

ANALISIS SENTIMEN PENGGUNA *TWITTER* TERHADAP FITUR GRATIS ONGKOS KIRIM PADA APLIKASI *SHOPEE* INDONESIA MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Alysa Salsabila, Nina Sulistiyowati, Mohamad Jajuli

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl.HS. RonggoiWaluyo, Telukjambe Timur, Karawang, Indonesia
1910631170159@student.unsika.ac.id

ABSTRAK

Shopee merupakan tempat transaksi perdagangan *online* yang populer di Indonesia. *Shopee* memiliki keunggulan adanya fitur gratis ongkos kirim ke seluruh wilayah di Indonesia yang dapat menarik perhatian sebagian besar masyarakat. Semakin berkembangnya *Shopee*, semakin banyak juga pengguna aplikasi *Shopee* yang mengeluh bahwa kebijakan baru terkait gratis ongkos kirim yang dikeluarkan oleh *Shopee* sedikit memberatkan pengguna. Oleh karena itu, untuk mengetahui opini tentang fitur gratis ongkos kirim pada aplikasi *Shopee* yang diberikan oleh pengguna *Twitter* dapat dilakukan dengan analisis sentimen. Proses analisis sentimen ini menggunakan metodologi *Knowledge Discovery in Databases* (KDD) dengan tahap awal melakukan pengambilan data dari cuitan di *Twitter* yang didapat melalui teknik *crawling*. Kemudian, data tersebut diseleksi untuk masuk ke tahap *preprocessing*. Setelah itu data masuk ke tahap *transformation* yaitu pemberian label positif dan negatif yang kemudian divalidasi oleh pakar bahasa. Selanjutnya tahap *data mining* menggunakan algoritma *K-Nearest Neighbor*, dengan melakukan pembagian *dataset* menjadi 4 skenario. Terakhir tahap *evaluation* untuk menguji model algoritma yang digunakan menggunakan *confusion matrix*. Berdasarkan data yang sudah diklasifikasi, menghasilkan sentimen positif sebanyak 130 data dan sentimen negatif sebanyak 1.045 data. Dengan pembagian data 80:20, penelitian ini menghasilkan nilai akurasi sebesar 91%. Dengan demikian sentimen pengguna *Twitter* terhadap fitur gratis ongkos kirim pada aplikasi *Shopee* diketahui lebih banyak mengandung sentimen negatif.

Kata Kunci: Analisis Sentimen, *Shopee*, *K-Nearest Neighbor*, *Knowledge Discovery in Databases*, *Confusion Matrix*.

1. PENDAHULUAN

Perkembangan teknologi saat ini mampu memberikan perubahan yang cukup besar bagi kehidupan manusia, contohnya pada perilaku jual beli [1]. Sebelum adanya perkembangan teknologi, manusia melakukan transaksi jual beli dengan cara mendatangi toko atau pasar. Namun setelah adanya perkembangan teknologi, cukup dengan menggunakan ponsel pintar manusia sudah dapat melakukan transaksi jual beli. *Shopee* merupakan salah satu tempat transaksi perdagangan *online* yang berbasis aplikasi *mobile* maupun *website*, yang populer di Indonesia [2]. Melalui *Shopee* pengguna dapat melakukan transaksi baik sebagai penjual maupun pembeli.

Shopee memiliki keunggulan dimana banyak harga yang dijual lebih murah dibandingkan harga yang biasa dijual di pasaran, serta adanya fitur pengiriman gratis ke seluruh wilayah di Indonesia yang tentu saja menarik perhatian sebagian besar masyarakat [3]. Semakin berkembangnya *Shopee*, semakin banyak juga masyarakat yang mengalami beberapa keluhan ataupun komentar terhadap fitur dan pelayanan yang diberikan oleh pihak *Shopee*. Seperti yang sedang ramai diperbincangkan saat ini, banyak pengguna aplikasi *Shopee* yang mengeluh dan berpendapat bahwa kebijakan baru terkait gratis ongkos kirim yang dikeluarkan oleh *Shopee* sedikit

memberatkan. Beberapa pengguna menyuarakan pendapatnya berupa pendapat positif dan negatif terhadap kebijakan baru fitur gratis ongkos kirim ini yang mereka tuangkan dalam cuitan pada aplikasi *Twitter*.

Selain sebagai wadah untuk berdiskusi, *Twitter* juga digunakan sebagai tempat bagi sebagian orang untuk mengungkapkan bentuk sentimen mereka, baik dalam sentimen positif, negatif, ataupun. Analisis sentimen dilakukan untuk mengetahui seberapa besar kepuasan pengguna terhadap layanan yang diberikan [4].

2. TINJAUAN PUSTAKA

2.1. *Twitter*

Twitter digunakan oleh banyak kalangan sebagai tempat untuk mengekspresikan opini seseorang dan juga wadah untuk memberikan tanggapan netral, positif, atau bahkan negatif terhadap isu yang sedang ramai diperbincangkan [5].

2.2. Analisis Sentimen

Analisis sentimen dilakukan dengan mengolah data berbentuk teks dengan tujuan memperoleh informasi berupa sentimen positif ataupun negatif [6]. Analisis sentimen berfokus pada pernyataan yang berisikan opini positif atau opini negatif, sehingga

analisis sentimen juga dapat diartikan *opinion mining* [7].

2.3. Shopee

Shopee adalah *platform marketplace online* di Indonesia yang mempermudah transaksi antar penjual dan pembeli melalui *gadget*. Aplikasi *Shopee* dirilis pada tahun 2015 di beberapa wilayah Asia Tenggara seperti, Malaysia, Vietnam, Singapura, Thailand, dan juga Filipina [8].

2.4. Python

Python merupakan bahasa pemrograman yang populer dan mudah dipahami dan salah satu keunggulannya adalah dapat membaca kode yang tinggi, sehingga membuat kode atau sintaks mudah untuk dipahami, *python* juga memiliki *library* yang beragam [9].

2.5. Google Colab

Google Colab merupakan tools gratis yang diciptakan dengan *environment jupyter* dan menunjang hampir semua *library* yang dibutuhkan dalam mengolah data [10]. *Google Colab* menggunakan *Python* sebagai bahasa pemrograman dengan format “*notebook*” yang dapat digunakan melalui *Mozilla Firefox*, *Chrome*, atau penjelajah website lainnya [11].

2.6. Data Mining

Data Mining ialah tahapan untuk mendapatkan koneksi, bentuk, dan tren baru yang bermakna dengan cara membagi data dalam jumlah yang besar kemudian disimpan dalam database, dengan bantuan teknologi pengenalan bentuk serta teknik statistik dan matematika [12].

2.7. K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* (KNN) ialah algoritma dengan cara kerjanya yaitu dengan menemukan jarak terdekat berdasarkan nilai *k* dalam *data training* yang paling mendekati dengan *data testing* [13].

2.8. Knowledge Discovery in Database

Knowledge Discovery in Databases atau KDD merupakan suatu metode dalam *data mining* yang digunakan untuk melakukan proses pengolahan data yang sudah tersedia pada *database* dan kemudian diolah untuk mendapatkan informasi yang baru dan lebih mudah dipahami [14].

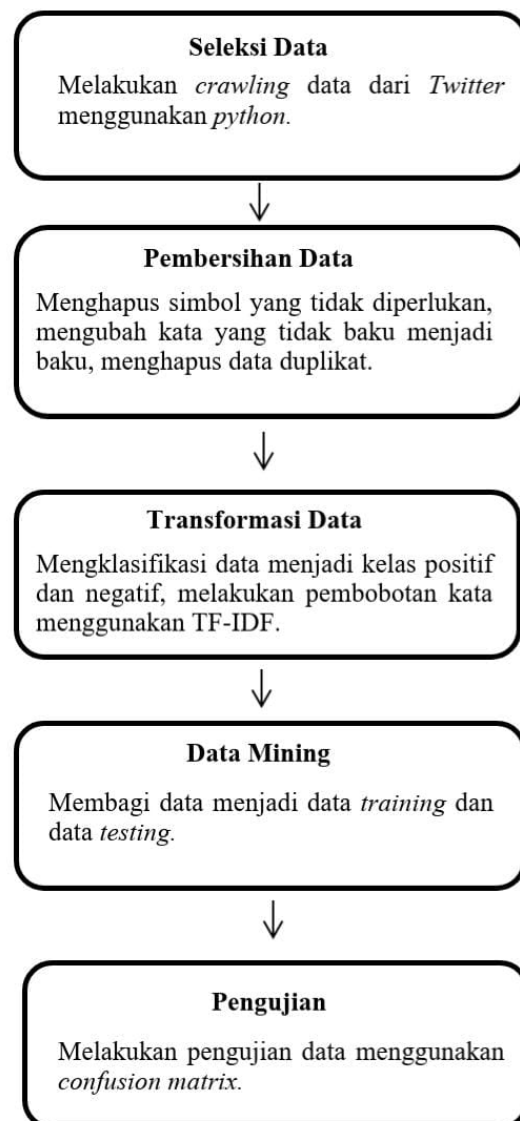
2.9. Term Frequency – Inverse Document Frequency (TF-IDF)

Metode Term Frequency Inverse Document Frequency (TF-IDF) mengombinasikan konsep frekuensi kebalikan kata pada suatu dokumen dan frekuensi kebalikan dokumen yang berisikan kata tersebut. Penghitungan bobot dengan bantuan TF-IDF

melibatkan perhitungan nilai TF per kata dengan bobot individu untuk setiap kata [15].

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam pengolahan dan proses penyaringan data adalah *Knowledge Discovery in Databases* (KDD), dengan 5 tahapan yaitu *Data Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation*.



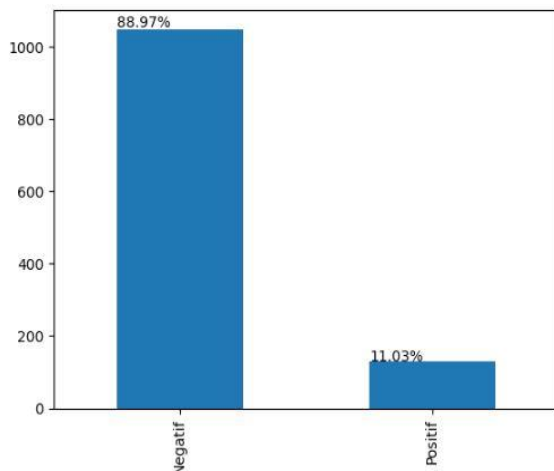
Gambar 1. Alur penelitian

4. HASIL DAN PEMBAHASAN

Dalam penelitian ini digunakan metode *K-Nearest Neighbor* sebagai algoritma untuk melakukan analisis sentimen pengguna *Twitter* terhadap fitur gratis ongkos kirim pada aplikasi *Shopee*. Dalam penelitian ini menggunakan bahasa pemrograman *Python* sebagai bahasa utama dengan bantuan tools *Google Colab*. Data yang digunakan yaitu berupa data *tweet* yang diperoleh menggunakan teknik *crawling* dengan bantuan *API Twitter*. Proses

Crawling dilakukan per bulan November 2022 sampai dengan bulan Maret 2023, dan memperoleh data sebanyak 1.180 data. Data tersebut selanjutnya diolah menggunakan metodologi *Knowledge Discovery in Databases* (KDD), serangkain proses yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, dan diakhiri dengan *evaluation*.

Hasil dari proses *preprocessing* berupa data yang telah bersih dari karakter yang tidak dibutuhkan, perubahan huruf kapital menjadi huruf kecil, pembagian kalimat menjadi kata, merubah kata yang tidak sesuai menjadi kata yang sesuai kaidah baku, dan menghapus kata yang memiliki imbuhan. Hasil dari tahapan *transformation* yaitu pemberian kelas positif dan negatif serta proses pemberian bobot pada kata menggunakan TF-IDF. Data yang sudah dibagi menjadi kelas positif dan negatif tersebut divalidasi oleh ahli bahasa. Klasifikasi hasil validasi oleh ahli bahasa mendapatkan 130 sentimen positif dan 1.045 sentimen negatif.



Gambar 2. Diagram hasil analisis sentimen

Berdasarkan Gambar 2 jumlah data sentimen negatif pada dataset lebih banyak dibandingkan dengan jumlah dataset sentimen positif. Dari total 1.175 dataset, sentimen positif berjumlah 130 data atau sama dengan 11,03% sedangkan sentimen negatif sebanyak 1.045 data atau sama dengan 88,97%.

Data yang sudah diberi label positif dan negatif kemudian masuk pada proses pembobotan kata dengan bantuan TF-IDF (*Term Frequency – Inverse Document Frequency*). Setiap kata sudah bernilai TF-IDF, dan nilai nol muncul untuk kata yang jarang ditemukan dalam dokumen.

Tahap selanjutnya merupakan tahap data mining menggunakan algoritma *K-Nearest Neighbor* melalui 4 skenario dengan perbandingan data latih dan data uji yaitu 90:10, 80:20, 70:30 dan 60:40. Dan diakhiri dengan tahapan evaluasi menggunakan *confusion matrix*.

Berikut penjelasan dari 4 skenario tersebut.

4.1. Skenario 1

Pada skenario pertama, data dibagi dengan rasio 90:10, di mana 90% merupakan data latih dan 10% merupakan data uji. Dengan kata lain, terdapat 1.061 data latih dan 118 data uji. Pembagian data 90:10 ini ditunjukkan dalam Gambar 3.

```
Data Uji: ((118, 2881), (118,)) Data Latih: ((1061, 2881), (1061,))
Jumlah data uji dengan sentimen positif: 10
Jumlah data uji dengan sentimen negatif: 108
Jumlah data latih dengan sentimen positif: 120
Jumlah data latih dengan sentimen negatif: 941
Negatif 1049
Positif 130
Name: Sentiment, dtype: int64
```

Gambar 3. Pembagian data 90:10

Berdasarkan Gambar 4, hasil evaluasi pada pembagian data 90:10 mendapat tingkat akurasi sebesar 90%.

```
Akurasi: 0.9067796610169492
=====
Confusion Matrix:
{confusion_matrix(Y_test, y_pred_sm)}
=====
```

	precision	recall	f1-score	support
Negatif	0.9292	0.9722	0.9502	108
Positif	0.4000	0.2000	0.2667	10
accuracy			0.9068	118
macro avg	0.6646	0.5861	0.6084	118
weighted avg	0.8844	0.9068	0.8923	118

Gambar 4. Hasil akurasi data 90:10

4.2. Skenario 2

Pada skenario kedua, data dibagi dengan rasio 80:20 di mana 80% merupakan data latih dan 20% data uji dengan kata lain terdapat 943 data latih dan 236 data uji. Gambar 5 menunjukkan pembagian data 80:20.

```
Data Uji: ((236, 2881), (236,)) Data Latih: ((943, 2881), (943,))
Jumlah data uji dengan sentimen positif: 22
Jumlah data uji dengan sentimen negatif: 214
Jumlah data latih dengan sentimen positif: 108
Jumlah data latih dengan sentimen negatif: 835
Negatif 1049
Positif 130
Name: Sentiment, dtype: int64
```

Gambar 5. Pembagian Data 80:20

Berdasarkan Gambar 6, hasil evaluasi pada pembagian rasio 80:20 menghasilkan tingkat akurasi sebesar 91%.

```

Akurasi: 0.9194915254237288
=====

Confusion Matrix:
{confusion_matrix(Y_test, y_pred_sm)}
=====

              precision    recall  f1-score   support

Negatif      0.9333      0.9813      0.9567      214
Positif      0.6364      0.3182      0.4242       22

accuracy      0.9195      0.9195      0.9195      236
macro avg     0.7848      0.6497      0.6905      236
weighted avg  0.9056      0.9195      0.9071      236

```

Gambar 6. Hasil akurasi data 80:20

4.3. Skenario 3

Pada skenario ketiga membagi data 70:30 dengan persentase 70% data latih dan 30% data uji atau sama dengan menggunakan 825 data latih dan 354 data uji. Gambar 7 adalah pembagian data 70:30.

```

Data Uji: ((354, 2881), (354,)) Data Latih: ((825, 2881), (825,))
Jumlah data uji dengan sentimen positif: 39
Jumlah data uji dengan sentimen negatif: 315
Jumlah data latih dengan sentimen positif: 91
Jumlah data latih dengan sentimen negatif: 734
Negatif 1049
Positif 130
Name: Sentiment, dtype: int64

```

Gambar 7. Pembagian data 70:30

Berdasarkan Gambar 8, hasil evaluasi pada pembagian data 70:30 menghasilkan nilai akurasi sebesar 90%.

```

Akurasi: 0.9011299435028248
=====

Confusion Matrix:
{confusion_matrix(Y_test, y_pred_sm)}
=====

              precision    recall  f1-score   support

Negatif      0.9094      0.9873      0.9467      315
Positif      0.6667      0.2051      0.3137       39

accuracy      0.9011      0.9011      0.9011      354
macro avg     0.7880      0.5962      0.6302      354
weighted avg  0.8826      0.9011      0.8770      354

```

Gambar 8. Hasil akurasi data 70:30

4.4. Skenario 4

```

Data Uji: ((472, 2881), (472,)) Data Latih: ((707, 2881), (707,))
Jumlah data uji dengan sentimen positif: 54
Jumlah data uji dengan sentimen negatif: 418
Jumlah data latih dengan sentimen positif: 76
Jumlah data latih dengan sentimen negatif: 631
Negatif 1049
Positif 130
Name: Sentiment, dtype: int64

```

Gambar 9. Pembagian data 60:40

Skenario keempat membagi data 60:40 dengan persentase 60% data latih dan 40% data uji atau sama dengan menggunakan 707 data latih dan 472 data uji.

Berdasarkan Gambar 10, hasil evaluasi pada pembagian data 60:40 menghasilkan nilai akurasi sebesar 88%.

```

Akurasi: 0.8898305084745762
=====

Confusion Matrix:
{confusion_matrix(Y_test, y_pred_sm)}
=====

              precision    recall  f1-score   support

Negatif      0.8910      0.9976      0.9413      418
Positif      0.7500      0.0556      0.1034       54

accuracy      0.8898      0.8898      0.8898      472
macro avg     0.8205      0.5266      0.5224      472
weighted avg  0.8749      0.8898      0.8455      472

```

Gambar 10. Hasil akurasi data 60:40

Hasil dari tahapan data mining berupa pembagian data ke dalam 4 skenario yaitu 90:10, 80:20, 70:30 dan 60:40, dari ke empat skenario tersebut memiliki nilai akurasi yang berbeda.

Tabel 1. Perbandingan hasil akurasi

Rasio Perbandingan	Akurasi
90:10	90%
80:20	91%
70:30	90%
60:40	88%

Pada Tabel 1 hasil evaluasi pada skenario pertama (90:10) menghasilkan nilai akurasi sebesar 90%, pada skenario kedua (80:20) menghasilkan akurasi sebesar 91%. Kemudian pada skenario ketiga (70:30) menghasilkan akurasi 90%, dan untuk skenario terakhir (60:40) menghasilkan akurasi 88%. Berdasarkan perbandingan tersebut data yang diklasifikasi dengan menggunakan 80% data latih, 20% data uji memperoleh nilai akurasi tertinggi yaitu, sebesar 91%.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, maka dapat ditarik beberapa kesimpulan yaitu: Hasil analisis dan klasifikasi sentimen berdasarkan cuitan pengguna media sosial *Twitter* terhadap fitur gratis ongkos kirim aplikasi *Shopee* kedalam dua kelas yaitu kelas positif dan negatif. Metodologi yang digunakan adalah *Knowledge Discovery in Database* (KDD) dimana memiliki 5 tahapan, yaitu *data selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation*. Tahap *labeling* dilakukan secara manual yang kemudian divalidasi oleh ahli bahasa, menghasilkan sentimen positif sebanyak 130 dan sentimen negatif sebanyak 1.045. Berdasarkan *labeling* tersebut, sebagian besar pengguna *Twitter*

yang menggunakan aplikasi *Shopee* memberikan sentimen negatif terhadap fitur gratis ongkos kirim yang disediakan *Shopee*. Hasil analisa dan evaluasi terkait opini atau pendapat masyarakat pada media sosial *Twitter* terhadap fitur gratis ongkos kirim pada aplikasi *Shopee* dengan menggunakan algoritma *K-Nearest Neighbor*, hasil nilai akurasi tertinggi adalah data pembagian 80:20, dengan nilai akurasi yang didapat sebesar 91%

Penelitian ini mengakui adanya kelemahan dan batasan tertentu, sehingga penting untuk meningkatkan penelitian dengan pendekatan serupa di masa depan. Saran untuk pengembangan penelitian mendatang adalah sebagai berikut: Dapat membandingkan lebih dari satu algoritma atau metode untuk mengetahui mana yang lebih baik untuk hasil akurasi. Menambah jumlah dataset yang digunakan sehingga jumlah atribut lebih beragam. Untuk penelitian selanjutnya dapat menggunakan beberapa kelas sentimen, seperti sentimen positif, negatif, dan netral.

DAFTAR PUSTAKA

- [1] Nur Ainiyah, "Remaja Milenial Dan Media Sosial : Media Sosial Sebagai Media Informasi Pendidikan Bagi Remaja Millennial," *Jurnal Pendidikan Islam Indonesia*, pp. 221-236, 2018.
- [2] Widya Sastika, "Analisis Kualitas Layanan dengan Menggunakan E-service Quality untuk Mengetahui Kepuasan Pelanggan Belanja Online Shoppe (Studi Kasus: Pelanggan Shopee di Kota Bandung 2017)," *Ikraith-Humaniora*, vol. 2, no. 2, pp. 69-74, 2018.
- [3] Romadhonna Siregar, Muhammad Isa, and Adanan Murroh Nasution, "Pengaruh Penggunaan Tagline Gratis Ongkir Dan Online Customer Rating Terhadap Minat Beli Produk Pakaian Muslim (Studi Pada Mahasiswa FEBI IAIN Padangsidimpuan)," *Profetik Jurnal Ekonomi Syariah*, pp. 133-149, 2022.
- [4] Brata Mas Pintoko and Kemas Muslim L, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier.," *e-Proceeding of Engineering*, pp. 8121-8130, 2018.
- [5] Debby Alita, Yusra Fernando, and Heni Sulistiani, "Implementasi Algoritma Multiclass SVM Pada Opini Publik Berbahasa Indonesia Di Twitter," *Jurnal TEKNOKOMPAK*, pp. 86-91, 2020.
- [6] Dwi Normawati and Surya Allit Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *Jurnal Sains Komputer & Informatika*, pp. 697-711, 2021.
- [7] Charles Bayu Saputra, Ari Muzakir, and Devi Udariansyah, "Analisis Sentimen Masyarakat Terhadap #2019GANTIPRESIDEN Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Clasifier," *Bina Dharma Conference on Computer Science*, pp. 403-413, 2020.
- [8] Heru Dwi Permana, Anindya Ananda Hapsari, Dian Nugraha, and Arief Jaenal, "Evaluasi Kinerja Sistem Aplikasi E-commerce Shopee menggunakan Metode PIECES Framework," *Jurnal ICT : Information Communication & Technology*, pp. 202-209, 2021.
- [9] Trie Maya Kadarina and Muhammad Hafidz Ibnu Hajar, "Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wiliayah Kembangan Utara," *Jurnal Abdi Masyarakat (JAM)*, pp. 11-16, 2019.
- [10] Cecilia Tania Emanuella, Musfita, and Armin Lawi, "Klasifikasi Suara Kucing dan Anjing Menggunakan Convolutional Neural Network," *Konferensi Nasional Ilmu Komputer (KONIK)*, pp. 321-327, 2021.
- [11] Khamzul Rifki, Joko Priambodo, and Arif Mustofa, "Pengenalan Plat Nomor dan Wajah Pengendara Menggunakan Convolutional Neural Network dan Metode Absolute Difference pada Sistem Gerbang Otomatis," *Jurnal Teknik ITS*, pp. 386-392, 2021.
- [12] Zulfa Nabila, Auliya Rahman Isnain, Permata, and Zaenal Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *Jurnal Teknologi dan Sistem Informasi (JTISI)*, pp. 100-108, 2021.
- [13] Syafaat Rizki Amardita, Adiwijaya, and Dwifabri Mahendra Purbolaksono, "Analisis Sentimen terhadap Ulasan Paris Van Java Resort Lifestyle Place di Kota Bandung Menggunakan Algoritma KNN," *Jurnal Riset Komputer*, pp. 62-68, 2022.
- [14] Dito Putro Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, pp. 437-444, 2020.
- [15] Nur Arifin, Ultach Enri, and Nina Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) Dengan TF-IDF N-GRAM Untuk Text Classification," *Satuan Tulisan Riset dna Inovasi Teknologi*, vol. VI, pp. 129-136, 2021.