

Klasifikasi Opini Masyarakat terhadap Jasa Ekspedisi J&T Express pada Media Sosial Twitter dengan *Naïve Bayes*

Bramantyo Ridho Ansyahry¹, Imam Husni Al Amin²

Teknik Informatika, Universitas Stikubank Semarang, Jl. Tri Lomba Juang, Mugassari, Semarang Selatan, Kota Semarang, Indonesia, 50241

e-mail: ¹bramantyoridhoansyahry@mhs.unisbank.ac.id, ²Imam@edu.unisbank.ac.id

Submitted Date: June 20th, 2023

Reviewed Date: June 26th, 2023

Revised Date: June 28th, 2023

Accepted Date: June 30th, 2023

Abstract

With the rise of online sales transactions through e-commerce and social media, the impact of changing consumer behavior. This has resulted in many sentiment assessments from users of these expedition services. Expedition services are now one of the most popular services. One of the shipping service companies operating in Indonesia is J&T Express. Through social media, especially Twitter with the number of followers up to 154,439 and the number of tweets up to 103,100, a user can form an opinion on the performance of J&T Express and get data as many as 1694 tweets. Machine learning algorithms are needed to enable sentiment analysis to be classified. One of them is the Naive Bayes algorithm. Before running the classification process, a preprocessing process is needed so that the dataset can be recognized by the system. Based on the tests conducted, the results obtained accuracy of 84%, precision of 76%, and recall of 87%. The results of this research show that the data can be used as a basis for evaluating business decisions.

Keywords: J&T Express; Naïve Bayes; Sentiment analysis; Classification

Abstrak

Dengan maraknya transaksi penjualan online melalui e-commerce dan media sosial, membuat dampak yang berefek pada berubahnya perilaku konsumen. Hal ini mengakibatkan banyaknya penilaian sentimen dari para pengguna jasa ekspedisi tersebut. Jasa ekspedisi kini menjadi salah satu jasa yang paling diminati. Perusahaan jasa pengiriman yang beroperasi di Indonesia salah satunya adalah J&T Express. Melalui media sosial, khususnya Twitter dengan jumlah pengikut hingga 154.439 dan jumlah tweet hingga 103.100, seorang pengguna dapat membentuk opini terhadap kinerja J&T Express dan mendapatkan data sebanyak 1694 tweet. Algoritma pembelajaran mesin diperlukan untuk memungkinkan analisis sentimen dapat diklasifikasikan. Salah satunya adalah algoritma Naive Bayes. Sebelum menjalankan proses klasifikasi, diperlukan proses preprocessing agar dataset dapat dikenali oleh sistem. Berdasarkan pengujian yang dilakukan, didapatkan hasil akurasi sebesar 84%, precision sebesar 76%, dan recall sebesar 87%. Hasil dari riset ini menampilkan bahwa data dapat digunakan sebagai dasar mengevaluasi keputusan bisnis.

Keywords: J&T Express; Naïve Bayes; Analisis sentimen; Klasifikasi

1 Pendahuluan

Di era digitalisasi global yang terus berkembang, media sosial telah berevolusi menjadi platform *microblogging* untuk berbagi pendapat dan pengalaman (Irawan et al., 2022). Dalam konteks tersebut, penelitian ini bertujuan untuk mengklasifikasikan opini publik mengenai layanan

ekspedisi J&T Express pada platform media sosial yakni Twitter menggunakan metode algoritma Naive Bayes.

Twitter telah berkembang pesat menjadi platform online di mana orang dapat membuat, mengirim, memperbarui, dan membaca pesan teks pendek yang disebut Tweet. Platform Twitter

bahkan secara tidak langsung dapat memengaruhi budaya tradisional, terutama ketika jurnalis mengumpulkan informasi melalui tweet dan merepost pesan-pesan berharga dari pengguna tentang peristiwa penting. (Ruz et al., 2020). Studi ilmiah tentang opini masyarakat dari tweet ini disebut analisis sentimen (Diamantini et al., 2019).

J&T Express sendiri merupakan perusahaan yang menyediakan jasa pengiriman dan logistik. Dengan semakin berkembangnya e-commerce di Indonesia, jumlah pengguna jasa pengiriman barang pun meningkat pesat. Dengan banyaknya pengguna J&T Express, tidak mengherankan jika kini banyak bermunculan jasa pengiriman baru di Indonesia. J&T Express menyediakan media komunikasi dengan pelanggan, termasuk media sosial seperti Twitter. (Ika et al., 2021).

Ketidakpercayaan pelanggan terhadap penjual menjadi tantangan tersendiri bagi para pebisnis ketika mengevaluasi kinerja layanan terhadap konsumennya. Para pebisnis perlu menyadari bahwa masalah muncul dari opini pelanggan. Namun, membaca dan mengklasifikasikan opini satu per satu memakan waktu dan dianggap tidak efisien. Penelitian ini didasarkan pada data Twitter-nya dan menganalisis berbagai pendapat pengguna tentang pengalaman mereka dengan J&T Express. Salah satu teknik klasifikasi yang paling populer, yaitu klasifikasi Naive Bayes, digunakan untuk menentukan opini positif, negatif, dan netral.

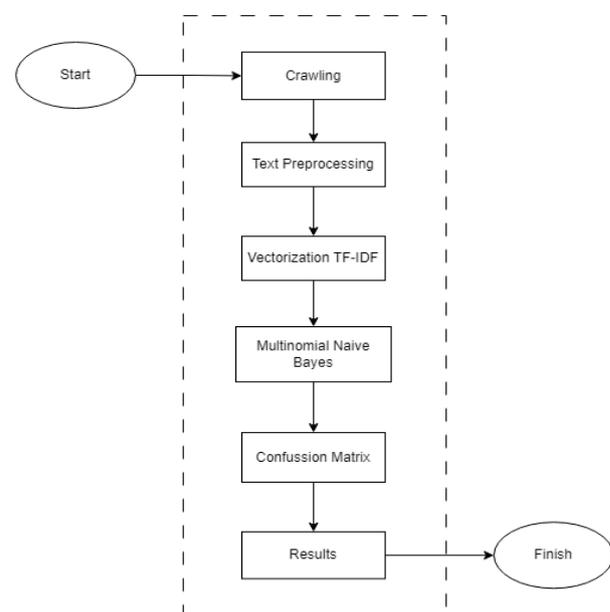
Hasil penelitian yang diharapkan dapat memberikan kontribusi yang lebih baik tentang persepsi masyarakat kepada layanan J&T Ekspres dan membantu pengusaha untuk meningkatkan kualitas layanan mereka. Selain itu, penelitian ini diharapkan dapat bermanfaat dalam pengembangan teknik analisis sentimen dalam konteks jasa pelayanan.

Mengingat pentingnya informasi yang terkandung dalam opini konsumen dan keunggulan metode Naive Bayes dalam mengekstraksi informasi dari opini konsumen, maka analisis sentimen yang dilakukan dengan menganalisis opini media sosial pengguna jasa ekspedisi J&T Ekspres perlu dilakukan. Twitter dianalisis dengan menggunakan Naive Bayes. Analisis sentimen digunakan untuk mengetahui kepuasan pengguna terhadap jasa pengiriman J&T Ekspres agar dapat meningkatkan kualitas pelayanan.

Dengan latar belakang di atas, tujuan dari penelitian ini yakni untuk melakukan analisis sentimen terhadap opini publik pengguna layanan J&T Express Expedition di media sosial Twitter dan mengukur performa algoritma Naive Bayes dalam melakukan klasifikasi analisis sentimen.

2 Metode Penelitian

Klasifikasi opini masyarakat mengenai jasa ekspedisi J&T Ekspres yang diperoleh melalui media sosial twitter. Setelah itu diletakkan pada sentimen untuk memperoleh sentimen positif, negatif dan netral. Ada beberapa tahapan untuk melakukan analisis sentimen berikut:



Gambar 1. Tahapan analisis sentimen

2.1 Crawling

Crawling adalah proses pengumpulan data secara otomatis dari berbagai situs web atau platform. Metode ini menggunakan program komputer atau bot yang secara sistematis memindai platform tertentu dan mengakses kontennya untuk mengekstrak informasi yang relevan. Data yang *crawling* dapat dianalisis, diproses, atau digunakan untuk berbagai tujuan tergantung pada kebutuhan Anda. (Handoko et al., 2022)

2.2 Text Preprocessing

Text preprocessing adalah mekanisme persiapan dari text mining. Tweet dapat berisi banyak pendapat yang diungkapkan oleh pengguna yang berbeda dengan cara yang berbeda. (Ruz et al., 2020). Tujuan papemrosesan teks adalah untuk

menghilangkan noise dari data untuk menguranginya dan membuatnya lebih terstruktur. Langkah-langkah dalam proses Text preprocessing:

1. Cleaning data
 Menghapus karakter khusus seperti tanda baca, angka, dan karakter non-teks yang tidak memberikan informasi yang relevan.
2. Stemming data
 Mengubah kata menjadi bentuk dasarnya (stem) dengan mengkonjugasikan akhiran dan menghilangkan akhiran turunan. Tujuan dari stemming adalah untuk menempatkan kata-kata yang berbeda dalam bentuk yang sama sehingga kata-kata dengan stem yang sama dianggap identik.
3. Tokenisasi data
 Proses membagi teks (kalimat, paragraf, dokumen, dll.) menjadi token.

2.3 Vectorization TF-IDF

Pembobotan TF-IDF merupakan tipe pembobotan yang selalu digunakan pada *information retrieval* (Yutika & Faraby, 2021). Teknik untuk mengganti teks berupa dokumen atau corpus menjadi representasi numerik berbasis vektor dengan menggunakan model *TF-IDF*. *TF-IDF* merupakan singkatan dari “*Term Frequency-Inverse Document Frequency*”. metode *TF-IDF* memberikan skor yang memperlihatkan tingkat kepentingan kata dalam dokumen dan korpus. Nilai *IDF* terlihat dari teks atau term itu sendiri. Jika tidak sering muncul, nilainya lebih besar dari nilai *IDF* yang sering muncul. Jika seluruh dokumen berisi term tertentu, nilai term *IDF* adalah 0 (Verawati & Audit, 2022) Hasil representasi dokumen sebagai vektor numerik menggunakan skor sebagai nilai untuk setiap vektor (Puh & Bagić Babac, 2022).

2.4 Multinomial Naïve Bayes

Algoritma *Multinomial Naive Bayes* adalah teknik pengkajian probabilistik berdasarkan teorema Bayes dan digunakan pada machine learning untuk *Natural Language Processing* (NLP) (Nurul Hidayah & Sahibu, 2021). Algoritma ini didasarkan pada rancangan term frequency, artinya berapa kali kata tersebut muncul dalam sebuah dokumen. Model ini memperhitungkan dua fakta, yakni apakah suatu kata tampak atau tidak

serta frekuensinya pada dokumen (Lesmana & Andarsyah, 2022). *Multinomial Naive Bayes* dirumuskan sebagai berikut:

Rumus 1.

$$P(p|n) \propto P(p) \prod_{1 \leq k \leq n} P(t_k|p) \text{ (Singh et al., 2019)}$$

Keterangan:

$P(t_k|p)$ = Probabilitas kemunculan dokumen teks
 n = jumlah total dokumen
 p = Dokumen yang menunjukkan polaritas atau kesamaan

Rumus 2.

$$P(t_k|p) = \frac{\text{count}(t_k|p)+1}{\text{count}(t_p)+|v|} \text{ (Singh et al., 2019)}$$

Keterangan:

$(t_k|p)$ = Jumlah tk yang muncul dalam dokumen dengan polaritas
 (t_p) = Jumlah token dengan polaritas dalam dokumen.

2.5 Confusion Matrix

Confusion matrix adalah tabel untuk mengevaluasi kinerja model klasifikasi. Tabel ini menunjukkan jumlah observasi yang diklasifikasikan dengan benar dan salah oleh model untuk setiap kategori klasifikasi. Mengklasifikasikan dengan tiga kelas atau label, *confusion matrix* 3x3 sering digunakan (Handayani & Sulistiyawati, 2021). Tabel ini terdiri dari tiga kolom dan tiga baris yang mewakili tiga kategori berbeda.

Tabel 1. *Confusion Matrix*

Actual	Predicted		
	Positive	Neutral	Negative
Positive	TP	FN	FN
Neutral	FP	TP	FN
Negative	FN	FP	TP

Penjelasan True Positif (TP), merupakan total dari satu kelas TRUE yang dapat diklasifikasikan dengan benar pada kelas TRUE. True Negative (TN), merupakan total dari satu kelas FALSE yang dapat diklasifikasikan dengan benar pada kelas FALSE. False Positif (FP), merupakan contoh keadaan dimana kelas TRUE yang diklasifikasikan salah dikelas FALSE. Sedangkan False Negatif (FN), merupakan keadaan kelas FALSE yang diklasifikasikan salah pada kelas TRUE.



3 Hasil Penelitian

Penelitian yang dilakukan menggunakan dataset hasil yang diperoleh dari komentar pengguna Twitter terhadap layanan ekspedisi J&T Express Indonesia. dataset yang dihasilkan menggunakan web scraping www.vicinitas.io dengan menghasilkan sebanyak 1694 dataset menggunakan keyword “jntexpressid”. Data tersebut akan diberi label sesuai sentimennya menggunakan tiga kelas positif, netral dan negatif.

Sebelum dataset diberi label, data terlebih dahulu melalui proses preprocessing yang terdiri atas cleaning, normalisasi, tokenisasi, dan stemming untuk mendapatkan hasil data yang bebas dari noise.

Tabel 2. Text preprocessing

Tweet	Proses	Hasil
Min paket JP08800676 83 kok gk bergerak sejak 3 hari lalu	Cleaning	min paket JP kok gk bergerak sejak hari lalu
https://t.co/Sdhulc01jr	Normalisasi	min paket JP kok gak bergerak sejak hari lalu
	Tokenisasi	“min”, “paket”, “JP”, “kok”, “gak”, “bergerak”, “sejak”, “hari”, “lalu”
	Stemming	“min”, “paket”, “JP”, “kok”, “gak”, “gerak”, “sejak”, “hari”, “lalu”

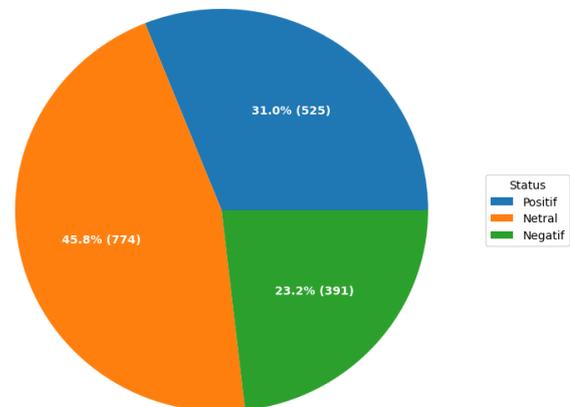
Semua data yang lolos tahap preprocessing diberi label sentimen, yaitu positif, negatif, dan netral. Hasil klasifikasi sentimen ditunjukkan pada tabel 3 dan divisualisasikan dalam diagram lingkaran sebagai berikut:

Tabel 3. Hasil klasifikasi sentimen

	Klasifikasi sentimen		
	Positif	Netral	Negatif
Jumlah	525	774	392

Dari Gambar 2, dapat dilihat bahwa diagram lingkaran, yang membagi total 1694 set

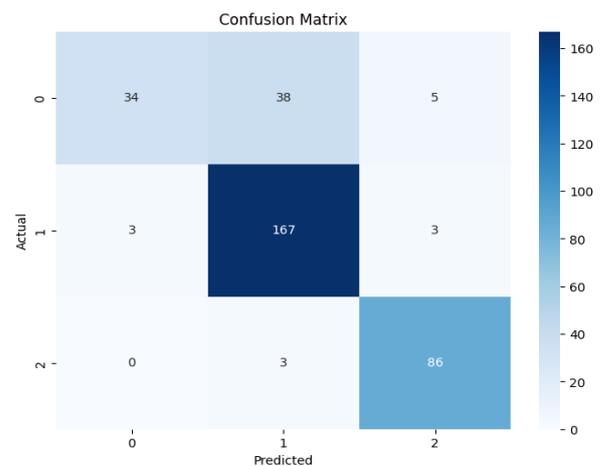
data ke dalam 3 kelas sentimen, memiliki 31,0% sentimen positif, 45,8% sentimen netral, dan 23,3% sentimen negative



Gambar 2. Diagram klasifikasi sentimen

Proses selanjutnya adalah membagi dua dataset menjadi data pelatihan dan data uji. Proporsinya adalah 80% data pelatihan dan 20% data uji. Data pelatihan digunakan untuk membangun acuan dan data uji digunakan untuk menguji data pelatihan selama konstruksi. Gunakan perintah `X_train, X_test, y_train, y_test = train_test_split(x,y,test_size=0.2,random_state=42)` untuk memproses.

Setelah menerapkan proses splitting, dilakukan training *Multinomial Naive Bayes* untuk memprediksi data testing untuk membuat *confusion matrix* dan menghitung hasil akurasi, presisi, dan recall. Proses ini menghasilkan *confusion matrix* seperti yang ditunjukkan di bawah ini.



Gambar 3. Confusion Matrix

- SAKTI), 6(2), 790–799.
- Ika, N., Kalingara, P., Pratiwi, O. N., & Anggana, H. D. (2021). Analisis Sentimen Review Customer Terhadap Layanan Ekspedisi Jne Dan J & T Express Menggunakan Metode Naïve Bayes Sentiment Analysis Review Customer of Jne and J & T Express Expedition Services Using Naïve Bayes Method. *E-Proceeding of Engineering*, 8(5), 9035–9048.
- Irawan, F. R., Jazuli, A., Khotimah, T., Studi, P., Informatika, T., Kudus, U. M., & Neighbor, K. (2022). Analisis Sentimen terhadap Pengguna Gojek Menggunakan Metode K-Nearest Neighbors. *JIKO (Jurnal Informatika Dan Komputer)*, 5(1), 62–68. <https://doi.org/10.33387/jiko>
- Lesmana, R., & Andarsyah, R. (2022). Model Klasifikasi Multinomial Naïve Bayes Untuk Analisis Sentiment Terkait Non-Fungible Token. *Jurnal Teknik Informatika*, 14(3), 135–139.
- Nurul Hidayah, Y., & Sahibu, S. (2021). Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 820–826. <https://doi.org/10.29207/resti.v5i4.3146>
- Puh, K., & Bagić Babac, M. (2022). Predicting sentiment and rating of tourist reviews using machine learning. *Journal of Hospitality and Tourism Insights*, 6(3), 1188–1204. <https://doi.org/10.1108/JHTI-02-2022-0078>
- Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Generation Computer Systems*, 106, 92–104. <https://doi.org/10.1016/j.future.2020.01.005>
- Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification. *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*, 593–596. <https://doi.org/10.1109/ICACTM.2019.8776800>
- Verawati, I., & Audit, B. S. (2022). Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By . u. 6, 1411–1417. <https://doi.org/10.30865/mib.v6i3.4132>
- Yutika, C. H., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *5(April)*, 422–430. <https://doi.org/10.30865/mib.v5i2.2845>