



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK IV - Surabaya, 27 April 2024

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : [10.31284/p.snestik.2024.5906](https://doi.org/10.31284/p.snestik.2024.5906)

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Penggunaan Metode Naive Bayes dengan Pembobotan Term Menggunakan R-Kuadrat untuk Analisis Sentimen Ulasan Pengguna Aplikasi Shopee

Anindya Berlinani Santoso*, Maftahatul Hakimah, Septiyawan Rosetya Wardhana

Institut Teknologi Adhi Tama Surabaya^{1,2,3}

e-mail: *anindyabs54@gmail.com

ABSTRACT

The rapid development of e-commerce has occurred in Indonesia, especially with the Shopee application. E-commerce has become popular due to the ease and convenience of shopping online, including discounts, promotions, and free shipping. However, there are also issues with fraud and long delivery times that make some people doubt e-commerce. Therefore, this research aims to conduct a sentiment analysis of the Shopee application on the Google Play Store using the Naive Bayes method with R-Square weighting. This method was employed to classify user opinions as positive or negative. The test results showed that the Naive Bayes method with R-Square weighting had a better accuracy of 81% than the Naive Bayes method without weighting of 42%. In conclusion, R-Square weighting in the Naive Bayes method could improve the performance of sentiment analysis for the Shopee application on the Play Store. With R-Square weighting, accuracy increased by 39% by 71%, and F1-score value by 47%. However, it should be noted that the precision decreased by 15%.

Keywords: E-Commerce; Google Play Store; Naive Bayes; R-Square; Sentiment analysis, Shopee

ABSTRAK

Perkembangan pesat *e-commerce* di Indonesia, khususnya aplikasi Shopee. *E-commerce* menjadi populer karena kemudahan dan kenyamanan dalam berbelanja secara online, termasuk diskon, promo, dan pengiriman gratis. Namun, ada juga masalah penipuan dan waktu pengiriman yang lama yang membuat beberapa orang meragukan *e-commerce*. Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap aplikasi Shopee di Google Play Store menggunakan metode Naive Bayes dengan pembobotan R-

Kuadrat. Metode ini digunakan untuk mengklasifikasikan opini pengguna menjadi positif atau negatif. Hasil pengujian menunjukkan bahwa metode Naïve Bayes dengan pembobotan R-Kuadrat memiliki akurasi yang lebih baik 81% dibandingkan dengan metode Naïve Bayes tanpa pembobotan 42%. Dari hasil pengujian tersebut, dapat disimpulkan bahwa penggunaan pembobotan R-Kuadrat pada metode Naïve Bayes mampu meningkatkan performa analisis sentimen terhadap aplikasi Shopee di Play Store. Dengan pembobotan R-Kuadrat, akurasi meningkat sebesar 39%, *recall* meningkat sebesar 71%, dan nilai *F1-score* meningkat sebesar 47%. Namun, perlu diperhatikan bahwa *precision* mengalami penurunan sebesar 15%.

Kata kunci: Analisis Sentimen; *E-Commerce*; Google Play Store; Naive Bayes; R-Kuadrat; Shopee

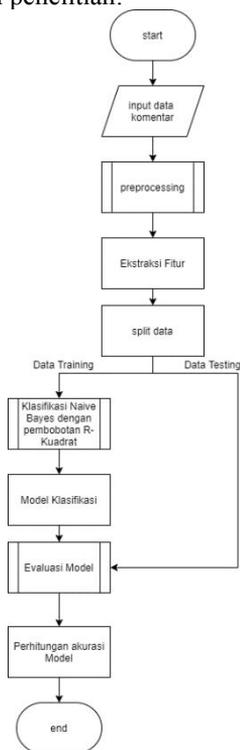
PENDAHULUAN

Situs belanja online *e-commerce* sedang mengalami perkembangan pesat. Salah satu alasan berkembangnya *e-commerce* adalah karena kemudahan dan kenyamanan dalam penggunaan serta memiliki banyak kelebihan yaitu dapat membeli barang tanpa harus datang ke toko sehingga dapat membeli di mana saja tanpa ada batasan dan dapat dilakukan selama 24 jam. Selain itu, keuntungan lain menggunakan *e-commerce* adalah adanya banyak diskon, promo bahkan gratis ongkos kirim ke pilihan masyarakat yang paling banyak diminati untuk berbelanja [1]. Dalam satu dekade terakhir ada berbagai *e-commerce* yang memasuki Indonesia seperti Tokopedia, Shopee, Bukalapak, Lazada, Blibli dan lain sebagainya [2]. Banyaknya *e-commerce* tersebut membuktikan bahwa kegiatan jual beli *online* kini sudah menjadi bagian dari kehidupan masyarakat Indonesia pada satu dekade terakhir. Tantangan yang dihadapi pada *e-commerce* sebagai bisnis ekonomi digital adalah permintaan konsumen yang menuntut pengiriman tepat waktu dan juga kepuasan terhadap keaslian barangnya [3]. Tantangan tersebut jika tidak dipenuhi maka opini masyarakat terhadap platform belanja online tersebut akan menjadi tidak menguntungkan terhadap penilaian *e-commerce* tersebut. Oleh karena itu penelitian ini akan melakukan analisis sentimen terhadap ulasan aplikasi *e-commerce* agar dapat memberikan informasi kepada para calon pengguna. Aplikasi *e-commerce* yang dibahas penelitian ini adalah aplikasi Shopee karena menurut data App Annie, aplikasi tersebut paling banyak diunduh di Indonesia.

Metode Naïve Bayes merupakan pengklasifikasian probabilitistik sederhana yang menghitung sekumpulan nilai probabilitas. Kelebihan metode ini adalah model yang sederhana namun dapat bersaing dengan model algoritma yang lain. Naïve Bayes memiliki asumsi bahwa setiap atribut pada sebuah dataset saling independen. Asumsi tersebut menjadi kelemahan dari Naïve Bayes karena jarang terjadi pada sebuah permasalahan riil [4]–[6]. Untuk mengatasi kelemahan dari Naïve Bayes tersebut, penelitian ini akan menggunakan pembobotan R-Kuadrat untuk melakukan klasifikasi ulasan pengguna aplikasi *Shopee*. Pada penelitian sebelumnya pembobotan R-Kuadrat pada Naïve Bayes bisa meningkatkan akurasi sebesar 3%, dibandingkan dengan Naïve Bayes standar [7]. R-Kuadrat digunakan untuk menilai seberapa besar pengaruh variabel independen tertentu terhadap variabel dependen. Manfaat metode R-Kuadrat adalah dapat menjelaskan seberapa jauh data dependen dapat dijelaskan oleh data independen. Kelebihan dari R-Kuadrat adalah dapat mengidentifikasi sekuat apa pengaruh yang diberikan variabel independen terhadap variabel dependen. Sedangkan kelemahan R-Kuadrat adalah bias terhadap jumlah variabel bebas yang dimasukkan dalam model regresi di mana setiap penambahan satu variabel bebas dan jumlah pengamatan dalam model akan meningkatkan nilai R-Kuadrat meskipun variabel yang dimasukkan tersebut tidak memiliki pengaruh yang signifikan terhadap variabel terikatnya [8].

METODE

Dalam penelitian ini akan menggunakan tahapan secara urut dan benar agar proses penelitian dapat tersusun dengan baik sehingga dapat menghasilkan alur penelitian yang diharapkan. Berikut gambaran umum penelitian:



Gambar 1. Tahapan Penelitian

Pengumpulan dan Pelabelan Data

Pengumpulan data yang digunakan dalam penelitian ini menggunakan teknik scraping. Web scraping merupakan teknik gabungan yang digunakan untuk mengumpulkan informasi dari website secara otomatis tanpa harus menyalinnya secara manual. Tujuan dari web scraping adalah pencarian informasi tertentu dan kemudian mengumpulkannya dalam website yang baru. Jumlah data yang diperoleh dari proses tersebut sebanyak 400 ulasan dengan 106 review negatif dan 294 review positif. Pengambilan data tersebut diambil pada range tanggal 08 Oktober – 22 Desember 2022. Berikutnya, setiap ulasan akan diberi label sesuai dengan rating bintang secara manual dengan ketentuan bintang 1 dan 2 untuk kelas negatif, serta bintang 4 dan 5 untuk kelas positif.

Tabel 1. Contoh Dataset Penelitian

No.	Ulasan	Kelas
1	Setelah update malah makin gajelas, fitur toko hilang yang ada di pojok kiri atas gak tau kemana, mau akses toko juga kesusahan.	Negatif
2	tolong dong shopee untuk filter pencarian di perbaiki lagi, masak uda di filter yang muncul di atas tetep aja iklan semua, sama produk2 yang gajelas, terlalu banyak barang kw/palsu, dan ongkir mahal.	Negatif
3	Aplikasi nya sangat bagus sangat membantu untuk kaum mager seperti saya dan ongkir murah □ tinggal tunggu dirumah barang sampai.	Positif

Pre-processing

Tahapan *pre-processing* adalah tahapan pertama sebelum dilakukannya proses klasifikasi yaitu melakukan pembersihan data dan mengurangi atribut yang kurang berpengaruh agar hasil klasifikasi mempunyai model performance yang baik. Berikut pada Tabel 2 merupakan contoh data setelah *pre-processing*:

Tabel 2. Contoh Hasil *Pre-processing* Data

Masukan	Keluaran
saya edit ulasan saya full karena sudah diverifikasi pembayaran terimakasih shopee sudah merespon	['saya', 'edit', 'ulasan', 'saya', 'full', 'karena', 'sudah', 'diverifikasi', 'pembayaran', 'terimakasih', 'shopee', 'sudah', 'merespon']

Perhitungan TF

TF (*term frequency*) yaitu proses menghitung frekuensi kemunculan fitur pada setiap dokumen setelah proses *pre-processing*. Berikut Contoh hasil perhitungan TF pada Tabel 3.

Tabel 3. Contoh Hasil Perhitungan TF

No.	Fitur	Term						DF
		D1	D2	D3	D4	D5	D6	
1	update	1	0	0	0	0	0	1
2	banyak	1	1	0	0	1	0	3
3	fitur	1	0	0	0	0	0	1
4	toko	2	0	0	0	0	0	2
5	akses	1	0	0	0	0	0	1

Naive Bayes dengan Pembobotan R-Kuadrat

Keakuratan pengklasifikasi Naïve Bayes ditingkatkan dengan mempertimbangkan bobot setiap atribut yang diperoleh dari nilai R-Kuadrat setiap term di kelas. Diawali dengan proses split data. *Dataset* di bagi menjadi 2, yaitu *training* dan *testing*. Pembagian dataset menggunakan 60% data *training* dan 40% sebagai data *testing*. Dengan proses *training* Naïve Bayes, probabilitas sebuah term ke-*j*, X_j , di kelas ke-*i*, Y_i , menggunakan Persamaan 1 berikut [9].

$$P(X_j|Y_i) = \frac{P(Y_i|X_j)P(X_j)}{P(Y_i)} \tag{1}$$

Nilai R-Kuadrat untuk term ke-*i*, R_i^2 diberikan oleh Persamaan 2 berikut ini [10]:

$$R_i^2 = \frac{((n)(\sum \mathbf{XY}) - (\sum \mathbf{X})(\sum \mathbf{Y}))^2}{(n(\sum \mathbf{X}^2) - (\sum \mathbf{X})^2)(n(\sum \mathbf{Y}^2) - (\sum \mathbf{Y})^2)} \tag{2}$$

dengan, n merupakan jumlah data training. Sehingga, bobot term ke-*i*, w_i , menggunakan Persamaan 3 berikut ini.

$$w_i = \frac{R_i^2}{\sum_{j=1}^m R_j^2} \tag{3}$$

dengan, m merupakan jumlah term. Klasifikasi sebuah ulasan menggunakan Naïve Bayes dengan pembobotan R-Kuadrat diberikan menggunakan Persamaan 4 berikut ini.

$$Y(X_k) = \arg \max_y P(Y_i) \prod_{j=1}^m P(X_j|Y_i) * w_j \tag{4}$$

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan 400 *review* yang diambil dari Google Play, khususnya pada aplikasi Shopee. Gambar 2 menunjukkan sampel data ulasan yang digunakan pada penelitian ini.

id_review	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedAtVersion	at	replyContent	repliedAt
1	bd41fc28-a	Anis Anisa	https://play-lh.g	Lemot anj*ng	1	0		2023-0		NaT
2	cc86d89a-6	ABCDE FGHIJK	https://play-lh.g	Mantap setiap hari ada promo dan diskon 50%	5	0	3.05.11	2023-0		NaT
3	11691094-	Ahyar Nasution	https://play-lh.g	Mksh shopee	5	0	2.99.23	2023-0	Hallo ka	2022-1
4	3f261a66-8	Kall Boyz	https://play-lh.g	good	5	0		2023-0		NaT
5	d44fe9c4-5	Dedi Setyawan	https://play-lh.g	Smoga sllu amanah n is the best	5	0	3.06.11	2023-0		NaT

Gambar 2. Hasil Pengumpulan Data

Tahapan selanjutnya yaitu proses *pre-processing* data dan perhitungan TF. Selanjutnya akan masuk ke dalam tahap menghitung R-Kuadrat dan menghitung bobot dari setiap term. Selanjutnya yaitu klasifikasi menggunakan metode Naïve Bayes diimplementasikan untuk memprediksi sentimen teks. Penelitian ini menggunakan rasio data training dan data testing 60:40. Data training sebesar 60% dari dataset digunakan untuk mengembangkan model, sementara 40% data testing digunakan untuk mengukur kinerja model. Probabilitas kelas dan probabilitas kata dalam setiap kelas dihitung. Prediksi sentimen dilakukan terhadap data testing. Probabilitas kata dalam setiap kelas dihitung dengan mengalikan probabilitas kemunculan kata dengan probabilitas kelas. Hasil dari sistem adalah sentimen untuk setiap komentar yang memiliki probabilitas tertinggi. Evaluasi menggunakan confusion matrix untuk klasifikasi hasil testing dan menghitung metrik seperti akurasi, presisi, *recall*, dan *F1-score*. Selain itu, pengukuran waktu yang dibutuhkan juga menjadi pertimbangan penggunaan metode.

Hasil evaluasi metode menunjukkan bahwa Nilai *F1-score* yang dihasilkan pada skenario *testing* dengan menggunakan metode *Naïve Bayes* dengan pembobotan R-Kuadrat bernilai 89% dan pada skenario testing dengan menggunakan metode *Naïve Bayes* menghasilkan nilai *F1-score* sebesar 42%. Skenario testing dengan menggunakan metode *Naïve Bayes* dengan pembobotan R-Kuadrat menghasilkan nilai akurasi, *recall* dan *F1-score* lebih besar namun menghasilkan nilai *precision* lebih kecil dan membutuhkan waktu *running* yang lebih lama dibandingkan dengan skenario *testing* menggunakan metode *Naïve Bayes*. Berikut merupakan rangkuman perbandingan kinerja *Naïve Bayes* dengan Pembobotan R-Kuadrat dengan *Naïve Bayes* tanpa pembobotan.

Tabel 4. Perbandingan Kinerja *Naïve Bayes* dengan Pembobotan R-Kuadrat dan Tanpa Pembobotan

Metode	Akurasi	Recall	Precision	F1-Score	Running Time
Naïve Bayes dengan R-Kuadrat	0,8125	0,9762	0,8200	0,8913	0,014
Naïve Bayes	0,4188	0,2698	0,9714	0,4224	0,005

KESIMPULAN

Naïve Bayes telah diimplementasikan pada analisis sentimen terhadap aplikasi Shopee di Play Store. Berdasarkan hasil pengujian metode *Naïve Bayes* dengan pembobotan R-Kuadrat ketika melakukan proses *testing*, diperoleh akurasi sebesar 81%, *recall* sebesar 98%, *precision* sebesar 82%, dan nilai *F1-score* sebesar 89%. Sedangkan hasil pengujian berdasarkan metode *Naïve*

Bayes menghasilkan nilai akurasi sebesar 42%, *recall* sebesar 27%, *precision* sebesar 97%, dan nilai *F1-score* sebesar 42%. Berdasarkan waktu, metode Naïve Bayes lebih cepat dibandingkan Naïve Bayes dengan pembobotan R-Kuadrat. Dengan hasil tersebut, penambahan bobot berdasarkan R-Kuadrat pada Naïve Bayes bisa meningkatkan akurasi sebesar 39%, *recall* meningkat sebesar 71%, *F1-score* meningkat sebesar 47% sedangkan *precision* menurun sebesar 15%. Penelitian berikutnya, pembobotan R-Kuadrat bisa diterapkan pada metode klasifikasi yang lain untuk analisis sentimen ulasan pengguna.

DAFTAR PUSTAKA

- [1] N. Ika, Z. Fitriyah, and N. C. Dewi, "Impulse Buying Di E-Commerce Shopee," *J. Ilmu Adm. dan Manaj.*, no. 3, pp. 57–61, 2020.
- [2] D. Witro, M. Izazi Nurjaman, D. Ayu, and C. Sofwan Al-Rasyid, "Analisis Strategi Promosi Tokopedia, Shopee, Bukalapak, Lazada, dan Blibli dalam Pemulihan Ekonomi Nasional," *Oikon. J. Kaji. Ekon. dan Keuang. Syariah*, vol. 2, pp. 33–42, 2021.
- [3] A. B. Setiawan, "Revolusi Bisnis Berbasis Platform Sebagai Penggerak Ekonomi Digital Di Indonesia," *Masy. Telemat. Dan Inf. J. Penelit. Teknol. Inf. dan Komun.*, vol. 9, no. 1, p. 61, 2018, doi: 10.17933/mti.v9i1.118.
- [4] A. Chaudhuri and T. P. Sahu, "Feature weighting for naïve Bayes using multi objective artificial bee colony algorithm," *Int. J. Comput. Sci. Eng.*, vol. 24, no. 1, pp. 74–88, 2021.
- [5] et al. Chen, Hong, "Improved naïve Bayes classification algorithm for traffic risk management," *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, pp. 1–12, 2021.
- [6] L. K. Foo, S. L. Chua, and N. Ibrahim, "Attribute weighted naïve bayes classifier," *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 1945–1957, 2022, doi: 10.32604/cmc.2022.022011.
- [7] R. Indransyah, Y. H. Chrisnanto, P. N. Sabrina, and S. Kom, "Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naive Bayes Classifier," *INFOTECH J.*, vol. 8, no. 2, pp. 60–66, 2022, [Online]. Available: <https://doi.org/10.31949/infotech.v8i2.3103>.
- [8] Henry TogarManurung, "Analisis Pengaruh Return On Equity, Earning Per Share, Net Profit Margin dan Market Value Added Terhadap Harga Saham (Studi Kasus Pada Perusahaan Manufaktur Go Public Sektor Food dan Beverages di BEI Tahun 2009–2013)," *Univ. Diponegoro Ekon. dan Bisnis / Manaj. Judul*, vol. 1, no. 1, pp. 1–72, 2015.
- [9] M. Hakimah and R. R. Muhimah, "Klasifikasi Penderita Penyakit Jantung Menggunakan Metode Naive Bayes dengan Chi-Square untuk Pemilihan Atribut," *Semin. Nas. Tek. Elektro, Sist. Inf. dan Tek. Inform.*, no. 1, pp. 257–262, 2021.
- [10] B. A. Mukhtar, "Analisis Performa Algoritme Weighted Naive Bayes Classifier," *Teknomatika*, vol. 10, no. 1, pp. 29–40, 2017.