

ANALISIS SENTIMEN *TWITTER* TERHADAP OPINI MASYARAKAT PADA SEA GAMES KAMBOJA 2023 MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

ANALYSIS OF TWITTER SENTIMENT ON PUBLIC OPINION AT SEA GAMES CAMBODIA 2023 USING SUPPORT VECTOR MACHINE ALGORITHM

Yuma Akbar¹, Aldino Nur Ihsan²

^{1,2}Sekolah Tinggi Ilmu Komputer Cipta Karya, Jakarta
yuma.pjj@gmail.com, aldinonurihsan.sdjcell@gmail.com

ABSTRACT

SEA Games is the biggest sporting event in Southeast Asia which is held every two years. In 2023, the SEA Games will be held in Cambodia, and public opinion about this event will have a huge impact on the perception and success of the event. This study aims to conduct a sentiment analysis of public opinion about the 2023 Cambodia SEA Games using the SVM (Support Vector Machine) algorithm. The data used are tweets with the words "Sea Games Cambodia" and "Sea Games Kamboja National Team Day". The data is 1220 tweets taken with the help of jupyter notebook. The data used has 921 positive sentiments and 299 negative sentiments. with the Support Vector Machine Algorithm produces the best performance with an accuracy value of 92.08%, precision of 91.55%, recall of 86.50%, f1-score of 88.65%. The results of this sentiment analysis will provide valuable insight into public opinion regarding the Cambodian SEA Games 2023.

Keywords: *Analysis, Sentiment, Twitter, Support Vector Machine, Sea Games.*

ABSTRAK

SEA Games merupakan ajang olahraga terbesar di kawasan Asia Tenggara yang diadakan setiap dua tahun sekali. Pada tahun 2023, SEA Games diadakan di Kamboja, dan opini masyarakat terhadap acara ini memiliki dampak besar terhadap persepsi dan keberhasilan penyelenggaraan. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap opini masyarakat tentang SEA Games Kamboja 2023 menggunakan algoritma SVM (Support Vector Machine). Data yang digunakan yaitu tweet dengan kata "Sea Games Kamboja" dan "Timnas Day Sea Games Kamboja". Data berjumlah 1220 tweet yang diambil dengan bantuan jupyter notebook. Data yang digunakan Memiliki 921 sentimen positif dan 299 sentimen negatif. dengan Algoritma Support Vector machine menghasilkan performa terbaik dengan nilai accuracy 92,08%, precision 91,55%, recall 86,50%, f1-score 88,65%. Hasil analisis sentimen ini akan memberikan wawasan yang berharga tentang opini masyarakat terkait SEA Games Kamboja 2023.

Kata Kunci: Analisis, Sentimen, Twitter, Support Vector Machine, Sea Games

PENDAHULUAN

Pada era digital saat ini, media sosial menjadi platform yang sangat populer untuk berbagi informasi dan pendapat. Salah satu media sosial yang paling banyak digunakan adalah Twitter, di mana pengguna dapat dengan mudah mengungkapkan opini mereka tentang berbagai topik, termasuk acara olahraga seperti Sea Games. Sea Games merupakan ajang olahraga regional yang diikuti oleh negara-negara Asia Tenggara, termasuk Kamboja, dan menjadi sorotan publik karena prestisinya.

Dalam menghadapi SEA Games Kamboja 2023, penting bagi pemerintah dan penyelenggara untuk memahami sentimen masyarakat terkait acara tersebut. Analisis sentimen Twitter dapat menjadi alat yang berguna untuk melacak dan menganalisis opini publik secara real-time. Dengan menggunakan algoritma Support Vector Machine (SVM), penelitian ini bertujuan untuk menganalisis sentimen Twitter terhadap opini masyarakat mengenai SEA Games Kamboja 2023.

Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang kuat dalam klasifikasi dan

analisis sentimen. SVM bekerja dengan membangun model yang memisahkan dua kelas data berdasarkan fitur-fitur yang diambil dari teks atau data lainnya. Dalam konteks ini, SVM dapat digunakan untuk mengklasifikasikan tweet yang berisi opini positif dan negatif, terkait SEA Games Kamboja 2023.

Dengan melakukan analisis sentimen terhadap opini masyarakat di Twitter, penelitian ini akan memberikan wawasan yang berharga bagi pemerintah dan penyelenggara SEA Games Kamboja 2023. Hasil analisis sentimen dapat membantu mereka dalam merumuskan strategi komunikasi, mengidentifikasi isu-isu yang penting, dan merespon masukan masyarakat dengan lebih efektif. Selain itu, penelitian ini juga dapat memberikan kontribusi terhadap bidang ilmu data dan pengenalan pola, khususnya dalam menganalisis sentimen di media sosial.

Dengan memanfaatkan algoritma Support Vector Machine (SVM) dan data tweet yang tersedia, penelitian ini bertujuan untuk memberikan pemahaman yang lebih mendalam tentang opini masyarakat terkait SEA Games Kamboja 2023, sehingga dapat meningkatkan pengelolaan acara dan memastikan partisipasi dan dukungan publik yang lebih baik.

Di zaman modern ini, media sosial telah menjadi alat komunikasi, interaksi, dan ekspresi antara individu dan kelompok. Media sosial telah menjadi forum opini tentang banyak subjek dan masalah. Menurut studi yang dilakukan oleh We Are Social Agency bekerja sama dengan Hootsuite, terdapat 202,6 juta pengguna internet di Indonesia, atau 73,7% dari total penduduk Indonesia. Pengguna aktif jejaring sosial Indonesia mencapai 170 juta orang atau 61,8% dari total penduduk Indonesia. Media sosial yang populer saat ini adalah Instagram, Facebook, Whatsapp, Youtube, Tiktok dan Twitter. Menurut Riyanto (2021), Twitter merupakan platform media sosial aktif kelima di Indonesia dengan pangsa pengguna 63,6 persen. Twitter dipertimbangkandaripada

media sosial, yang terkait erat dengan isu-isu viral yang sedang berlangsung pada saat tertentu. Berbagai topik dibahas di Twitter, mulai dari isu politik, layanan, produk hingga acara yang didiskusikan secara publik.

Pemanfaatan media sosial Twitter sebagai wadah masyarakat untuk menyampaikan pendapatnya dapat dijadikan sebagai alat untuk mendapatkan informasi penting tentang suatu permasalahan. Salah satu topik yang bisa dicari informasinya di Twitter adalah opini Masyarakat Pada SEA Games Kamboja 2023. Salah satu analisis yang akan dilakukan adalah mengumpulkan berbagai kicauan berbahasa Indonesia di Twitter pada topik tertentu dan mengklasifikasikannya berdasarkan opini positif dan negatif, opini Masyarakat Pada SEA Games Kamboja 2023. Analisis opini publik ini disebut analisis sentimen. Analisis sentimen digunakan untuk mengetahui apakah opini atau opini publik cenderung membentuk perasaan positif atau negatif terhadap topik tertentu

METODE

Data Penelitian

Pada penelitian ini, dataset yang digunakan diambil dari data teks public pada media sosial twitter berbahasa Indonesia. Data teks berupa komentar atau tweet dari pengguna media sosial twitter. Kemudian diproses oleh aplikasi jupyter notebook Studio dengan cara crawling data. Data dibagi menjadi 2 sentimen yaitu sentimen positif dan sentiment negatif.

← Tweet



FOURTH
@tawarrannn

Ikuti

Keluarin Kamboja dari ASEAN. Cabut hak-haknya sebagai negara ASEAN. Persiapan SEA Games 2023 yang al kadarnya memang masih dimaklumi, tapi kalau sudah menyangkut sportivitas dan menghalalkan segala cara untuk menang, this country is a damn retarded.

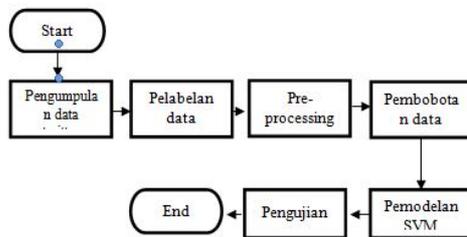
Gambar 1. Tweet Negatif



Gambar 2. Tweet Positif

Penerapan Metodologi

Dalam penelitian ini penulis menggunakan beberapa metode untuk memproses data. Dapat dilihat pada gambar berikut.



Gambar 3. Tahapan Metodologi

Pengumpulan data

Pengumpulan data twitter ini menggunakan teknik crawling data yaitu proses pengambilan data dari twitter yang diproses dengan Jupyter Notebook, berfungsi untuk mengumpulkan dataset yang akan diolah serta merupakan tahap awal dalam penelitian ini. Data yang dikumpulkan merupakan data tweet dengan keyword “Sea Games Kamboja” dan ”Timnas Day Sea Games Gamboja”. Dari hasil crawling tersebut berhasil mendapatkan data berjumlah 1220 datatweet yang selanjutnya akan dilakukan pengolahan agar menjadi data yang berkualitas dalam proses analisis sentiment.

Pelabelan data

Data yang telah di ambil dari twitter akan dilakukan pelabelan secara Otomatis. Data tweet hanya dibagi menjadi kategori positif (1) dan negative(-1). Evaluasi parameter label positif meliputi ungkapan-ungkapan yang baik seperti pujian,

dorongan, dukungan, dll. Dan kalimat negatif adalah kalimat yang mengandung kalimat yang tidak baik, seperti hinaan, putus asa, hinaan, ejekan, dll.

Data Preparation (Preprocessing)

Data yang diambil merupakan data unstruktur atau data yang tidak terstruktur sehingga perlu dilakukan proses preprocessing. Dalam proses ini, kumpulan data tweet melalui proses preprocessing untuk melakukan transformasi data. Pada tahap ini dilakukan penyiapan data teks yang memiliki beberapa permasalahan seperti data duplikat, noise, tidak lengkap.(Setyohadi et al., 2017) Proses tersebut meliputi beberapa tahap yaitu:

a. Casefolding

Tahap Casefolding merupakan proses dalam text preprocessing yang dilakukan untuk menyeragamkan karakter pada data. Proses case folding adalah proses merubah semua huruf menjadi huruf kecil

b. Tokenizing

Tokenisasi adalah proses memecah teks menjadi bagian-bagian dalam bentuk token, yang dapat berupa huruf, kata atau kalimat, sebelum dianalisis lebih lanjut. Entitas yang dapat disebut sebagai token misalnya Kata, angka, simbol, tanda baca, dll.

c. Stopwords

Filtering (stopwords removal). Stop word didefinisikan sebagai term yang tidak berhubungan (irrelevant) dengan subyek utama dari database meskipun katater sebut sering kali hadir di dalam dokumen. Contoh stop words adalah ada, adalah, adanya, adapun, agak, dll.

d. Stemming

Stemming adalah metode untuk mencari kata dasar dari sebuah kata. Stemming merupakan salah satu tahapan dalam pre processing. Proses stemming memiliki pengaruh dalam tingkat akurasi temu kembali informasi. Stemming dilakukan dengan cara menghilangkan imbuhan yang terdapat pada kata.

Pembobotan Data

Metode TF-IDF merupakan metode penghitungan bobot setiap kata yang paling umum digunakan dalam temu kembali informasi. Cara ini juga dikenal efektif, mudah, dan memberikan hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) untuk setiap token (word) pada setiap dokumen dalam korpus. (Amrizal, 2018) Cara ini akan menghitung bobot setiap token pada dokumen dengan rumus:

$$tf_{ij} = \frac{f_d(i)}{\max_{j \in d} f_d(j)}$$

Term Frequency of term mendefinisikan frekuensi (level frekuensi) dari term dalam dokumen. Sedangkan frekuensi dokumen adalah jumlah dokumen yang mengandung term tersebut. Seperti yang Anda lihat pada rumus di atas, nilai yang digunakan dalam perhitungan adalah nilai frekuensi dari suku frekuensi dokumen. N disebut jumlah total dokumen dan df menyatakan nilai frekuensi dokumen dari term yang akan dicari inversnya. Nilai inverse document frequency (idf) diperoleh dari perhitungan berikut:

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

N adalah jumlah total document dalam corpus, $N = |D|$. $|\{d \in D : t \in d\}| = df(t)$, adalah jumlah dokumen yang mengandung term t. IDF juga dapat dituliskan dalam bentuk

$$idf(t, D) = \log \left(\frac{N}{df(t) + 1} \right)$$

Penambahan 1 untuk menghindari pembagian terhadap 0 jika $df(t)$ tidak ditemukan pada corpus. Setelah itu apply L1 normalization dengan fungsi `normalize()` pada library `scikitlearn`. Preprocessing sehingga formulanya tepat sama dengan definisi TF diawal tulisan. L1 Normalization ditulis dalam bentuk:

$$\|x\|_1 = \sum_i |x_i|^1$$

Selanjutnya lakukan perkalian antara TF dan IDF dengan fungsi `multiply()`, sehingga didapatkan TFIDF vector. Penggunaan `smooth_idf = False` seperti dibahas sebelumnya agar fungsi IDF berbentuk, $idf(t) = \log [n / df(t)] + 1$. (Ryansyah, n.d.)

Metode Support Vector Machine (SVM)

Dalam proses klasifikasi ini digunakan algoritma Support Vector Machine (SVM) guna mengenali pola data training, proses klasifikasi ditentukan menjadi 2 kelas yaitu positif dan negatif. Kemudian dataset dibagi menjadi data training dan data testing.

Support Vector Machine ini digunakan untuk mengenali pola data latih yang sebelumnya telah diberi label sentiment, kemudian proses ini akan menghasilkan machine learning model. Model yang telah dihasilkan selanjutnya akan digunakan untuk mengenali data uji yang belum di beri label sehingga pada proses pengujian akan menghasilkan sebuah prediksi kelas sentimen pada data tweet/ data uji apakah termasuk sentimen positif atau sentimen negatif (Lutfianti, 2023)

Rancangan Pengujian

Salah satu teknik yang dapat digunakan untuk mengukur keefektifan suatu model, khususnya sistem klasifikasi (Seperti: analisis sentimen) adalah confusion matrix. Confusion Matrix merupakan metode evaluasi yang dapat digunakan untuk menghitung efisiensi atau akurasi dari proses klasifikasi. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada Confusion Matrix yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). (Arsi & Waluyo, 2021)

Confusion Matrix kemudian dapat dilakukan penghitungan nilai akurasi, presicion, dan recall. (Azhari et al., 2021)

a. Akurasi, adalah metode pengujian berdasarkan kedekatan nilai prediksi dengan nilai sebenarnya. Keakuratan hasil prediksi dapat ditentukan jika jumlah informasi yang diklasifikasikan dengan benar diketahui. Persamaan akurasi ditunjukkan pada persamaan berikut.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

b. Precision, Presisi adalah metode pengujian yang membandingkan sekumpulan informasi relevan yang diterima oleh sistem dengan total informasi yang diambil oleh sistem, terlepas dari relevan atau tidaknya. Persamaan precision ditunjukkan pada persamaan berikut.

$$Precision = \frac{TP}{TP + FP}$$

c. Recall adalah metode pengujian yang membandingkan jumlah informasi yang relevan yang diterima oleh sistem dengan jumlah total informasi yang relevan untuk kumpulan data (baik yang terambil atau tidak terambil oleh sistem). Persamaan recall ditunjukkan pada persamaan berikut

$$Recall = \frac{TP}{TP + FN}$$

HASIL DAN PEMBAHASAN PENELITIAN

Pengumpulan data

Pengumpulan data twitter ini menggunakan teknik crawling data yaitu proses pengambilan data dari twitter yang diproses dengan Jupyter Notebook, berfungsi untuk mengumpulkan dataset yang akan diolah serta merupakan tahap awal dalam penelitian ini. Data yang dikumpulkan merupakan data tweet dengan keyword “Sea Games Kamboja” dan “Timnas Day Sea Games Gamboja”. Dari hasil crawling tersebut berhasil mendapatkan data berjumlah 1220

datatweet yang selanjutnya akan dilakukan pengolahan agar menjadi data yang berkualitas dalam proses analisis sentiment. Contoh data yang dihasilkan dari Crawling adalah sebagai berikut:

Tabel 1. Data Hasil Crawling

Data Hasil Crawling
Kamboja gausah jadi tuan rumah sea games deh. Orang tempat gathering OI jadi admin jokud dijadiin tuan rumah. Kalo di sea games ada mahjong, great rhino sama bonanza baru dah gapapa adain sea games di Kamboja. Negara goblok, satu satunya Kamboja yg gua doyan cuma asinan Kamboja.

Pelabelan data

Data yang telah di ambil dari twitter akan dilakukan pelabelan secara Otomatis. Data tweet hanya dibagi menjadi kategori positif (1) dan negative(-1). Berikut adalah teknik pelabelan data.

```
from textblob import TextBlob
import re
def clean_tweet(tweet):
    return ' '.join(re.sub("([@-#%&()*+,-.:;{ } ]+)|([0-9A-Za-z ]+)|(\w+:\w+/\w+)", " ", tweet).split())
def analyze_sentiment(tweet):
    analysis = TextBlob(clean_tweet(tweet))
    if analysis.sentiment.polarity > 0:
        return 1
    elif analysis.sentiment.polarity == 0:
        return 0
    else:
        return -1
def analyze_subjectorobject(tweet):
    analysis = TextBlob(clean_tweet(tweet))
    if analysis.sentiment.polarity > 0:
        return 'Positif'
    elif analysis.sentiment.polarity == 0:
        return 'Netral'
    else:
        return 'Negatif'
```

Gambar 4. Pelabelan Data

Casefolding

Tahap Casefolding merupakan proses dalam text preprocessing yang dilakukan untuk menyeragamkan karakter pada data. Proses case folding adalah proses merubah semua huruf menjadi huruf kecil (Okfalisa & Harahap, 2016). Contoh data yang dihasilkan dari Casefolding adalah sebagai berikut:

Tabel 1. Casefolding

Sebelum Casefolding	Hasil Casefolding
"Timnas Indonesia Mobile Legends (MLBB) wanita berhasil meraih medali emas usai menang 3-2 di Grand Final SEA Games 2023.Babak Final	timnas indonesia mobile legends wanita mlbb wanita berhasil meraih medali emas usai menang di grand final sea games babak final mobile

Mobile Legends legends wanita
 Wanita berlangsung berlangsung pada
 pada Kamis, 11 Mei kamis mei di naba
 2023 di Naba Theater theater di
 di NagaWorld 2, nagaworld phnom
 Phnom Penh, penh kamboja
 Kamboja.
<https://t.co/0sjUslwpm3>"

Tokenizing

Pada tahap ini dilakukan pemotongan kata pada kalimat menjadi katayangterpisah . sehingga setiap kata dapat dilakukan pengecekan satu persatu padalangkah selanjutnya.(Widyatama &Suprpty, 2018) Contoh data yang dihasilkan dari Tokenizing adalah sebagai berikut:

Tabel 3. Tokenizing

Sebelum Tokenizing	Hasil Tokenizing
timnas indonesia	'timnas', 'indonesia',
mobile legends mlbb	'mobile', 'legends',
wanita berhasil meraih	'mlbb', 'wanita',
medali emas usai	'berhasil', 'meraih',
menang di grand final	'medali', 'emas',
sea games babak final	'usai', 'menang', 'di',
mobile legends wanita	'grand', 'final', 'sea',
berlangsung pada	'games', 'babak',
kamis mei di naba	'final', 'mobile',
theater di nagaworld	'legends', 'wanita',
phnom penh kamboja	'berlangsung', 'pada', 'kamis', 'mei', 'di', 'naba', 'theater', 'di', 'nagaworld', 'phnom', 'penh', 'kamboja'

Stopwords

Filtering (stopwords removal). Stop word didefinisikan sebagai term yang tidak berhubungan (irrelevant) dengan subyek utama dari database meskipun katater sebut sering kali hadir di dalam dokumen. Contoh stop words adalah ada, adalah, adanya, adapun, agak, dll.(Jasa, 2000) Contoh data yang dihasilkan dari Stopwords adalah sebagai berikut:

Tabel 2. Stopwords

Sebelum Stopwords	Hasil Stopwords
-------------------	-----------------

'timnas', 'indonesia', 'timnas',
 'mobile', 'legends', 'indonesia',
 'mlbb', 'wanita', 'mobile',
 'berhasil', 'meraih', 'legends', 'mlbb',
 'medali', 'emas', 'wanita',
 'usai', 'menang', 'di', 'berhasil',
 'grand', 'final', 'sea', 'meraih',
 'games', 'babak', 'medali', 'emas',
 'final', 'mobile', 'menang',
 'legends', 'wanita', 'grand', 'final',
 'berlangsung', 'sea', 'games',
 'pada', 'kamis', 'babak', 'final',
 'mei', 'di', 'naba', 'mobile',
 'theater', 'di', 'legends',
 'nagaworld', 'wanita', 'kamis',
 'phnom', 'penh', 'mei', 'naba',
 'kamboja' 'theater',
 'nagaworld',
 'phnom', 'penh',
 'kamboja'

Stemming

Stemming merupakan proses tahap pencarian kata dasar dari hasil proses filtering dengan membuang kata imbuhan. (Magriyanti, 2018) Contoh data yang dihasilkan dari Stemming adalah sebagai berikut:

Tabel 3. Stemming

Sebelum Stemming	Hasil Stemming
'timnas', 'indonesia',	'timnas',
'mobile', 'legends',	'indonesia',
'mlbb', 'wanita',	'mobile',
'berhasil', 'meraih',	'legends', 'mlbb',
'medali', 'emas',	'wanita', 'hasil',
'menang', 'grand',	'raih', 'medali',
'final', 'sea', 'games',	'emas', 'menang',
'babak', 'final',	'grand', 'final',
'mobile', 'legends',	'sea', 'games',
'wanita', 'kamis',	'babak', 'final',
'mei', 'naba',	'mobile',
'theater',	'legends',
'nagaworld',	'wanita', 'kamis',
'phnom', 'penh',	'mei', 'naba',
'kamboja']	'theater', 'nagaworld', 'phnom', 'penh', 'kamboja']

Pembobotan Data

Metode TF-IDF merupakan metode penghitungan bobot setiap kata yang paling umum digunakan dalam temu kembali informasi. Cara ini juga dikenal efektif, mudah, dan memberikan hasil yang akurat. Contoh data yang dihasilkan dari pembobotan data adalah sebagai berikut:

sekolah	0.04166666666666664	0.2672274565486399
ajar	0.04166666666666664	0.22145194452088197
baca	0.04166666666666664	0.20946519150197776
tulis	0.04166666666666664	0.2383463240253088
uda	0.04166666666666664	0.259333077044133
serta	0.04166666666666664	0.22145194452088197
kamboja	0.04166666666666664	0.01550060323976476
laos	0.04166666666666664	0.2298486760538834
brunei	0.04166666666666664	0.2383463240253088
myanmar	0.04166666666666664	0.1255108989793834
timur	0.04166666666666664	0.2672274565486399
leste	0.04166666666666664	0.10422649798913378
sangkut	0.04166666666666664	0.2672274565486399
paut	0.04166666666666664	0.2672274565486399
indonesia	0.0033333333333333333	0.04452696625277454
sea	0.04166666666666664	0.006371746952365223
games	0.04166666666666664	0.006331969701920786
tuan	0.04166666666666664	0.12990092046512614
rumah	0.04166666666666664	0.12913637302061798
atur	0.04166666666666664	0.1832731640260455
gk	0.04166666666666664	0.2045575650162951
main	0.04166666666666664	0.07984666603691171
ajang	0.04166666666666664	0.16546299674158968

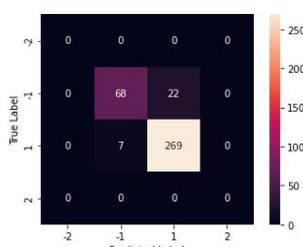
Gambar 5. TF IDF

Hasil Pemodelan dan Pengujian Metode Support Vector Machine (SVM)

Dengan Algoritma Support Vector machine menghasilkan performa terbaik dengan nilai accuracy 92,08%, precision 91,55%, recall 86,50%, f1-score 88,65%. Berikut adalah hasil dari Pemodelan dan pengujian metode algoritma Support Vector Machine

Accuracy :0.9207650273224044
 Precision :0.9155326460481099
 Recall :0.865096618357488
 F1-Score :0.8865480198813533

	precision	recall	f1-score	support
-1	0.9067	0.7556	0.8242	90
1	0.9244	0.9746	0.9489	276
accuracy			0.9208	366
macro avg	0.9155	0.8651	0.8865	366
weighted avg	0.9200	0.9208	0.9182	366



Gambar 6. Hasil akurasi

SIMPULAN

Berdasarkan hasil dari proses pengujian yang telah dilakukan terhadap tweets sentimen terhadap opini masyarakat pada sea games kamboja 2023 dari media sosial twitter sebanyak 1.220 tweets (921

positif dan 299 negatif) menggunakan. Berdasarkan hasil uraian yang telah dibahas, kesimpulan yang dapat ditarik pada penelitian ini adalah sentimen analisis dapat dilakukan menggunakan jupyter notebook untuk membantu memberikan label pada data tweet yang telah di crawling. Model SVM juga bisa digunakan untuk tugas sentimen analisis. Berdasarkan hasil accuracy, precision dan recall. widget test & F1 score yang didapat menggunakan mampu menghasilkan nilai diatas 80 % sehingga dapat dikatakan algoritma support vector machine mampu melakukan analisis sentimen pada opini masyarakat pada sea games kamboja 2023.

DAFTAR PUSTAKA

Amrizal, V. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *Jurnal Teknik Informatika*, 11(2), 149–164. <https://doi.org/10.15408/jti.v11i2.8623>

Arsi, P., & Waluyo, R. (2021). Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM). *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(1), 147. <https://doi.org/10.25126/jtiik.0813944>

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>

Jasa, K. (2000). Bab 2 Landasan Teori. *Aplikasi Dan Analisis Literatur Fasilkom UI*, m(1998), 7–34. <http://elib.unikom.ac.id/files/disk1/6>

- 55/jbptunikompp-gdl-supriadini-32740-6-12.unik-i.pdf
- Lutfianti, N. (2023). Penerapan Sentimen Analisis Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023. *Sisfotenika*, 13(1), 53–64.
- Magriyanti, A. A. (2018). Analisis Pengembangan Algoritma Porter Stemming Dalam Bahasa Indonesia. *Sekolah Tinggi Elektronika Dan Komputer PAT*, 1(1), 1–5. <http://kompas.com>
- Okfalisa, & Harahap, A. H. (2016). Implementasi Metode Term Frequency–Inverse Document Frequency (TF-IDF) Dan Maximum Marginal Relevance Untuk Monitoring Diskusi Online. *Jurnal Sains, Teknologi Dan Industri*, 13(2), 151–159.
- Ryansyah, A. (n.d.). *Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen. 1.*
- Setyohadi, D. B., Kristiawan, F. A., & Ernawati, E. (2017). Perbaikan Performansi Klasifikasi Dengan Preprocessing Iterative Partitioning Filter Algorithm. *Telematika*, 14(01), 12–20. <https://doi.org/10.31315/telematika.v14i01.1960>
- Widyatama & Suprpty. (2018). Klasifikasi Sentimen Terhadap Bukalapak Dengan Menggunakan Metode Naive Bayes Classifier. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.