perhitungan sentimen

Sentimen akhir adalah kolom yang menjadi acuan untuk perhitungan persentase semua sentimen dengan batas sentimen positif $> 0,7$, sentimen negatif $< 0,5$ dan sisanya bersentimen netral.



Gambar 8. Presentase tweet berlabel

Pada gambar diatas menunjukkan bahwa sentimen positif lebih dominan dibandingkan dengan sentimen lainnya, dengan persentase sebesar 71,2% atau 640 tweet, untuk sentimen negatif sebesar 25,5% atau 229 tweet dan untuk sentimen netral sebesar 3,3% atau 30 tweet saja, kemudian hasilnya disimpan dengan nama hasil pembobotan.csv.

Setelah data diberi label, langkah selanjutnya adalah pembagian semua data menjadi data latih dan data uji.

Dengan rasio sebagai berikut:

Tabel 1. Rasio data latih dan data tes

Rasio	Data Latih	Data Uji
90:10	809	90
80:20	719	180
70:30	629	270

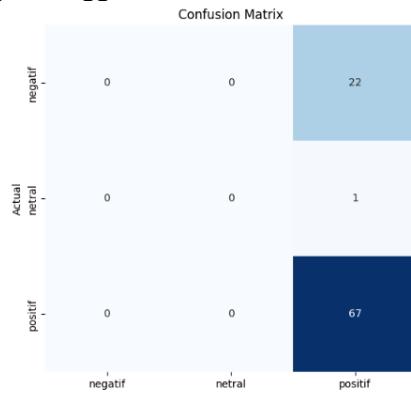
Kemudian data uji dihitung dengan metode klasifikasi *multinomial naïve bayes* dan *support vector machine* (SVM) untuk mengetahui accuracy yang didapat. Untuk metode *multinomial naïve bayes* diperoleh nilai *accuracy* terbaik pada rasio data 90:10 dengan nilai sebesar 0,744 atau 74%. Sedangkan pada metode *support vector machine* (SVM) diperoleh nilai *accuracy* terbaik pada rasio 90:10 dengan nilai sebesar 0,86 atau 86%. Sebelum perhitungan *support vector machine* (SVM) terlebih dahulu menentukan *parameter* yang digunakan dalam proses perhitungan dengan menggunakan teknik *GridSearch*, adapun *parameter*-nya adalah sebagai berikut :

```
'C': [0.1, 1, 10, 100],
'kernel': ['linear', 'poly', 'rbf', 'sigmoid'],
'gamma': ['scale', 'auto', 0.01, 0.1, 1]
```

Gambar 9. Parameter yang akan dicari

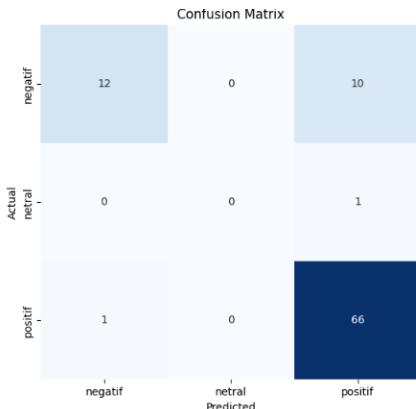
Assess

Evaluasi performa model dilakukan dengan menggunakan rasio terbaik masing-masing metode, dikarenakan rasio yang mendapat nilai terbaik sama yaitu 90:10 maka keduanya menggunakan rasio tersebut.



Gambar 10. Hasil confusion matrix metode *multinomial naïve bayes*

Pada gambar diatas didapatkan data untuk prediksi benar pada 68able68ent positif atau *true* positif sebanyak 67, untuk prediksi benar pada 68able68ent 68able68en atau *true* 68able68en sebanyak 0 dan untuk prediksi benar pada 68able68ent netral atau *true* netral sebanyak 0.



Gambar 11. Hasil *confusion matrix* metode *support vector machine* (SVM)

Dari gambar diatas didapatkan data untuk prediksi benar pada 68able68ent netral atau *true* netral sebanyak 0, untuk prediksi benar pada 68able68ent 68able68en atau *true* 68able68en sebanyak 12 dan untuk prediksi benar pada 68able68ent positif atau *true* positif sebanyak 66.

Kemudian dilakukan perhitungan 10-fold cross validation pada kedua metode yang digunakan. Adapun hasilnya adalah sebagai berikut:

Tabel 2. Perhitungan 10-fold cross validation dengan metode *naïve bayes*

Akurasi	Presisi	Recall	F1-Score
74%	55%	74%	64%
70%	77%	70%	59%
67%	44%	67%	53%
68%	74%	68%	56%
73%	54%	73%	62%
80%	81%	80%	72%
70%	71%	70%	59%
68%	72%	68%	56%
78%	69%	78%	69%
71%	50%	71%	59%

Pada 68able diatas menunjukkan bahwa nilai *accuracy* tertinggi dari 10-fold cross validation dengan metode *multinomial naïve bayes* adalah 80% yang terdapat pada fold ke 6.

Tabel 3. Perhitungan 10-fold cross validation dengan metode *support vector machine*

Akurasi	Presisi	Recall	F1-Score
83%	82%	83%	81%
83%	82%	83%	82%

79%	76%	79%	76%
84%	82%	84%	83%
83%	83%	83%	83%
87%	88%	87%	87%
77%	75%	77%	75%
84%	86%	84%	83%
88%	84%	88%	86%
81%	85%	81%	83%

Pada tabel diatas menunjukkan bahwa nilai *accuracy* tertinggi dari *10-fold cross validation* dengan metode *support vector machine* adalah 88% yang terdapat pada fold ke 9.

Pengujian Akhir

Dilakukan dengan membuat sebuah kolom masukkan atau *tweet*, kolom tersebut dapat diisi dengan *tweet* atau komentar secara bebas. Hasil komentar dan sentimennya akan ditampilkan dibawah dengan memuat jumlah komentar, persentase sentimen dan jumlah komentar berdasarkan sentimennya. Jika tidak ada data yang dimasukkan maka akan muncul tulisan “Tidak ada komentar yang diinputkan”. Kemudian data disimpan dengan nama Laporan_input.csv. Untuk menghentikan program dapat mengetikkan kata “selesai” kemudian program akan berhenti dengan sendirinya.

teks	prediksi_nb	prediksi_svm
kalian semua luar biasa gapapa kalah kita balas dikandang	positif	positif
haduh lawan cina aja kalah dasar pemain tolol	negatif	negatif
eliano face bapak bapak greenflag	positif	positif
tidak bisa dibiarkan klo hasilnya begini terus	positif	negatif

Gambar 12. Contoh komentar baru dan prediksi sentimennya

KESIMPULAN

- Hasil pelabelan dengan metode *lexicon-based* dan TF-IDF menunjukkan bahwa sentimen positif memiliki nilai persentase lebih besar dari kedua sentimen lainnya yaitu sebesar 71,2% atau 640 *tweet* berlabel positif, untuk sentimen berlabel negatif memiliki persentase 25,5% atau 299 *tweet* dan sentimen berlabel netral memiliki persentase 3,3% atau 30 *tweet*.
- Hasil pencarian akurasi terbaik dari kedua metode menunjukkan bahwa keduanya mencapai nilai akurasi tertinggi ketika data

dibagi dengan rasio 90% untuk data *training* dan 10% untuk data *testing*. Metode *support vector machine* memiliki nilai *accuracy* lebih baik dibandingkan metode *naïve bayes* dengan nilai *accuracy* sebesar 0,86 atau 86% dengan menggunakan parameter terbaik yaitu C: 100, gamma: 0,1 dan kernel: rbf dibanding dengan metode *naïve bayes* yang hanya memiliki nilai *accuracy* sebesar 0,74 atau 74%. Dan setelah dilakukan proses *10-fold cross validation* diperoleh hasil *accuracy* yang maksimal untuk metode *support vector machine* pada *fold* ke 9 dengan nilai 88% dan untuk metode *naïve bayes* diperoleh hasil *accuracy* yang maksimal pada *fold* ke 6 dengan nilai 80%.

- Hasil penggabungan semua data baik positif, negatif dan netral diperoleh kata yang sering muncul yaitu kata meskipun dengan frekuensi sebanyak 442 kata, indonesia 341 kata dan tim 329 kata. Pada sentimen netral didominasi oleh *tweet* yang berisi komentar tentang keseharian dua pemain timnas yang sedang diteliti, seperti contoh *tweet* tentang Mees Hilgers yang sedang menari poco-poco atau Eliano Reinjders yang sedang berfoto dengan buah hatinya. Pada sentimen positif didominasi oleh *tweet* yang berisi dukungan kepada tim nasional sepak bola Indonesia, seperti adanya kata “semangat”, “selamat”, “dukung” dan “menang”. Sedangkan pada sentimen negatif didominasi oleh *tweet* yang berisi kata-kata hinaan, ataupun makian dan juga kata yang mengandung makna ketidak sukaan kepada tim nasional sepak bola Indonesia, seperti kata “jelek”, “tidak”, “tolol” dan “anjing”.
- Disarankan dapat mencoba algoritma atau model *state-of-the-art* seperti BERT atau transformer lainnya untuk meningkatkan akurasi dan memahami konteks sentimen dengan lebih baik. Fokus pada analisis aspek tertentu

DAFTAR PUSTAKA

- [1] We Are Social (2024). Digital 2024 Indonesia. Diakses pada tanggal 27 Agustus 2024 dari <https://wearesocial.com/id/blog/2024/01/digital-2024>

- [2] Donny, D. P. (2023). Analisis Sentimen Opini Publik Terhadap Chatgpt di Twitter Menggunakan Metode Naive Bayes. *Jurnal Nasional Ilmu Komputer*, 4(4), 35–44.
- [3] Hasri, C. F., & Alita D.(2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, 3(2), 145–160.
- [4] Tineges, R., & Triayudi, A.(2020). Analisis Sentimen Terhadap Layanan Indihome Berbasis Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 3(10), 650-658.
- [5][8]Fikri, M. I., et al.(2020). Perbandingan Metode Naïve Bayes Dan Support Vector Machine Pada Analisis Sentimen Twitter. *STIKI Informatika Jurnal (SMATIKA)*, 10(2), 71-76.
- [6] Putri, D. D., et al.(2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1), 34-40.
- [7] Liedfray, T., et al.(2022). Peran Media Sosial Dalam Mempererat Interaksi Antar Keluarga Di Desa Esandom Kecamatan Tombatu Timur Kabupaten Minahasa Tenggara. *Jurnal Ilmiah Society*, 2(1), 1-13.
- [9] N. Hotz. “What is SEMMA?. Data Science Process Alliance”, 12 Agustus 2023. Diakses pada 24 Agustus 2024, dari laman <https://www.datascience-pm.com/semma/>
- [10][12] Devi, D. L., et al.(2024). Analisis Sentimen Ulasan Pengguna *Access by KAI* Menggunakan Metode *Word2Vec* dan Algoritma SVM. *Jurnal Elektronika Dan Teknik Elektro Terapan (JITET)*, 12(3), 2724-2733.
- [11] Franko, B. et al (2024). Analisis Sentimen terhadap Naturalisasi Pemain pada Youtube Menggunakan *Decision tree* dan *Naive bayes*. *Jurnal Session (Software Development, Digital Business Intelligence, and Computer Engineering)*. 3(1).8-16.
- [12] Cindo, M., et al.(2019). Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen Maskapai Penerbangan Menggunakan Support Vector Machine dan Maximum Entropy. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)* 3(3), 402-407.
- [14] Hikmah, N. et al.(2022). Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF. *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 4(2), 133–148.
- [15] Alfudola, M. et al (2023). Klasifikasi Pemilihan Tipe Hero Mobile Legends Terhadap Minat Pemain Menggunakan Algorima Naive Bayes Studi Kasus: Komunitas Game Mobile Legends Kota Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1269–1273.
- [16] Riyanto, U.(2018). Analisis Perbandingan Algoritma *Naive Bayes* Dan *Support Vector Machine* Dalam Mengklasifikasikan Jumlah Pembaca Artikel Online. *Jurnal Teknik Informatika (JIKA)* Universitas Muhammadiyah Tangerang, Tangerang, 62-72.