

Emotion Recognition Tentang Ditutupnya TikTok Shop Menggunakan Metode Naive Bayes

Rama Ariya Candra

Universitas Pembangunan Nasional “Veteran” Jawa Timur

Alamat: Jl. Rungkut Madya No. 1 Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur

Korespondensi penulis: ramacandra46@gmail.com

Abstract. *The policy of shutting down TikTok Shop has sparked both pros and cons. On one side, it eliminates jobs for content creators whose income relies on TikTok Shop, while on the other side, it saves UMKM from predatory pricing wars that harm them. Utilizing the Naive Bayes algorithm, a classification method capable of predicting the likelihood of a class and making decisions based on learning data, the Emotion Recognition research on YouTube comments related to the closure of TikTok Shop is conducted. Data will be classified into five classes: happy, angry, sad, afraid, and surprised. The objective of this research is to find the best emotional model using the Naive Bayes method. The results of user testing with Naive Bayes and Tf-Idf show that the precision values for sad, happy, afraid, and surprised emotions are high, while for anger, the percentage is 59%. The percentages for afraid, happy, sad, and surprised emotions are 91%, 87%, 84%, and 79%, respectively. The overall accuracy is 82%.*

Keywords: *Naive Bayes, Emotion Recognition, the closure of TikTokshop, TF-IDF*

Abstrak. Kebijakan ditutupnya TikTok Shop menimbulkan pro dan kontra, satu sisi menghilangkan pekerjaan para konten kreator yang pendapatannya berasal dari TikTok Shop dan sisi lainnya menyelamatkan para umkm dari perang harga (predatory pricing) yang merugikan umkm. Dengan menggunakan algoritma Naive Bayes, metode klasifikasi yang dapat memprediksi kemungkinan sebuah kelas dan membuat keputusan berdasarkan data pembelajaran, penelitian Emotion Recognition pada komentar YouTube yang terkait ditutupnya TikTok Shop. Data akan diklasifikasikan ke dalam lima kelas: senang, marah, sedih, takut, dan terkejut. Tujuan dari penelitian ini adalah untuk menemukan model emosional terbaik dengan menggunakan metode naive Bayes. Hasil tes pengguna dengan Naive Bayes dan Tf-Idf. Dengan nilai *precision* sedih, senang, takut, dan terkejut mempunyai jumlah persentase yang tinggi, sementara marah untuk pengujian jumlah persentasenya= 59% sedangkan emosi takut= 91%, senang=87%, sedih=84%, dan terkejut=79%. Serta akurasi keseluruhan sebesar 82%.

Kata kunci: Naive Bayes, Emotion Recognition, ditutupnya tiktokshop, TF-IDF

LATAR BELAKANG

Pada bulan Juli 2023, dalam laporan dari CNBC Indonesia, Menteri Koperasi dan UKM, Teten Masduki, mengungkapkan harapannya agar Kementerian Perdagangan mempercepat revisi Peraturan Menteri Perdagangan Nomor 50/2020 tentang Perizinan Usaha, Periklanan, Pembinaan, dan Pengawasan Pelaku Usaha dalam Perdagangan Melalui Elektronik (PPMSE). Tujuannya adalah untuk mencegah gangguan terhadap bisnis UMKM yang disebabkan oleh munculnya Proyek S TikTok Shop. Proyek S TikTok Shop merupakan inisiatif dari perusahaan induk di China yang memperluas penetrasi ritel online-nya dengan menjual produk langsung melalui platform TikTok Shop.

Hal ini menimbulkan kekhawatiran di kalangan pedagang, terutama di Pasar Tanah Abang, seperti yang dilaporkan oleh CNNIndonesia.com. Pedagang di pasar tersebut mengalami penurunan pengunjung yang signifikan. Salah satu pedagang, Soleh,

mengungkapkan bahwa omzetnya turun drastis sejak adanya TikTok Shop. Sebelumnya, pada tahun 2019, ia bisa mendapatkan puluhan juta per hari, namun kini hanya mendapat satu pembeli dalam sehari. Soleh menyatakan bahwa kehadiran platform seperti Lazada dan Shopee tidak begitu berdampak pada pasar, tetapi dengan adanya TikTok, situasinya menjadi berbeda.

Ketua Komisi VI DPR RI Fraksi PDIP, Darmadi Duriyanto, juga turut menyuarakan keluhan pedagang Pasar Tanah Abang kepada Ketua KPPU M Afif Hasbullah. Menurutnya, penurunan omzet penjualan pedagang mencapai 60 persen sebagai dampak langsung dari kehadiran TikTok Shop (Bestari 2023).

Berdasarkan konteks yang telah diuraikan, penelitian ini bertujuan untuk mengidentifikasi tahapan dalam pengenalan emosi pada komentar YouTube yang terkait dengan penutupan TikTok Shop menggunakan algoritma Naïve Bayes. Naïve Bayes merupakan metode klasifikasi yang mampu memperkirakan probabilitas suatu kelas, dan dengan demikian menghasilkan keputusan berdasarkan data pembelajaran. Penelitian ini akan menggunakan metode Naïve Bayes bersama dengan teknik TF-IDF untuk menganalisis komentar pengguna YouTube.

Tujuan utama dari penelitian ini adalah untuk mengevaluasi tingkat akurasi dan hasil yang diperoleh dari prediksi emosi menggunakan metode Naïve Bayes dan TF-IDF pada komentar pengguna YouTube yang terkait dengan penutupan TikTok Shop. Dengan demikian, penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang bagaimana metode ini dapat diterapkan dalam mengenali emosi dalam konteks komentar online, serta seberapa efektif metode tersebut dalam menghasilkan prediksi yang akurat.

Ada tiga penelitian terdahulu yang relevan dalam domain analisis sentimen menggunakan algoritma machine learning. Penelitian pertama oleh Sudianto pada tahun 2022 berjudul "Analisis Kinerja Algoritma Machine Learning untuk Klasifikasi Emosi pada Media Sosial Twitter", menyoroti bahwa algoritma Naive Bayes memiliki tingkat akurasi yang mencolok sebesar 79.1%, dengan Precision 0.79, Recall 0.77, dan F1-score 0.76. Sedangkan SVM mencapai akurasi 76.6%, dengan Precision 0.76, Recall 0.75, dan F1-score 0.75. Penelitian kedua, dilakukan oleh Burhanudin dan tim pada tahun 2018, membahas "Klasifikasi Komentar Spam pada YouTube Menggunakan Metode Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors." Hasilnya menunjukkan bahwa Naive Bayes juga mencapai tingkat akurasi yang tinggi, mencapai 88.24%, dengan Precision 0.89, Recall 0.88, dan F1-score 0.88. Penelitian ketiga, yang dilakukan oleh Robi Kurniawan dan rekan pada tahun 2022, berjudul "Klasifikasi Emosi pada Twitter terkait Penerapan New Normal menggunakan Algoritma

Naïve Bayes." Penelitian ini menegaskan bahwa Naive Bayes kembali menunjukkan kinerja yang baik, dengan akurasi sebesar 70%, Precision 0.75, Recall 0.71, dan F1-score 0.73.

KAJIAN TEORITIS

A. Deteksi Emosi

Rona Nisa Sofia Amriza dan Didi Supriyadi dalam studi mereka mengartikan deteksi emosi sebagai proses menggali atau mengenali berbagai jenis emosi manusia, seperti kebahagiaan, kesedihan, kemarahan, ketakutan, keterkejutan, dan rasa jijik, yang termanifestasi dalam bahasa alami. Identifikasi emosi dalam teks, khususnya yang tersebar luas di media sosial, dianggap sebagai sumber informasi berharga untuk memahami reaksi individu terhadap berbagai situasi dan peristiwa. Beragam faktor seperti manajemen dan pemasaran, interaksi pengguna, aspek keuangan, politik, kesehatan, dan pendidikan, semuanya dapat dipengaruhi oleh deteksi emosi ini. (Nisa, Sofia, Amriza and Supriyadi 2021). Data komentar akan diklasifikasikan kedalam 5 kelas yaitu senang, marah, sedih, takut dan terkejut.

B. *TF-IDF*

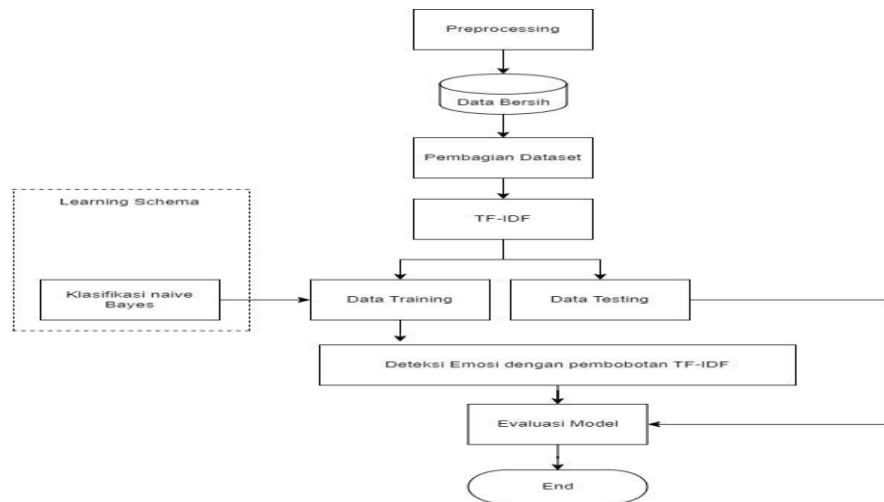
Teknik pembobotan term TF-IDF menghasilkan bobot term t dari dokumen dengan mengalikan nilai Frekuensi Term dengan Frekuensi Inverse. Teknik ini adalah yang paling umum digunakan sebagai perbandingan dengan metode pembobotan term lainnya. (Handayani, Kusrini, and Fatta 2020).

C. *Naive Bayes*

Teorema Bayes merupakan fondasi dari metode klasifikasi Naïve Bayes, yang memungkinkan pengkategorian kelas dengan memanfaatkan statistik dasar. Teorema Bayes mengoperasikan asumsi bahwa kemungkinan keberadaan atau ketiadaan suatu kelas bergantung pada keberadaan atau ketiadaan fitur-fitur lainnya. (Fanesya, Wihandika, and Indriati 2019).

METODE PENELITIAN

Langkah-langkah yang diambil untuk mencapai tujuan dari penelitian ini menggunakan metode algoritma Naive Bayes adalah sebagai berikut:



Gambar 1. Alur Penelitian

1. *Preprocessing Data*

Data preprocessing dilakukan untuk menghilangkan gangguan karena sebelum tahap selanjutnya diproses, data harus bersih dan terstruktur. Tahap-tahap ini termasuk pembersihan, pengolahan case, penghapusan stopwords, normalisasi, lemmization, dan imbalance.

2. **Pembagian Dataset Untuk Training dan Testing Model**

Dalam penerapan metode holdout untuk pembagian dataset, komposisi yang direkomendasikan untuk instruksi dan pengujian model Naive Bayes adalah 90% untuk data instruksi dan 10% untuk data pengujian.

3. **TF-IDF**

Tahap ini bertujuan untuk mengkonversi setiap kata dalam kumpulan data latihan menjadi representasi numerik, sehingga dapat diproses menggunakan algoritma Naïve Bayes.

4. **Klasifikasi Emosi Model Naive Bayes**

Model Naïve Bayes dilatih menggunakan data yang telah diberi label pada langkah sebelumnya. Proses pelatihan ini ditujukan untuk meningkatkan kinerja model agar dapat mengelompokkan data secara akurat.

5. **Evaluasi Model**

Dalam penelitian ini, dilakukan evaluasi terhadap model untuk mendeteksi emosi. Proses evaluasi model ini mencakup tahap di mana hasil dari uji data testing akan dipresentasikan dalam bentuk nilai tertentu yang mencerminkan kinerja model terhadap atribut yang telah disiapkan sebelumnya.

HASIL DAN PEMBAHASAN

1. *Preprocessing Data*

Data komentar yang digunakan berjumlah 1.149 data. Dari data tersebut lalu dilakukan pelabelan menjadi 5 kelas yaitu marah, sedih, senang, takut dan terkejut. Dengan jumlah tiap kelasnya adalah

ada 374 data berlabel marah, 265 data berlabel senang, 183 berlabel terkejut, 168 data berlabel takut, dan 159 data berlabel sedih. Setelah itu dilakukan preprocessing data, hasil dari preprocessing data yang telah dilakukan terdapat pada tabel 1.

Tabel 1. Hasil Preprocessing Text

Komentar	Emosi
[Harus, jangan, tutup, karena bukan, solusi ,mungkin ,bisa ,batasi, dari, jual, misal, boleh, jualan, produk, pakaiyan, jadi, ,mau, blanja, pakaiyan, blanja, secara, offline, otomatis, penjual, selain, pakainyan, pasar, offline, juga, laku, penjual, online, selain, produk]	marah
[kalo, jutaan, orang, rugi, gulung, tikar, menikmati, seberapa]	takut
[kecewa, cinta, negara, Indonesia]	marah
[pemerintah, modal, main, nutup, tanpa, solusi]	marah
[harus, pemerintah, membuat, tik tok, shop, tandingan, sebelum, memblokir]	marah
[salah, tiktaknya, lagi, izinnya, social, media, malah, pelan, meleset, jadi, ecommerce]	marah

Preprocessing dilakukan dengan menghapus reply komentar, data duplikat, link atau tautan, eliminasi kata yang tidak penting (stopwords) dan melakukan lemmization. Setelah data set sudah melalui tahap preprocessing, dilakukan imbalance data untuk menyeimbangkan data tiap kelasnya dengan melakukan oversampling. Hasil dari oversampling adalah jumlah tiap kelasnya adalah ada 376 data berlabel marah, 376 data berlabel senang, 376 berlabel terkejut, 376 data berlabel takut, dan 376 data berlabel sedih.

2. Klasifikasi Emosi

Setelah data bersih melalui preprocessing, lalu dataset akan melalui tahapan td-idf, sebelum itu data akan di split atau dibagi menjadi data train dan data test. Dalam klasifikasi terdapat 5 skenario yang digunakan untuk memperoleh model yang optimal untuk prediksi emosi. Skenario pertama algoritma multinomial naive bayes hyperparameter ($\alpha=0.1$, $\text{fit_prior}=\text{True}$), skenario kedua Bernoulli Naive Bayes dengan hyperparameter ($\alpha=1.0$, $\text{binarize}=0.0$, $\text{fit_prior}=\text{True}$), skenario ketiga Bernaoulli Naive Bayes tanpa hyperparameter, skenario keempat Bernaoulli Naive Bayes dengan hyperparameter ($\alpha=1.0$, $\text{binarize}=0.5$, $\text{fit_prior}=\text{False}$), dan skenario kelima Multinomial Naive Bayes tanpa hyperparameter. Dari 5 skenario tersebut yang memiliki peforma yang optimal dan akurasi terbaik adalah skenario pertama. Berikut hasil dari klasifikasi emosi skenario pertama pada gambar 2 Confusion Matrix dan Classication Report.

```
print(confusion_matrix(y_test, y_pred_test))
```

```
[[10  5  2  2  5]
 [ 1 37  0  0  2]
 [ 4  0 27  0  2]
 [ 0  0  2 31  0]
 [ 2  2  0  1 33]]
```

```
# evaluation
from sklearn.metrics import classification_report, confusion_matrix
print(classification_report(y_test, y_pred_test))
```

	precision	recall	f1-score	support
marah	0.59	0.42	0.49	24
sedih	0.84	0.93	0.88	40
senang	0.87	0.82	0.84	33
takut	0.91	0.94	0.93	33
terkejut	0.79	0.87	0.82	38
accuracy			0.82	168
macro avg	0.80	0.79	0.79	168
weighted avg	0.81	0.82	0.81	168

Gambar 2. Confusion Matrix dan Classification Report

Setelah melihat hasil Confusion Matrix dan Classification Report, diketahui bahwa nilai precision sedih, senang, takut, dan terkejut mempunyai jumlah persentase yang tinggi, sementara marah untuk pengujian jumlah persentasenya= 59% sedangkan emosi takut= 91%, senang=87%, sedih=84%, dan terkejut=79%. Hal ini dapat dijelaskan pada confusion matrix dimana pada kelas marah di baris pertama menunjukkan prediksi yang dihasilkan lebih condong ke kelas sedih dan terkejut yaitu 10 tebakan benar bahwa itu marah, 5 tebakan salah yang harusnya marah tapi di prediksi sedih, 2 tebakan salah yang harusnya marah tapi di prediksi senang, 2 tebakan salah yang harusnya marah tapi di prediksi takut, dan 5 tebakan salah yang harusnya marah tapi di prediksi terkejut.

KESIMPULAN DAN SARAN

Hasil Confusion Matrix dan Classification Report, diketahui bahwa nilai *precision* sedih, senang, takut, dan terkejut mempunyai jumlah persentase yang tinggi, sementara marah untuk pengujian jumlah persentasenya= 59% sedangkan emosi takut= 91%, senang=87%, sedih=84%, dan terkejut=79%. Hal ini dapat dijelaskan pada *confusion matrix* dimana pada kelas marah di baris pertama menunjukkan prediksi yang dihasilkan lebih condong ke kelas sedih dan terkejut. Adapun saran yang dapat diberikan untuk pengembangan penelitian ini lebih lanjut adalah Menambahkan kata stopwords dalam bentuk file dengan menyesuaikan kondisi perkembangan media social, menggunakan pembobotan kata selain *TF-IDF* untuk meningkatkan akurasi pada setiap prediksi label emosi, dalam *emotion recognition* harus lebih memperhatikan tanda baca agar pelabelan dan prediksi emosi lebih akurat, menerjemahkan emoji dan emoticon, melakukan analisis kata dan ulasan sarkasme, dan memperhatikan kata berulang, karena dalam *emotion recognition* bisa terdapat emosi dalam kata yang berulang..

DAFTAR REFERENSI

- B Uarbat, B. E., and R. Cretulscu. 2003. "AISB 2008 Convention Communication, Interaction and Social Intelligence." *Technology* 25–35.
- Bestari, Novina Putri. 2023. "TikTok Monopoli, Dilarang Sekaligus Medsos-Ecommerce." *CNBC INDONESIA*.
- Candra Ardiansyah, Indriati, Marji. 2020. "Klasifikasi Emosi Pada Komentar Youtube Menggunakan Metode Modified K-Nearest Neighbor Dengan BM25 Dan Seleksi Fitur Chi-Square." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 4(4):1027–32.
- Dhiya Abirawa, Irfan, Andrew Brian Osmond, and Casi Setianingsih. 2018. "Klasifikasi Emosi Pada Lirik Lagu Menggunakan Metode Support Vector Machine Emotion Classification in a Song Lyrics Using Support Vector Machine Method 1." 5(3):6203–9.

- Ekman, Paul, and Daniel Cordaro. 2011. "What Is Meant by Calling Emotions Basic." *Emotion Review* 3(4):364–70. doi: 10.1177/1754073911410740.
- Fadli, Nur. 2023. "Klasifikasi Emosi Kompleks Yang Negatif Pada Anak Menggunakan Metode Naïve Bayes Classification of Negative Complex Emotions in Children Using Naïve Bayes Method." 10(1):460–66.
- Fanesya, Fera, Randy Cahya Wihandika, and Indriati. 2019. "Deteksi Emosi Pada Twitter Menggunakan Metode Naive Bayes Dan Kombinasi Fitur." *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 3(7):3.
- Handayani, Riska Dwi, Kusri Kusri, and Hanif Al Fatta. 2020. "Perbandingan Fitur Ekstraksi Untuk Klasifikasi Emosi Pada Sosial Media." *Jurnal Ilmiah SINUS* 18(2):21. doi: 10.30646/sinus.v18i2.457.
- KEMP, SIMON. 2023. "DIGITAL 2023: INDONESIA." *Datareportal*.
- Komorowski, Matthieu, Dominic C. Marshall, Justin D. Saliccioli, and Yves Crutain. 2016. "Secondary Analysis of Electronic Health Records." *Secondary Analysis of Electronic Health Records* (September):1–427. doi: 10.1007/978-3-319-43742-2.
- Nisa Sofia Amriza, Rona, and Didi Supriyadi. 2021. "Komparasi Metode Machine Learning Dan Deep Learning Untuk Deteksi Emosi Pada Text Di Sosial Media." *Jurnal JUPITER* 13(2):130–39.
- Nurfauzan, Azward, and Warih Maharani. 2021. "Klasifikasi Emosi Pada Pengguna Twitter Menggunakan Metode Klasifikasi Decision Tree." *Universitas Telkom*.
- Puspitarini, Dinda Sekar, and Reni Nuraeni. 2019a. "Pemanfaatan Media Sosial Sebagai Media Promosi." *Jurnal Common* 3(1):71–80. doi: 10.34010/common.v3i1.1950.
- Puspitarini, Dinda Sekar, and Reni Nuraeni. 2019b. *PEMANFAATAN MEDIA SOSIAL SEBAGAI MEDIA PROMOSI (Studi Deskriptif Pada Happy Go Lucky House)*. Vol. 3.
- Riani, Adella Putri, Nina Sulistyowati, Taufik Ridwan, and Apriade Voutama. 2023. "Klasifikasi Emosi Publik Terhadap Larangan Penggunaan Obat Sirup Menggunakan Algoritma Naive Bayes." *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi* 7(2):325–39. doi: 10.46880/jmika.vol7no2.pp325-339.
- Sudianto, Sudianto. 2022a. "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi." *Building of Informatics, Technology and Science (BITS)* 4(2):1027–34. doi: 10.47065/bits.v4i2.2261.
- Sudianto, Sudianto. 2022b. "Analisis Kinerja Algoritma Machine Learning Untuk Klasifikasi Emosi." *Building of Informatics, Technology and Science (BITS)* 4(2):1027–34. doi: 10.47065/bits.v4i2.2261.
- Yanwardhana, Emir. 2023. "Menkominfo: Social Commerce Tidak Diatur, Bisa Monopoli." *CNBC INDONESIA*.