

## ANALISIS SENTIMEN KEPUASAN CUSTOMER TERHADAP EKSPEDISI TIKI, SICEPAT EXPRESS DAN NINJA EXPRESS MENGGUNAKAN ALGORITMA NAIVE BAYES

### ANALYSIS OF CUSTOMER SATISFACTION SENTIMENT ON TIKI, SICEPAT EXPRESS AND NINJA EXPRESS EXPEDITION USING NAIVE BAYES

Nurhaliza Bin Aras<sup>1</sup>, Risawandi<sup>2</sup>, dan Lidya Rosnita\*<sup>3</sup>

<sup>1</sup>Mahasiswa Prodi Teknik Informatika, Universitas Malikussaleh

<sup>2,3</sup>Dosen Prodi Teknik Informatika, Fakultas Teknik Universitas Malikussaleh

Jalan Batam, Blang Pulo, Muara Satu, Lhokseumawe, Aceh, 24352

Corresponding Author: [lidyarosnita@unimal.ac.id](mailto:lidyarosnita@unimal.ac.id)

**Abstrak**—Perkembangan ilmu pengetahuan dan teknologi informasi pada masa ini sangat pesat. Adanya pemasaran produk secara global tersebut menjadikan perkembangan ekspedisi barang juga mengalami kemajuan yang signifikan. Kebutuhan penggunaan jasa ekspedisi barang yang dipergunakan masyarakat untuk memenuhi berbagai kebutuhannya sangat meningkat pesat. Hadirnya berbagai jasa ekspedisi barang tidak hanya mempermudah masyarakat namun juga para pengusaha atau seller. Penelitian ini bertujuan untuk menganalisis sentimen terhadap kepuasan customer ekspedisi yaitu tiki, sicepat express dan ninja express pada *twitter* dengan menggunakan metode Algoritma Naïve Bayes. Beberapa proses dalam melakukan klasifikasi sentimen, yang pertama melakukan koleksi data di *twitter* menggunakan *scraping* setelah itu pemberian labelling, kemudian dilakukan *text pre-processing* pada data yang meliputi *cleansing data*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Selanjutnya dilakukan proses klasifikasi pada data. Data yang digunakan dalam penelitian ini berjumlah 3000, setiap objeknya dengan jumlah 1000 data kemudian dibagi menjadi 3 kelas yaitu positif, negatif dan netral. Dari 3000 data dibagi menjadi 2 bagian yaitu 70% data training dan 30% data testing. Berdasarkan hasil evaluasi klasifikasi dengan algoritma Naïve Bayes menghasilkan akurasi yang sangat tinggi. Akurasi Sicepat Express sebesar 89,73%, presisi sebesar 58,81%, recall sebesar 40,1% dan f1-score sebesar 42,6%. Akurasi Ninja Express sebesar 80,66%, presisi sebesar 49,4%, recall sebesar 40,8% dan f1-score sebesar 41,5%. Akurasi Tiki sebesar 74,48%, presisi sebesar 65,42%, recall sebesar 57,14% dan f1-score sebesar 56,81%.

**Kata kunci:** Ekspedisi, Sentimen, Data, Naïve Bayes

**Abstract**— The development of science and information technology at this time is very rapid. The existence of global product marketing has made the development of freight forwarding also experience significant progress. The need for the use of freight forwarding services that are used by the community to meet their various needs is increasing rapidly. The presence of various freight forwarding services not only makes it easier for the community but also entrepreneurs or sellers. This study aims to analyze sentiment on customer satisfaction on expeditions, namely tiki, sicepat express and ninja express on *twitter* using the Naïve Bayes algorithm. There are several processes in classifying sentiments, the first is to collect data on *twitter* using *scraping* after that labeling, then *text pre-processing* is carried out on the data which includes *data cleansing*, *case folding*, *tokenizing*, *stopword removal*, and *stemming*. Furthermore, the classification process is carried out on the data. The data used in this study amounted to 3000, each object with a total of 1000 data was then divided into 3 classes, namely positive, negative and neutral. Of the 3000 data is divided into 2 parts, namely 70% training data and 30% testing data. Based on the results of the classification evaluation with the Naïve Bayes algorithm, it produces a very high accuracy. The accuracy of Sicepat Express is 89.73%, precision is 53.5%, recall is 40.1% and f1-score is 42.6%. Ninja Express accuracy is 80.66%, precision is 49.4%, recall is 40.8% and f1-score is 41.5%. Tiki's accuracy is 74.48%, precision is 65.42%, recall is 57.14% and f1-score is 56.81%.

**Keywords:** Expedition, Sentiment, Data, Naïve bayes

#### I. PENDAHULUAN

Ilmu pengetahuan dan teknologi informasi berkembang sangat pesat hingga saat ini. Peningkatan kemajuan informasi telah membawa dampak baru yaitu membuka peluang bisnis untuk menjual produk dalam skala global. Pengusaha menggunakan ini untuk mengembangkan bisnis

mereka melalui internet Karena itu, e-commerce menjadi tempat atau platform yang sangat bermanfaat bagi mereka yang ingin memulai dan mulai memasuki dunia bisnis melalui internet.

Adanya pemasaran produk global telah membuat kemajuan besar dalam perkembangan pengiriman barang.

Di era globalisasi, manusia dituntut untuk menjadi sangat dinamis. Selama masih ada aktivitas pasar bebas di Indonesia, menjadikan tidak terbatasnya laju perdagangan barang serta jasa. Kebutuhan penggunaan jasa ekspedisi barang yang dipergunakan masyarakat untuk memenuhi berbagai kebutuhannya sangat meningkat pesat. Hadirnya berbagai jasa ekspedisi barang tidak hanya mempermudah masyarakat namun juga para pengusaha atau seller. Para pengusaha atau seller Bisnis online ini menggunakan jasa kurir untuk mengirimkan produk ke pembeli di seluruh Indonesia dan luar negeri [1].

Untuk meningkatkan kinerja perusahaan, Kepuasan pelanggan merupakan aspek yang sangat penting. Perusahaan perlu fokus pada apa yang menurut pelanggan penting untuk mencapai kepuasan pelanggan. Oleh karena itu, Perusahaan sebaiknya melakukan penilaian untuk menentukan faktor apa saja yang mempengaruhi kepuasan pelanggan dan bagaimana memenuhi harapan pelanggan. Jika pembeli puas dengan manfaat pelayanan perusahaan maka loyalitas pelanggan akan terbangun. Namun, mempertahankan pelanggan setia merupakan salah satu tantangan yang dihadapi perusahaan[2].

Analisis sentimen yang merupakan proses dari penggunaan analisis teks agar mendapatkan dari bermacam sumber data diinternet dan platform lainnya. Analisis sentimen digunakan untuk menganalisis opini masyarakat terhadap jasa pengiriman ekspedisi tiki, sicepat express dan ninja express.

## II. STUDI PUSTAKA

### A. Analisis Sentimen

Analisis sentimen adalah proses identifikasi mengenali dan mengklasifikasikan polaritas teks dalam menentukan dokumen atau kumpulan sentimen positif, negatif atau netral. Pada Saat ini, banyak peneliti yang menggunakan analisis sentimen sebagai salah satu cabang penelitian ilmu computer seperti Twitter, sering digunakan untuk analisis sentimen dalam menentukan persepsi publik. Analisis sentimen juga bisa disamakan dengan opinion mining. Dalam analisis sentimen, data mining dilakukan untuk menganalisis, mengolah, dan mengubah data tekstual dalam suatu entitas, seperti layanan, produk, orang, fenomena, atau obyek. Analisis dapat berupa ulasan teks, forum, tweet atau blog, dengan preprocessing data termasuk enkripsi, kata kunci, penghapusan, radikalisasi, identifikasi sentimen, dan analisis jenis 'emosional. [3].

### B. Kepuasan

Kepuasan pelanggan adalah perasaan puas atau kecewa seseorang yang diakibatkan perbandingan antara kinerja produk yang diterima dengan harapan seseorang. Setia pelanggan terhadap bisnis asli harus dipertahankan dan ditingkatkan. Itu artinya kepuasan pelanggan adalah awal dari terciptanya kepercayaan pelanggan terhadap bisnis. Untuk meningkatkan kepuasan, perusahaan perlu menambahkan nilai yang dapat membuat pelanggan mereka mendapatkan apa yang mereka bayar atau lebih dari apa yang mereka inginkan. mereka menunggu, sehingga mereka dapat bertahan dan melakukan lebih banyak pembelian. Referensi ini sangat diperlukan, karena mengetahui tingkat kepuasan pelanggan, pelanggan tersebut secara otomatis akan menjadi pelanggan setia

perusahaan dan menjadi pelanggan aktif perusahaan. Dengan mengetahui pelanggan aktif, perusahaan dapat mengetahui pendapatan tetap untuk setiap periode. Untuk itu diperlukan pengolahan data tentang pelanggan aktif untuk mengetahui jumlah pelanggan aktif dalam kurun waktu tertentu. Jika jumlah pelanggan aktif bertambah, diperlukan strategi untuk mempertahankan dan meningkatkan jumlah tersebut. Jika jumlah pelanggan aktif berkurang, diperlukan strategi lain untuk menambah jumlahnya pelanggan tersebut[4].

### C. Data mining

Data mining adalah ilmu yang digunakan untuk menganalisis sebuah data, mengkategorikan, mengelompokkan dan menyimpulkannya, dalam beberapa proses tersebut terdapat teknik mengelompokkan data pada data mining yang yaitu klasifikasi [5].

### D. Text Mining

Text mining adalah data yang berupa *text* dengan sumber data diperoleh dari dokumen. Teks mining juga merupakan salah satu teknik yang dilakukan untuk klasifikasi yang bertujuan untuk mencari kembali kata yang bisa mewakili isi dari sebuah dokumen yang nantinya dapat dilakukan proses analisa dan mencari hubungan antar dokumen [6]

### E. Confusion matrix

Confusion matrix adalah metode untuk mengevaluasi dengan menggunakan tabel matriks. Hasil pengujian metode dianalisis dengan menggunakan tabel (confusion matrix) untuk mendapatkan skor/nilai akurasi, presisi dan recall.

*Confusion Matriks* disajikan dengan tabel yang menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan dengan salah.[7].

Tabel 1. Confusion Matrix

| Data Aktual | Data Prediksi |         |        |
|-------------|---------------|---------|--------|
|             | Positif       | Negatif | Netral |
| Positif     | TP            | FP      | PN     |
| Negatif     | FN            | TNg     | NgN    |
| Netral      | NP            | NNg     | TN     |

Keterangan:

|                               |   |
|-------------------------------|---|
| <i>True Positive (Tp)</i>     | =kelas kata prediksi benar bernilai positif |
| <i>False Negative (FN)</i>    | =kelas kata negatif terprediksi positif     |
| <i>Neutral Positive (NP)</i>  | =kelas kata netral terprediksi positif      |
| <i>False Positive (FP)</i>    | =kelas kata positif terprediksi negatif     |
| <i>True Negative (TNg)</i>    | =kelas kata prediksi benar bernilai negatif |
| <i>Neutral Negative (NNg)</i> | =kelas kata netral terprediksi negatif      |
| <i>Positive Neutral (PN)</i>  | =kelas kata positif terprediksi netral      |
| <i>Negative Neutral (NgN)</i> | =kelas kata negatif terprediksi             |

netral  
*True Neutral (TN)* = kelas kata terprediksi netral

Selain itu kita dapat menghitung nilai tersebut di antaranya *Accuracy, Precision, Recall*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(1)$$

Akurasi: ini adalah ntuk mengukur tingkat akurasi dari pengklasifikasi data yang akan dievaluasi.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

Presisi: Ini adalah rasio jumlah sentimen yang diprediksi secara akurat dengan jumlah total sentimen yang diprediksi.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(3)$$

Recall: Ini adalah rasio jumlah sentimen yang diprediksi secara akurat dengan jumlah total sentimen aktual.

$$F-1 \text{ Score} = \frac{2 (Recall \times precision)}{recall+Precision} \dots\dots\dots(4)$$

F-1 Score: itu menggambarkan perbandingan presisi dan rata-rata presisi dan recall. Akurasi sebagai tolak ukur kinerja algoritma jika data kita memiliki jumlah salah negatif dan salah positif yang sangat dekat (simetris). Tetapi jika angkanya tidak mendekati, gunakan fl-score sebagai referensi.

#### F. Algoritma Naïve Bayes

Metode Naive Bayes adalah metode klasifikasi untuk analisis sentimen dalam text mining. Pendekatan ini mungkin bagus dalam klasifikasi dalam hal akurasi dan komputasi data. Naive Bayes banyak digunakan dalam teknik klasifikasi, khususnya Twitter menggunakan beberapa metode seperti Unigram Naive Bayes, Multinomial Naive Bayes, dan Maximum Entropy Classification. Fitur utama klasifikasi Naive Bayesian adalah untuk mendapatkan hipotesis yang kuat dari kondisi atau peristiwa apa pun. Perhitungan kategori probabilitas pada Naive Bayes menggunakan metode algoritma Naïve Bayes, dengan menggunakan persamaan [3].

Persamaan dari teorema Bayes adalah:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \dots\dots\dots(5)$$

Keterangan:

- X = Data pada kelas yang belum diketahui
- Y = Hipotesis data x merupakan suatu class spesifik
- P(Y|X) = Probabilitas hipotesis berdasarkan kondisi
- P(Y) = Probabilitas hipotesis
- P(X|Y) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas Y

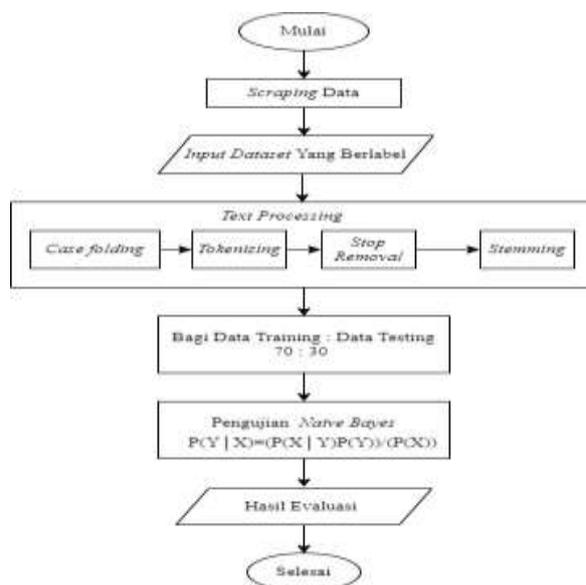
#### G. Text processing

*Text preprocessing* merupakan langkah awal pengurangan noise pada data teks agar dapat diproses lebih lanjut untuk mengurangi noise. *Preprocessing* teks mencakup semua rutinitas dan proses untuk menyiapkan data untuk digunakan dalam fungsi pengambilan data dari sistem penambahan teks. Data teks diproses dalam beberapa tahapan *text pre-processing* yaitu *case folding, tokenization, filtering dan Stemming*. Kumpulan data yang bersih dibentuk dengan pra-pemrosesan teks,

kumpulan data yang dibentuk oleh proses ini memfasilitasi pemrosesan sistem.[8].

#### H. Skema sistem

Adapun skema dari Analisis Sentimen Pada *Twitter* Terkait Kepuasan *Customer* Tiki, Sicepat Express dan Ninja Express Menggunakan Metode *Naïve Bayes* yaitu:



### III. METODE PENELITIAN

#### A. Pengumpulan Data

Dalam penelitian ini pengumpulan data dilakukan dengan cara mengambil komentar-komentar atau opini dari *customer* pada media sosial *Twitter*. Data yang telah dikumpulkan akan dibagi menjadi dua yaitu data *training* dan data *testing*. Pengambilan data yang dibutuhkan pada *Twitter* menggunakan teknik *scraping*, kemudian disimpan ke *database* sistem. Setelah itu data kan dipisah menjadi data *training* dan data *testing*. Semua data harus diproses sesuai dengan tahapan dari *text preprocessing* untuk mendapatkan teks yang diperlukan dan menghilangkan teks yang tidak penting.

1. Penglabelan Manual  
 memberi label pada semua komentar atau opini menjadi tiga kelas yaitu positif, negatif dan netral. kemudian dataset tersebut dibagi menjadi dua yaitu data *training* dan data *testing*
2. Text processing  
*text preprocessing* mempunyai tujuan untuk merangkup data menjadi dimensi yang lebih kecil sehingga proses komputasi menghasilkan keakuratan yang lebih tinggi. Tahapan *text preprocessing* terdiri *cleaning, case folding, tokenizing, filtering dan stemming*.
3. Data Training  
 Tahapan ini melakukan pengklasifikasian terhadap data training yang telah masuk ke dalam tahapan text processing. Tahapan ini bertujuan untuk menghasilkan kelompok data *training* yang nantinya sistem *corpus*, yaitu kamus data teks yang menjadi pembanding dengan data uji.

#### 4. Data Testing

Data *testing* merupakan tahapan terakhir dalam penelitian ini dengan membandingkan data *testing* yang telah melewati tahapan *text preprocessing* dengan data *training*. Pada tahap ini data *testing* diukur jarak kemiripannya dengan *corpus* dari hasil klasifikasi pada data *training*.

#### IV. HASIL DAN PEMBAHASAN

Tabel 2. Pengujian Confusion Matrix Sicepat Express

| aktual  | Prediksi |         |        |
|---------|----------|---------|--------|
|         | Positif  | Negatif | Netral |
| Positif | 0        | 7       | 0      |
| Negatif | 0        | 265     | 1      |
| Netral  | 0        | 23      | 6      |

Tabel 3. Hasil evaluasi performansi

|           | Precision | Recall | F-1 score |
|-----------|-----------|--------|-----------|
| Positif   | 0         | 0      | 0         |
| Negatif   | 0,89      | 0,99   | 0,94      |
| Netral    | 0,85      | 0,20   | 0,34      |
| Rata-rata | 58,51%    | 40,1%  | 42,67%    |
| Accuracy  | 89,73%    |        |           |

Tabel 4. Pengujian Confusion Matrix Ninja Express

| aktual  | Prediksi |         |        |
|---------|----------|---------|--------|
|         | Positif  | Negatif | Netral |
| Positif | 10       | 31      | 1      |
| Negatif | 3        | 232     | 0      |
| Netral  | 2        | 21      | 0      |

Tabel 5. Hasil Evaluasi Performansi

|           | Precision | Recall | F-1 score |
|-----------|-----------|--------|-----------|
| Positif   | 0,666     | 0,238  | 0,3495    |
| Negatif   | 0,816     | 0,987  | 0,8929    |
| Netral    | 0         | 0      | 0         |
| Rata-rata | 49,4%     | 40,8%  | 41,5%     |
| Accuracy  | 80,66%    |        |           |

Tabel 6. Pengujian Confusion Matrix Tiki

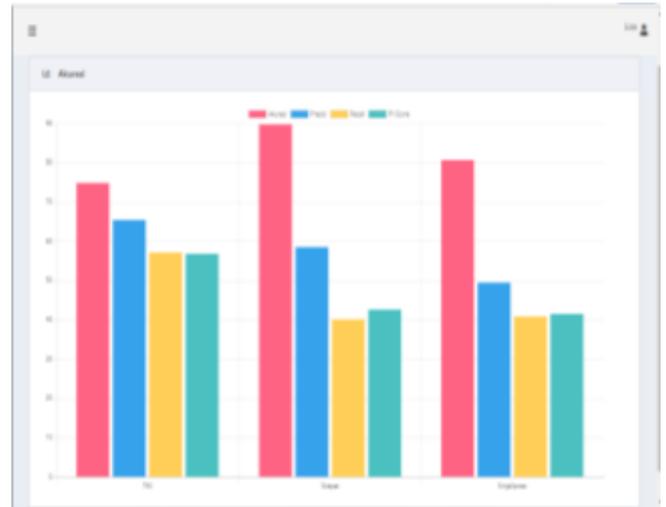
| aktual  | Prediksi |         |        |
|---------|----------|---------|--------|
|         | Positif  | Negatif | Netral |
| Positif | 56       | 23      | 0      |
| Negatif | 1        | 180     | 5      |
| Netral  | 1        | 50      | 2      |

Tabel 7. Hasil Evaluasi Performansi

|           | Precision | Recall | F-1 score |
|-----------|-----------|--------|-----------|
| Positif   | 0,9655    | 0,7088 | 0,81741   |
| Negatif   | 0,7114    | 0,9677 | 0,81998   |
| Netral    | 0,2857    | 0,0377 | 0,06660   |
| Rata-rata | 65,42%    | 57,14% | 56,81%    |
| Accuracy  | 74,84%    |        |           |

Berdasarkan hasil evaluasi klasifikasi dengan algoritma Naïve Bayes menghasilkan akurasi yang sangat tinggi. Akurasi Sicepat Express sebesar 89,73%, presisi sebesar 58,81%, recall sebesar 40,1% dan f1-score sebesar 42,6%. Akurasi Ninja Express sebesar 80,66%, presisi sebesar 49,4%, recall sebesar 40,8% dan f1-score sebesar 41,5%. Akurasi Tiki sebesar 74,48%, presisi sebesar 65,42%, recall sebesar 57,14% dan f1-score sebesar 56,81%.

Dari perhitungan confusion matrix di atas dibuat grafik tingkat keakurasian sebagai berikut:



#### V. KESIMPULAN

Berikut ini kesimpulan dari hasil penelitian metode *Naïve Bayes* dalam menganalisis sentimen terhadap Tiki, Sicepat Express dan Ninja Express:

1. Algoritma Naïve Bayes terbukti dapat digunakan dalam menganalisis sentimen terhadap Tiki, Sicepat Express dan Ninja Express. Penelitian ini juga menghasilkan sebuah sistem yang dapat mengkategorikan tweet terkait kedalam sentimen positif, negatif dan netral.
2. Penelitian ini menggunakan 3000 *tweet*. Kemudian menggunakan komposisi data sebesar 70% data training dan 30% data testing,
3. Berdasarkan hasil evaluasi klasifikasi dengan algoritma Naïve Bayes menghasilkan akurasi yang sangat tinggi. Akurasi Sicepat Express sebesar 89,73%, presisi sebesar 58,81%, recall sebesar 40,1% dan f1-score sebesar 42,6%. Akurasi Ninja Express sebesar 80,66%, presisi sebesar 49,4%, recall sebesar 40,8% dan f1-score sebesar 41,5%. Akurasi Tiki sebesar 74,48%, presisi sebesar 65,42%, recall sebesar 57,14% dan f1-score sebesar 56,81%.

#### REFERENSI

- [1] M. tania Nitami and H. Februariyanti, "Analisis Sentimen Ulasan Ekpedisi J&T Expres Menggunakan Algoritma Naive Bayes," 5(1),20-29, 2022, <https://doi.org/10.36595/misi.v5i1.396>
- [2] Yunitasari, H. S. Hopipah, and R. Mayasari, "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes," *Technomedia J.*, vol. 6, no. 1, pp. 99–110, 2021, doi: 10.33050/tmj.v6i1.1531.
- [3] W. A. Prabowo and C. Wiguna, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [4] A. Sumbaryadi, N. Iriadi, and F. I. Yuliasari, "Analisa Pengaruh Pelayanan Terhadap Kepuasan Customer Dalam Menggunakan Jasa Pengiriman Barang Di Jne Pondok Kelapa," *J. Infortech*, vol.

- 3, no. 1, pp. 59–64, 2021, doi: 10.31294/infortech.v3i1.10481.
- [5] Bustami, Fadlisyah, and S. Eliyanda, “Pengelompokan Siswa Penyandang Disabilitas Berdasarkan Tingkat Tunagrahita Menggunakan Metode Naïve Bayes,” *J. Teknol. Terap. Sains 4.0*, vol. 2, no. 1, pp. 1–17, 2021, doi: <https://doi.org/10.1976/tts%204.0.v2i1.3703>.
- [6] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, “Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC),” *Sistemasi*, vol. 10, no. 2, p. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [7] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, “Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN),” *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.
- [8] L. Ardiani, H. Sujaini, and T. Tursina, “Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak,” *J. Sist. dan Teknol. Inf.*, vol. 8, no. 2, p. 183, 2020, doi: 10.26418/justin.v8i2.36776.