Vol. 4, No. 2, Agustus 2023, hal. 376-384

CoSciTech Computer Science and Information Technology

p-ISSN: 2723-567X
e-ISSN: 2723-5661

Jurnal Computer Science and Information Technology (CoSciTech)

http://ejurnal.umri.ac.id/index.php/coscitech/index

Analisis Sentimen Twitter Terhadap *Cyberbullying* Menggunakan Metode *Support Vector Machine* (SVM)

Rismi Nurlaely[1], Dwi Sartika Simatupang[2], Kamdan[3], Ivana Lucia Kharisma[*4]
rismi.nurlaely_ti19@nusaputra.ac.id, dwi.simatupang@nusaputra.ac.id, kamdan@nusaputra.ac.id, ivana.lucia@nusaputra.ac.id
[1234]Program Studi Teknik Informatika, Fakultas Teknik Dan Desain, Universitas Nusa Putra Sukabumi

Diterima: 12 Juni 2023 | Direvisi: 09 Agustus 2023 | Disetujui: 28 Agustus 2023

Abstrak

Perkembangan teknologi membantu manusia untuk berkomunikasi satu sama lain. Pesatnya perkembangan teknologi telah menyebabkan lahirnya banyak platform media baru. Seiring meningkatnya kemajuan teknologi di era digital ini. Keberadaan teknologi informasi memberikan dampak yang besar bagi peradaban manusia. salah satu bentuknya adalah penggunaan media social. Namun, kemudahan berbagi informasi melalui media sosial tidak luput dari penyalahgunaan oleh para penggunanya. Salah satu bentuk penyalahgunaan adalah cyberbullying yang dilakukan di media sosial terutama twitter. Survei Penetrasi Internet dan Perilaku Pengguna Internet di Indonesia tahun 2018 yang diterbitkan oleh Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) menunjukkan bahwa 49 persen pengguna internet pernah mengalami perundungan dalam bentuk ejekan atau pelecehan di media sosial. Berkaitan dengan hal itu maka dilakukan analisis sentimen melalui media sosial twitter untuk mengukur nilai akurasi menggunakan algoritma SVM (support vector machine). Analisis sentimen dengan melakukan crawling pada data twitter sebanyak 1000 data tweet yang berbahasa Indonesia menggunakan tools RapidMiner. Dan Klasifikasi dengan algoritma SVM (Support Vector Machine) ini mendapatkan akurasi sebesar 92% dengan pengujian pada proposi 80:20 yaitu 80% data latih dan 20% data testing. Dari 998 data yang sudah dilakukan preprocessing, mendapatkan 625 data pada kelas positif dan 374 data pada kelas negatif.

***Kata kunci**: analisis sentiment, support vector machine, twitter, cyberbullying*

Twitter sentiment analysis of cyberbullying using Support Vector Machine (SVM) method

Abstract

The development of technology helps humans to communicate with each other. The rapid development of technology has led to the birth of many new media platforms. As technology advances in this digital era. The existence of information technology has a great impact on human civilization. One form is the use of social media. However, the ease of sharing information through social media has not escaped abuse by its users. One form of abuse is cyberbullying carried out on social media, especially Twitter. The 2018 Internet Penetration and Internet User Behavior Survey in Indonesia published by the Indonesian Internet Service Providers Association (APJII) shows that 49 percent of internet users have experienced bullying in the form of ridicule or harassment on social media. In this regard, sentiment analysis was carried out through social media twitter to measure the accuracy value using the SVM (support vector machine) algorithm. Sentiment analysis by crawling twitter data as many as 1000 tweet data in Indonesian using the RapidMiner tool. And this classification with SVM (Support Vector Machine) algorithm gets an accuracy of 92% by testing at 80:20 proportion, which is 80% training data and 20% testing data. Of the 998 data that have been preprocessed, 625 data on the positive class and 374 data on the negative class have been obtained.

Jurnal *Computer Science and Information Technology* (CoSciTech) Vol. 4, No. 2, Agustus 2023, hal. 376-384

***Keywords**: sentiment analysis, support vector machine, twitter, cyberbullying*

1. PENDAHULUAN

Perkembangan teknologi membantu manusia untuk berkomunikasi satu sama lain. Pesatnnya perkembangan teknologi telah menyebabkan lahirnya banyak platform. Hal ini tentu membantu manusia dalam berkomunikasi menjadi lebih mudah. Komunikasi dapat terjadi tidak hanya dapat dilakukan secara langsung, tetapi juga dapat melalui dunia maya[1]. Media sosial adalah *platform* media yang berfokus pada kehadiran pengguna dan memfasilitasi aktivitas kolaborasi mereka. Semakin maju, semakin mudah informasi tersebar dengan mudah dan cepat sehingga mempengaruhi cara pandang, gaya hidup, serta budaya suatu bangsa[2]. Baru-baru ini, yang paling sering digunkan adalah: *Instagram, Facebook, YouTube*, dan *Twitter*. Media sosial telah menjadi media favorit anak muda[3]. Namun kemudahan berbagi informasi melalui media social tidak luput dari penyalahgunaan oleh para penggunanya. Salah satu bentuk penyalahgunaan adalah *cyberbullying*.

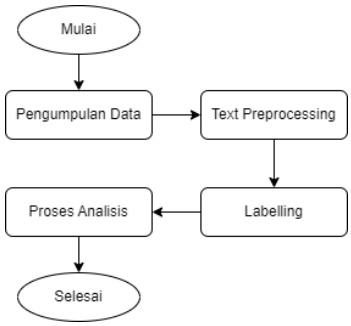
Cyberbullying adalah ketika anak atau remaja lain mengolok-olok, mempermalukan, mengintimidasi, atau dipermalukan oleh anak atau remaja lain menggunakan internet, teknologi digital atau ponsel. Survei Penetrasi Internet dan Perilaku Pengguna Internet di Indonesia tahun 2018 yang dirilis oleh Asosiasi Penyelenggara Jasa Internet di Indonesia (APJII) menunjukkan bahwa 49 persen pengguna internet pernah mengalami pelecehan dalam bentuk ejekan atau pelecehan di media sosial. hal ini membuat orang dewasa, remaja dan anak-anak rentan terhadap perlakuan negatif bahkan menjadi pelaku[4]. Salah satu media sosial yang sangat populer saat ini yaitu *Twitter*, perkembangan *Twitter* sebagai *platform* yang digandrungi masyarakat terbukti dengan data yang menunjukkan bahwa Indonesia peringkat lima di dunia dengan banyak pengguna mencapai 18.4 juta[4]. *Twitter* menjadi media sosial yang mendorong penggunanya untuk berpikir kreatif dalam berkicau atau menggugah *tweet,* sehingga penggunaannya dibebaskan dalam memberikan tanggapan atau opini yang ingin disampaikan melaui *tweet.*

Oleh karena itu, tujuan dari penelitian ini adalah untuk melakukan "Analisis Sentimen Twitter Terhadap *Cyberbullying* Menggunakan Metode *Support Vector Machine* (SVM)". Maka penelitian ini merumuskan masalah yang akan diteliti yaitu bagaimana cara menerapkan algoritma *Support Vector Machine* untuk sentiment di media sosial *twitter,* dan peneliti juga menetapkan batasan masalah agar penelitian lebih terfokus berdasarkan rumusan masalah dan sesuai batasan kemampuan penulis yaitu, data yang digunakan/dikumpulkan untuk penelitian ini adalah data pada media sosial *twitter* berbahasa Indonesia. Dan Metode yang dipakai dalam penelitian ini menggunakan algoritma *Support Vector Machine.*

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Langkah-langkah yang dilakukan dalam penelitian kali ini adalah sebagai berikut:



Gambar 1. Tahapan Penelitian

Pada Gambar 1 diatas menunjukkan tahapan-tahapan berdasarkan penelitian, berikut penjelasan setiap tahapannya:

2.2 Pengumpulan Data

a. Studi Pustaka

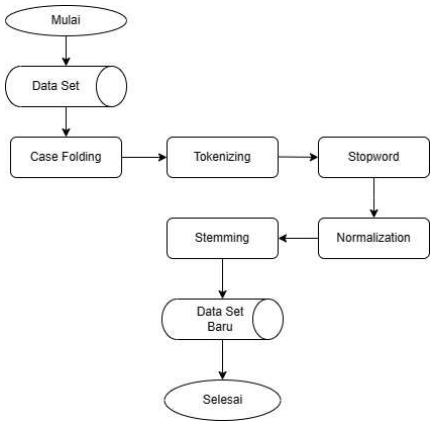
Studi kepustakaan berkaitan dengan kajian teoritis dan referensi lain yang berkaitan dengan nilai, budaya, dan norma yang berkembang dalam situasi sosial yang dikaji. Teori yang mendasari penelitian memiliki tiga kriteria yaitu relevansi, kompleksitas, dan orisinalitas. Relevansi berarti bahwa teori yang disajikan konsisten dengan masalah yang sedang dipelajari. Jika melihat topik kepemimpinan, teori yang disajikan berkaitan dengan dengan kepemimpinan, bukan teori sikap atau motivasi. Kemutakhiran berarti terkait dengan kebaruan teori atau referensi yang digunakan[5].

2.3 Crawling data

Proses pengumpulan data secara otomatis dari internet dengan menjelajahi halaman *web* secara sistematis. *Crawling* sering digunakan sebagai metode untuk mengumpulkan data yang relevan dengan topik penelitian. Pada proses pengumpulan *dataset* dilakukan dengan cara *crawling* data *Twitter* melalui akses token dari *Twitter API* menggunakan *tools RapidMiner.*

2.4 Text Processing

Preprocessing adalah langkah terpenting dalam melakukan analisis sentimen. *Preprocessing* atau pengolahan kata, menyiapkan teks untuk digunakan, siap untuk diproses lebih lanjut dengan tujuan untuk menghilangkan *noise*[4]. Dan mengubahnya menjadi bentuk yang lebih terstruktur. Berikut tahapan pada proses *text preprocessing*:



Gambar 2. Tahapan text preprocessing

Proses pada tahapan kali ini menggunakan bahasa pemrograman *Python* dengan *Google Colab* sebagai *coding environment*.

2.5 Pelabelan Sentimen

Setelah melakukan pemrosesan teks, langkah berikutnya adalah melakukan pelabelan sentimen. Pelabelan sentimen sangat penting dalam analisis sentimen, dan dalam penelitian ini, peneliti menggunakan pemrograman *Python* untuk membuat sistem pelabelan sentimen. Mereka juga membuat kamus kata positif dan negatif dalam Bahasa Indonesia untuk melatih sistem yang dibuat. Dalam penelitian ini, sistem pelabelan dibagi menjadi dua kategori: sentimen positif dan sentimen negatif.

2.6 Pembobotan TF-IDF

Pembobotan kata dilakukan dengan menghitung jumlah kata dalam dokumen atau percakapan. Dengan menggunakan metode TF-IDF, vector multi-term dapat dibuat dalam proses pembobotan, dimana setiap kata dapat diidentifikasi yang dianggap sebagai satu fitur[6]. Pembobotan yang paling populer digunakan adalah TF-IDF, pendekatan ini berasal dari TF, istilah analisis berdasarkan frekuensi dari sebuah teks, dimana IDF adalah definisi konsep dalam ruang dokumen atau spesifisitas dari dokumen tersebut, istilah pengukuran pertama kali diusulkan oleh Sparck Jones dan kemudian sekarang dikenal sebagai IDF[7].

2.7 Visualisasi Hasil Analisis

Setelah melalui tahap text preprocessing selanjutnya adalah visualisasi hasil analisis. Pada langkah ini, dimana dilakukan proses visualisasi dan asosiasi teks untuk informasi data sentimen positif dan negatif[8].

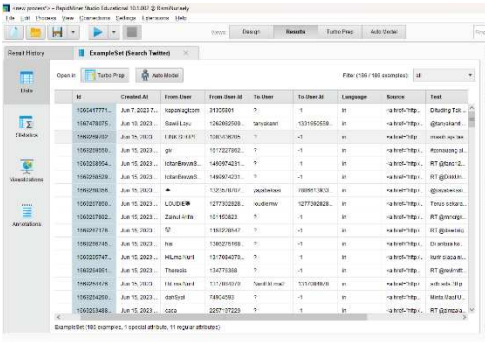
2.8 Klasifikasi *Support Vector Machine*

Pada tahap ini, setelah melalui proses *preprocessing*, dataset dipecah menjadi dua bagian yaitu data latih dan data uji, untuk perbandingan yang ditentukan. Perincian ini dapat bervariasi, mulai dari menggunakan 90% data latih dan 10% data uji, hingga ditemukan kombinasi yang memberikan nilai akurasi tertinggi. Data latih merupakan subset dari data awal yang kita miliki, sedangkan data uji merupakan subset lain yang diambil secara acak dari data awal. Setelah itu, dilakukan validasi dengan menguji sistem menggunakan data uji untuk mendapatkan hasil akurasi. Dengan melakukan berbagai perbandingan data latih dan data uji, kita dapat menemukan kombinasi yang memberikan kinerja dan akurasi terbaik untuk sistem. kelebihan SVM mampu mengatasi permasalahan klasifikasi teks yang menentukan jarak pemisah kelas positif dan negatif yang optimal[9].

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Crawling adalah Teknik pengumpulan data *website* dengan cara memasuki *Uniform Resource Locator* (URL). URL ini berfungsi sebagai referensi untuk menemukan semua hyperlink aktif di situs web[5]. Pada tahap *crawling* data dilakukan meggunakan *tools RapidMiner,* dengan mencari *keyword* "tasyi" kemudian didapatkan hasil data *tweet* seperti pada Gambar 3 hasil dari *crawling* data yang telah dilakukan.



Gambar 3. Data Crawling

Pada gambar di atas adalah hasil data *crawling* dengan *tweet* berbahasa Indonesia dengan pencarian dari yang terbaru hingga yang popular.

3.2 Text preprocessing

Pada tahap *preprocessing*, data akan melalui serangkaian langkah pengolahan teks. Langkah-langkah ini bertujuan untuk membersihkan data dari informasi yang tidak relevan dan *noise*, serta mengubahnya menjadi bentuk yang lebih terstruktur.

3.2.1 *Case folding*

Case folding adalah tahap pertama di *preprocessing*, pada tahapan ini, dilakukan untuk mengubah kata menjadi sama. Masalah ini dilakukan dengan merubah kata menjadi huruf kecil[10]. Sehingga pada saat melakukan klasifikasi data, akan memudahkan model untuk membaca dataset. Hasil sebelum dan sesudahnya dilakukan case folding terdapat pada Tabel 1 berikut.

Tabel 1. Hasil Case Folding

Sebelum	Sesudah
tasyi kalo kurus cantik bet gilaa kalo skrg malah kayak bulldozer wkwk maap	tasyi kalo kurus cantik bet gilaa kalo skrg malah kayak bulldozer wkwk maap
Tadi ngeliat ss'an mantan karyawan tasyi yg ga dibayar ckckckck Lagi2 tasyi dah	tadi ngeliat ssan mantan karyawan tasyi yg ga dibayar ckckckck lagi tasyi dah

3.2.2 *Tokenizing*

Selanjutnya adalah tahapan *tokenizing*, Tokenisasi adalah proses membagi teks menjadi unit yang lebih kecil yang disebut token. Tanda dapat berupa kata-kata, frasa, karakter, atau unit-unit lainnya tergantung pada tujuan analisis atau konteks pemrosesan teks yang dilakukan. Tokenisasi penting dalam pemrosesan bahasa alami dan analisis teks karena dapat memberikan representasi yang lebih terstruktur dan terperinci dari teks yang akan diproses. Pada tabel 2 adalah hasil dari tokenisasi pada suatu data, sehingga data menjadi sebuah token-token atau kata yang independen.

Tabel 2. Hasil dari Tokenizing

Sebelum	Sesudah

tasyi kalo kurus cantik bet gilaa kalo skrg malah kayak bulldozer wkwk maap	['tasyi', 'kalo', 'kurus', 'cantik', 'bet', 'gilaa', 'kalo', 'skrg', 'malah', 'kayak', 'bulldozer', 'wkwk', 'maap']
Tadi ngeliat ss'an mantan karyawan tasyi yg ga dibayar ckckckck Lagi2 tasyi dah	['tadi', 'ngeliat', 'ssan', 'mantan', 'karyawan', 'tasyi', 'yg', 'ga', 'dibayar', 'ckckckck', 'lagi', 'tasyi', 'dah']

3.2.3 *Stopwords*

Banyaknya tweet yang digunakan dalam penelitian menghasilkan banyak frase dan fitur. Memilih kata yang bermakna dalam pembuatan model dengan menghilangkan kata yang kurang bermakna dapat meningkatkan akurasi sistem klasifikasi[11]*, stopwords* adalah langkah yang umum dilakukan sebagai bagian dari pra-pemrosesan teks. Pada Tabel 3 berikut adalah hasil yang telah dilakukan stopwords atau filtering pada data.

Tabel 3. Hasil Stopwords

Sebelum	Sesudah
tasyi kalo kurus cantik bet gilaa kalo skrg malah kayak bulldozer wkwk maap	['tasyi', 'kurus', 'cantik', 'bet', 'gilaa', 'skrg', 'kayak', 'bulldozer', 'wkwk', 'maap']
Tadi ngeliat ss'an mantan karyawan tasyi yg ga dibayar ckckckck Lagi2 tasyi dah	['ngeliat', 'ssan', 'mantan', 'karyawan', 'tasyi', 'ga', 'dibayar', 'ckckckck', 'tasyi', 'dah']

3.2.4 Normalisasi

Pada tahapan selanjutnya adalah normalisasi, Normalisasi adalah proses mengubah teks menjadi bentuk standar atau normal yang lebih terstruktur. Normalisasi dilakukan dalam pemrosesan bahasa alami dan analisis teks. Hasil sebelum dan sesudah dilakuknnya normalisasi diberikan dalam tabel 4 berikut.

Tabel 4 Hasil Normalisasi

Sebelum	Sesudah
tasyi kalo kurus cantik bet gilaa kalo skrg malah kayak bulldozer wkwk maap	['tasyi', 'kurus', 'cantik', 'banget', 'gilaa', 'sekarang', 'kayak', 'bulldozer', 'haha', 'maap']
Tadi ngeliat ss'an mantan karyawan tasyi yg ga dibayar ckckckck Lagi2 tasyi dah	['ngeliat', 'screenshot', 'mantan', 'karyawan', 'tasyi', 'ga', 'dibayar', 'haha', 'tasyi', 'dah']

3.2.5 Stemming

Tahapan berikutnya adalah *stemming* merupakan proses dalam pemrosesan bahasa alami dimana kata-kata diubah kedalam bentuk dasar atau kata dasar. Tujuan utama dari *stemming* adalah untuk mengurangi variasi kata yang terkait secara morfologis menjadi bentuk dasarnya. Dengan melakukan *stemming*, kata-kata dengan akar kata yang sama akan diubah menjadi satu bentuk dasar yang seragam, sehingga mempermudah pengolahan dan analisis teks. Hasil dari *stemming* terdapat pada Tabel 5 berikut:

Tabel 5. Hasil Stemming

Sebelum	Sesudah
Kaya gni ya keselnya Tasya ma Tasyi sekarang maybeee 'Pencitraan' kata Tasya https://t.co/q9EeJzP9PP	'kaya', 'gini', 'kesel', 'tasya', 'sama', 'tasyi', 'mungkin', 'citra',

3.3 Pelabelan Sentimen

Tahapan *preprocessing* data sudah dilakukan, selanjutnya adalah dengan memberikan label pada data. Proses pelabelan data dengan cara melakukan perbandingan kata di kamus Bahasa Indonesia dengan menghikung skor sentimen[12] pada setiap data *tweet.* Terdapat pada Gambar 4 adalah perintah untuk melakukan label pada data dengan memberikan list kata-kata yang positif dan negatif.

```
import pandas as pd

positive_words = ['famous','jutawan','nikah','baru','sadar','usaha','dukungan','ga','belain','damai','sadar','bener',

negative_words = ['jelek','ketar','ketir','anjing','ghibahin','fitnah','salah','aneh','caper','fitnah','angkuh','nge

data = pd.read_excel("data_preprocessed.xlsx")
data['sentiment'] = ""

for i in range(len(data)):
    text = data.iloc[1]['Text']

    if isinstance(text, str):
        num_pos_words = sum(text.count(word) for word in positive_words)
        num_neg_words = sum(text.count(word) for word in negative_words)

        if num_pos_words > num_neg_words:
            sentiment = 'positif'
        elif num_pos_words < num_neg_words:
            sentiment = 'negatif'


    data.at[i, 'sentiment'] = sentiment

print(data.head(5))
```

Gambar 4. Tahap Labelling

Pada Gambar 5 berikut adalah hasil data yang telah diberikan label atau sentimen.

Text	tokens	ext_token	kens_nor	okens_ste	sentimen
iyaaa dulu	['iyaaa', 'd	['iyaaa', 'p	['iya', 'pas	['iya', 'pas	negatif
beneran g	['beneran	['beneran	['beneran	['beneran	negatif
aku baru s	['aku', 'bai	['sadar', 'n	['sadar', 'n	['sadar', 'n	negatif
rt dia sela	['rt', 'dia',	['nyalahin	['nyalahin	['nyalahin	negatif
rt pdhl di	['rt', 'pdhl'	['pdhl', 'ta	['padahal'	['padahal'	negatif
tapi benei	['tapi', 'be	['bener', '[	['bener', '	['bener', '	negatif
rt keshwa	['rt', 'kesh	['aneh', 'ta	['aneh', 'ta	['aneh', 'ta	positif
rt kalo tas	['rt', 'kalo'	['tasya', 'si	['tasya', 'si	['tasya', 'si	positif

Gambar 5. Hasil Labelling

3.4 Pembobotan Kata TF-IDF

Tahapan selanjutnya adalah TF-IDF *vectorizer* untuk mengubah dokumen teks menjadi representasi numerik berdasarkan bobot TF-IDF setiap kata dalam dataset. Representasi ini memungkinkan pemodelan dan analisis lebih lanjut menggunakan metode pembelajaran mesin yang akan dilakukan dengan algoritma SVM *(Support Vector Machine*). Dengan menggunakan TF-IDF *vectorizer*, dapat menggambarkan dokumen teks menjadi vektor numerik yang merepresentasikan kepentingan kata-kata dalam dokumen tersebut.

```
[ ]  #tf idf
     from sklearn.feature_extraction.text import TfidfVectorizer
     import pickle
     vectorizer = TfidfVectorizer()
     X_train= vectorizer.fit_transform(X_train)
     pickle.dump(vectorizer, open('tfidf6.pkl', 'wb'))
     X_test = vectorizer.transform(X_test)
```

Gambar 6. Tahap TF-IDF Vectorizer

Pada Gambar 6 merupakah perintah untuk dilakukannya tdidf *vectorizer* sebelum melakukan klasifikasi dengan model SVM.

3.5 Klasifikasi Support Vector Machine

Klasifikasi menggunakan algoritma *Support Vector Machine*, sebelum dilakukannya klasifikasi, maka data dibagai menjadi proposi 80:20 dengan 80% data latih dan 10% data *testing.* Seperti pada Gambar 7 merupakan perintah untuk melakukan pembagian data.

```
] #pembagian data training dan data test
  from sklearn.model_selection import train_test_split
  X_train, X_test, y_train, y_test = train_test_split(df['Text'], df['label'], test_size=0.2, stratify=df['label'], random_state=30
  #membagi data  menjadi data training dan data testing, test_size= 0.2 yang artinya bayaknya data testing adalah 20%
```

Gambar 7. Pembagian Data

Setelah data dibagi, selanjutnya adalah pemodelan menggunakan algoritma SVM, pada Gambar 8 berikut adalah perintah untuk melakukan klasifikasi SVM.

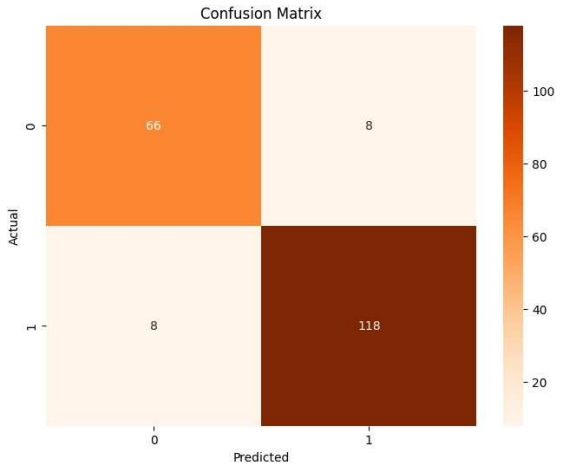
```
#suport vector machine
from sklearn import svm
from sklearn.model_selection import cross_val_score
clf = svm.SVC(kernel = 'linear').fit(X_train,y_train)
#menyimpan model svm
pickle.dump(clf, open('svm6.pkl', 'wb'))
#prediksi data test
prediksi = clf.predict(X_test)
```

Gambar 8. Tahap Pemodelan SVM

Gambar 8 menunjukkan perintah saat melakukan klasifikasi menggunakan SVM, dengan mencari tingkat *hyperplane* terbaik yang dapat memisahkan kedua *class* tersebut dengan margin yang maksimal. *Hyperplane* ini merupakan batas keputusan yang membagi data dengan kategori yang berbeda. Data yang berada pada sisi yang berbeda dari *hyperplane* akan diklasifikasikan ke dalam kelas yang sesuai.

3.6 Confusion Matrix

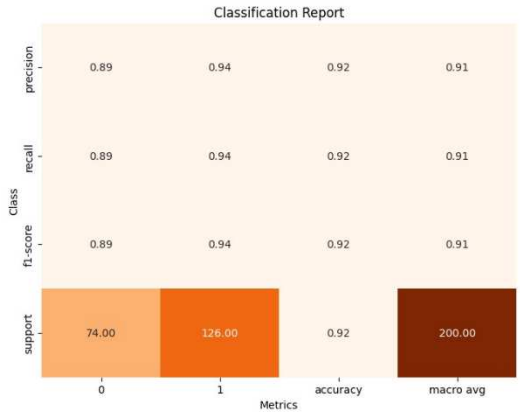
Klasifikasi data menggunakan SVM kemudian evaluasi tes dilakukan dengan menggunakan *confusion matrix*. Pada Gambar 9 berikut adalah hasil dari *confusion matrix* dengan menguji *true positive, false positive, true negative, dan false negative*.



Gambar 9. Hasil Confusion Matrix

Pada Gambar 9 menunjukkan bahwa prediksi pada data testing dengan 200 data, model memprediksi pada 126 data yang merupakan kelas positif, namun pada saat evaluasi dengan prediksi yang sebenarnya terdapat 118 data *true* pada kelas positif, dan 8 data yang *false positive*. Sedangkan untuk kelas *negative* model memprediksi 74 data yang negatif, namun prediksi yang sebenarnya hanya 66 data yang terdapat *true negative* serta 8 data yang *false negative*.

Selanjutnya adalah hasil dari *precision, recall* dan *f1-score* terdapat pada Gambar 10 berikut.



Gambar 10. Hasil Precisio, Recall dan F1-Score

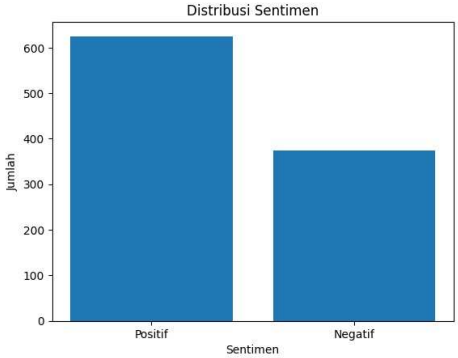
Dalam hasil evaluasi klasifikasi yang diberikan, terdapat beberapa metrik yang digunakan untuk mengukur performa model klasifikasi SVM, antara lain:

1. *Precision*: Menggambarkan sejauh mana hasil prediksi positif (kelas 1) akurat. Nilai precision untuk kelas 0 adalah 0.89, yang berarti sekitar 89% dari prediksi kelas 0 benar, sedangkan untuk kelas 1 adalah 0.94, yang berarti sekitar 94% dari prediksi kelas 1 benar.
2. *Recall*: Dikenal juga sebagai *True Positive Rate* atau *Sensitivity*, menggambarkan sejauh mana model dapat mengenali dan mengklasifikasikan dengan benar anggota kelas yang benar (positif). Nilai recall untuk kelas 0 adalah 0.89, yang berarti sekitar 89% dari total anggota kelas 0 berhasil ditemukan oleh model, sedangkan untuk kelas 1 adalah 0.94, yang berarti sekitar 94% dari total anggota kelas 1 berhasil ditemukan.
3. *F1-score:* Adalah *harmonic mean* dari *precision* dan *recall. F1-score* digunakan untuk menyatukan kedua metrik tersebut menjadi satu angka untuk memberikan gambaran keseluruhan performa model. Nilai *F1-score* untuk kelas 0 adalah 0.89, sedangkan untuk kelas 1 adalah 0.94.
4. Support: Menunjukkan jumlah sampel yang ada dalam masing-masing kelas pada data test.

Selain itu, terdapat juga metrik lainnya seperti *accuracy*, yang menggambarkan seberapa baik model dapat memprediksi secara keseluruhan. Pada hasil tersebut, akurasi *(accuracy)* yang diperoleh menggunakan algoritma SVM adalah 0.92, yang berarti sekitar 92% dari total data berhasil diprediksi dengan benar oleh model.

3.7 Visualisasi Hasil Analisis

Pada Gambar 11 merupakan visualisai dari hasil analisis yang telah dilakukan dengan model SVM, dan topik mengenai *cyberbullying*. Terdapat 625 data pada kelas positif dan 374 data pada kelas negatif.



Gambar 11. Visualisasi Hasil Analisis

KESIMPULAN

Penelitian yang telah dilakukan terdapat kesimpulan yang didapatkan dari hasil analisis sentimen twitter terhadap *cyberbullying* menggunakan algoritma SVM adalah sebagai berikut:

1. Analisis sentimen dengan melakukan *crawling* pada data twitter sebanyak 1000 data *tweet* yang berbahasa Indonesia menggunakan tools RapidMiner.
2. Klasifikasi menggunakan algoritma SVM (*Support Vector Machine)* ini mencapai akurasi sebesar 92% ketika pengujian pada proposi 80:20 yaitu 80% data latih dan 20% data *testing*.

Dari 998 data telah dilakukan *preprocessing,* terdapat 625 data pada kelas positif dan 374 data pada kelas negatif.

DAFTAR PUSTAKA

[1] E. O. Afifah and T. Kusuma, "Analisis Komunikasi Antar Penggemar Seventeen Sebagai Cyberfandom Di Twitter," *Mediator: Jurnal Komunikasi*, vol. 12, no. 1, Jun. 2019, doi: 10.29313/mediator.v12i1.4624.

[2] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 3, no. 1, p. 21, May 2020, doi: 10.24014/ijaidm.v3i1.9145.

[3] R. I. Marchellia and C. Siahaan, "PERANAN MEDIA SOSIAL INSTAGRAM SEBAGAI MEDIA KOMUNIKASI REMAJA PENGGEMAR KPOP," *JRK (Jurnal Riset Komunikasi)*, vol. 13, no. 1, p. 65, Jun. 2022, doi: 10.31506/jrk.v13i1.14737.

[4] A. Putri and A. Muzakir, "ANALISIS SENTIMEN CYBERBULLYING KPOP DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES," *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 7, no. 9, 2022, doi: https://doi.org/10.36418/syntax-literate.v7i9.9334.

[5] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, p. 147, Feb. 2021, doi: 10.25126/jtiik.0813944.

[6] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA," *Edutic - Scientific Journal of Informatics Education*, vol. 7, no. 1, Nov. 2020, doi: 10.21107/edutic.v7i1.8779.

[7] Ash Shiddicky and Surya Agustian, "Analisis Sentimen Masyarakat Terhadap Kebijakan Vaksinasi Covid-19 pada Media Sosial Twitter menggunakan Metode Logistic Regression," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 99–106, Aug. 2022, doi: 10.37859/coscitech.v3i2.3836.

[8] S. Hasna Kamila, "Analisis Sentimen Data Ulasan Menggunakan Algoritma Support Vector Machine," Universitas Islam Indonesia, 2021.

[9] D. Angraina and A. Putri, "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 472–478, Dec. 2022, doi: 10.37859/coscitech.v3i3.4260.

[10] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[11] N. Fitriyah, B. Warsito, D. Asih, and I. Maruddani, "ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM)," *JURNAL GAUSSIAN*, vol. 9, no. 3, pp. 376–390, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[12] R. Nooraeni, A. Fikri Fadhilah I, H. Dwi, S. Fatimatul, S. Pertiwi, and Y. Ronaldias, "Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM)," *Paradigma – Jurnal Informatika dan Komputer*, vol. 22, no. 1, 2020, doi: 10.31294/p.v21i2.