Jurnal Ilmiah Wahana Pendidikan, November 2023, 9 (21), 314-323
DOI: https://doi.org/10.5281/zenodo.10077023
p-ISSN: 2622-8327 e-ISSN: 2089-5364
Accredited by Directorate General of Strengthening for Research and Development
Available online at https://jurnal.peneliti.net/index.php/JIWP

Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik *Lifestyle*

Fadlila Nurwanda[1], Jelita Ratu Rizkiani[2]

Program Studi Statistika, Universitas Sebelas Maret, Surakarta, Indonesia

Received: 20 Oktober 2023
Revised : 27 Oktober 2023
Accepted: 01 November 2023

Abstract

Technology is currently developing rapidly thanks to the widespread growth of the internet worldwide. This growth has triggered an increasing demand for diverse information, especially in textual form. One way to fulfill this information demand is through social media platforms, which enable communication and interaction among individuals. Twitter has become a popular social media platform in Indonesia, providing a space for people to express their opinions on various topics, including lifestyle. These opinions can range from positive to negative or even neutral. Sentiment analysis is needed to provide a general overview of the sentiment expressed by the Indonesian public regarding lifestyle topics. This research utilizes the Naive Bayes Classifier (NBC) and Support Vector Machine (SVM) classification methods to compare which method is most effective in analyzing sentiment towards lifestyle topics in Indonesian society. The study found that the SVM method achieved the highest accuracy of 61% and produced consistent prediction results.

Keywords: *Twitter; lifestyle; Indonesia; Naive Bayes Classifier; Support Vector Machine*

(*) Corresponding Author: jelitaratur@student.uns.ac.id

How to Cite: Nurwanda, F., & Rizkiani, J. R. (2023). Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik Lifestyle. https://doi.org/10.5281/zenodo.10077023

PENDAHULUAN

Teknologi saat ini berkembang karena konsekuensi dari pertumbuhan internet yang meluas di seluruh dunia. Selaku konsekuensi dari meningkatnya kebutuhan warga dengan bermacam tipe informasi yang nantinya akan digunakan untuk kebutuhan yang berbeda - beda. Pesatnya perkembangan dunia teknologi informasi dan komunikasi tidak terlepas dari *web service provider* yang menyediakan berbagai informasi. Informasi yang menyebabkan bertambahnya data, terutama data teks dapat digunakan sebagai sumber yang sangat potensial untuk penemuan lebih lanjut. Setiap tahunnya data mengalami peningkatan mencapai 80% terdiri dari data yang sifatnya tidak terstruktur, khususnya seperti data dalam bentuk teks.

Internet adalah jaringan besar yang saling menghubungkan mulai dari jaringan-jaringan komputer yang satu ke jaringan- jaringan komputer diseluruh dunia melalui satelit. Salah satu manfaat dari perkembangan teknologi internet ini adalah sarana untuk berkomunikasi. Sarana komunikasi yang sangat populer saat ini adalah media sosial. Media sosial adalah medium di internet yang memungkinkan pengguna merepresentasikan dirinya maupun berinteraksi, bekerja sama, berbagi, berkomunikasi dengan pengguna lain membentuk ikatan sosial secara virtual. Selain digunakan untuk melakukan komunikasi dan interaksi dengan orang lain terkadang medial sosial digunakan untuk tindakan yang kurang baik.

Utamanya media sosial mempunyai akibat yang signifikan pada cara bangsa beroperasi.

Twitter merupakan jenis media sosial cukup popular dan diminati oleh seluruh masyarakat dunia termasuk Indonesia. Media sosial ini yang dikembangkan oleh Jack Dorsey dengan tujuan untuk mengirimkan pesan antar pengguna melalui tweet atau kicauan. Pada tahun 2020, penggunaan twitter mencapai 19,5 juta pengguna dengan mendapatkan peringkat ke-5 sebagai media sosial yang sering digunakan. Dilihat dari data tersebut, memberikan informasi bahwa twitter ialah media sosial yang cukup memiliki potensi bagi pengguna khususnya di Indonesia. Twitter sebagai salah satu jejaring sosial memudahkan masyarakat dalam berpendapat ataupun memberikan kritik suatu hal secara realtime. Salah satunya mengenai gaya hidup anak remaja milenial yang mengutarakan opininya dalam hal konser, drama korea, perokok, dan lain sebagainya yang sering menjadi konsumsi masyarakat Indonesia. Hal ini menghasilkan pro dan kontra dalam bentuk saran maupun keluhan.

Terkadang, kesamaan informasi ini juga menimbulkan suatu kebudayaan dari si pembawa informasi. Dan dalam jangka panjang, kebudayaan yang terbawa juga tersebar ke beberapa negara yang mana informasi itu masuk. Salah satu aspek yang paling mempengaruhi yaitu aspek budaya suatu negara dengan negara lainnya. Salah satu budaya yang sedang berkembang di Indonesia saat ini adalah gelombang Korea atau yang sering kita dengar dengan istilah Korean Wave. Tersebarnya Korean wave melalui berbagai cara juga dipermudah dengan akses internet dan banyaknya media sosial yang memudahkan siapapun dapat mengakses informasi yang tersedia dalam berbagai bahasa. Keterbatasan akan perbedaan bahasa akhirnya juga dapat diatasi dengan makin banyaknya fans ataupun non-fans yang membuka jasa terjemahan *subtitle* drama Korea ataupun musik k-pop. Dengan begitu, orang-orang yang menaruh minat pada budaya Korea semakin mencintainya.

Analisis sentimen sangat berguna dalam pemantauan media sosial karena dapat memungkinkan untuk mendapatkan *insight* dan gambaran umum opini publik yang lebih luas dibalik topik tertentu. Sentimen cukup mudah untuk dimengerti. Itu merupakan wujud perasaan atau emosi, sikap atau opini. Sentimen positif adalah reaksi atau sikap yang meningkatkan nilai seseorang atau sesuatu, sedangkan sentimen negatif adalah reaksi atau sikap yang menurunkan nilai seseorang atau sesuatu. Di media sosial, sentimen sebuah postingan bisa dilihat dengan nada atau emosi yang disampaikan terhadap sebuah brand. Namun, analisis sentimen pada praktiknya dirasa memiliki beberapa tantangan antara lain, keberagaman bahasa yang digunakan oleh masyarakat Indonesia, yang mengakibatkan masalah sekaligus tantangan terhadap proses klasifikasi tulisan atau opini yang berada di media sosial yang menggunakan bahasa Indonesia. Analisis sentimen jika digunakan tepat pada sasaran, maka dapat digunakan terhadap bisnis dan lebih proaktif dalam melihat perubahan dinamika pasar dan pelanggan. Oleh karena itu, diperlukan alat yang dapat membantu proses analisis sentimen, sehingga analisisnya dapat dilakukan dapat lebih cepat dan efisien. Oleh karena itu, maka dapat diperlukan sebuah solusi berupa analisis sentimen terhadap opini masyarakat yang telah diberikan mengenai gaya hidup anak remaja milenial yang mengutarakan opininya dalam hal konser, drama korea, perokok, dan lain sebagainya yang sering menjadi konsumsi masyarakat Indonesia dengan

mengelompokkan komentar ke dalam suatu kelas positif, netral, dan negatif menggunakan metode Naive Bayes Classifier dan Support Vector Machine.

Naive Bayes adalah suatu metode yang digunakan untuk pengklasifikasian probabilitas *simple* model dan untuk menghitung kumpulan dari probabilitas dengan menjumlahkan seluruh frekuensi serta campuran nilai dari *dataset*. Kelebihan menggunakan Naive Bayes yaitu telah teruji mempunyai akurasi serta kecepatan yang besar jika digunakan dalam *database* dengan informasi yang besar. Metode klasifikasi Naive Bayes juga merupakan salah satu metode yang cukup populer untuk klasifikasi teks dan memiliki performa yang baik. Algoritma SVM juga merupakan salah algoritma *machine learning* yang dikenal cukup baik dalam melakukan klasifikasi berdasarkan pembobotan yang diproses dalam algoritma tersebut. Dalam penelitian ini dilakukan perbandingan model algoritma SVM serta Naive Bayes Classifier untuk melakukan analisis sentimen pada postingan Twitter berdasarkan kata kunci *lifestyle* seperti opini dalam hal konser, film drama, perokok, dan lain sebagainya yang sering menjadi konsumsi masyarakat Indonesia

Tujuan dari penelitian ini adalah untuk membandingkan metode Naive Bayes Classifier dan Support Vector Machine dalam analisis sentimen opini masyarakat Indonesia terhadap topik *lifestyle* melalui Twitter menggunakan bahasa pemrograman python pada Google Colab. Enam tahapan utama yang dilakukan dalam analisis penelitian ini yaitu pengumpulan data, *preprocessing* data, visualisasi, pembobotan, klasifikasi, dan evaluasi. Hasil evaluasi akan menentukan metode mana yang paling baik digunakan dalam analisis sentimen Twitter topik *lifestyle*.

METODOLOGI PENELITIAN

Pengumpulan Data

Penelitian ini menggunakan data primer yang berisi *tweet* dari 11.000 pengguna Twitter dengan topik *lifestyle* yang dikumpulkan selama bulan Februari 2023. Pengumpulan data dilakukan dengan cara *crawling data* pada media sosial Twitter dalam bentuk *tweet*. Kata kunci yang digunakan untuk mengumpulkan *tweet* dengan topik *lifestyle* yaitu konser, merokok, liburan, film, otomotif, dan permainan. Data yang telah terkumpul selanjutnya disimpan dalam bentuk file Excel untuk diolah lebih lanjut.

Preprocessing Data

Proses pengumpulan data teks yang berupa opini dari media sosial seperti Twitter dan Facebook memiliki tantangan tersendiri karena bahasa yang digunakan tidak sesuai dengan kata baku, disingkat, menggunakan bahasa daerah, bahkan kata-kata yang tidak ada di dalam kamus. Tahap awal dalam pengolahan data untuk mengolah data teks yang masih mentah tersebut adalah tahap *preprocessing* data. *Preprocessing* data disebut juga normalisasi, diperlukan untuk mengatasi data teks dengan format yang masih belum sesuai untuk diolah. Tahap ini bertujuan untuk mengembalikan sebanyak mungkin bahasa pesan ke bahasa alami dengan menghilangkan ekspresi atipikal sehingga dapat meminimalkan *noise* pada tahapan selanjutnya. Proses *preprocessing* data dilakukan melalui beberapa tahapan, yaitu *case folding*; penghapusan simbol-simbol atau tanda baca; tokenisasi; konversi *slang word*; penghapusan *stopword*; serta *stemming* (mengubah kata imbuhan menjadi kata dasar).

Word Cloud

Word cloud merupakan salah satu metode visualisasi data yang menampilkan data dalam bentuk susunan kata yang kemunculannya berkaitan dengan frekuensi kata tersebut di dalam data. Di tahap ini, data hasil *preprocessing* akan ditampilkan menggunakan *word cloud* dengan tujuan untuk memudahkan proses analisis. Dengan menggunakan *word cloud*, dapat diketahui kata-kata yang sering muncul dalam data karena semakin sering kata tersebut muncul, maka akan ditampilkan dengan ukuran huruf yang semakin besar.

Term Weighting

Term Weighting merupakan cara atau strategi yang dilakukan untuk menentukan bobot pada suatu kata sehingga dapat meningkatkan kinerja analisis sentimen dan proses teks *mining* lainnya pada kategorisasi teks. Pada tahap ini dilakukan teknik pembobotan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). *Term Frequency* (tf(w,d)) merupakan berapa kali sebuah kata ada didalam sebuah dokumen. Sedangkan ukuran TF-IDF merupakan statistik yang menunjukkan pentingnya sebuah kata tersebut di dalam sebuah *dataset* atau dokumen.

Analisis Sentimen

Menurut Liu, analisis sentimen merupakan suatu bidang studi yang menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, suatu masalah, peristiwa, topik, dan atributnya. Analisis sentimen juga merupakan sebuah bidang penelitian baru di dalam *Natural Language Processing* (NLP) dengan tujuan untuk mendeteksi subjektivitas pada teks dan/atau mengekstraksi dan melakukan klasifikasi terhadap pendapat dan sentimen. Data yang sudah melalui tahap *preprocessing* dan pembobotan selanjutnya dianalisis menggunakan analisis sentimen. Dalam metode klasifikasi sentimen terdapat tiga teknik yaitu *machine learning, lexicon based*, dan *hybrid approach*. Analisis sentimen yang dilakukan pada penelitian ini menggunakan pendekatan *machine learning* karena pendekatan ini digunakan untuk memprediksi polaritas sentimen berdasarkan data yang sudah dilatih sesuai dengan data tes.

Naive Bayes Classifier (NBC)

Pada tahap ini, data hasil *preprocessing* dan pembobotan akan dibagi (*splitt*) menjadi dua yaitu, data *training* dan data *testing* dengan bantuan Google Colab. Data *training* digunakan dalam melatih model yang akan dibuat dengan menggunakan metode *naive bayes*. *Naive Bayes Classifier* (NBC) merupakan salah satu metode pembelajaran probabilitas pada *machine learning* yang beranggapan bahwa setiap kata terjadi secara independen. NBC memodelkan distribusi dokumen di setiap kelas menggunakan sebuah model probabilitas dengan asumsi independensi tentang distribusi dari kata-kata yang berbeda. Metode ini didasari oleh Thomas Bayes, seorang ilmuwan asal Inggris yang memperkenalkan Teorema Bayes yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa lalu. Dalam penerapannya pada analisis sentimen, pengalaman dimasa lalu dianalogikan sebagai data *training* dan masa depan sebagai data *testing*.

Support Vector Machine (SVM)

Selain menggunakan metode NBC, pemodelan data juga dilakukan dengan menggunakan metode *Support Vector Machine* (SVM). SVM merupakan salah satu

teknik pembelajaran *supervised* dengan kualitas dan akurasi baik yang membuatnya menjadi algoritma populer di antara algoritma lainnya. "Klasifikasi SVM berupaya untuk mempartisi ruang data dengan menggunakan penggambaran linier atau non-linier antara kelas yang berbeda". Dalam geometri, SVM dapat dilihat sebagai *hyperplane* pada fitur ruang yang memisahkan titik-titik yang mewakili kategori dari hal positif dan dari titik-titik yang mewakili hal negatif. Prinsip utama dari SVM adalah untuk menentukan ruang pemisah di ruang pencarian yang dapat memisahkan kelas-kelas yang berbeda. Proses SVM dimulai dari mengubah data *text* ke dalam bentuk vektor data dan dikombinasikan dengan nilai TF-IDF untuk pembobotan.

Evaluasi

Evaluasi dilakukan untuk mengetahui performa yang paling baik di antara kedua model yang sudah dibentuk. Tahap ini dilakukan dengan cara membandingkan nilai akurasi dan ketepatan prediksi model. Model dengan performa paling baik adalah model yang memiliki nilai akurasi tertinggi di antara kedua model dan menghasilkan prediksi sentimen yang sesuai ketika data dimasukkan ke dalam model.

HASIL DAN PEMBAHASAN

Dalam penelitian ini pengumpulan data dilakukan dengan cara *crawling data* dari media sosial Twitter. Data yang diambil yaitu kumpulan *tweet* berbahasa Indonesia selama bulan Februari 2023 dengan kata kunci konser, merokok, liburan, film, otomotif, dan permainan. Data yang berhasil diambil yaitu sebanyak 11.000 data yang disimpan dalam bentuk file Excel.

Data yang telah dikumpulkan diberi label untuk menentukan sentimennya. Sentimen yang digunakan dalam penelitian ini adalah sentimen positif, negatif, dan netral. Sentimen positif diberikan pada data yang berisi ungkapan positif dimana ungkapan tersebut dapat meningkatkan nilai seseorang atau sesuatu. Sebaliknya, pada sentimen negatif, data berisi ungkapan negatif yang dapat menurunkan nilai seseorang atau sesuatu. Sedangkan, sentimen netral diberikan pada data yang berisi ungkapan yang tidak bersifat positif maupun negatif. Hasil dari tahap ini ditunjukkan melalui Tabel 1 dengan sentimen positif berjumlah 4134 *tweet*, sentimen negatif berjumlah 3646 *tweet*, dan sentimen netral berjumlah 3220 *tweet*.

Tabel 1. Data Hasil Pelabelan

Clean Tweet	Sentimen
Aneh bgt ini konser aneh bgt masa project bisa beda2 per section	negatif
Oh dia pengen permainan ini cepat selesai ternyata	netral
⋮	⋮
Maunya ikut apa Konser rock n roll	netral

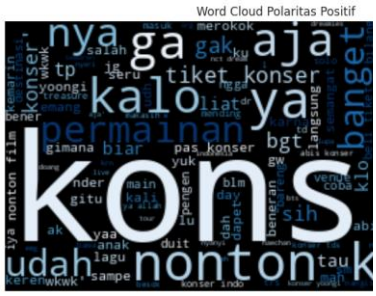
Sebelum dilakukan pengolahan data, diperlukan tahapan *preprocessing* untuk mempersiapkan data agar sesuai dengan kebutuhan analisis sehingga dapat

menghasilkan hasil analisis yang baik. Tahapan *preprocessing* yang dilakukan dalam penelitian ini yaitu *case folding*, *tokenizing*, *stopword removal*, *filtering*, dan *lemmatization*. *Case folding* adalah tahap pengubahan ukuran huruf yang semula berukuran besar (*uppercase*) menjadi huruf berukuran kecil (*lowercase*). *Tokenizing* adalah tahap memecah sebuah kalimat yang terdiri dari sekumpulan kata menjadi kata-kata terpisah yang berdiri sendiri. *Stopword removal* adalah tahap penghilangan atau penghapusan kata hubung (konjungsi). *Filtering* merupakan tahap yang dilakukan untuk mengambil kata yang sifatnya dianggap penting di dalam sebuah kalimat. *Lemmatization* adalah tahap perubahan kata dari kata berimbuhan menjadi kata dasar. Data yang telah melalui tahap *preprocessing* dapat digunakan untuk pengolahan data lebih lanjut. Hasil dari tahap ini ditunjukkan melalui Tabel 2.

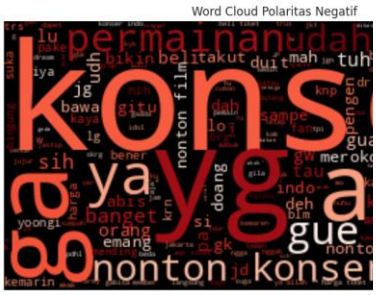
Tabel 2. Data Hasil *Preprocessing*

Clean Tweet	Sentimen
aneh bgt konser aneh bgt project beda2 section	negatif
oh pengen permainan cepat selesai	netral
⋮	⋮
maunya konser rock n roll	netral

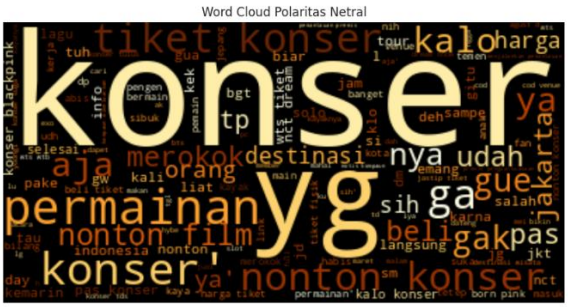
Visualisasi data dengan menggunakan *word cloud* bertujuan untuk mengetahui kata-kata yang sering muncul pada suatu data. *Word cloud* merupakan suatu gambar yang terdiri dari kumpulan kata, dimana besarnya kata merepresentasikan frekuensi kemunculan atau tingkat kepentingan kata tersebut. Semakin besar suatu kata muncul, maka semakin sering kata tersebut disebutkan dalam suatu dokumen teks. Pada penelitian ini *word cloud* dari ketiga sentimen ditunjukkan melalui gambar berikut.



Gambar 1. *Word Cloud* Sentimen Positif



Gambar 2. *Word Cloud* Sentimen Negatif



Gambar 3. *Word Cloud* Sentimen Netral

Dari ketiga *word cloud* di atas diketahui bahwa kata "konser" menjadi kata yang paling sering muncul pada ketiga sentimen ditunjukkan oleh besarnya kata pada *word cloud*. Hal ini berarti bahwa pada bulan Februari 2023 banyak pengguna Twitter yang sedang membahas konser yang akan diadakan di Indonesia dalam beberapa bulan ke depan. Meskipun sekilas terlihat sama, terdapat beberapa perbedaan pada ketiga *word cloud* tersebut. Pada *word cloud* sentimen positif, kata yang paling sering muncul selain kata "konser" adalah kata "nonton" yang mengindikasikan bahwa pengguna Twitter banyak yang sedang membahas rencana untuk menonton konser yang akan datang. Pada *word cloud* sentimen negatif, kata yang paling sering muncul selain kata "konser" adalah kata "ga" yang mengindikasikan bahwa pengguna Twitter banyak yang sedang membahas ketidakmampuannya untuk bisa menonton konser dikarenakan beberapa alasan. Sedangkan pada *word cloud* sentimen netral, kata yang paling sering muncul selain kata "konser" adalah kata "permainan" yang mengindikasikan bahwa pengguna Twitter banyak yang sedang membahas mengenai konser dan juga permainan sepak bola dari klub yang beberapa waktu lalu baru saja bertanding.

Pembobotan atau vektorisasi dilakukan untuk menentukan bobot atau nilai suatu kata pada data sehingga kata tersebut dapat diolah dan dicari polanya dalam tahap klasifikasi. Pada penelitian ini, pembobotan dilakukan dengan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) yaitu dengan menganalisis sampel data kemudian melakukan perhitungan pada data tersebut. Sebelumnya, telah dilakukan *splitting* atau pembagian data menjadi dua, yaitu data *training* dan data *testing*. Pembagian data ini dilakukan dengan bantuan Google Colab untuk memisahkan data yang akan digunakan untuk melatih (*training*) dan menguji (*testing*) model yang dibentuk.

Setelah dilakukan pembobotan dan *splitting*, tahap berikutnya yaitu membentuk model untuk melakukan analisis sentimen. Model yang dibentuk pada penelitian ini menggunakan metode Naive Bayes Classifier dan Support Vector Machine (SVM). Pembentukan model dilakukan dengan menggunakan *library python* sklearn yang ada pada Google Colab. Data *training* dihitung menggunakan kedua model hingga diperoleh hasil ketepatan klasifikasi yang memuat nilai akurasi. Nilai akurasi digunakan sebagai salah satu bahan pembanding performa antara kedua model. Proses penghitungan pada data *training* ini akan menjadi sebuah sarana pelatihan bagi model yang nantinya akan menjadi acuan model dalam menentukan hasil klasifikasi data *testing*. Hasil ketepatan klasifikasi analisis sentimen pada model dengan metode Naive Bayes Classifier diperoleh akurasi

sebesar 59%. Sedangkan, hasil ketepatan klasifikasi pada model dengan metode SVM diperoleh akurasi sebesar 61%. Perbandingan hasil ketepatan klasifikasi kedua model ditunjukkan melalui gambar di bawah ini.

	precision	recall	f1-score	support
negatif	0.59	0.70	0.64	1093
netral	0.75	0.31	0.43	983
positif	0.55	0.73	0.63	1224
accuracy			0.59	3300
macro avg	0.63	0.58	0.57	3300
weighted avg	0.62	0.59	0.57	3300

Gambar 4. Hasil Ketepatan Klasifikasi Model Naive Bayes Classifier

	precision	recall	f1-score	support
negatif	0.61	0.66	0.63	1093
netral	0.59	0.57	0.58	983
positif	0.62	0.59	0.60	1224
accuracy			0.61	3300
macro avg	0.61	0.61	0.61	3300
weighted avg	0.61	0.61	0.61	3300

Gambar 5. Hasil Ketepatan Klasifikasi Model Support Vector Machine (SVM)

Tahap akhir dari proses analisis pada penelitian ini yaitu melakukan prediksi menggunakan kedua model yang telah dibentuk. Prediksi dilakukan dengan cara memasukkan data *testing* ke dalam model. Data *testing* akan dihitung menggunakan kedua model yang sebelumnya sudah dilatih hingga diperoleh hasil prediksi berupa sentimen data tersebut. Hasil prediksi kedua model ditunjukkan melalui gambar berikut ini.

```
data = ["Semangat cari duit yuk mas agus mau konser di jakarta 3 bulan la
vect = tfidf.transform(data).toarray()

my_prediction = clf.predict(vect)
print(my_prediction)
```

```
['positif']
```

Gambar 6. Hasil Prediksi Model Naive Bayes Classifier

```
data = ["capek bgt nonton film hush"]
vect = cv.transform(data).toarray()

my_prediction = clf.predict(vect)
print(my_prediction)
```

```
['negatif']
```

Gambar 7. Hasil Prediksi Model Support Vector Machine (SVM)

Perbandingan hasil prediksi menunjukkan bahwa kedua model sama-sama mampu melakukan prediksi data secara tepat. Namun, jika membandingkan nilai akurasi pada hasil ketepatan klasifikasi menunjukkan bahwa performa model

dengan metode SVM lebih baik dibandingkan model dengan metode Naive Bayes Classifier. Oleh karena itu, dari kedua model yang dibentuk pada penelitian ini, model dengan metode Support Vector Machine (SVM) menjadi model terbaik dalam melakukan analisis sentimen data Twitter topik *lifestyle*.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan diperoleh beberapa kesimpulan sebagai berikut.

1. Hasil ketepatan klasifikasi analisis sentimen dengan menggunakan metode Naive Bayes Classifier pada data Twitter topik *lifestyle* diperoleh akurasi sebesar 59%
2. Hasil ketepatan klasifikasi analisis sentimen dengan menggunakan metode Support Vector Machine pada data Twitter topik *lifestyle* diperoleh akurasi sebesar 61%
3. Metode Naive Bayes Classifier dan Support Vector Machine mampu melakukan prediksi pada analisis sentimen data Twitter topik *lifestyle* secara tepat.
4. Perbandingan performa dari kedua metode tersebut menunjukkan hasil bahwa performa metode Support Vector Machine (SVM) dalam analisis sentimen data Twitter topik *lifestyle* lebih baik dibandingkan dengan metode Naive Bayes Classifier.
5. Penelitian selanjutnya diharapkan dapat melakukan analisis yang lebih dalam dengan menggunakan data yang lebih banyak untuk meningkatkan keakuratan prediksi sentimen mengingat nilai akurasi yang dihasilkan model terbaik dalam penelitian ini masih tergolong rendah.

DAFTAR PUSTAKA

Fhutuh, I. Analisis sentimen K-popers terhadap dunia hiburan tanah air dengan algoritma Convolutional Neural Network (CNN). Bandung: Doctoral dissertation, Universitas Islam Negri Sunan Gunung Djati, 2022.

Naufal, M. F., Arifin, T., & Wirjawan, H. Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan NaiveBayes untuk Klasifikasi Cyberbullying pada Media Sosial. Jurasik (Jurnal Riset Sistem Informasi dan Teknik Informatika), 8(1), 82-90, 2023.

Amelia, R., Darmansah, D., Prastiwi, N. S., & Purbaya, M. E. Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter. JURIKOM (Jurnal Riset Komputer), 9(2), 338-343, 2022.

Larasati, M. A. Z., Winarsih, N. A. S., Rohman, M. S., & Saraswati, G. W. Penerapan Metode K-Means Clustering Dalam Menganalisis Sentimen Masyarakat Terhadap K-Popers Pada Twitter. Progresif: Jurnal Ilmiah Komputer, 18(2), 201-210, 2022.

Setiyawati, D., & Cahyono, N. Analisis Sentimen Pengguna Sosial Media Twitter Terhadap Perokok Di Indonesia. Indonesian Journal of Computer Science, 12(1), 2023.

Fitri, E. Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine. Jurnal Transformatika, 18(1), 71-80, 2020.

Fitriana, F., Utami, E., & Al Fatta, H. Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. Jurnal Komtika (Komputasi Dan Informatika), *5*(1), 19-25, 2021.

Tuhuteru, H., & Iriani, A. Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier. Jurnal Informatika, *3*(03), 2018.