



ANALISIS SENTIMEN PRODUK *MIXUE* DENGAN METODE *NAÏVE BAYES CLASSIFIER* DAN *SUPPORT VECTOR MACHINE*(SVM) PADA MEDIA SOSIAL X

Abdi Muhaimin¹, Bambang Kurniawan²

^{1,2}Universitas Hang Tuah Pekanbaru

^{1,2}Jl. Mustafa Sari No.5, Pekanbaru, 28288

E-mail : abdi.muhaimin86@gmail.com

ABSTRAK

Media sosial saat ini memiliki peran penting dalam perkembangan sebuah bisnis dan sebuah organisasi. Media sosial seperti X dapat mempengaruhi persepsi orang terhadap suatu orang, bisnis, ataupun organisasi. Sebuah perusahaan yang bernama Mixue Ice Cream & Tea menjadi topik perbincangan di media sosial dikarenakan ekspansi yang mereka lakukan di Indonesia. Tetapi, Mixue Ice Cream & Tea tidak diterima langsung oleh orang Indonesia dikarenakan caranya ekspansidan pengiklanan yang kontroversial. Namun akan sangat sulit untuk mendeteksi emosi atau sentimen dari pengguna media sosial X dengan jumlah yang sangat banyak. Penelitian ini bertujuan untuk melakukan analisis sentimen yaitu untuk menganalisis sentimen dari pengguna media sosial X termasuk sentimen positif, negatif dan netral. Penelitian ini menggunakan 2 metode klasifikasi yaitu Metode Naïve Bayes Classifier dan Support Vector Machine. Penelitian ini juga membandingkan akurasi kedua metode tersebut dengan melakukan k-fold cross validation dengan nilai k = 10. Hasil dari perbandingan kedua metode menunjukkan bahwa Support Vector Machinememiliki performa paling baik dengan nilai akurasi sebesar 73% dari pada Naïve Bayes Classifier dengan nilai akurasi sebesar 61%.

Kata Kunci: Analisis Sentimen, Mixue, X, Naïve Bayes, Support VectorMachine

ABSTRACT

Social media currently plays a crucial role in the development of businesses and organizations. Social media platforms like X can influence people's perceptions of individuals, businesses, or organizations. A company called Mixue Ice Cream & Tea became a hot topic on social media due to its expansion in Indonesia. However, Mixue Ice Cream & Tea was not immediately well-received by Indonesians due to its controversial expansion and advertising methods. However, detecting the emotions or sentiments of the vast number of users on social media platform X is very difficult. This study aims to conduct sentiment analysis, analyzing the sentiments of social media platform X users, including positive, negative, and neutral sentiments. This study used two classification methods: the Naïve Bayes Classifier and the Support Vector Machine. This study also compared the accuracy of the two methods using k-fold cross-validation with a k value of 10. The results of the comparison of the two methods showed that the Support Vector Machine performed better with an accuracy of 73%, compared to the Naïve Bayes Classifier with an accuracy of 61%.

Keywords: Sentiment Analysis, Mixue, X, Naïve Bayes, Support Vector Machine

1. PENDAHULUAN

Perkembangan teknologi digital yang sangat pesat telah mengubah pola interaksi dan komunikasi masyarakat. Media sosial, khususnya platform X (sebelumnya *Twitter*), memainkan peran penting sebagai sarana penyampaian opini publik secara *real-time*. Menurut laporan *We Are Social*, pengguna media sosial X mencapai 556 juta secara global pada awal tahun 2023, dengan

24 juta pengguna berasal dari Indonesia (Q. A. Xu et al., 2022). Data ini menunjukkan potensi besar media sosial X sebagai sumber informasi untuk menganalisis opini publik terhadap berbagai topik, termasuk produk komersial.

Mixue Ice Cream & Tea, yang berasal dari Tiongkok dan mulai berekspansi ke Indonesia sejak tahun 2021, menjadi topik perbincangan di media sosial karena popularitasnya yang tinggi dan strategi



pemasaran yang agresif. Namun, ekspansi ini juga menimbulkan berbagai respons dari masyarakat, baik positif maupun negatif. Untuk itu, diperlukan suatu pendekatan yang sistematis untuk memahami persepsi publik terhadap brand *Mixue* secara objektif.

Analisis sentimen merupakan metode yang umum digunakan untuk mengidentifikasi dan mengklasifikasikan opini publik dari data teks. Metode ini mengkategorikan opini menjadi sentimen positif, negatif, atau netral. Dalam penelitian ini digunakan dua metode machine learning yaitu *Naïve Bayes Classifier* dan *Support Vector Machine*(SVM), yang telah terbukti efektif dalam pengolahan teks (Maryati, U. H, 2024).

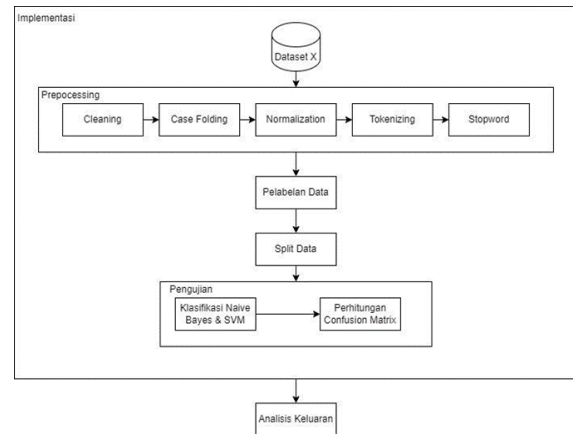
Salah satu metode penelitian yang paling sering digunakan untuk mengetahui pikiran dan emosi orang dari data teks adalah analisis sentimen (Leelawat et al., 2022). Pengguna internet menggunakan analisis sentimen di media sosial untuk berbagi pendapat atau penilaian pribadi mereka. Dengan demikian, analisis sentimen pada subjek *Mixue* memungkinkan Anda untuk melihat bagaimana perasaan publik tentang *Mixue*. Studi ini membandingkan akurasi dua metode: *Naïve Bayes Classifier* dan algoritma *Support Vector Machine*(SVM). Salah satu teknik yang sangat efektif dan sering digunakan dalam analisis sentimen adalah pendekatan *Naïve Bayes Classifier* (Fitri et al., 2019). Sementara itu, *Supervised Learning* (SVM) memanfaatkan algoritma pembelajaran untuk menganalisis data dalam konteks analisis regresi. Emosi dan sentimen pengguna dapat terdeteksi dari *tweet* dan balasannya dan membentuk sebuah jaringan emosi. Dari jaringan emosi ini, emosi dapat dideteksi sebagai emosi positif dan negatif. Metode ini melakukan tugas klasifikasi dengan memaksimalkan margin yang memisahkan kelas sambil meminimalkan kesalahan klasifikasi (Ahmad et al., 2017).

Beberapa penelitian mengatakan metode *Support Vector Machine*(SVM) memiliki performa yang terbaik jika diukur dari akurasi, tetapi tidak sedikit juga penelitian yang mengatakan metode *Naïve Bayes Classifier* memiliki akurasi yang lebih tinggi.

2. METODE PENELITIAN

Metode penelitian bertujuan untuk mengungkap kebenaran secara ilmiah dengan mengumpulkan fakta secara sistematis, ilmiah, dan rasional. Tujuannya adalah untuk menemukan dan memecahkan masalah serta mengembangkan pengetahuan baru yang kemudian menghasilkan kebenaran ilmiah. Kebenaran ilmiah adalah hasil

penelitian yang akan terus diuji, dikritik, dan bahkan direvisi. Dalam karya ini, penulis menerapkan tahap pra-pemrosesan bersama dengan metode *Naïve Bayes Classifier* dan *Support Vector Machine* untuk melakukan analisis sentimen.



Gambar 1. Metode Implementasi

Berikut adalah penjelasan dari metode implementasi :

1. Dataset Twitter

Data dari media sosial X digunakan dalam dataset untuk penelitian ini. Data yang digunakan adalah *tweets* atau komentar mengenai *Mixue Ice Cream & Tea* menggunakan kata kunci “*Mixue*”. Bahasa pemrograman *Python* dan modul *snsrcape* yang banyak digunakan digunakan dalam strategi perayapan data untuk mengumpulkan data dari X. Dataset lalu disimpan ke dalam format CSV yang akan dilakukan split data lalu akan masuk ke dalam proses *pre-processing*.

2. Pre-processing

Tahap *pre-processing* bertujuan untuk menghilangkan noise yang tidak diperlukan dalam dataset. Proses ini mirip dengan tahap pengolahan data. Dalam penelitian ini, tahap pra-pemrosesan meliputi *Cleaning*, *Case Folding*, *Normalization*, *Tokenization*, dan *Stopword*. Proses-proses ini dilakukan dengan menggunakan *Google Colab*.

3. Split Data

Tahap berikutnya adalah membagi dataset menjadi dua bagian: data latih dan data uji, yang akan digunakan untuk pelabelan sentimen. Pelabelan sentimen dilakukan secara otomatis dengan menggunakan 30% dari data latih dan 70% dari data uji.

4. Klasifikasi dengan *Naïve Bayes Classifier* dan *Support Vector Machine*.

Setelah tahap *pre-processing*, langkah berikutnya adalah pengujian data, yang melibatkan klasifikasi menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine*. Proses



klasifikasi ini akan dilakukan dengan bantuan *Google Colab* untuk mempermudah pelaksanaannya.

5. Perhitungan *Confusion Matrix*

Hasil klasifikasi tersebut akan masuk ke tahap perhitungan *confusion matrix*, dimana tahap ini adalah tahap untuk menghitung keakuratan, presisi, dan *recall* dari kedua metode yang telah diuji.

6. Analisis Keluaran

Hasil dari analisis sentimen ini akan dianalisis untuk mempermudah dalam interpretasi data. Hasil analisis akan berupa perbandingan dari kedua metode yang telah digunakan yang akan menjadi kesimpulan akhir dari penelitian ini.

3. HASIL DAN PEMBAHASAN

Data penelitian ini diperoleh dari media sosial X. Data yang digunakan meliputi *tweets* atau komentar tentang *Mixue Ice Cream & Tea* dengan kata kunci "*Mixue*". Pengambilan data dilakukan dengan teknik *crawling* menggunakan bahasa pemrograman *Python* dan *library snsrape*, yang sering digunakan untuk mengumpulkan data dari X. Total data yang diambil selama periode yang telah ditentukan sebanyak 4678 data *tweets*. Dataset lalu disimpan kedalam format CSV yang akan dilakukan split data lalu akan masuk ke dalam proses *pre-processing*. Berikut adalah tahapan dari *pre-processing*.

1. *Pre-processing* Data

```

full_text
0      Bener2 kek di hospital playlist Ga salah kalo sakit berobatnya disini
1  @kegb1gnunfaedh Mana gua kalo kemana Mkn sendiri takut du jadiin bahan bersukur
2      Bm liburan
3      Gimana enggak yaa kerjaan minum mixue lemon tea + kopi dalam sehari tuh
4      Minta buatin outlet mixue
    
```

Gambar 2. Hasil *Split* Data

Pada tahapan ini, dataset yang telah dilakukan *crawling* akan dimasukkan kedalam *google collab* untuk dilakukan pengolahan data. Tahapan ini bertujuan agar *machine learning* dapat mudah mengenali pola sehingga dapat mempermudah mempelajari dataset tersebut.

2. *Cleaning*

Proses *cleaning* merupakan proses membersihkan data yang telah di kumpulkan. Komponen yang dibersihkan yaitu komponen yang tidak relevan atau tidak berpengaruh untuk proses klasifikasi data. Seperti, komponen angka, tanda baca, *hashtag*, *mention*, *link*, dan emoji. Selain itu,

pada tahap *cleaning* juga membersihkan data *duplicate*.

```

full_text
0      Bener kek hospital playlist salah kalo sakit berobatnya disini
1      Mana gua kalo kemana Mkn sendiri takut jadiin bahan bersukur
2      liburan
3      Gimana enggak yaa kerjaan minum mixue lemon tea kopi dalam sehari tuh
4      Minta buatin outlet mixue
...
4673     aku adaa aja nderr
4674     JKT JKT WKWKWKWK MIXUE
4675     aku ada apart louvin nder
4676     paha bekas lukanya gede bgt soalnya dijahit itu grgr jatuh pas main bola trs kemasukan kerikil memang bokem
4677     ayo bantu aku nemu pacar kek gini
4678 rows x 1 columns
    
```

Gambar 3. Hasil *Cleaning*

3. *Case Folding*

Tahap *case folding* adalah bagian dari *text preprocessing* yang bertujuan untuk mengubah semua huruf besar menjadi huruf kecil (*lowercase*) di seluruh dokumen.

```

full_text
0      bener kek hospital playlist salah kalo sakit berobatnya disini
1      mana gua kalo kemana mkn sendiri takut jadiin bahan bersukur
2      liburan
3      gimana enggak yaa kerjaan minum mixue lemon tea kopi dalam sehari tuh
4      minta buatin outlet mixue
...
4673     aku adaa aja nderr
4674     jkt jkt wkwkwkwk mixue
4675     aku ada apart louvin nder
4676     paha bekas lukanya gede bgt soalnya dijahit itu grgr jatuh pas main bola trs kemasukan kerikil memang bokem
4677     ayo bantu aku nemu pacar kek gini
4678 rows x 1 columns
    
```

Gambar 4. Hasil *Case Folding*

4. *Normalization*

Pada tahap ini, tujuannya adalah untuk memperbaiki kata-kata non-standar, singkatan, atau bahasa slang menjadi bentuk baku sesuai dengan KBBI. Selain itu, normalisasi bertujuan untuk menyamakan penulisan kata-kata yang memiliki makna yang sama tetapi ditulis secara berbeda. Proses normalisasi melibatkan beberapa langkah yang perlu dilakukan:

1. Menyiapkan terlebih dahulu kamus dataset normalisasi yang akan digunakan sebagai kamus normalisasi.
2. Melakukan proses normalisasi dengan mengubah kata yang tidak baku atausingkatan dengan berdasar dari kamus yang telah diinput sebelumnya.



	full_text
0	bener seperti hospital playlist salah sakit berobatnya disini
1	mana saya kalau kemana makan sendiri takut jadiin bahan bersukur
2	liburan
3	bagaimana enggak yaa kerjaan minum mixue lemon tea kopi dalam sehari tuh
4	mintak buatin outlet mixue
...	...
4673	saya adaa saja nderr
4674	jakarta jakarta wkwkwkwk mixue
4675	saya ada apart louvin nder
4676	paha bekas lukanya besar banget soalnya dijahit itu grgr jatuh pas main bola terus kemasukan kerikil memang bokem
4677	ayo bantu saya menemukan pacar seperti begini

4678 rows x 1 columns

Gambar 5. Hasil Normalization

5. Tokenizing

Langkah berikutnya adalah *tokenizing*, yaitu proses memecah teks menjadi kata, frasa, atau bagian yang bermakna untuk mengetahui asal-usul kata. Proses ini juga dikenal sebagai segmentasi teks. *Tokenizing* dilakukan dengan menggunakan operator *Process Documents from Data* yang mengaplikasikan vektorisasi TF-IDF, serta mencakup tahap pre-processing seperti *Case Folding*, *Cleansing*, dan *Normalization*. Hasil dari proses ini berhasil memecah seluruh teks menjadi 3.174 atribut kata. Berikut adalah gambar yang menunjukkan proses dan hasil tokenisasi pada dataset.

	full_text
0	[bener, seperti, hospital, playlist, salah, sakit, berobatnya, disini]
1	[mana, saya, kalau, kemana, makan, sendiri, takut, jadiin, bahan, bersukur]
2	[liburan]
3	[bagaimana, enggak, yaa, kerjaan, minum, mixue, lemon, tea, kopi, dalam, sehari, tuh]
4	[mintak, buatin, outlet, mixue]
...	...
4673	[saya, adaa, saja, nderr]
4674	[jakarta, jakarta, wkwkwkwk, mixue]
4675	[saya, ada, apart, louvin, nder]
4676	[paha, bekas, lukanya, besar, banget, soalnya, dijahit, itu, grgr, jatuh, pas, main, bola, terus, kemasukan, kerikil, memang, bokem]
4677	[ayo, bantu, saya, menemukan, pacar, seperti, begini]

4678 rows x 1 columns

Gambar 6. Hasil Tokenizing

6. Stopword Removal

Tahapan ini melibatkan *stopword removal*, yaitu menghapus kata-kata yang tidak mempengaruhi sentimen, seperti konjungsi, preposisi, dan artikel. Pada penelitian ini, kamus atau *corpus stopwords* yang digunakan adalah *stopword* berbahasa Indonesia yang berasal dari Kaggle (Indonesian Stoplist | Kaggle, n.d.). Pada proses ini juga dilakukan *filtering* kata berdasarkan jumlah karakter sehingga kata yang dihasilkan dataset tidak memiliki kata yang terlalu panjang atau pendek. Filter yang digunakan adalah penghapusan kata kecuali kata dengan minimal 4 karakter dan maksimal 25 karakter.

	full_text
0	bener hospital playlist salah sakit berobatnya
1	kemana makan takut jadiin bahan bersukur
2	liburan
3	yaa kerjaan minum mixue lemon tea kopi sehari tuh
4	mintak buatin outlet mixue
...	...
4673	adaa nderr
4674	jakarta jakarta wkwkwkwk mixue
4675	apart louvin nder
4676	paha bekas lukanya banget dijahit grgr jatuh pas main bola kemasukan kerikil bokem
4677	ayo bantu menemukan pacar

[4678 rows x 1 columns]

Gambar 7. Hasil Stopwards

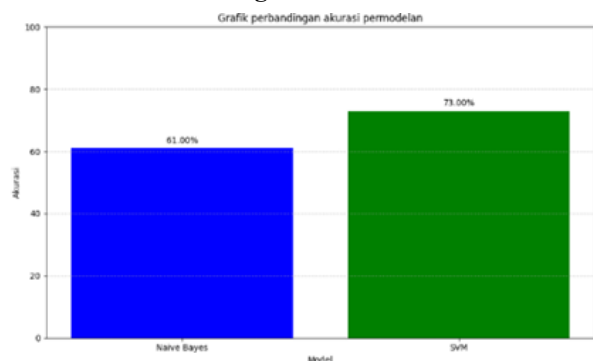
7. Pelabelan Dataset

Tahapan ini adalah melakukan pelabelan secara manual dan otomatis. Pelabelan dilakukan untuk menentukan data tersebut termasuk sentimen positif, negatif atau netral yang nantinya akan masuk ke dalam pengujian data. Pelabelan ini bertujuan untuk menentukan data latih sehingga dapat dilakukannya pengujian terhadap dataset. Pelabelan manual dilakukan dengan melakukan pelabelan kelas sentimen melalui pendapat sendiri atau ahli bahasa.

Tabel 1 . Pelabelan Dataset

No	Mention	Sentimen
1.	bener hospital playlist salah sakit berobatnya	Negative
2.	kemana makan takut jadiin bahan bersukur	Negative
3.	yaa kerjaan minum <i>Mixue</i> lemon tea kopi sehari tuh	Positive
4.	mintak buatin outlet <i>Mixue</i>	Netral
5.	jakarta jakarta wkwkwkwk <i>Mixue</i>	Netral
...	...	
4677.	ayo bantu menemukan pacar	Positif

8. Grafik Perbandingan Model



Gambar 8. Grafik Perbandingan

ini menunjukkan performa dari dua model yang berbeda dalam analisis sentimen terhadap produk *Mixue*. Model *Naive Bayes Classifier* dan SVM (*Support Vector Machine*) dasar memiliki akurasi sebesar 61% dan 73%, yang menunjukkan kemampuan dasar model dalam mengklasifikasikan



sentimen dengan cukup baik.

Hasil analisis menunjukkan bahwa *Naïve Bayes Classifier* menghasilkan akurasi 61%, dengan kelemahan dalam mengenali sentimen negatif akibat distribusi data yang tidak seimbang. Sementara itu, *Support Vector Machine* mencapai akurasi 73% dengan presisi dan F1-score lebih tinggi. *Confusion matrix* menunjukkan SVM mampu memprediksi kelas positif dan netral lebih baik dibandingkan *Naïve Bayes*.

4. KESIMPULAN

Berdasarkan rumusan masalah dan hasil penelitian yang telah dilakukan, kesimpulan yang dapat diambil adalah sebagai berikut:

1. Hasil analisis sentimen menggunakan *Naïve Bayes Classifier* memiliki nilai akurasi sebesar 61,00% dengan nilai presisi untuk prediksi positif sebesar 82,00% , presisi untuk prediksi negatif sebesar 53,00% dan presisi untuk prediksi netral sebesar 77,00%, dengan nilai *recall* untuk data positif sebesar 44,00% dan *recall* untuk data negatif sebesar 94,00%, dan *recall* untuk data netral sebesar 31,00%, Sedangkan untuk hasil analisis sentimen menggunakan *Support Vector Machine* memiliki nilai akurasi sebesar 73,00% dengan nilai presisi untuk prediksi positif sebesar 79,00% , presisi untuk prediksi negatif sebesar 72,00%, dan presisi untuk prediksi netral sebesar 69,00%, dengan nilai *recall* untuk data positif sebesar 63,00,% , *recall* untuk data negatif sebesar 79,00%, dan *recall* untuk data netral sebesar 73,00%.
2. Hasil Perbandingan menunjukkan bahwa metode *Support Vector Machine* memiliki performa terbaik dibandingkan dengan metode *Naïve Bayes Classifier*, dengan akurasi dan *recall* yang lebih tinggi, terutama dalam mendeteksi sentimen negatif.

5. REFERENSI

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan *Support Vector Machine* pada analisis

sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71-76.

Kirana, A. S., Rusdah, R., Roeswidiah, R., & Pudoli, A. (2025). ANALISIS SENTIMEN PADA MEDIA SOSIAL TERHADAP LAYANAN SAMSAT DIGITAL NASIONAL DENGAN SUPPORT VECTOR MACHINE. *IDEALIS: InDonEsiA journal Information System*, 8(1), 53-63.

Mauliza, R. N., & Sipayung, Y. R. (2024). Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes. *vol, 9*, 1-16.

MARYATI, U. H. (2024). *Perbandingan Klasifikasi Algoritma Naïve Bayes Classifier Dan Support Vector Machine(Svm) Terhadap Sentimen Analisis Event Piala Afc U-23 Tahun 2024 Pada Media Sosial Twitter (X)* (Doctoral Dissertation, Universitas Mercu Buana-Menteng).

Mufidah, F. S., Winarno, S., Alzami, F., Udayanti, E. D., & Sani, R. R. (2022). Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma *Naïve Bayes Classifier*. *JOINS (Journal of Information System)*, 7(1), 14-25

Susilawati, P. (2025). *Analisis Peran Akun Media Sosial Instagram@ Smansaku Sebagai Sarana Diseminasi Informasi Kepada Siswa Sma Negeri 1 Tegal* (Doctoral dissertation, Universitas Pancasakti Tegal).

Tarigan, V. T., & Yusupa, A. (2024). Perbandingan Algoritma Maching Learning dalam Analisis Sentimen Mobil Listrik di Indonesia pada Media Sosial Twitter/X. *Jurnal Informatika Polinema*, 10(4), 479-490.

Wijonarko, P., & Putri, E. L. (2024, November). Penerapan Machine Learning dalam Analisis Sentimen terhadap Ridwan Kamil sebagai Kandidat Gubernur DKI Jakarta. In *SEMINAR NASIONAL KONSORSIUM UNTAG SE INDONESIA* (pp. 69-78).