

Analisis Sentimen Komentar Pengguna Instagram Terhadap @Jntexpress_Bali dengan Menggunakan Metode Naïve Bayes

Josephine Kayla Riwanda¹, Helen², Aneesa Joenice Putri³, Sylvia Herman⁴, Achmad Yasser⁵, Alvin Nicolas Gunadi⁶, Viny Christanti Mawardi⁷

^{1,2,3,4,5,6,7}Universitas Tarumanegara

e-mail: ¹josephine.915210124@stu.untar.ac.id, ²helen.915210049@stu.untar.ac.id,

³aneesa.915210094@stu.untar.ac.id, ⁴sylvia.915210081@stu.untar.ac.id,

⁵achmad.915210130@stu.untar.ac.id, ⁶alvin.825210015@stu.untar.ac.id, ⁷viny@fti.untar.ac.id

Diterima	Direvisi	Disetujui
30-09-2023	28-11-2023	01-12-2023

Abstrak - Analisis sentimen komentar pengguna Instagram @jntexpress_bali dengan metode Naïve Bayes adalah penelitian yang bertujuan untuk memahami persepsi dan tanggapan pengguna terhadap layanan pengiriman J&T Express di wilayah Bali. Dalam penelitian ini, kami mengumpulkan dan menganalisis komentar yang ditemukan di postingan Instagram resmi @jntexpress_bali, dengan fokus pada komentar yang menyatakan ketidakpuasan terkait pengiriman paket yang tidak berhasil. Metode Naïve Bayes Digunakan untuk mengklasifikasikan sentimen dari komentar-komentar ini, yang dapat dibagi menjadi tiga kategori: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa sebagian besar komentar mengandung sentimen negatif, yang mencerminkan ketidakpuasan pelanggan terhadap pengiriman paket yang tidak dapat terkirim dengan baik. Penelitian ini memberikan wawasan berharga bagi J&T Express Bali untuk meningkatkan layanan mereka. Dengan mengidentifikasi sentimen negatif ini, perusahaan dapat memfokuskan upaya perbaikan pada area yang diperlukan, seperti meningkatkan kehandalan pengiriman dan komunikasi dengan pelanggan. Dengan demikian, diharapkan bahwa hasil penelitian ini mampu memberikan peningkatan terhadap perusahaan untuk memperoleh kepuasan dari pelanggan dan meningkatkan reputasi mereka dalam industri logistik di wilayah Bali.

Kata Kunci: JNT Express, Naïve Bayes, Komentar.

Abstract - Sentiment analysis of comments from Instagram users @jntexpress_bali using the Naïve Bayes method is research that aims to understand user perceptions and responses to J&T Express delivery services in the Bali area. In this research, we collected and analyzed comments found on the official Instagram posts of @jntexpress_bali, focusing on comments expressing dissatisfaction regarding unsuccessful package delivery. The Naïve Bayes method was used to classify the sentiment of these comments, which can be divided into three categories: positive, negative and neutral. The analysis results show that the majority of comments contain negative sentiment, reflecting customer dissatisfaction with packages that cannot be delivered properly. This research provides valuable insight for J&T Express Bali to improve their services. By identifying this negative sentiment, companies can focus improvement efforts on necessary areas, such as improving delivery reliability and communication with customers. Thus, it is aimed that the outcome of this research can assist companies to increase customer satisfaction and improve their reputation in the logistics industry in the Bali region.

Keyword: JNT Express, Naïve Bayes, Comments

PENDAHULUAN

Perkembangan yang terus berlangsung di bidang Teknologi Informasi (TI) secara signifikan terus mendukung pertumbuhan internet, memungkinkan para pengguna terutama pengusaha, untuk memimpin dan mengembangkan kegiatan komersial serta bisnis dengan lebih mudah. Kecepatan Internet telah memungkinkan banyak teknologi berbeda untuk memfasilitasi publikasi beragam informasi dalam bentuk elektronik, tidak

hanya dengan cara yang efisien dan mudah digunakan tetapi juga dengan menyediakan database terpusat dan sistem penentuan posisi umum. Siapapun bisa menggunakannya dengan cara yang tidak terlalu sulit. Menurut data yang dirilis WeAreSocial, 88,1 juta orang Indonesia menggunakan Internet dari total populasi 259 juta jiwa. Data yang dikutip dalam laporan Digital, Society, dan Mobile tahun 2016 menunjukkan bahwa pengguna aktif di Indonesia meningkat sebesar 15% dibandingkan Januari 2015. Yang jelas, tahun 2016 tingkat penetrasi internet

mencapai 34%, sedangkan tahun 2015 masih 28%. Pada saat yang sama, saat ini terdapat sekitar 3,4 miliar pengguna internet di dunia. Mengingat jumlah penduduk dunia yang berjumlah 7,4 miliar orang, hal ini berarti bahwa internet digunakan oleh hampir 50% penduduk dunia. Dengan semakin meluasnya penggunaan Internet, perdagangan elektronik (e-commerce) semakin banyak dilakukan oleh bisnis dengan berbagai ukuran, mulai dari perusahaan raksasa hingga toko online rumahan (Rehatalanit,2016). Dari data diatas yang menunjukkan bahwa pengguna pengguna e-commerce semakin banyak setiap tahunnya dan didukung oleh perkembangan internet yang semakin maju, penting adanya jasa e kspedisi yang mendukung mobilitas pengiriman barang. Beberapa ekspedisi yang terkenal di Indonesia saat ini adalah: JNT, Tiki, Anteraja, SiCepat, NinjaExpress dan ShopeeExpress.

J&T Express merupakan suatu perusahaan yang bergerak dibidang jasa pengiriman barang yang beroperasi di Indonesia. JNT didirikan pada tanggal 20 Agustus 2015. Layaknya perusahaan pada umumnya, perusahaan ekspedisi juga harus terus berinovasi agar dapat bersaing dengan kompetitor serta harus mempertahankan konsumen. Semacam inovasi teknologi pada sistem pengiriman, dan pelayan yang mengutamakan tanggung jawab serta meningkatkan loyalitas konsumen supaya dapat mempertahankan kepuasan konsumennya. Kualitas pelayanan jasa pengiriman memiliki kedudukan penting dalam menciptakan kepuasan pelanggan. Kualitas pelayanan harus dilakukan oleh perusahaan agar dapat bertahan dan tetap memperoleh kepercayaan konsumen. Menurut Tjiptono (2014), kualitas pelayanan berfokus pada upaya memenuhi kebutuhan dan keinginan, serta memberikan pasokan yang menyeimbangkan harapan pelanggan.

Menurut A.J. Burkart dan S. Medik (1981) seperti yang diartikan oleh Soekadijo (2000), pariwisata adalah aktivitas berpindah sementara dengan maksud melibatkan diri di luar lingkungan sehari-hari tempat tinggal dan pekerjaan. Bali, sebagai destinasi utama pariwisata di Indonesia, menawarkan berbagai tempat wisata alam, budaya, dan kuliner yang memiliki ciri khasnya. Keindahan pesona pariwisata yang dimiliki oleh Provinsi Bali diakui dengan penghargaan sebagai pulau terbaik di dunia dalam Trip Advisor Traveller's Choice Award 2017.

Maulana, Abdullah Kusuma (2022) menyatakan bahwa jumlah wisatawan domestik maupun asing dinyatakan sebagai pelaku ekonomi dalam aktivitas pariwisata dan mempunyai peranan penting sebagai tolak ukur penentuan bagaimana pengaruhnya kegiatan bepergian tersebut terhadap Pendapatan Asli Daerah Bali. Struktur perekonomian Bali yang bergeser dari tumpuannya di sektor Primer menjadi sektor Tersier yaitu sektor Pariwisata (Fahlevvi, 2022). Bali pun dikenal sebagai salah satu bantuan bagian perekonomian dikarenakan kegiatan

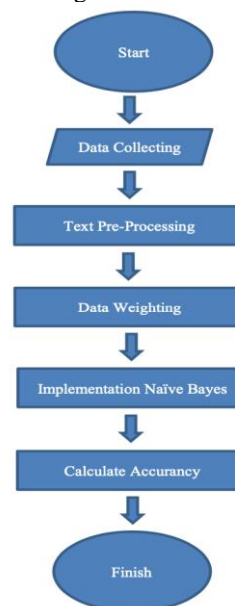
pariwisatanya (Suhendra, 2006). Perekonomian Bali pada triwulan IV 2022 pun tercatat masih meningkat dari triwulan sebelumnya. Ekonomi bali tercatat mengalami pertumbuhan sebesar 8,10% di mana pada triwulan sebelumnya hanya mencapai 6,61% (bi.go.id). Tujuan dari penelitian ini adalah melihat serta mempertimbangkan di mana salah satu industri terbesar di Bali adalah pariwisata, peneliti ingin melihat serta meneliti bagaimana industri pariwisata dan logistik serta perputaran ekonomi Bali yang tinggi tersebut bisa berpengaruh terhadap ekspedisi J&T.

METODE PENELITIAN

Berikut adalah Metode Penelitian yang di gunakan dalam Penelitian ini:

1. Skema Alur Penelitian

Analisis sentimen adalah sebuah kegiatan berbasis digital yang digunakan untuk menganalisis suatu pendapat atau opini yang diberikan oleh seseorang terhadap sebuah topik. Analisis sentimen memiliki keunggulan dalam mengklasifikasikan beberapa teks (tulisan) dari dokumen, kalimat, atau fitur. Kalimat tersebut dapat diklasifikasikan menjadi tiga sifat, yaitu: netral, positif dan negatif (Fahlevvi, 2022). Tim peneliti menggunakan bahasa pemrograman yaitu Python dengan berbagai *library* yang dipakai untuk menganalisa data secara lanjut . Python telah menjadi bahasa pemrograman yang populer dalam bidang analisis data dan kecerdasan buatan. Python memiliki sejumlah pustaka dan kerangka kerja yang kuat, seperti *scikit-learn* dan *matplotlib*, yang menyediakan algoritma klasterisasi yang efisien dan beragam.

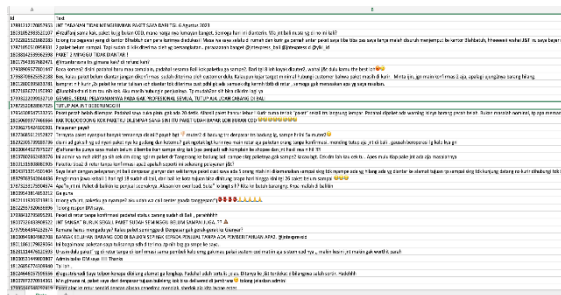


Sumber: Nitami Et All 2022

Gambar 1. Alur *Text Processing*

2. Pengumpulan Data

Proses pengumpulan data yang kami lakukan pada penelitian ini adalah mengambil data menggunakan extensions Google Chrome yaitu: InsC - Export Instagram Comments, pengumpulan data juga dapat dilakukan secara manual oleh tim peneliti. Data yang terkumpul berjumlah 120 dataset dan data tersebut berhasil dikumpulkan mulai dari tanggal 8 Agustus sampai 28 September 2023. Kemudian data tersebut disalin dan disimpan pada platform Microsoft Excel, dengan format penyimpanan xlsx.



Sumber: Hasil Penelitian

Gambar 2. Data yang Tersimpan pada Excel

3. Text Preprocessing

Text Preprocessing merupakan sebuah tahapan dalam proses text mining yang bertujuan untuk membuat text berbentuk data. Hal ini dilakukan agar text tersebut dapat menjalani pengolahan data pada tahapan berikutnya. Pada tahapan test preprocessing, contoh kegiatan yang dilakukan adalah pemecahan kalimat menjadi sekumpulan kata dan merubah kata yang bersifat kompleks menjadi lebih mudah dipahami. Contoh katanya adalah part of speech, tagging, dan parse tree.

Saat menggunakan text mining, peneliti harus menyiapkan data berupa kumpulan kata atau kalimat yang terstruktur. Tahapan yang terdapat pada pengimplementasian text mining yaitu tahap text preprocessing. Tahap text preprocessing merupakan tahapan seleksi data yang akan merubah data peneliti menjadi lebih terstruktur pada dokumen yang telah terlampirkan (Nitami et al, 2022).

Pada tahap preprocessing dokumen terdapat 4 langkah pemrosesan, yaitu:

- Case Folding, case folding berguna untuk mengubah seluruh huruf pada data menjadi huruf kecil.
- Normalisasi atau standardisasi, merupakan proses menghilangkan tanda baca, angka, dan karakter khusus dari teks atau dokumen.
- Tokenizing atau tokenisasi, merupakan proses pemisahan suatu kalimat menjadi satu kata dan biasanya dipisahkan dengan karakter seperti spasi, titik (.) dan koma (,).

- Stopword Removal merupakan langkah untuk memilah kata-kata yang berharga dari hasil token dengan menyimpan kata-kata yang berharga dan membuang kata-kata yang tidak berharga (Nitami et al, 2022).

4. Pembobotan Data

Pembobotan data adalah frekuensi kemunculan suatu kata dalam suatu kata. Semakin tinggi jumlah kemunculan suatu kata dalam suatu dokumen maka semakin besar bobotnya, sehingga nilai kepatuhannya semakin besar. Misalnya, jika muncul empat (4) kali maka kata atau istilah tersebut bernilai empat (4) kali.

Tabel 1. Contoh Pembobotan Data

Kata	Frekuensi
JNT	2
Pengiriman	1
Lambat	5

Sumber: Hasil penelitian

5. Naïve Bayes

Pengklasifikasi *Naïve Bayes* adalah pengklasifikasi generatif yang membangun model tentang bagaimana kelas menghasilkan data masukan tertentu (Jurafsky dan Martin, 2020). Kemudian temukan kelas-kelas yang paling mungkin dan bentuklah menjadi observasi untuk diproses.

Pengklasifikasi *Naïve Bayes* mewakili dokumen *text* sebagai sekumpulan kata tanpa memikirkan posisi setiap kata, misalnya menggunakan kata-kata yang diurutkan dalam kalimat seperti "Saya menyukai cerita filmnya" dan "Saya pikir filmnya bagus", yang mana menyederhanakan kata "saya" yang muncul dua kali di seluruh dokumen dengan mengutip kata "film" dua kali. Dan semua kata akan dihitung berdasarkan jumlah kata (Hakim, 2021). Tim peneliti menggunakan bahasa pemrograman Python untuk melakukan metode *Naïve Bayes* karena algoritma yang relatif sederhana dan cepat. Sehingga cocok untuk pemrosesan teks yang memungkinkan untuk analisis *real-time* atau analisis besar.

```
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_vec, y_train)
```

Sumber: Hasil Penelitian

Gambar 4. Penerapan Metode Naïve Bayes

Pengklasifikasi probabilistik akan mengembalikan kelas dengan probabilitas posterior maksimum dalam dokumen tertentu. Notasi untuk memperkirakan kelas yang diklasifikasikan dengan benar adalah sebagai berikut: $c = \text{argmaks } P(c | d)$. Hal ini bermula dari gagasan inferensi Bayesian yang diterapkan pada klasifikasi *text* pertama dalam sejarah pada tahun 1964.

Aturan Bayes disajikan untuk memberikan cara menguraikan probabilitas bersyarat $P(x|y)$ menjadi tiga probabilitas lainnya: $P(x|y) = (P(y|x)P(x))/(P(y))$ dan ditentukan *argmax* yang dihasilkan (Jauhainen, 2018).

6. Hitung Akurasi

Perhitungan akurasi mengenai hasil prediksi sentimen pada klasifikasi *Naive Bayes* dihitung menggunakan persamaan dibawah ini:

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

Keterangan : TP (*True Positive*), TN (*True Negative*), FN (*False Negative*), and FP (*False Positive*).

Data hasil klasifikasi dihitung berdasarkan banyaknya data benar dan banyak nya data yang salah. Dibawah ini merupakan contoh dari asumsi hasil klasifikasi dengan data uji sebanyak 50 data yang diperoleh dari tabel *confusion matrix* (Nitami et al, 2022).

HASIL DAN PEMBAHASAN

Pada penelitian yang dilakukan, dapat menghasilkan klasifikasi data ulasan pada pengguna jasa ekspedisi J&T Express. Tahapan yang dilakukan untuk proses klasifikasi yaitu data ulasan melewati beberapa tahapan pemrosesan. Untuk melakukan pemrosesan klasifikasi pada penelitian ini menggunakan *tools* yaitu Rapidminer untuk pemilihan data dan bahasa pemrograman Python khususnya *Naive Bayes*.

1. Text Preprocessing

Langkah-langkah pada text preprocessing untuk membersihkan dan memproses komentar. Ini bisa melibatkan langkah-langkah seperti menghapus karakter khusus, mengubah huruf menjadi huruf kecil, menghapus *stop words*, *stemming*, dan tokenisasi. Seperti di Gambar.5

```

import nltk
from nltk.corpus import stopwords
stop_words = set(stopwords.words("Indonesian"))
stemmer = PorterStemmer()

0.0s

#Text Preprocessing
def preprocess_text(text):
    if isinstance(text, str):
        text = text.lower() # Mengubah huruf menjadi huruf kecil
        text = re.sub(r"[^\w\s]", "", text) # Menghapus karakter khusus
        words = text.split()
        words = [word for word in words if word not in stop_words] # Menghapus stop words
        words = [stemmer.stem(word) for word in words] # Melakukan stemming
        return " ".join(words)
    else:
        return ""

data["text"] = data["text"].apply(preprocess_text)
    
```

Sumber: Hasil Penelitian

Gambar 5. Proses Melakukan *Text Preprocessing*

2. Akurasi

Berdasarkan hasil prediksi sentimen pada klasifikasi *Naive Bayes*, selanjutnya dihitung keakuratan algoritma *Naive Bayes* dalam mengklasifikasikan data review pengguna pengiriman J&T Express. Gambar 5 menunjukkan hasil sentimen positif dan negatif yang dilakukan oleh peneliti.

Akurasi Sentimen Positif: 0.16666666666666666
Akurasi Sentimen Negatif: 1.0

Sumber: Hasil Penelitian

Gambar 6. Akurasi Sentimen Positif dan Sentimen Negatif

Akurasi Sentimen Positif: 0.16667, Ini berarti bahwa model yang diteliti dengan *Naive Bayes* berhasil memprediksi sentimen positif dengan akurasi sekitar 16.67%. Dalam konteks analisis sentimen, ini menunjukkan bahwa model ini hanya memprediksi dengan benar sekitar 16.67% dari seluruh sentimen positif dalam data pengujian. Ini adalah tingkat akurasi yang rendah.

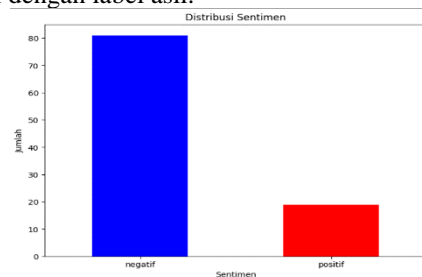
Akurasi Sentimen Negatif: 1.0, Ini berarti bahwa model Anda berhasil memprediksi sentimen negatif dengan akurasi 100%. Dalam konteks analisis sentimen, ini menunjukkan bahwa model ini memprediksi dengan benar seluruh sentimen negatif dalam data pengujian. Akurasi sebesar 1.0 (100%) adalah hasil yang sangat baik dan menunjukkan bahwa model sangat baik dalam mengenali sentimen negatif.

Akurasi Naive Bayes: 0.75

Sumber: Hasil Penelitian

Gambar 6. Akurasi pada *Naive Bayes*

Pada gambar 7 menunjukkan akurasi keseluruhan sebesar 0.75 (atau 75%) menunjukkan bahwa *Naive Bayes* berhasil memprediksi data dengan benar yaitu 75% dari seluruh data dalam set pengujian, termasuk sentimen positif dan negatif. Ini berarti dari seluruh data pengujian yang digunakan, 75% dari prediksi yang dibuat oleh model benar dan sesuai dengan label asli.



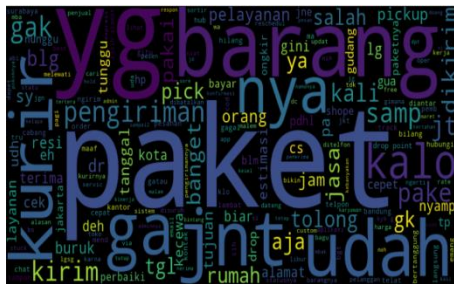
Sumber: Hasil Penelitian

Gambar 8. Distribusi Sentimen

Pada gambar 8 hasil distribusi yang sudah dilatih oleh Naïve Bayes terlihat hasil sentimen negatif lebih besar dari sentimen positif yang dimana sentimen negatif berkisar di angka 80 dan sentimen positif berkisar di angka 20. Oleh karena itu tim peneliti menyimpulkan bahwa komentar dari Instagram @jntexpress_bali lebih dominan komentar negatif dibandingkan komentar positif.

2. Visualisasi

Proses visualisasi dilakukan untuk memahami hasil kata atau istilah yang frekuensinya paling tinggi pada data penelitian yang dilakukan. Visualisasi data membantu dalam pengambilan keputusan yang berdasarkan bukti. Ketika data disajikan dengan cara yang visual, keputusan dapat diambil dengan lebih baik karena pemangku kepentingan dapat melihat gambaran lengkap dari situasi. (Few, *et all* 2007). Tim peneliti menggunakan visualisasi *word cloud* pada sentimen negatif dan positif.



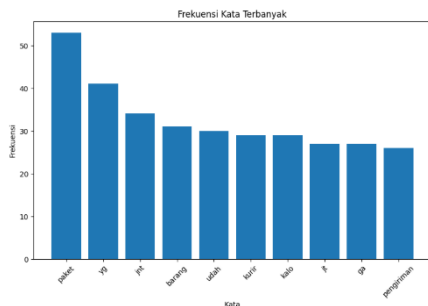
Sumber: Hasil Penelitian

Gambar 9. Visualisasi Word Cloud Negatif



Sumber: Hasil Penelitian

Gambar 10. Visualisasi Word Cloud Positif



Sumber: Hasil Penelitian

Gambar 11. Frekuensi Kata Terbanyak

Berdasarkan gambar 11 dari data visualisasi 10 frekuensi kata terbanyak yang paling sering muncul di dalam data tersebut yaitu paket, yg, JNT, barang, udah, kurir, kalo, JNT, ga, dan pengiriman.

KESIMPULAN

Dari hasil penelitian yang telah dilakukan oleh tim peneliti, dapat disimpulkan bahwa dalam pengujian klasifikasi sentimen menggunakan metode *Naïve Bayes* pada komentar pengguna Instagram (dengan memperhatikan nama pengguna Instagram yang tidak disebutkan), terdapat kecenderungan yang kuat untuk mengkategorikan komentar-komentar tersebut sebagai sentimen negatif. Hasil evaluasi performa model *Naïve Bayes* menunjukkan bahwa akurasi sentimen positif hanya mencapai 0.1666, mengindikasikan keterbatasan model dalam mengenali sentimen positif. Sebaliknya, akurasi sentimen negatif mencapai tingkat yang sangat baik, yakni 1.0 atau 100%, menggambarkan kemampuan model dalam mengenali sentimen negatif secara akurat.

Dalam distribusi sentimen, ditemukan bahwa sebanyak 20% dari total data merupakan sentimen positif, sedangkan 80% lainnya adalah sentimen negatif. Dengan demikian, berdasarkan mayoritas sentimen yang terdeteksi, model dapat dengan kuat diklasifikasikan sebagai model dengan kemampuan mendeteksi sentimen negatif. Kesimpulan ini menyoroti bahwa model yang digunakan dalam penelitian cenderung lebih baik dalam mengidentifikasi komentar dengan sentimen negatif daripada sentimen positif, dan demikianlah model ini cenderung masuk ke dalam kategori klasifikasi sentimen negatif.

Saran untuk penelitian ini melibatkan penggunaan metode seperti moving average, exponential smoothing, atau ARIMA (*Autoregressive Integrated Moving Average*) untuk meramalkan harga pangan di masa depan, perolehan data secara real-time melalui API, dan pemanfaatan berbagai pustaka yang relevan.

REFERENSI

- jet.co.id (2023, September). Improving Delivery Service Quality, J&T Express Wins Indonesia Original Brand Award 2023. Diakses pada 29 September 2023, dari <https://www.jet.co.id/news/news?show=197>
- Fahlevvi, M. R. (2022). Analisis Sentimen Terhadap Ulasan Aplikasi Pejabat Pengelola Informasi dan Dokumentasi Kementerian Dalam Negeri Republik Indonesia di Google Playstore Menggunakan Metode Support Vector Machine. *Jurnal Teknologi Dan Komunikasi Pemerintahan*, 4.
- Nitami, M. T., & Februariyanti, H. (2022). Analisis Sentimen Ulasan Ekspedisi J&T Express Menggunakan Algoritma *Naïve Bayes*. *Jurnal Manajemen Informatika & Sistem Informasi*, 5.

- <https://doi.org/10.36595/misi.v5i1>
- Jauhiainen, T., Lui, M., Zampieri, M., Baldwin, T., and Linden, K. (2018). Automatic language identification in texts: A survey.
- Jurafsky, D., Martin, J. H. (2020). Speech and Language Processing, Chapter: *Naïve Bayes and Sentiment Classification*.
- Few, S., & Edge, P. (2007). Data visualization: past, present, and future. IBM Cognos Innovation Center, 1-12.
- Hakim, B. (2021). Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning . *Journal of Business and Audit Information Systems*, 4. <https://doi.org/http://dx.doi.org/10.30813/jbase.v4i2.3000>
- Rehatalanit, Y. L. R. (2016). Peran E-Commerce Dalam Pengembangan Bisnis . *Jurnal Teknologi Industri* <https://doi.org/https://doi.org/10.35968/jti.v5i0.764>