

Analisis Sentimen Terhadap Pemain Diaspora Timnas Indonesia Pada Media Sosial Instagram Menggunakan Metode Support Vector Machine (SVM)

Sendi Ramdani^(b¹), Azahari^(b²), Tommy Bustomi^(b³)

Program Studi Sistem Informasi, STMIK Widya Cipta Dharma
Jl. Prof. M. Yamin No 25, Samarinda 75123
E-mail: 2041080@wicida.ac.id⁽¹⁾, azahari@wicida.ac.id⁽²⁾, tbustomi@wicida.ac.id⁽³⁾

ABSTRAK

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pemain diaspora Tim Nasional Indonesia pada platform media sosial Instagram dengan menggunakan metode Support Vector Machine (SVM). Pemain diaspora, yang merupakan pemain keturunan Indonesia yang bermain di luar negeri. Hasil penelitian, yang dilakukan di Dinas Pemuda dan Olahraga Provinsi Kalimantan Timur, menunjukkan bahwa metode Support Vector Machine Mampu mengklasifikasikan sentiment dengan Tingkat akurasi sebesar 82%, persisi sebesar 81%, dan recall sebesar 80% penelitian ini dapat diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis sentiment yang lebih canggih dan efektif serta menjadi landasan bagi penelitian lebih lanjut dalam memahami dinamika opini publik dalam konteks politik modern yang semakin kompleks.

Kata Kunci: Analisis Sentimen, Support Vector Machine, Pemain Diaspora, Instagram

Sentiment Analysis of Diaspora Players of the Indonesian National Team on Instagram Using the Support Vector Machine (SVM) Method

ABSTRACT

This study aims to analyze public sentiment towards diaspora players of the Indonesian National Team on the social media platform Instagram using the Support Vector Machine (SVM) method. Diaspora players are Indonesian-descendant players who play abroad. The research, conducted at the Youth and Sports Office of East Kalimantan Province, shows that the Support Vector Machine method can classify sentiment with an accuracy rate of 82%, a precision of 81%, and a recall of 80%. This research is expected to contribute to the development of more sophisticated and effective sentiment analysis methods and serve as a foundation for further research in understanding the dynamics of public opinion in the increasingly complex context of modern politics

Keywords: Sentiment Analysis, Support Vector Machine, Diaspora Players, Instagram

1. PENDAHULUAN

Diaspora adalah perantau yaitu orang yang meninggalkan tanah kelahirannya untuk pergi ke daerah atau ke negara lain untuk mencari kehidupan yang lebih baik, ketimbang di daerah atau negaranya sendiri. Istilah Diaspora berbeda dengan imigrasi. Istilah diaspora digunakan untuk merujuk pada penyebaran kelompok agama atau kelompok etnis dari tanah air mereka, baik dipaksa maupun dengan sukarela. Kata ini juga digunakan untuk merujuk pada penyebaran orang-orang sebagai kelompok kolektif dan masyarakat. Diaspora mengharuskan anggota suatu masyarakat pergi bersama dalam periode waktu yang singkat, bukan pergi perlahan-lahan dalam waktu lama meninggalkan kampung halaman. Masyarakat yang melakukan diaspora juga dicirikan dengan usaha mereka untuk mempertahankan

budaya, agama, dan kebiasaan lainnya di tempat baru. Mereka biasanya hidup berkelompok dengan sesamanya, dan kadang tidak mau berinteraksi dengan warga lokal. Dalam beberapa tahun terakhir, PSSI (Persatuan Sepak Bola Seluruh Indonesia) telah aktif merekrut pemain-pemain Diaspora, yaitu pemain yang memiliki hubungan keluarga dengan Indonesia namun lahir atau dibesarkan di luar negeri. Keterlibatan atlet Diaspora ini telah menarik perhatian publik, dan opini tentang kontribusi mereka dalam meningkatkan kualitas Timnas Indonesia menjadi topik yang menarik untuk diteliti. Pemain Diaspora merujuk pada pemain yang memiliki keturunan dari suatu negara tetapi lahir atau besar di negara lain. Pemain pemain ini sering kali memiliki kewarganegaraan ganda atau memenuhi syarat untuk memperoleh kewarganegaraan dari negara asal

leluhurnya, sehingga mereka dapat dipanggil untuk bermain di tim nasional negara tersebut. Pemain Diaspora dapat membawa pengalaman dan keterampilan yang diperoleh dari sistem pelatihan sepak bola di negara tempat mereka dibesarkan. Dengan berkembangnya teknologi dan media sosial, platform seperti Instagram telah menjadi medium utama bagi penggemar untuk mengekspresikan pendapat mereka. Berbagai komentar, baik positif, negatif dan netral, mengenai pemain Diaspora di Tim Nasional Indonesia dapat ditemukan di berbagai akun media sosial Instagram. Analisis sentimen terhadap pemain diaspora di Timnas Indonesia ini penting untuk memahami bagaimana penerimaan dan persepsi masyarakat terhadap kontribusi mereka di tim nasional. Support Vector Machine (SVM) adalah salah satu metode yang efektif dalam klasifikasi teks dan analisis sentimen. SVM bekerja dengan memisahkan data ke dalam kelas-kelas yang berbeda dengan margin maksimal, serta sangat cocok digunakan pada data yang memiliki dimensi tinggi seperti teks. Metode SVM memiliki keunggulan dalam menangani kompleksitas dan variasi bahasa yang sering ditemukan dalam analisis teks media sosial. Penggunaan SVM dalam analisis sentimen terhadap pemain Diaspora Timnas Indonesia di media sosial Instagram dapat memberikan wawasan yang berharga mengenai opini publik dan membantu para pengambil keputusan dalam mengelola tim dan komunikasi publik. Penelitian ini diharapkan dapat memberikan informasi tentang opini publik mengenai pemain PSSI Diaspora di Timnas Indonesia. Informasi ini dapat digunakan untuk meningkatkan pemahaman tentang persepsi publik terhadap pemain diaspora dan meningkatkan dukungan publik untuk pemain diaspora. Dengan demikian, partisipasi pemain diaspora dalam Timnas Indonesia dapat didorong, dan diharapkan mereka dapat memainkan peran yang lebih besar dalam memajukan sepak bola Indonesia di kancah internasional.

2. RUANG LINGKUP

Dalam penelitian ini permasalahan mencakup:

2.1 Cakupan permasalahan

Cakupan permasalahan Berdasarkan latar belakang maka rumusan masalah yang di kemukakan adalah “Bagaimana melakukan analisis sentimen terhadap pemain Diaspora Timnas Indonesia pada media Instagram menggunakan metode Support Vector Machine?”

2.2 Batasan-batasan penelitian

Agar tidak meluasnya permasalahan maka dibuatlah batasan masalah sebagai berikut :

1. Analisis akan berfokus pada identifikasi pola-pola sentimen positif, negatif, atau netral terhadap pemain PSSI (Persatuan Sepak Bola Seluruh Indonesia) Diaspora Timnas Indonesia menggunakan metode Support Vector Machine (SVM) dalam pengolahan teks.
2. Penelitian terbatas pada analisis opini, pandangan, dan tanggapan masyarakat terhadap pemain PSSI

Diaspora Timnas Indonesia, khususnya yang terungkap melalui teks seperti caption, komentar, dan berbagai bentuk interaksi di Instagram.

2.3 Rencana hasil yang didapatkan

Penelitian ini bertujuan untuk menganalisis sentimen dari komentar di Instagram dengan memanfaatkan algoritma *Support Vector Machine* dan selanjutnya untuk mengevaluasi kinerja dari *Support Vector Machine* tersebut.

3. BAHAN DAN METODE

Dengan menggunakan metode ini, diharapkan dapat diperoleh pemahaman yang mendalam tentang sentimen publik terhadap pemain diaspora Timnas Indonesia di Instagram, serta memberikan kontribusi pada pengembangan metode analisis sentimen yang lebih canggih dan efektif.

3.1 Analisis Sentimen

Menurut Liu (Rahmawati, 2023) Analisis sentimen atau opinion mining merupakan salah satu bidang Natural Language Processing (NLP) yang menganalisis pendapat, perilaku, penilaian, dan emosi seseorang terkait dengan produk, topik, layanan, organisasi, individu, atau kegiatan lainnya yang diekspresikan ke dalam bentuk teks

3.2 Diaspora

Menurut Sari, B. R. (2022). dalam era globalisasi yang ditandai dengan meningkatnya mobilitas manusia dan konektivitas, diaspora menjadi tema penting dalam dialog global tentang migrasi. Diaspora merupakan aset sekaligus kekuatan ekonomi baru atas sebuah negara karena berpotensi untuk memberikan kontribusi dalam pembangunan. Di beberapa negara, diaspora dengan latar belakang kehidupan sosial, budaya, ekonomi, dan politik semakin diakui sebagai aktor kunci, tidak hanya tingkat nasional, bilateral, atau internasional, tetapi juga dalam diskursus saling keterkaitan antara migrasi dan pembangunan..

3.3 Support Vector Machine

Lee dkk (2022) Support Vector Machine termasuk dalam supervised learning yang berarti model atau mesin mempelajari terlebih dahulu untuk melakukan klasifikasi dengan membagi data menjadi dataset, yaitu data training dan data testing.

3.4 Metode

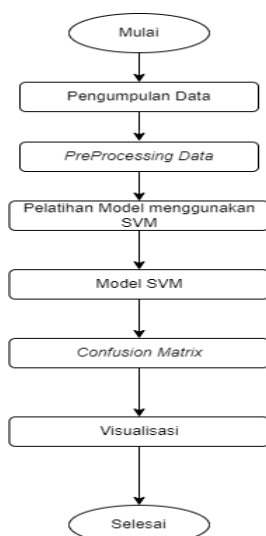
Metode penelitian merupakan sebuah cara yang digunakan untuk mendapatkan dan menganalisa data untuk kegunaan dan tujuan tertentu. Di dalam metode penelitian ini akan ada beberapa hal yang perlu diperhatikan meliputi pemilihan metode yang sesuai, teknik pengumpulan data yang tepat, serta cara analisis yang mampu mengungkap informasi.

3.1 Tempat dan Waktu Penelitian

Tempat penelitian dilakukan di Dinas Pemuda dan Olahraga Provinsi Kalimantan Timur. Waktu penelitian ini dilaksanakan selama satu bulan, mulai dari bulan juni hingga Juli 2024.

3.2 Tahapan Penelitian

Diagram Tahapan Penelitian yang dibuat adalah rencana yang akan dilakukan oleh peneliti dalam melakukan analisis sentimen. Pada diagram tahapan penelitian terdapat rincian dalam proses pencarian awal data sampai dengan akhir, pada masing-masing proses yang ditulis pada diagram tahapan akan dijelaskan satu-persatu guna memperjelas diagram yang diolah tersebut. Bisa dilihat pada gambar. Penjelasan rinci ini mencakup semua langkah penting, mulai dari pengumpulan data mentah hingga tahap analisis dan interpretasi hasil, memastikan bahwa setiap tahapan penelitian diuraikan dengan jelas untuk memudahkan pemahaman keseluruhan proses penelitian. Langkah-langkah ini meliputi pengumpulan data, pra-pemrosesan data, metode analisis, dan lain – lain secara terstruktur dalam diagram tersebut.



Gambar 3. 1 Tahapan Penelitian

3.2.1 Pengumpulan Data

Penelitian ini menggunakan data kuantitatif untuk menganalisis data dari teks berbahasa Indonesia berupa komentar dari postingan PSSI yang di peroleh dari Instagram terkait pemain Diaspora Timnas Indonesia metode – metode yang digunakan dalam pengumpulan data selama penelitian adalah:

1. Observasi

Observasi dilakukan sebagai studi pendahuluan, peneliti mengamati secara langsung postingan hasil proses Diaspora dari Timnas Indonesia dengan melihat pendapat atau komentar yang ada pada media sosial Instagram. Observasi dilakukan peneliti pada bulan Juli 2024. Beberapa hal yang diperoleh peneliti dari observasi yang telah dilakukan adalah memperoleh informasi bahwa terdapat tanggapan atau komentar yang beragam dari postingan PSSI terkait pemain Diaspora Timnas Indonesia pada Instagram.

2. Studi Pustaka

Metode pengumpulan studi pustaka ini dilakukan untuk mencari, mempelajari, dan menggunakan berbagai literatur seperti, buku, jurnal, paper, ebook, atau literatur

lain yang berhubungan dengan tema penelitian ini. Informasi yang terkumpul akan dijadikan data pendukung dan pembanding untuk penelitian.

3. Crawling Data

Pada tahapan ini dilakukan pengumpulan dan pencarian dataset Instagram Bahasa Indonesia. Data yang dikumpulkan adalah data komentar masyarakat dari Instagram yang merupakan pesan ataupun opini dari masyarakat terkait pemain Diaspora Timnas Indonesia.

3.2.2 Pre-Processing Data

Pre-processing merupakan tahapan awal yang akan dilalui dalam memproses teks. Pada penelitian ini akan dilakukan tahapan Pre-processing dengan tahapan case folding, data cleaning, tokenisasi, stopword removal/filtering, dan stemming. Alur proses pada tahapan ini akan ditunjukkan pada Gambar 3.2



Gambar 3. 2 Tahapan Text Processing

1. Case Folding

Case folding adalah proses mengubah semua karakter dalam sebuah dokumen menjadi huruf kecil atau huruf besar yang sama, untuk mempercepat perbandingan selama proses pemrosesan data.

2. Cleansing

Cleansing merupakan tahapan yang bertujuan menghilangkan karakter atau symbol link url (<http://situs.com>), username atau mention (@username), hastage(#), retweet, dan emoticon. Selain itu, pada tahap ini juga membersihkan duplikasi data.

3. Tokenizing

Pada proses tokenizing dilakukan untuk memisahkan teks menjadi bagian-bagian yang lebih terstruktur sehingga mempermudah proses analisis selanjutnya, seperti penghapusan karakter khusus, menghitung frekuensi kata, atau membangun representasi vektor dari teks.

4. Stopword Removal/Filtering

Tahap stopword removal membuang kata-kata tidak bermakna yang tidak mempengaruhi klasifikasi sentimen, seperti kata penghubung dan kata depan, berdasarkan kamus stopwords. Hal ini akan membantu masyarakat memahami dinamika dan keragaman pendapat yang ada di tengah-tengah mereka.

5. Stemming

Tahap ini bertujuan untuk mengembalikan suatu kata ke dalam bentuk dasarnya. Dengan menghilangkan kata imbuhan awalan, akhiran, sisipan atau pun kombinasi imbuhan. Pada penelitian ini akan digunakan library sastrawi untuk melakukan proses stemming. Sastrawi merupakan library pada Bahasa pemrograman python

yang dibangun dengan algoritma NA. Algoritma tersebut berdasarkan aturan Bahasa Indonesia.

6. Pelabelan

Pelabelan adalah memberikan kategori atau tanda pada data agar atribut tertentu dapat diidentifikasi. Dengan pelabelan ini memudahkan menentukan kalimat sentimen yang terdiri dari negatif, netral, positif.

3.2.3 Pelatihan Model Menggunakan SVM

1. Pembobotan TF-IDF

Setelah tahap preprocessing berikutnya adalah tahap pembobotan kata. Pembobotan bertujuan untuk memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. TF-IDF (Term Frequency-Inverse Document Frequency) memberi nilai penting pada kata dalam satu dokumen dengan mempertimbangkan seberapa sering kata tersebut muncul di dokumen itu dan seberapa jarang kata itu muncul di seluruh kumpulan dokumen. Rumus pada TF-IDF seperti pada dibawah berikut ini :

$$TF_{(t,d)} = \frac{\text{Number of times } t \text{ appear in document } d}{\text{Total number of terms in document } d}$$

$$IDF_{(t,d)} = \log \left(\frac{\text{Total number of documents } D}{\text{Number of document the term in it}} \right)$$

$$TFIDF(t, d, D) = TF(t,d) * IDF(t,d)$$

dimana : t_i = Dokumen ke - i ,

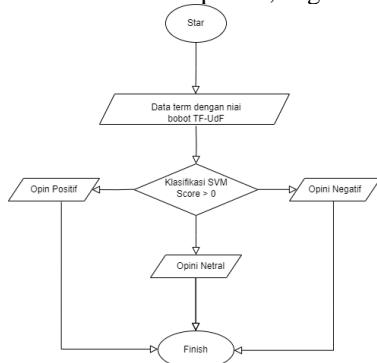
df = Document frequency,

n = Banyaknya data

idf = Inverse document frequency.

3.2.4 Pemodelan Dengan Metode Support Vector Machine

Metode Support Vector Machine (SVM) digunakan dalam penelitian ini untuk analisis sentimen. Hasilnya adalah klasifikasi teks menjadi kelas positif, negatif dan netral berdasarkan bobot fitur masing-masing dokumen. Proses ini melibatkan pembobotan fitur dengan SVM, menghasilkan pola yang digunakan dalam pengujian. Evaluasi dilakukan dengan memberi skor pada dokumen, dan analisis keputusan didasarkan pada tingkat akurasi serta jumlah dokumen di kelas positif, negatif dan netral..



Gambar 3. 3 Tahap Proses Metode SVM Sumber :

3.2.5 Confusion Matrix

Setelah hasil pengujian menggunakan Support Vector Machine selesai, akan dilakukan confusion matrix untuk menentukan 3 macam hasil performa machine learning yakni akurasi, presisi, dan recall. Hasil klasifikasi yaitu berupa jumlah sentimen positif, negatif, netral komentar masyarakat terhadap postingan terkait pemain Diaspora Timnas Indonesia di instagram, sehingga akan diperoleh kesimpulan mengenai tanggapan masyarakat indonesia terhadap hasil proses Naturalisasi di indonesia.

3.2.6 Visualisasi

Tahap ini merupakan tahap memvisualisasikan kata-kata yang sering muncul pada data komentar ke dalam bentuk wordcloud. Hal ini untuk mengetahui kata-kata apa saja yang paling banyak terdapat pada data komentar dan berapa frekuensi kata tersebut muncul dalam sebuah komentar. Visualisasi juga digunakan untuk menampilkan jumlah emosi pada data komentar dalam bentuk barplot yang terdiri atas beberapa parameter emosi.

4. HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Pada penelitian ini dataset yang digunakan ialah data yang berasal dari media sosial Instagram, pada gambar 4.1 adalah objek penelitiannya sendiri ialah postingan dari akun Persatuan SepakBola Seluruh Indonesia (PSSI). Adapun tahapan proses pengolahannya hingga menjadi dataset yang akan digunakan untuk pengujian penelitian ini, berikut tahapan proses pengambilan dataset:



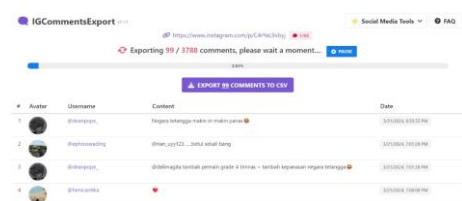
Gambar 4. 1 Instagram PSSI

4.2 Pembahasan

4.2.1 Pengumpulan Data

1. Crawling Data

Proses crawling dilakukan menggunakan Export Comment dari Google Chrome yang secara otomatis melakukan pengumpulan data. Data yang dikumpulkan kemudian disimpan dalam format yang terstruktur untuk memudahkan analisis lebih lanjut. Selain itu, alat ini memastikan bahwa setiap komentar dan metadada yang relevan diambil dengan akurasi tinggi, sehingga memberikan gambaran menyeluruh tentang interaksi pengguna



Gambar 4. 2 Proses Pengumpulan Data

Tahap ini melakukan pengumpulan data komentar pada salah satu postingan pada akun sosial PSSI. Pada gambar 4.2 adalah contoh mengambil data nama pada komentar yang akan di ambil dan menunjukkan jumlah nama yang di dapat pada komentar pada suatu postingan.

User Id	Username	Comment Id	Comment Text	Profile U
0	10992373544	dearpopo_18074687959462604	Negara tetangga makin iri makin panas	https://www.instagram.com/dearpopo
1	30133367497	epohswading_1805583988562790	@rian_uyy123betul sekali bang	https://www.instagram.com/epohswadi
2	10992373544	dearpopo_17916032099806484	@delimaglia tambah pemain grade A timnas = tam...	https://www.instagram.com/dearpopo
3	1018795501	henicandika_17989534766441895		https://www.instagram.com/henicandi
4	51182184997	hiisu.id_18044862001632633	@meeshligers @kevindika2 @riedewald @million...	https://www.instagram.com/hiisu

Gambar 4. 3 Hasil Crawling Data

Pada gambar 4.3 adalah hasil *crawling* data bentuk soft file csv jumlah keseluruhan data yang direkam berjumlah 3476 data yang berjumlah 7 atribut diantaranya User Id, Username, Comment Id, Comment Text, Profil URL, Avatar URL, dan Date dari 7 atribut yang ditampilkan setelah itu proses pembersihan data dan menyeleksi atribut teks.

4.2.2 Pre-Processing Data

Sebelum melakukan tahapan *preprocessing* terlebih dahulu import library yang dibutuhkan untuk memanggil dataset yang akan di proses di *google collab*. Setelah library berhasil di import, selanjutnya memanggil dataset yang telah didapat dari proses *crawling* data. Pada gambar 4.3 merupakan hasil pemanggilan dataset yang berformat CSV dengan nama *comments2.csv*.

1. Cleansing

Comment Text	cleansing
0 Negara tetangga makin iri makin panas	Negara tetangga makin iri makin panas
1 @rian_uyy123betul sekali bang	betul sekali bang
2 @delimaglia tambah pemain grade A timnas = tam...	tambah pemain grade A timnas tambah kepanasan ...
3	
4 @meeshligers @kevindika2 @riedewald @million...	m r hope u next guys
...	...
3470 Leg 1 bisa main ga klo begini	Leg 1 bisa main ga klo begini
3471 @marcelindo ferdinan udah tgl 12 kemarin oi	ferdinan udah tgl 12 kemarin oi
3472 @wilman_wilman bagus juga kak Jersey yang baru timnas	bagus juga kak Jersey yang baru timnas
3473 Dapat KTP nya pake ngantri di discapil gak ya.	Dapat KTP nya pake ngantri di discapil gak ya
3475 gas terus pak, sampe 11x2 kan target nya	gas terus pak sampe 11x2 kan target nya

Gambar 4. 4 Pembersihann Data

Pada gambar 4.4 menunjukkan hasil pembersihan data pada atribut teks untuk menghilangkan atau menghapus bagian yang tidak diperlukan seperti URL, hashtag, metion, emoticon, emoticon dan karakter spesial lainnya. Proses pembersihan ini sangat penting untuk memastikan data yang digunakan dalam analisis lebih akurat dan relevan. Setelah data dibersihkan, langka selanjutnya adalah melakukan *pre-processing* yang lebih mendetail, seperti case folding, tokenisasi, stemming, dan penghapusan *stop words*.

2. Case Folding

Comment Text	cleansing	case_folding
0 Negara tetangga makin iri makin panas	Negara tetangga makin iri makin panas	negara tetangga makin iri makin panas
1 @rian_uyy123betul sekali bang	betul sekali bang	betul sekali bang
2 @delimaglia tambah pemain grade A timnas = tam...	tambah pemain grade A timnas tambah kepanasan ...	tambah pemain grade a timnas tambah kepanasan ...
3		
4 @meeshligers @kevindika2 @riedewald @million...	m r hope u next guys	m r hope u next guys
...
3470 Leg 1 bisa main ga klo begini	Leg 1 bisa main ga klo begini	leg 1 bisa main ga klo begini
3471 @marcelindo ferdinan udah tgl 12 kemarin oi	ferdinan udah tgl 12 kemarin oi	ferdinan udah tgl 12 kemarin oi
3472 @wilman_wilman bagus juga kak Jersey yang baru timnas	bagus juga kak Jersey yang baru timnas	bagus juga kak jersey yang baru timnas

Gambar 4. 5 Hasil Case Folding

Pada gambar 4.5 menunjukkan hasil yang setelah dilakukan *cleansing* dilanjut dengan *Case Folding* pada atribut teks untuk merubah semua huruf besar menjadi huruf kecil. Contoh kalimat yang ada pada gambar “Dapat KTP nya pake ngantri di discapil gak ya, Negara tetangga makin iri makin panas” menjadi “dapat ktp pake ngantri di discapil gak ya, negara tetangga makin iri makin panas.”

3. Tokenizing

Comment Text	cleansing	case_folding	tokenize
0 Negara tetangga makin iri makin panas	Negara tetangga makin iri makin panas	negara tetangga makin iri makin panas	[negara, tetangga, makin, iri, makin, panas]
1 @rian_uyy123betul sekali bang	betul sekali bang	betul sekali bang	[betul, sekali, bang]
2 @delimaglia tambah pemain grade A timnas = tam...	tambah pemain grade A timnas tambah kepanasan ...	tambah pemain grade a timnas tambah kepanasan ...	[tambah, pemain, grade, a, timnas, tambah, kep...
3			
4 @meeshligers @kevindika2 @riedewald @million...	m r hope u next guys	m r hope u next guys	[m, r, hope, u, next, guys]
...
3470 Leg 1 bisa main ga klo begini	Leg 1 bisa main ga klo begini	leg 1 bisa main ga klo begini	[leg, 1, bisa, main, ga, klo, begini]
3471 @marcelindo ferdinan udah tgl 12 kemarin oi	ferdinan udah tgl 12 kemarin oi	ferdinan udah tgl 12 kemarin oi	[ferdinan, udah, tgl, 12, kemarin, oi]

Gambar 4. 6 Hasil Tokenizing

Pada gambar 4.6 merupakan tahapan tokenisasi setelah dilakukan *case folding* kalimat “negara tetangga makin iri makin panas” yang ditokenisasikan menjadi “[‘negara’, ‘tetangga’, ‘makin’, ‘iri’, ‘makin’, ‘panas’]” dari kalimat tersebut telah terbagi menjadi kata token yang terpisah.

4. Stopword removal

Comment Text	cleansing	case_folding	tokenize	filtering/stopword removal
0 Negara tetangga makin iri makin panas	Negara tetangga makin iri makin panas	negara tetangga makin iri makin panas	[negara, tetangga, makin, iri, makin, panas]	[negara, tetangga, iri, panas]
1 @rian_uyy123betul sekali bang	betul sekali bang	betul sekali bang	[betul, sekali, bang]	[bang]
2 @delimaglia tambah pemain grade A timnas = tam...	tambah pemain grade A timnas tambah kepanasan ...	tambah pemain grade a timnas tambah kepanasan ...	[tambah, pemain, grade, a, timnas, tambah, kep...	[pemain, grade, a, timnas, kepanasan, negara, ...]
3				
4 @meeshligers @kevindika2 @riedewald @million...	m r hope u next guys	m r hope u next guys	[m, r, hope, u, next, guys]	[m, r, hope, u, next, guys]

Gambar 4. 7 Hasil Stopword removal

Pada gambar 4.7 merupakan tahapan *stopword* menghilangkan kata yang tidak berguna seperti dan, pada, sebagai dll. Seperti hasil dari kalimat berikut “negara tetangga makin iri makin panas” yang jika ditokenisasikan menjadi “[‘negara’, ‘tetangga’, ‘makin’, ‘iri’, ‘makin’, ‘panas’]” kata yang dihilangkan yaitu ‘makin’ karena dianggap tidak terlalu memiliki arti penting dalam proses *text mining* dan kalimat tersebut telah terbagi menjadi kata token yang terpisah..

5. Stemming

Comment Text	cleansing	case_folding	tokenize	filtering/stopword removal	stemming_data
0 @septianboengoe nahhh lu tuh	nahhh lu tuh	nahhh lu tuh	[nahhh, lu, tuh]	[nahhh, tuh]	nahhh tuh
1 @septianboengoe yok cari fakta krp gbrn bisa lolos	yok cari fakta krp gbrn bisa lolos	yok cari fakta krp gbrn bisa lolos	[yok, cari, fakta, krp, gbrn, bisa, lolos]	[yok, cari, fakta, krp, gbrn, bisa, lolos]	yok cari fakta krp gbrn lolos
2 @otra_melati2 tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	tp mungkin mba nya bim paham politik gpp sih k...	[tp, mungkin, mba, nya, bim, paham, politik, gpp, sih, k...]	[tp, mba, nya, bim, paham, politik, gpp, sih, ...]	tp mba nya bim paham politik gpp sih lar aja ...
3 @septianboengoe yudh deh aku jelsin kma aku juga tau anda ml...	yudh deh aku jelsin kma aku juga tau anda ml...	yudh deh aku jelsin kma aku juga tau anda ml...	[yudh, deh, aku, jelsin, kma, aku, juga, tau, anda, ml...]	[yudh, deh, jelsin, kma, tau, mis, cari, fa...]	yudh deh jelsin kma tau mis cari fakta znyk...
4 @otra_melati2 mana ada aku gak nerima curang	mana ada aku gak nerima curang	mana ada aku gak nerima curang	[mana, ada, aku, gak, nerima, curang]	[gak, nerima, curang]	gak nerima curang
5 @septianboengoe lahhh coba lu scroll keatas ini	lahhh coba lu scroll keatas ini	lahhh coba lu scroll keatas ini	[lahhh, coba, lu, scroll, keatas, ini]	[lahhh, coba, lu, scroll, keatas, ini]	lahhh coba lu scroll atas ini

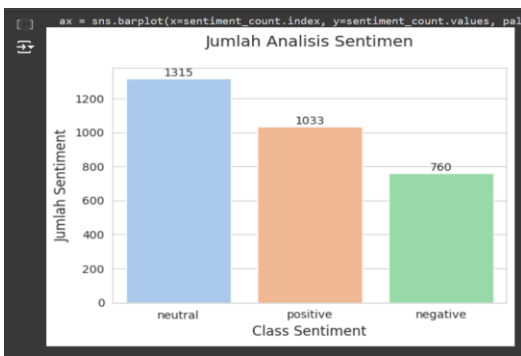
Gambar 4. 8 Hasil Tahapan Stemming

Pada gambar 4.8 di atas, ditunjukkan proses hasil *stemming* di mana kata- kata dengan imbuhan diubah menjadi kata dasar pada dataset. Sebagai contoh, kata 'keatas' setelah proses *stemming* menjadi 'atas'. Proses ini

bertujuan untuk menyederhanakan kata-kata dan mengurangi variasi dalam dataset agar analisis sentimen atau pemrosesan teks menjadi lebih efektif dan akurat.

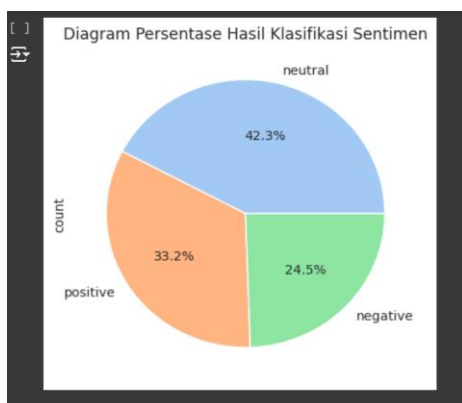
6. Pelabelan

Setelah pelaksanaan tahapan pre-processing dilaksanakan dan menghasilkan pre-processing dengan variabel `stemming_data` yang merupakan data tidak terstruktur dan tidak memiliki kelas/label, agar dapat digunakan untuk proses mining dengan algoritma SVM, maka perlu dilakukan proses pelabelan data untuk menentukan kata-kata yang merepresentasikan kata positif, netral, negatif dan menghitung jumlah kata positif, netral, negatif pada setiap satu kalimat dengan bantuan dari kamus inset lexicon positif, netral, negatif yang bersumber dari penelitian dan repository. Pelabelan diklasifikasikan kedalam 3 kelas yaitu positif, netral dan negatif..



Gambar 4. 9 Hasil Jumlah Sentimen

Pada gambar 4.9 di atas menunjukkan bahwa jumlah sentimen netral lebih banyak, dengan total 1315 dokumen, sentimen positif berjumlah 1033 dokumen sementara sentimen negatif berjumlah 760 dokumen. Hal ini menunjukkan bahwa sentimen netral lebih dominan dibandingkan dengan sentimen lainnya. Jika dipersentasikan, hasilnya adalah 42.3% netral, 33.2% positif, 24.5% negatif, seperti yang ditampilkan pada gambar 4.10. Dominasi sentimen netral ini mengindikasikan bahwa sebagian besar teks yang dianalisis cenderung memiliki pandangan atau opini netral..



Gambar 4. 10 Diagram Hasil Klasifikasi Sentimen Dalam Bentuk Persen

4.2.3 Pelatihan Model Menggunakan SVM

1. Pembobotan dengan TF-IDF

Setelah pelaksanaan proses pelabelan dilanjutkan dengan tahapan pembobotan Term Frequency-Inverse Document frequency (TF-IDF) yang menggunakan perhitungan modul `scikit-learn` yaitu `TfidfVectorizer`. Dimana hasil dari tahap TF-IDF yaitu dari 3476 dataset setelah dilakukan pembobotan kata atau kemunculan kata berjumlah 3978

```
{'negara': 2494,
'tetangga': 3603,
'iri': 1496,
'panas': 2755,
'bang': 369,
'main': 2153,
'grade': 1276,
'timnas': 3640,
'nan': 2459,
'hope': 1407,
'next': 2512,
'guys': 1299,
'idolo': 1433,
'aw57': 318,
'peci': 2802,
'adat': 105,
'layak': 2009,
'belangkom': 447,
'citayam': 754,
'lu': 2111,
'nyalain': 2633,
'rusuh': 3118,
'udah': 3741,
'hujat': 1421,
```

Gambar 4. 11 kamus (dictionary) yang menghubungkan kata-kata (terms) dengan indeks uniknya

pada gambar 4.11 menjelaskan kamus ini digunakan dalam `CountVectorizer` atau `TfidfVectorizer` dari library `scikit-learn`, yang mengubah koleksi dokumen teks menjadi matriks istilah-dokumen. Setiap kata atau term dalam teks diberikan indeks unik untuk mengidentifikasi kata tersebut didalam vektor atau matriks. Misalnya, kata “negara” memiliki indeks 2494, “tetangga” memiliki indeks 3603, dan seterusnya

```
#contoh memanggil salah satu kata pada kolom 1
stemming_data_tf = stemming_data[0]
print(stemming_data_tf)
```

```
['negara', 'tetangga', 'iri', 'panas']
```

Gambar 4. 12 Hasil TF

Pada gambar 4.12 merupakan hasil TF dari kalimat “negara tetangga iri panas”, terhadap ada 4 kata dengan masing-masing nomor indeksnya. Contoh jika dalam hasil yang diberikan (0, 1496) 1, ini berarti dalam dokumen pertama (indeks 0), kata “cari” muncul sekali dan begitupun seterusnya.

```
#perhitungan tf-idf dalam satu dokumen
from sklearn.feature_extraction.text import TfidfTransformer
dokumen = cv.transform(stemming_data)
tfidf_transformer = TfidfTransformer().fit(dokumen)
print(tfidf_transformer.idf_)

tfidf = tfidf_transformer.transform(term_frequency)
print(tfidf)

[8.34890923 8.34890923 8.34890923 ... 8.34890923 8.34890923 7.94344412]
(0, 3603) 0.48331054675643176
(0, 2755) 0.5119402586148847
(0, 2494) 0.3959832849415104
(0, 1496) 0.5895127861635017
```

Gambar 4. 13 Hasil TF-IDF

Pada gambar 4.13 menunjukkan hasil nilai yang diperoleh dari perhitungan tf-idf yang menggunakan tools google collab. [8.34890923 8.34890923 8...34890923 . 8.34890923 8.34890923 7.94344412] merupakan array nilai IDF untuk setiap term dalam koleksi dokument yang dihitung berdasarkan jumlah kata yang sering muncul yang dibagi dengan jumlah seluruh data. Sedangkan (0, 3603) 0.48331054675643176 dan seterusnya menunjukkan pasangan indeks term dan nilai TF-IDF-nya yang hasilnya dikalikan. Nilai TF-IDF memberikan bobot pada setiap term tersebut muncul di dokumen lain.

4.2.4 Pemodelan Dengan Metode Support Vector Machine

Setelah melewati pembobotan TF-IDF, selanjutnya tahapan pembagian data dilakukan untuk mengevaluasi performa model dari SVM, yang dibagi menjadi data training dan data testing dimana didalam penelitian ini menggunakan pembagian data.

	precision	recall	f1-score	support
negative	0.76	0.66	0.71	228
neutral	0.81	0.90	0.86	395
positive	0.87	0.83	0.85	310
accuracy			0.82	933
macro avg	0.81	0.80	0.80	933
weighted avg	0.82	0.82	0.82	933

Gambar 4. 14 Hasil dari pemodelan algoritma svm

Pada gambar 4.14 diatas mengindikasikan bahwa nilai akurasi dengan pembagian 90% data training dan 10% data testing adalah 81%. Pada kelas “negative” memiliki precision 0.76, recall 0.66, dan F1-score 0.71. kelas “neutral” precision 0.81, recall 0.90, dan F1-score 0.86. sedangkan kelas positif precision 0.87, recall 0.83, dan F1-score 0.85.

4.2.5 Confusion Matrix

Setelah didapatkannya hasil klasifikasi sentimen menggunakan svm perlu dilakukan pengujian menggunakan confusion matrix untuk mengetahui informasi yang diprediksi dan hasil yang sebenarnya. Perhitungan dari confusion matrix adalah accuracy,

recall, precision dan F1-score. Berikut merupakan hasil pengujian menggunakan Confusion Matrix.



Gambar 4. 15 Pengujian Confusion Matrix

dari gambar 4.15 diatas dapat dijelaskan bahwa memiliki 150 data sebagai true negatif, 357 data sebagai true netral dan 258 data sebagai true positif. Pada perhitungan manual akurasi, presisi dan matriks diatas sebagai berikut:

$$\text{Accuracy} = \frac{\text{Jumlah prediksi benar} (TN + TP)}{\text{Jumlah Prediksi} (TN + FN + FP + TP)} \times 100\%$$

$$\text{Accuracy} = \frac{150 + 357 + 258}{150 + 55 + 23 + 23 + 357 + 15 + 24 + 28 + 258} \times 100\%$$

$$\text{Accuracy} = \frac{765}{933} \times 100\%$$

$$\text{Accuracy} = 0.819 \times 100\% = 81.9\%$$

4.2.6 Visualisasi



Gambar 4. 16 Wordcloud ulasan negative

Pada gambar 4.16 merupakan hasil visualisasi dari ulasan negative. Terlihat jelas bahwa frekuensi kemunculan sebuah kata meningkat seiring dengan ukurannya di dalam wordcloud. Beberapa istilah yang sering muncul dalam ulasan negatif terhadap aplikasi yaitu “Jelek”, “kalah”, “salah”, dan lainnya.



Gambar 4. 17 Wordcloud ulasan positive

Gambar 4.17 menampilkan hasil visualisasi ulasan positif. Dimana kata "allamdulillah", "selamat", "welcome", sering digunakan dalam ulasan positif.



Gambar 4. 18 Wordcloud ulasan netral

Sementara itu, gambar 4.18 menampilkan hasil visualisasi ulasan netral. Kata "nyala", "indonesia", "timnas", "garuda" sering digunakan dalam ulasan netral..

5. KESIMPULAN

1. Dalam penelitian ini, menentukan sentimen masyarakat dengan cara mengambil data dari ulasan atau komentar instagram masyarakat terkait pemain diaspora timnas indonesia kemudian data itu di olah atau dianalisis menggunakan Algoritma Support Vector Machine (SVM) dan metodologi Knowledge Discovery in Database.
2. Sebanyak 3475 data komentar diperoleh melalui web crawling dari situs exportcomments.com. Setelah melalui proses preprocessing, 3108 data yang digunakan dan dibagi menjadi kelas : 1315 data kelas netral, 1033 data kelas positif, 760 data kelas negatif.
3. Penggunaan algoritma SVM dengan kernel linear pada tahapan data mining mampu mengklasifikasikan komentar dengan baik. Pengujian dilakukan dengan empat skenario pembagian data 90:10 menggunakan confusion matrix untuk menilai skor. Hal terbaik diperoleh pada skenario 90:10 dengan akurasi tertinggi

sebesar 82% serta nilai presisi, recall, dan F1-score yang signifikan

6. SARAN

Saran untuk pengembangan lebih lanjut dari penelitian ini sebagai berikut :

1. Mengaplikasikan metode analisis sentimen ini pada platform media sosial lainnya seperti Twitter, Facebook, dan YouTube untuk mendapatkan gambaran sentimen yang lebih luas dan holistik.
2. Menggunakan ekstraksi fitur lainnya pada proses klasifikasi sentimen.
3. Penelitian selanjutnya dapat melakukan jenis analisis sentimen lain seperti analisis sentimen berbasis aspek pada topik yang sama

7. DAFTAR PUSTAKA

- SAFITRI, Rahmi; ALI, Irfan; RAHANINGSIH, Nining. ANALISIS SENTIMEN TERHADAP TREN FASHION DI MEDIA SOSIAL DENGAN METODE SUPPORT VECTOR MACHINE (SVM). JATI
- AMALIA, A. (2018). *KLASIFIKASI FANS DAN HATERS BERDASARKAN KOMENTAR INSTAGRAM MENGGUNAKAN NAÏVE BAYES CLASSIFIER* (Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau). (Komentar)
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis sentimen cyberbullying pada komentar instagram dengan metode klasifikasi support vector machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704-4713