

# Analisis Sentimen Brand Tas Pada *Platform X* (Twitter) Menggunakan Metode *Klasifikasi Naive Bayes* dan *Support Vector Machine (SVM)*

Rita Rosita<sup>1</sup>, Hendra Sandhi Firmansyah<sup>2</sup>, Purnomo Yustianto<sup>3</sup>

Magister Teknik Informatika, Universitas Langlangbuana<sup>1,2,3</sup>

<sup>1</sup>ritaro0410@gmail.com

<sup>2</sup>hendra.sf@gmail.com

<sup>3</sup>purnomoyus@gmail.com

**Abstrak**—Perkembangan media sosial telah menghadirkan peluang besar bagi brand untuk memahami persepsi publik terhadap produknya. Twitter sebagai salah satu *platform* dengan interaksi tinggi, banyak digunakan untuk mengekspresikan opini, termasuk mengenai brand tas. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap brand tas melalui unggahan di *platform X* (Twitter) menggunakan metode klasifikasi *Naive Bayes Classifier* dan *Support Vector Machine (SVM)*. Data yang digunakan dikumpulkan melalui proses *crawling*, kemudian dilakukan tahap *preprocessing* meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya data dibagi menjadi data latih dan data uji untuk klasifikasi sentimen positif, negatif, dan netral. Hasil penelitian menunjukkan bahwa metode SVM memiliki tingkat akurasi lebih tinggi dibandingkan *Naive Bayes* dalam mengklasifikasikan sentimen publik. Temuan ini memberikan kontribusi dalam pemanfaatan metode machine learning untuk analisis opini konsumen sebagai dasar pengambilan keputusan strategis bagi brand tas.

**Kata kunci**— Analisis sentimen, Twitter, brand tas, *Naive Bayes*, *Support Vector Machine (SVM)*.

## I. Pendahuluan

Teknologi digital dan media sosial telah membawa perubahan besar dalam cara konsumen berinteraksi dengan brand. Twitter, sebagai salah satu platform media sosial populer, memungkinkan penggunaannya untuk menyampaikan opini, pengalaman, maupun keluhan secara cepat dan terbuka. Karakteristik Twitter yang singkat, padat, dan real-time menjadikannya sumber data yang relevan untuk memahami persepsi publik terhadap suatu produk, termasuk brand tas. Dalam konteks industri *fashion*, tas tidak hanya berfungsi sebagai kebutuhan fungsional, tetapi juga memiliki nilai simbolis yang mencerminkan gaya hidup, identitas, dan status sosial konsumen (Kapferer, 2012; Kotler & Keller, 2016).

Analisis sentimen hadir sebagai metode untuk menafsirkan opini publik dengan mengklasifikasikan teks ke dalam kategori positif, negatif, atau netral. Pendekatan ini dapat memberikan gambaran komprehensif mengenai

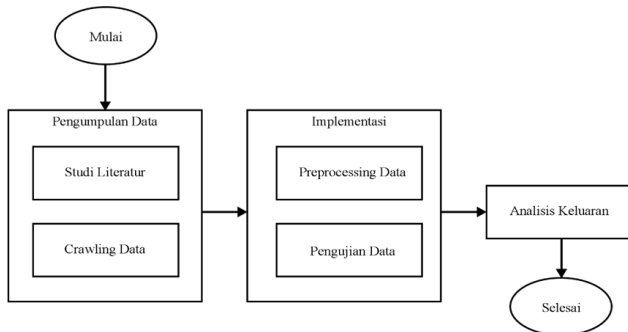
bagaimana konsumen memandang brand tas, baik dari aspek desain, kualitas, maupun harga. Dengan meningkatnya intensitas percakapan konsumen di media sosial, diperlukan metode analisis yang mampu mengolah data dalam jumlah besar dengan cepat dan akurat. Jika dilakukan secara manual, penilaian sentimen akan memakan waktu lama dan rawan bias subjektif. Oleh karena itu, penerapan machine learning menjadi solusi penting untuk memastikan hasil analisis lebih objektif dan efisien (Liu, 2012; Feldman, 2013).

Dua metode machine learning yang banyak digunakan dalam klasifikasi teks adalah *Naive Bayes Classifier* dan *Support Vector Machine (SVM)*. *Naive Bayes* dikenal dengan kesederhanaan serta efektivitasnya dalam mengklasifikasikan teks berbasis probabilitistik, sementara SVM lebih unggul dalam menemukan pemisah data yang optimal untuk meningkatkan akurasi prediksi. Dengan membandingkan kedua metode ini, penelitian bertujuan untuk mengetahui algoritma mana yang lebih tepat dalam menganalisis sentimen publik terhadap brand tas di Twitter. Hasil penelitian diharapkan tidak hanya memberikan kontribusi akademis, tetapi juga menjadi acuan praktis bagi *brand* dalam merumuskan strategi komunikasi digital dan pengembangan produk (Manning, Raghavan, & Schütze, 2008; Cortes & Vapnik, 1995)..

## II. Metode

Penelitian ini mengadopsi metode penelitian *experiment*. Penggunaan metode eksperimen sangat relevan pada penelitian yang melibatkan pengujian model atau algoritma *Naive Bayes* dan *Support Vector Machine (SVM)*. Pada konteks ini, dataset dibagi menjadi data latih dan data uji, kemudian kedua algoritma diuji kinerjanya berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan *F1-score*. Perbandingan hasil ini memungkinkan peneliti untuk menarik kesimpulan tentang metode mana yang lebih efektif.

A. Alur Penelitian

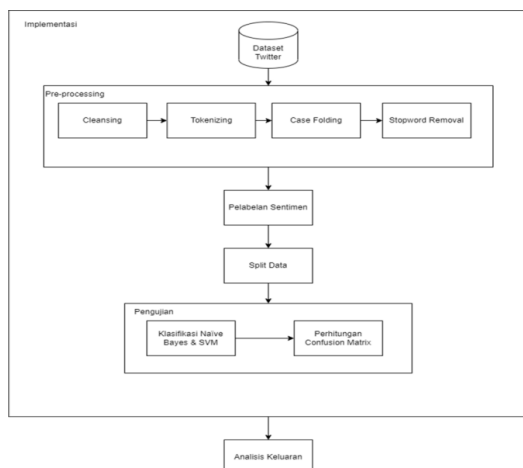


Gambar II-1 Alur Penelitian

Alur penelitian digunakan sebagai prosedur yang menjelaskan proses berjalannya sebuah penelitian, diantaranya :

1. *Studi Literatur* dalam meneliti mengumpulkan referensi dari buku, artikel, dan jurnal sebelumnya yang relevan dengan topik analisis sentimen, metode *Naïve Bayes Classifier* dan *Support Vector Machine (SVM)*, serta penerapan bahasa pemrograman Python dan aplikasi pendukung seperti RapidMiner.
2. *Crawling Data* dikumpulkan dari media sosial Twitter dengan kata kunci terkait Bodypack menggunakan library *snsrcape* pada Python. Pengambilan data dilakukan pada Juni–Juli 2025 saat berlangsung kampanye *Back to School*, dengan hasil sebanyak 451 *tweets*.
3. *Pre-processing* data mentah kemudian diolah melalui beberapa tahap, seperti *tokenizing*, *stopwords removal*, *stemming*, dan *konversi emotikon*. Proses ini bertujuan agar data siap untuk diklasifikasikan menjadi sentimen positif, negatif, atau netral.
4. Pengujuan Data merupakan tahap akhir dalam menguji data yang sudah diproses dengan metode klasifikasi *Naïve Bayes Classifier* dan *Support Vector Machine* untuk membandingkan hasil analisis sentimen dari kedua algoritma tersebut

B. Metode Implementasi



Gambar II-2 Metode Implementasi

1. *Dataset* Twitter ini menggunakan data dari media sosial Twitter dengan kata kunci “bodypack”. Data diambil berupa *tweets* atau komentar menggunakan teknik *crawling* dengan bahasa pemrograman Python.
2. *Pre-processing* Data yang sudah terkumpul diolah terlebih dahulu melalui beberapa tahap, yaitu *cleansing* (menghapus karakter tidak relevan), *tokenizing* (memecah teks menjadi kata-kata), *case folding* (mengubah semua huruf menjadi kecil), serta *stopword removal* (menghapus kata-kata umum yang tidak penting).
3. Pelabelan Sentimen setiap tweet diberi label sentimen berupa positif, negatif, atau netral, baik secara manual maupun otomatis menggunakan kamus kata (*lexicon*).
4. Split dataset kemudian dibagi menjadi dua, yaitu data latih (30%) dan data uji (70%), untuk proses pembelajaran dan pengujian model sentimen.
5. Klasifikasi data yang sudah diproses diuji menggunakan dua metode, yaitu *Naïve Bayes Classifier* dan *Support Vector Machine (SVM)*, dengan bantuan aplikasi *ritarosentimenanalisis.com* untuk memudahkan klasifikasi.
6. Perhitungan Confusion Matrix hasil klasifikasi dianalisis menggunakan confusion matrix untuk menghitung nilai akurasi, presisi, recall, dan *F1-Score* dari masing-masing metode.
7. Analisis Keluaran merupakan tahap akhir adalah menganalisis hasil perbandingan kedua metode untuk mengetahui metode mana yang lebih baik dalam mengklasifikasikan sentimen.

C. Parameter Evaluasi

Parameter evaluasi yang relevan dengan kebutuhan aplikasi—misalnya, di bidang medis lebih menekankan *recall* (*minim false negative*), sedangkan di *e-commerce* mungkin menekankan *accuracy* atau *F1-Score* tergantung pada target bisnis. G. James, D. Witten, T. Hastie, R. Tibshirani. *An Introduction to Statistical Learning*. Springer, 2021 & Tom M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

Metrik	Bobot Ideal*	Penjelasan	Referensi Akademik
Accuracy	= 33%	Mengukur proporsi prediksi yang benar terhadap seluruh data uji. Digunakan untuk menilai kinerja umum model.	Manning, R., Raghavan, P., & Schütze, H. (2008). <i>Introduction to Information Retrieval</i> . Cambridge University Press.
Recall	= 33%	Mengukur kemampuan model mendeteksi seluruh <i>positive class</i> dengan benar ( <i>sensitivity</i> ). Penting jika tujuan analisis adalah meminimalkan <i>false negative</i> .	Powers, D. M. W. (2011). <i>Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness &amp; Correlation</i> .
F1-Score	= 34%	Rata-rata harmonis antara <i>Precision</i> dan <i>Recall</i> . Cocok digunakan ketika data tidak seimbang ( <i>imbalanced dataset</i> ).	Sokolova, M., & Lapalme, G. (2009). <i>A systematic analysis of performance measures for classification tasks</i> . <i>Information Processing &amp; Management</i> , 45(4).
Precision	Tidak dibobotkan terpisah dalam AOA*, tetapi menjadi komponen penting F1-Score	Mengukur proporsi prediksi positif yang benar. Penting untuk menghindari <i>false positive</i> .	Manning et al., 2008; Powers, 2011

Gambar II-3 Parameter Evaluasi

Perhitungan :

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

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

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

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

*TP* : True Positive  
*TN* : True Negative  
*FP* : False Positive  
*FN* : False Negative

### III. Implementasi

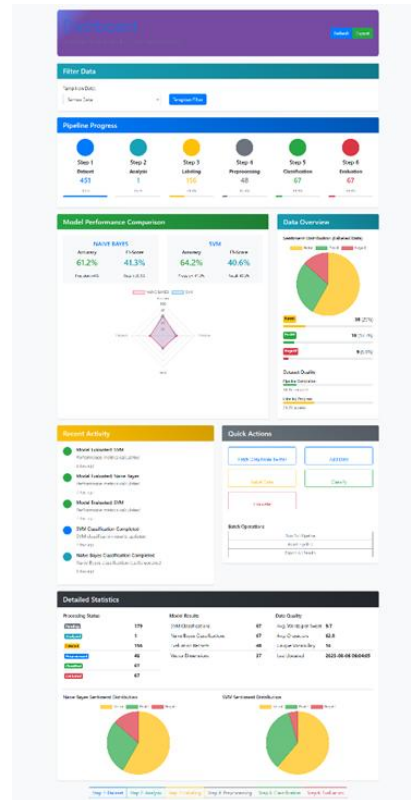
Penggunaan Aplikasi <https://ritarosentimenanalysis.com/>

#### A. Login



Fitur login di web ritarosentimenanalysis dirancang sebagai sistem terintegrasi untuk mendukung proses analisis sentimen secara aman dan personal. Melalui fitur login, setiap pengguna diwajibkan masuk menggunakan kredensial yang valid, sehingga data hasil analisis dapat tersimpan secara individual, terproteksi, dan mudah diakses kembali. Sistem ini memungkinkan pengguna untuk menyimpan riwayat penelitian, mengelola dataset, serta membandingkan hasil klasifikasi dari berbagai metode seperti *Naive Bayes* dan *Support Vector Machine (SVM)*. Dengan demikian, platform ini tidak hanya berfungsi sebagai media analisis, tetapi juga sebagai repositori pribadi yang mendukung konsistensi penelitian, keamanan data, dan kemudahan kolaborasi.

#### B. Dashboard



1. *Pipeline management & monitoring* yaitu menampilkan 6 tahap proses ML (*Dataset* → *Analysis* → *Labeling* → *Preprocessing* → *Classification* → *Evaluation*). *Status tracking* yaitu memantau berapa banyak data yang telah diproses di setiap tahap
2. *Model performance analysis* diantaranya menampilkan perbandingan performa antara *Naive Bayes* dan *SVM*, *Metrics evaluation* dalam menampilkan akurasi, *precision*, *recall*, dan *F1-score* serta menampilkan grafik radar untuk perbandingan visual antar model
3. *Data management & quality control* diantaranya sentiment distribution yaitu analisis sentimen (*Neutral*, *Positive*, *Negative*). *Data quality monitoring* yaitu status kelengkapan dataset dan proses labeling
4. *Operational features* diataranya batch operations yaitu menjalankan *pipeline* penuh, reset, atau export hasil. *Quick actions* yaitu *fetch data*, *add data*, *label data*, *classify*, dan *evaluate*
5. *Business Intelligence* diantaranya statistical insights yaitu *detailed statistics* untuk *decision making*. *Progress monitoring*: timeline dan status setiap tahap *processing*

#### IV. Hasil dan Pembahasan

##### A. Klasifikasi

Perbandingan Performa Model					
Model	Accuracy	Precision	Recall	F1-Score	Winner
SVM	64.36%	40.85%	38.99%	39.82%	-
NAIVE_BAYES	56.25%	41.21%	35.44%	37.97%	-

Accuracy Winner	Precision Winner	Recall Winner	F1 Winner
SVM 64.36%	NAIVE_BAYES 41.21%	SVM 38.99%	SVM 39.82%

Rekomendasi:  
 Berdasarkan evaluasi, model SVM menunjukkan performa terbaik dengan memenangkan 3 dari 4 metrik evaluasi.

Gambar IV-1 Klasifikasi

Hasil analisis sentimen menggunakan *Support Vector Machine* memiliki nilai akurasi sebesar 64.3% dengan nilai precision positif sebesar 40.7% dan presisi untuk prediksi negatif sebesar 0% sementara netral 81.8%, dengan nilai *recall* sebesar 38,9.63% untuk data positif sebesar 44.0% dan *recall* untuk data negatif sebesar 0% sementara netral 73%, dengan *F1-Score* sebesar 39.82% untuk data positif sebesar 41,5% dan *F1-Score* untuk data negatif sebesar 0% sementara netral 77.6%. Sedangkan untuk hasil analisis sentimen menggunakan *Naïve Bayes Classifier* memiliki nilai akurasi sebesar 54% dengan nilai presisi untuk precision positif sebesar 37.9% dan precision negatif sebesar 0% sementara netral 79.6%, dengan nilai *recall* 34,3% untuk data positif sebesar 44% dan *recall* untuk data negatif sebesar 0% sementara netral 58,9%, dengan *F1-Score* 36,9% untuk data positif sebesar 40,7% dan *F1-Score* untuk data negatif sebesar 0% sementara netral 67,7%

##### B Hasil Perbandingan Klasifikasi

Hasil Evaluasi				
Evaluasi SVM				
Hasil Evaluasi: SVM				
Metrik Keseluruhan				
Accuracy		64.36%		
Precision (Avg)		40.85%		
Recall (Avg)		38.99%		
F1-Score (Avg)		39.82%		
Confusion Matrix				
		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	11	3	11
	Negatif	1	0	1
	Netral	15	5	54
Metrik per Kelas				
Kelas	Precision	Recall	F1-Score	Support
Positif	40.7%	44%	42.3%	25
Negatif	0%	0%	0%	2
Netral	81.8%	73%	77.1%	74

Gambar IVI-2 Hasil Perbandingan dan Klasifikasi

Membuktikan bahwa Metode *Support Vector Machine* memiliki performa tertinggi dibandingkan metode *Naïve Bayes Classifier* dengan nilai *accuracy* 64.36%, *recall* 38.99% *F1* 39,82%. Oleh karena itu Sentimen Positif (40%) sebagian besar pengguna di *platform X* (Twitter) mengekspresikan pandangan yang mendukung, puas, atau memiliki kesan baik terhadap brand tas. Sentimen positif

ini mengindikasikan adanya potensi loyalitas pelanggan dan peluang besar untuk meningkatkan engagement. Sentimen Negatif (0%) : tidak ditemukan unggahan yang secara eksplisit mengekspresikan kekecewaan, kritik, atau keluhan terhadap brand tas. Ketidadaan sentimen negatif ini bisa menunjukkan bahwa brand memiliki reputasi yang cukup baik di mata konsumen.

Gambar IVI-2 Hasil Perbandingan dan Klasifikasi

##### I. Simpulan

Jurnal INFOSECURE berusaha memberikan yang terbaik untuk menjamin keseragaman tampilan format tulisan. Seluruh isi dokumen ini menggunakan tipe huruf *Times New Roman*. Tipe huruf yang lain boleh dipergunakan jika diperlukan untuk tujuan khusus.

Jurnal INFOSECURE berhak mendistribusikan dan merevisi *templates* ini sesuai dengan perkembangan dan kebutuhan. Informasi tentang jadwal penerimaan makalah dapat dilihat di *website* Jurnal INFOSECURE.

##### UCAPAN TERIMA KASIH / ACKNOWLEDGMENT

(JIKA DIPERLUKAN)

*Heading* dari ucapan terima kasih dan referensi tidak diberikan penomoran.

Jurnal INFOSECURE mengucapkan terima kasih dan penghargaan kepada semua kontributor pengembang *template* ini.

##### REFERENSI

- [1] Kapferer, J. N. (2012). *The New Strategic Brand Management: Advanced Insights and Strategic Thinking*. Kogan Page.  
 → Mendukung pembahasan tentang identitas brand dan persepsi konsumen.
- [2] Kotler, P., & Keller, K. L. (2016). *Marketing Management (15th ed.)*. Pearson Education.  
 → Digunakan untuk menjelaskan hubungan brand dengan perilaku konsumen.
- [3] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool.  
 → Menjadi dasar teori utama tentang analisis sentimen dan teknik klasifikasi opini.
- [4] Feldman, R. (2013). *Techniques and Applications for Sentiment Analysis*. *Communications of the ACM*, 56(4), 82–89.  
 → Mendukung penjelasan metode otomatis untuk analisis opini.
- [5] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.  
 → Digunakan untuk teori dasar pemrosesan teks dan klasifikasi data.
- [6] Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks*. *Machine Learning*, 20, 273–297.  
 → Referensi fundamental untuk teori SVM.
- [7] Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.  
 → Sebagai acuan umum algoritma klasifikasi termasuk Naïve Bayes.
- [8] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*. Springer.  
 → Mendukung pembahasan evaluasi model (*accuracy*, *precision*, *recall*, *F1-score*).