

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK MENGANALISIS PENGARUH KONTEN MEDIA SOSIAL TERHADAP KEPUTUSAN PEMBELIAN

IMPLEMENTATION OF THE NAÏVE BAYES ALGORITHM TO ANALYZE THE INFLUENCE OF SOCIAL MEDIA CONTENT ON PURCHASING DECISIONS

Hanina Assegaf¹⁾, Rike Andika Aris Susanti²⁾, Dila Wulandari³⁾, Siti Aisa⁴⁾, Santoso^{5*}

^{1,2,3,4,5}Jurusan Matematika, Universitas Abdurrahman Saleh Situbondo

*Email Korespondensi: santoso@unars.ac.id

Abstrak: Penelitian ini bertujuan untuk menganalisis pengaruh konten media sosial terhadap keputusan pembelian konsumen di Kabupaten Jember dengan menggunakan algoritma Naïve Bayes. Di era digital yang berkembang pesat, media sosial telah menjadi saluran penting dalam pemasaran produk dan jasa. Dengan metode penelitian kuantitatif deskriptif, data dikumpulkan melalui kuesioner online dari 55 responden yang aktif menggunakan media sosial dan pernah melakukan pembelian setelah melihat konten di platform tersebut. Hasil analisis menunjukkan bahwa konten media sosial, terutama jenis konten hiburan dan edukasi, memiliki dampak signifikan terhadap keputusan pembelian. Model Naïve Bayes yang diterapkan mencapai akurasi sekitar 68,5%, meskipun terdapat sejumlah klasifikasi yang tidak sesuai. Penelitian ini memberikan wawasan berharga bagi pelaku usaha lokal untuk mengembangkan strategi konten yang lebih efektif dan menarik bagi konsumen. Temuan ini juga menyoroti keterbatasan metode Naïve Bayes, terutama asumsi independensi antar fitur, yang dapat mempengaruhi hasil klasifikasi. Penelitian ini diharapkan dapat menjadi dasar bagi penelitian selanjutnya untuk mengeksplorasi metode analisis lain dan memperdalam pemahaman tentang perilaku konsumen di konteks yang lebih luas.

Kata Kunci: Media Sosial, Keputusan Pembelian, Konten Media Sosial, Algoritma Naïve Bayes

Abstract: This study aims to analyze the influence of social media content on consumer purchasing decisions in Jember Regency using the Naïve Bayes algorithm. In the rapidly evolving digital era, social media has become a crucial channel for marketing products and services. Using a descriptive quantitative research method, data was collected through an online questionnaire from 55 respondents who actively use social media and have made purchases after viewing content on these platforms. The analysis results indicate that social media content, particularly entertainment and educational content, has a significant impact on purchasing decisions. The applied Naïve Bayes model achieved an accuracy of approximately 68.5%, although there were several inconsistencies in classification. This study provides valuable insights for local businesses to develop more effective and engaging content strategies for consumers. The findings also highlight the limitations of the Naïve Bayes method, particularly the assumption of independence between features, which can influence classification results. This study is expected to serve as a foundation for further research exploring other analysis methods and deepening our understanding of consumer behavior in a broader context.

Keywords: Social Media, Purchasing Decisions, Social Media Content, Naïve Bayes Algorithm

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong transformasi signifikan dalam perilaku konsumen, khususnya dalam proses pengambilan keputusan pembelian. Media sosial tidak lagi berfungsi sebatas sarana komunikasi dan hiburan, melainkan telah berkembang menjadi platform strategis bagi perusahaan dan pelaku usaha dalam menyampaikan pesan pemasaran, membangun citra merek, serta memengaruhi sikap dan preferensi konsumen (Kaplan & Haenlein, 2010). Konten media sosial seperti ulasan pengguna, iklan digital, electronic word of mouth (e-WOM), dan aktivitas influencer marketing terbukti memiliki peran penting dalam membentuk persepsi konsumen terhadap suatu produk atau jasa (Mangold & Faulds, 2009).

Berbagai penelitian terdahulu menunjukkan bahwa kualitas dan kredibilitas konten media sosial berpengaruh signifikan terhadap keputusan pembelian konsumen. Konten yang informatif, persuasif, dan relevan cenderung meningkatkan kepercayaan serta minat beli, sementara konten negatif dapat menurunkan intensi pembelian secara signifikan (Erkan & Evans, 2016). Namun, sebagian besar penelitian tersebut masih menggunakan pendekatan analisis statistik konvensional seperti regresi linier atau Structural Equation Modeling (SEM), yang umumnya berfokus pada hubungan linier antar variabel dan mensyaratkan asumsi distribusi data tertentu (Hair et al., 2019).

Seiring dengan meningkatnya volume dan kompleksitas data media sosial yang bersifat tidak terstruktur, pendekatan statistik tradisional dinilai kurang optimal dalam menangkap pola laten dan dinamika perilaku konsumen secara komprehensif. Dalam konteks ini, penerapan teknik machine learning menjadi semakin relevan karena mampu mengolah data dalam jumlah besar serta mengenali pola klasifikasi yang kompleks secara lebih adaptif (Alpaydin, 2020). Salah satu algoritma yang banyak digunakan dalam klasifikasi teks dan analisis data digital adalah algoritma Naïve Bayes, yang dikenal memiliki efisiensi komputasi tinggi dan performa yang baik pada data berdimensi besar (Manning et al., 2008).

Penelitian-penelitian sebelumnya yang menerapkan algoritma Naïve Bayes pada data media sosial umumnya berfokus pada klasifikasi sentimen, seperti identifikasi opini positif, negatif, atau netral terhadap suatu produk atau layanan (Liu, 2012). Meskipun hasil klasifikasi sentimen tersebut memberikan gambaran umum mengenai persepsi konsumen, kajian tersebut sering kali belum mengaitkan hasil analisis secara langsung dengan implikasi

keputusan pembelian konsumen. Akibatnya, temuan penelitian cenderung berhenti pada level deskriptif sentimen tanpa memberikan kontribusi yang lebih aplikatif dalam konteks pengambilan keputusan pemasaran.

Berdasarkan celah penelitian tersebut, artikel ini menawarkan kebaruan (novelty) dengan mengimplementasikan algoritma Naïve Bayes untuk menganalisis pengaruh konten media sosial secara langsung terhadap keputusan pembelian. Penelitian ini tidak hanya mengklasifikasikan sentimen atau karakteristik konten media sosial, tetapi juga menghubungkan hasil klasifikasi tersebut dengan kecenderungan keputusan pembelian konsumen. Dengan pendekatan ini, analisis yang dihasilkan menjadi lebih kontekstual dan relevan bagi kebutuhan pemasaran digital berbasis data (data-driven marketing).

Selain itu, kebaruan penelitian ini terletak pada integrasi antara analisis konten media sosial dan model klasifikasi probabilistik berbasis Naïve Bayes sebagai alat bantu pengambilan keputusan. Pendekatan ini diharapkan mampu memberikan kontribusi empiris dalam pengembangan kajian pemasaran digital serta memperluas pemanfaatan algoritma machine learning dalam analisis perilaku konsumen. Hasil penelitian ini diharapkan tidak hanya memperkaya literatur akademik, tetapi juga memberikan rekomendasi praktis bagi pelaku usaha dalam merancang strategi konten media sosial yang lebih efektif dan berorientasi pada peningkatan keputusan pembelian konsumen.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan metode data mining untuk menganalisis pengaruh konten media sosial terhadap keputusan pembelian konsumen. Pendekatan kuantitatif dipilih karena data penelitian berupa data numerik dan kategorikal yang diperoleh melalui kuesioner, sedangkan data mining digunakan untuk menemukan pola dan hubungan antarvariabel dalam data (Sugiyono, 2019; Han et al., 2012).

Algoritma yang digunakan adalah Naïve Bayes, yaitu metode klasifikasi berbasis probabilitas yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antarfitur. Algoritma ini dipilih karena sederhana, efisien, dan memiliki performa yang baik pada data dengan jumlah sampel terbatas (Larose & Larose, 2015).

Populasi penelitian adalah konsumen di Kabupaten Jember yang aktif menggunakan media sosial dan pernah melakukan pembelian setelah melihat konten media sosial, dengan rentang usia 18–44 tahun. Sampel penelitian berjumlah 55 responden yang dipilih menggunakan metode purposive sampling berdasarkan kriteria aktif menggunakan media

sosial minimal 1 jam per hari, pernah melihat konten produk sebelum membeli, dan berada pada rentang usia yang ditentukan (Sugiyono, 2019). Data dikumpulkan menggunakan kuesioner online yang memuat profil responden, penggunaan media sosial, jenis konten yang dilihat, serta tingkat pengaruh terhadap keputusan pembelian.

Analisis data dilakukan melalui dua tahap, yaitu analisis deskriptif dan analisis klasifikasi. Analisis deskriptif digunakan untuk menggambarkan karakteristik responden dan variabel penelitian menggunakan statistik deskriptif seperti frekuensi dan persentase (Sugiyono, 2019). Analisis klasifikasi dilakukan menggunakan algoritma Naïve Bayes dengan fitur berupa karakteristik responden dan penggunaan media sosial, serta kelas berupa tingkat pengaruh terhadap keputusan pembelian.

Perhitungan Naïve Bayes didasarkan pada Teorema Bayes dengan rumus:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

Untuk data dengan banyak fitur, perhitungan dilakukan dengan mengalikan probabilitas setiap fitur terhadap kelas tertentu, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi (Han et al., 2012). Data dibagi menjadi data latih dan data uji untuk mengevaluasi kinerja model. Evaluasi dilakukan menggunakan akurasi dan confusion matrix, serta metrik tambahan seperti presisi dan recall jika diperlukan (Witten et al., 2017). Hasil klasifikasi selanjutnya dianalisis untuk mengidentifikasi jenis konten media sosial yang paling berpengaruh terhadap keputusan pembelian konsumen (Kotler & Keller, 2016).

HASIL DAN PEMBAHASAN

Hasil

Berdasarkan hasil perhitungan diperoleh data sebagai berikut:

Tabel 1.2 Tabel Probabilitas

Fitur	Jml Kls 1	Jml Kls 2	Jml Kls 3	Total Fitur	Fitur Kls 1	Fitur Kls 2	Fitur Kls 3	Kelas
Usia 18-24	7	13	26	46	1.00	0.93	0.84	0.13
Usia 25-34	0	0	4	4	0.00	0.00	0.13	0.27
Usia 35-44	0	0	1	1	0.0		0.03	0.60
P	6	13	25	44	0.86	0.93	0.78	-
L	1	1	7	9	0.14	0.07	0.22	-
SMA/K	6	11	23	40	0.86	0.79	0.74	-
Sarjana	1	3	6	10	0.14	0.21	0.19	-
Diploma	0	0	1	1	0.00	0.00	0.03	-
TT	4	6	9	19	0.57	0.43	0.29	-
IG	3	7	20	30	0.43	0.50	0.65	-
WA	0	0	3	3	0.00	0.00	0.10	-
FB	0	0	3	3	0.00	0.00	0.10	-
YT	0	0	2	2	0.00	0.00	0.06	-

F. SH	7	12	31	50	1.00	0.86	0.97	-
F. BKS	0	2	1	3	0.00	0.14	0.03	-
F. Jarang	0	0	3	3	0.00	0.00	0.10	
KH	2	3	11	16	0.29	0.21	0.34	-
KE	2	6	17	25	0.29	0.43	0.53	-
IP	0	0	6	6	0.00	0.00	0.19	-
UP	3	5	7	15	0.43	0.36	0.22	-
KR	0	0	1	1	0.00	0.00	0.03	-
Fashion/Kecantikan	6	12	32	50	0.86	0.86	1.00	-
Elektronik	1	1	3	5	0.14	0.07	0.09	-
Makanan/Minuman	0	1	2	3	0.00	0.07	0.06	-
Kategori Produk lainnya	0	0	2	2	0.00	0.00	0.06	-
Kelas Total	7	14	33	54	0.13	0.26	0.61	-

Tabel 1.3 Klasifikasi Metode Naïve Bayes untuk Semua Responden

No.	Inisial	Kelas Asli	P (K1)	P (K2)	P (K3)	Kelas Prediksi	Kecocokan
1	NS	1	0.016	0.023	0.107	3	Tidak
2	UR	1	0.020	0.020	0.068	3	Tidak
3	MR	2	0.002	0.001	0.008	3	Tidak
4	ADJ	3	0.000	0.000	0.012	3	Cocok
5	ABP	3	0.000	0.000	0.020	3	Cocok
6	AP	3	0.001	0.000	0.005	3	Cocok
7	NP	3	0.006	0.003	0.022	3	Cocok
8	RBL	3	0.000	0.000	0.020	3	Cocok
9	AZM	3	0.000	0.000	0.001	3	Cocok
10	WL	2	0.000	0.001	0.001	3	Tidak
11	MAZ	3	0.000	0.000	0.001	3	Cocok
12	ES	3	0.006	0.005	0.020	3	Cocok
13	HA	3	0.000	0.000	0.012	3	Cocok
14	NNI	3	0.000	0.000	0.020	3	Cocok
15	ND	3	0.002	0.001	0.004	3	Cocok
16	AM	3	0.000	0.000	0.007	3	Cocok
17	SJ	3	0.000	0.000	0.001	3	Cocok
18	SR	3	0.000	0.000	0.001	3	Cocok
19	FM	3	0.000	0.000	0.001	3	Cocok
20	EW	3	0.000	0.000	0.001	3	Cocok
21	HH	3	0.000	0.000	0.007	3	Cocok
22	SZ	3	0.006	0.005	0.020	3	Cocok
23	SU	2	0.006	0.003	0.022	3	Tidak
24	JA	2	0.000	0.000	0.001	3	Tidak
25	RH	1	0.020	0.001	0.003	1	Cocok
26	DB	2	0.000	0.000	0.001	3	Tidak
27	MF	1	0.000	0.000	0.001	3	Tidak
28	ANA	3	0.000	0.000	0.007	3	Cocok
29	ST	3	0.002	0.000	0.003	3	Cocok
30	LF	2	0.000	0.001	0.007	3	Tidak
31	KN	3	0.002	0.001	0.004	3	Cocok
32	RDA	3	0.006	0.003	0.022	3	Cocok
33	MY	3	0.000	0.000	0.020	3	Cocok
34	DL	3	0.000	0.000	0.004	3	Cocok
35	SU	3	0.000	0.000	0.001	3	Cocok
36	AR	3	0.000	0.000	0.001	3	Cocok
37	NR	2	0.006	0.005	0.020	3	Tidak

38	JO	3	0.002	0.001	0.004	3	Cocok
39	AI	3	0.006	0.005	0.020	3	Cocok
40	FS	3	0.002	0.001	0.007	3	Cocok
41	HA	2	0.006	0.005	0.020	3	Tidak
42	NF	3	0.000	0.000	0.001	3	Cocok
43	ES	2	0.003	0.003	0.022	3	Tidak
44	KA	3	0.000	0.000	0.004	3	Cocok
45	DW	2	0.000	0.000	0.001	3	Tidak
46	IUP	1	0.006	0.005	0.004	1	Cocok
47	DR	2	0.000	0.000	0.004	3	Tidak
48	FK	2	0.000	0.000	0.020	3	Tidak
49	NSP	2	0.006	0.005	0.020	3	Tidak
50	FA	1	0.000	0.000	0.001	3	Tidak
51	ZAW	3	0.000	0.000	0.020	3	Cocok
52	LD	3	0.002	0.001	0.004	3	Cocok
53	ND	3	0.002	0.000	0.003	3	Cocok
54	MH	3	0.000	0.000	0.012	3	Cocok

Hasil Akhir:

- Total responden : 54
- Jumlah Klasifikasi Cocok : 37
- Jumlah klasifikasi Tidak Cocok : 17
- Akurasi Model : ~68,5%

Tabel 1.3 Hasil klasifikasi Metode Metode Naïve Bayes

Jumlah Kelas	Probabilitas Prior (P(Kelas))
Kelas 1 (Sangat Berpengaruh): 7	$7/54 \cong 0.13$
Kelas 2 (Berpengaruh): 14	$14/54 \cong 0.26$
Kelas 3 (Cukup/Sedikit): 33	$33/54 \cong 0.61$

Pembahasan

Hasil pengujian model klasifikasi menggunakan algoritma Naïve Bayes menunjukkan bahwa sebanyak 37 responden berhasil diklasifikasikan sesuai dengan kelas aslinya dengan tingkat akurasi sekitar 68,5%, sedangkan 17 responden lainnya mengalami kesalahan klasifikasi, di mana sebagian besar kesalahan tersebut (15 responden) diprediksi masuk ke kelas 3. Kondisi ini terjadi karena algoritma Naïve Bayes mengambil keputusan berdasarkan perhitungan probabilitas dengan asumsi bahwa setiap fitur bersifat independen, sehingga kelas dengan probabilitas prior tertinggi dan nilai probabilitas kondisional terbesar, dalam hal ini kelas 3, cenderung lebih sering terpilih sebagai hasil prediksi. Selain itu, keterbatasan Naïve Bayes dasar juga terlihat dari adanya fitur tertentu yang memiliki probabilitas kondisional bernilai nol pada kelas tertentu, seperti platform YouTube pada kelas 1, yang menyebabkan nilai probabilitas total kelas tersebut menjadi sangat kecil atau nol, sehingga tidak mampu bersaing dengan kelas lain yang memiliki probabilitas lebih besar, khususnya kelas 3.

KESIMPULAN

Penelitian ini berhasil mengungkapkan peran signifikan media sosial dalam memengaruhi keputusan pembelian konsumen di Kabupaten Jember. Dengan menerapkan algoritma Naïve Bayes, ditemukan bahwa:

1. Dampak Konten Media Sosial: Konten yang dibagikan di platform media sosial, baik berupa hiburan maupun edukasi, terbukti menjadi faktor penting dalam proses pengambilan keputusan pembelian konsumen. Responden cenderung mencari informasi produk melalui media sosial sebelum melakukan pembelian.
2. Akurasi Model: Model Naïve Bayes yang digunakan dalam penelitian ini mencapai akurasi sekitar 68,5%. Meskipun hasil ini menunjukkan efektivitas model dalam memprediksi pengaruh konten, terdapat sejumlah responden yang tidak terkласifikasi dengan tepat, menunjukkan adanya ruang untuk perbaikan.
3. Identifikasi Jenis Konten: Jenis konten seperti konten hiburan dan edukasi memiliki pengaruh yang lebih besar dibandingkan dengan jenis konten lainnya. Hal ini menunjukkan pentingnya menciptakan konten yang menarik dan informatif untuk menarik perhatian konsumen.
4. Rekomendasi Strategis: Temuan ini memberikan wawasan bagi pelaku usaha lokal untuk merumuskan strategi pemasaran yang lebih efektif melalui media sosial, dengan fokus pada jenis konten yang paling berpengaruh terhadap keputusan pembelian.
5. Keterbatasan dan Arah Penelitian Selanjutnya: Penelitian ini juga mengakui keterbatasan metode Naïve Bayes, terutama dalam hal asumsi independensi antar fitur. Penelitian lebih lanjut dapat dilakukan dengan metode analisis lain untuk meningkatkan akurasi dan pemahaman tentang perilaku konsumen di berbagai konteks.

Secara keseluruhan, hasil penelitian ini menegaskan pentingnya memanfaatkan media sosial sebagai alat pemasaran yang strategis dan memberikan dasar yang kuat bagi pen-

DAFTAR PUSTAKA

- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior*, 61, 47–55. <https://doi.org/10.1016/j.chb.2016.03.003>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hutter, K., Hautz, J., Dennhardt, S., & Füller, J. (2013). The impact of user interactions in social media on brand awareness and purchase intention. *Journal of Business Research*, 66(9), 1390-1396. <https://doi.org/10.1016/j.jbusres.2013.02.008>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Kasali, R. (2021). *Transformasi Bisnis di Era Digital*. Jakarta: Penerbit Salemba Empat.
- Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson Education.
- Larose, D. T., & Larose, C. D. (2015). *Data mining and predictive analytics*. Wiley.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Mangold, W. G., & Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357–365. <https://doi.org/10.1016/j.bushor.2009.03.002>
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
- Pusat Penelitian Ekonomi dan Bisnis Universitas Jember. (2022). *Laporan Penelitian: Pengaruh Media Sosial terhadap Keputusan Pembelian di Kabupaten Jember*.
- Riyanto, B. (2023). *Analisis Algoritma Naïve Bayes dalam Studi Perilaku Konsumen di Jawa Timur*. Universitas Negeri Surabaya.
- Sugiyono. (2019). Metode penelitian kuantitatif, kualitatif, dan R&D. Alfabeta.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.