



## KLASIFIKASI NAÏVE BAYES DAN CONFUSION MATRIX PADA PENGGUNA APLIKASI E-COMMERCE DI PLAY STORE

Mohamad Rizki Humaidi <sup>1)</sup>, Alfi Maulani <sup>2)</sup>

<sup>1</sup> Program Studi Matematika, Universitas Pamulang

<sup>2</sup> Program Studi Matematika, Universitas Pamulang

email: <sup>1</sup> mohamadrizkihumaidi27@gmail.com, <sup>2</sup> dosen02330@unpam.ac.id

### ARTICLE INFO

#### Article History:

Received : 28 July 2023

Accepted : 25 December 2023

Published : 31 December 2023

#### Keywords:

Confusion Matrix

E-commerce

Naïve Bayes Classification

Play store

Shopee

#### IEEE style in citing this article:

M. R. Humaidi and A. Maulani, "Klasifikasi Naïve Bayes dan Confusion Matrix pada Pengguna Aplikasi E-Commerce di Play Store", *Jurnal.ilmiah.informatika*, vol. 8, no. 2, pp. 132-139, Dec. 2023.

#### Corresponding Author:

Mohamad Rizki Humaidi

Universitas Pamulang

### ABSTRACT

Shopee is an e-commerce application that is very popular among Indonesians. Shopee is an online shopping service center that is in great demand by Indonesians and offers many types of products. In addition, the shopee application has features, one of which is shopee food and shopee pay, which distinguishes the shopee application from other e-commerce applications. Although there are many Shopee application users, of course not all reviews and ratings given by users are good as a reference for improving the Shopee application features. In conducting research, a method is needed that can classify review data into positive and negative reviews. One that provides a review and rating feature is the Play Store application. In this study, using 1528 review data, Positive and negative class labels are the categories used.. The machine learning classification method used is Naïve Bayes classification. The outcomes of the method's classification accuracy testing are measured using a confusion matrix. So that the accuracy result using Naïve Bayes classification is 0.87 or 87%. Based on these results, using the Naïve Bayes classification gets high accuracy results in the process of classifying review data on Shopee application user research in the Play Store.

© 2023 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC licence

## 1. PENDAHULUAN

Aktivitas pemasaran dan penjualan produk dan layanan secara online dikenal sebagai *e-commerce*. Selain itu, *e-commerce* juga mencakup langkah-langkah yang diambil untuk mendukung transaksi pasar, seperti pemasaran, layanan pelanggan, pengiriman, dan pembayaran [1]. Perkembangan dalam *e-commerce* adanya inovasi dalam bidang teknologi yang dapat memenuhi kebutuhan dalam kehidupan dimasyarakat, salah satu diantaranya yaitu yang terjadi aktivitas dalam berbelanja. Perkembangan dalam teknologi telah banyak melahirkan inovasi yang sebelumnya belum pernah ada dalam dunia bisnis, salah satu inovasi yaitu teknologi internet yang telah mendapat peningkatan dalam transaksi *e-commerce*. *E-commerce* telah mengalami pertumbuhan yang signifikan dengan seiring fokusnya kenyamanan dan aksesibilitas yang menyebabkan belanja *online* semakin banyak. Indonesia merupakan negara yang banyak menggunakan layanan *e-commerce* sebanyak 88,1%, persentase tersebut merupakan yang tertinggi di dunia pada tahun 2021 [2].

Pertumbuhan teknologi informasi membawa dampak terhadap budaya perdagangan, seorang penjual dan pembeli melakukan pertukaran untuk barang, jasa, dan informasi dengan tatap muka, dengan adanya teknologi informasi saat ini penjual dan pembeli melakukan proses transaksi tanpa terhalang waktu dan tempat yang terbatas [3]. Indonesia merupakan negara kepulauan sehingga banyak instansi besar yang ada di Indonesia bergerak dalam bidang *e-commerce* untuk berinovasi dalam meningkatkan pelayanan. Aplikasi shopee merupakan aplikasi dari salah satu *e-commerce* yang terdapat pada aplikasi play store. Aplikasi shopee berada pada urutan kedua dalam penggunaan layanan *e-*

*commerce* di Indonesia dengan rata-rata 131,2 juta pengunjung pada tahun 2022 di Indonesia [4].

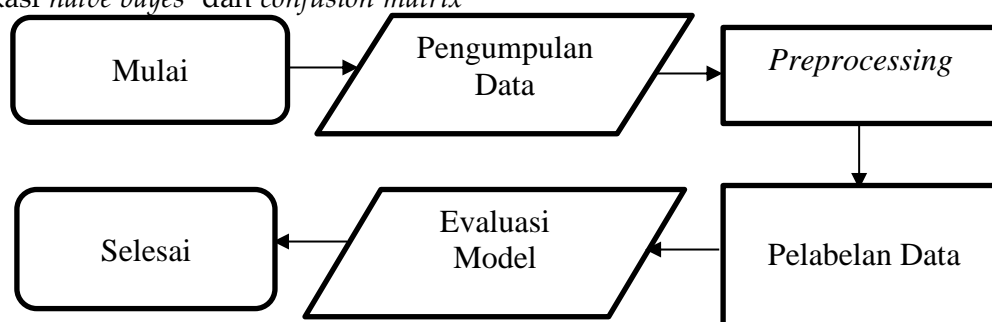
Aplikasi *google play store* terdapat fitur ulasan dan *rating* yang dapat digunakan oleh pengguna aplikasi yang di unduh pada aplikasi play store, sehingga dapat digunakan sebagai acuan untuk pengguna aplikasi yang tersedia di *google play store*. Ulasan dan *rating* yang diberikan oleh pengguna aplikasi shopee di play store pada umumnya berisikan saran yang positif maupun negatif.

Penelitian ini penulis akan mengumpulkan data ulasan dan *rating* yang diberikan pengguna aplikasi shopee di play store dengan *web scraping*. Metode *web scraping* memungkinkan untuk dapat menemukan data di situs web secara otomatis tanpa perlu mencari data secara manual [5]. Tahapan selanjutnya, dilakukannya penggunaan *machine learning* sebagai model klasifikasi yaitu dengan klasifikasi *naïve bayes* untuk mengklasifikasikan data pengguna menjadi negatif dan positif sebagai kelas, dengan mengukur akurasi sebagai tingkat hasil dari metode klasifikasi data yang digunakan menggunakan *confusion matrix*.

Adapun, penelitian yang telah dilakukan dengan digunakannya metode klasifikasi *naïve bayes* dan *confusion matrix* sebagai acuan penelitian yang telah dilakukan sebelumnya. Penelitian yang melakukan analisis pada banyaknya malware yang menyebabkan IDS (*Intrusion Detection System*), didapatkan hasil klasifikasi penelitian tersebut dengan akurasi sebesar 87% [6]. Penelitian lainnya, melakukan analisis terhadap banyaknya cuitan berasal dari twitter, didapatkan hasil klasifikasi penelitian dengan akurasi sebesar 82% [7]. Penelitian lainnya yang melakukan analisis terkait saat pandemi *covid-19* yang memberikan dampak yaitu transportasi umum, didapatkan hasil

klasifikasi penelitian dengan akurasi sebesar 83,8% [8].

Berdasarkan permasalahan dan rujukan dari hasil peneliti yang telah dilaksanakan sebelumnya, penulis akan melakukan penelitian dengan gunakan proses klasifikasi *naïve bayes* dan *confusion matrix* untuk dapat menyelesaikan permasalahan dalam penelitian ini, untuk bertujuan dapat menerapkan metode klasifikasi *naïve bayes* dan *confusion matrix*



Gambar 1. Diagram Alir Penelitian

### 2.1. Pengumpulan Data

Penelitian melakukan tahapan dalam mengumpulkan data dengan cara menggunakan *web scraping* sebanyak 1528 data, data yang digunakan yaitu data pada tanggal 10 Februari 2023 – 12 April 2023. Metode *scraping* data yang merupakan *extension* dari *google chrome*. Data yang digunakan yaitu data ulasan berbahasa Indonesia. Menggunakan alat bantu *python software* Anaconda.

### 2.2. Preprocessing Data

*Preprocessing* data melibatkan pengolahan data sebelum membuat model, dataset yang digunakan akan dibersihkan lebih baik menghindari kata-kata yang tidak relevan, seperti tanda baca, angka, simbol-simbol, dan *link url*. [9]. Dalam Penelitian ini terdapat 5 tahapan dalam *preprocessing* yaitu, *tokenizing*, *normalization*, *case folding*, *stemming*, dan *filtering*.

### 2.3. Pelabelan Data

Pelabelan data yaitu merupakan tahapan lanjutan setelah melakukan

dalam mengklasifikasi data pengguna aplikasi shopee di play store dalam kelas positif dan negatif.

## 2. METODE PENELITIAN

Penelitian ini terdapat Gambar 1 yang menunjukkan diagram alir metode penelitian yang akan digunakan dalam penelitian ini.

tahapan *preprocessing* data. Tahapan pelabelan kelas yaitu tahapan pembagian data ulasan dan *rating* yang diberikan oleh pengguna aplikasi shopee di play store yang telah diklasifikasikan menggunakan metode klasifikasi *naïve bayes* untuk menentukan kelas positif dan negatif. *Naïve bayes* merupakan salah satu teknik untuk klasifikasi data *mining* pada *machine learning* yang populer untuk melakukan klasifikasi data dan dapat digunakan untuk memprediksi kelas. *Naïve Bayes* yaitu merupakan salah satu algoritma pembelajaran mesin dan data *mining* yang paling efisien [10]. Penggunaan data aplikasi shopee yang telah diklasifikasikan menjadi kelas positif akan diberikan label angka 1 sedangkan untuk kelas negatif diberikan label angka 0.

### 2.4. Evaluasi Model

Tahapan Setelahnya, yaitu merupakan proses pembagian data latih dan uji selesai, evaluasi model dilakukan. Penulis penelitian ini, dalam melakukan untuk mengevaluasi keakuratan model, evaluasi

dilakukan dengan menggunakan *confusion matrix*, digunakan rumus *confusion matrix* digunakan. untuk mengetahui tingkat akurasi yaitu sebagai berikut [11].

$$\text{Akurasi} = \frac{(\text{TP} + \text{TN})}{(\text{TN} + \text{TP} + \text{FN} + \text{FP})} \quad (1)$$

### 3. HASIL DAN PEMBAHASAN

Penelitian ini, penulis akan menjelaskan berdasarkan langkah-langkah setiap langkah-langkah yang dipaparkan dimetode penelitian, yaitu pengumpulan data, *preprocessing* data, pelabelan data, dan evaluasi model. Sebagai berikut.

#### 3.1 Pengumpulan Data

Penelitian ini mengumpulkan data ulasan pengguna aplikasi shopee di play store sebanyak 1528 baris dengan 4 kolom. Metode pengumpulan data menggunakan *web scraping* data. Tahapan pertama dalam pengumpulan data yaitu dengan *install* pada program *python* yaitu *pip install google-play-scraper* selanjutnya menggunakan *library* yang digunakan seperti *google-play-scraper*, *pandas*, dan *numpy*, berikut merupakan hasil dalam pengumpulan data yang ditunjukkan pada Gambar 2.

1	userName	score	at	content
2	MASYIS GAMING	1	12/04/2023 12:01	Aneh banget shopee ini, padahal sinyal lancar tapi kok pas mau masuk flash sale malah loading terus, gw coba relog masih aja tetep, cobain sinyal di aplikasi lain lancar banget tuh padahal, suka aneh dah, perbaiki!!!!!!gbk
3	pranoto hadi	3	11/04/2023 22:17	Beberapa hari ini server shopee lambat, mohon ada perbaikan supaya transaksi lancar, masih sering lambat respon server, mohon ada perbaikan. Masih ditemukan review produk tidak bisa, hanya tanda muter tanpa terlihat gambar, coba cek HP teman juga sama,
4	fahmi azis	1	11/04/2023 10:48	Sangat mengecewakan setelah update terbaru malah gak karuan. Jika kita masuk ke halaman flash sale maka ada 3 kategori waktu misalkan jam 4, jam 12, jam 18. Saat pertama masuk otomatis kita ada di halaman yg jam 4. Saat kota coba melihat flashsale jam 18 dan masuk ke sub kategori pakaian pria lalu kita melihat salah satu item yg ditawarkan disitu. Gak cocok lalu ingin melihat item yg lain dengan klik tombol kembali. Yg terjadi adalah kita kembali jauh ke halaman awal flashsale di jam 4.
5	Dianta N	2	10/04/2023 20:29	Filter penjual kenapa tidak berfungsi ya? hanya disediakan 1 opsi penjual "premium", begitu dipilih, tidak ada produk yg muncul sama sekali. Mempersulit pencarian, sangat tidak membantu. Lebih baik dikembalikan seperti dulu ("star", "mall", dst) tolong segera diperbaiki
6	noppp	1	10/04/2023 19:29	Bagus bagus itemnya tapi lama banget pengirimannya saya bisa smpe 2minggu nunggu nya, mana dekat cuman Jakarta barat sama selatan. Pemakaian penyimpanan nya juga gede, suka ngelag juga kalo lgi scroll scroll cari cari barang.
7	...	...	....	....
8	M. Rayhan S.	2	10/02/2023 15:22	Saat masuk aplikasi kadang seperti logout padahal tidak dan ini terjadi di beberapa hp, suka lag padahal jaringan bagus, fitur chat eror tidak ada riwayat chat jadi harus bolak-balik cari dulu nama toko atau penjual. Tolong dari pihak shopee beberapa masalah ini diperbaiki.

Gambar 2. Hasil Pengumpulan Data Ulasan

#### 3.2 Preprocessing Data

Tahapan *preprocessing* data meliputi beberapa tahapan sehingga data menjadi bersih dan dapat data dapat digunakan untuk tahapan berikutnya, berikut merupakan tahapan pada *preprocessing* data yaitu.

*Case folding* merupakan tahapan untuk mengubah setiap teks ulasan menjadi bentuk *lowercase* atau tidak terdapat ulasan yang memiliki huruf kapital.

*Tokenizing* merupakan tahapan untuk menghapus pada data ulasan yang

mengandung *link url*, *@mention*, *hashtag*, angka dan simbol tanda.

*Normalization* merupakan tahapan untuk mengubah kata singkatan pada ulasan kedalam bentuk kata yang normal pada umumnya, seperti, "apk": "aplikasi", "bagusssss": "bagus", dan lain sebagainya. Penelitian ini menambahkan file *colloquial-indonesia-lexicon* atau kamus normalisasi dan normaliasasi secara manual untuk mengubah kata dalam bentuk normalnya.

*Filtering* merupakan tahapan untuk menghapus menghapus kata ulasan yang

tidak memiliki makna, dalam penelitian ini menggunakan *library* NLTK.

*Stemming* merupakan tahapan untuk mengubah dan mencari yang terdapat

Tabel 1. Hasil *Preprocessing* Data

Tahapan	Ulasan
Data awal	Aneh banget ga bisa isi saldo.. Bolak balik komplain tapi ga ada solusi.. Uda 3 hari masalah ga terpecahkan.. Saya bukan shopee pay later ya..
Case folding	aneh banget ga bisa isi saldo.. bolak balik komplain tapi ga ada solusi.. uda 3 hari masalah ga terpecahkan.. saya bukan shopee pay later ya..
Tokenizing	aneh banget ga bisa isi saldo bolak balik komplain tapi ga ada solusi uda hari masalah ga terpecahkan saya bukan shopee pay later ya
Normalization	aneh banget ga bisa isi saldo bolak balik komplain tapi ga ada solusi sudah hari masalah ga terpecahkan saya bukan shopee pay later ya
Filtering	aneh banget isi saldo bolak balik keluhan solusi terpecahkan shopee pay later ya
Stemming	aneh banget isi saldo bolak balik keluh solusi pecah shopee pay later ya
Hasil <i>Preprocessing</i>	aneh isi saldo bolak keluh solusi pecah shopee bayar spaylater

pada ulasan menjadi kata dasar (*root*). Berikut merupakan salah satu contoh dalam proses *preprocessing* dalam penelitian ini yang ditunjukkan pada Tabel 1.

### 3.3 Pelabelan Data

Setelah melakukan *preprocessing* data, selanjutnya yaitu dalam penelitian ini akan melakukan pelabelan data. Penelitian ini untuk data *rating* akan klasifikasikan untuk kelas positif yang memiliki *rating* 4

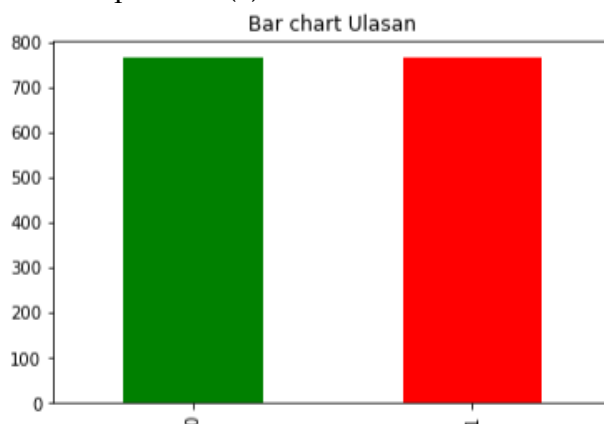
dan 5 sedangkan untuk kelas negatif yang memiliki *rating* 1, 2, dan 3. Tahapan dalam pelabelan data, maka akan didapatkan hasil label kelas dengan metode klasifikasi *naive bayes* yang telah membagi dan mengklasifikasikan data pengguna aplikasi shopee kedalam bentuk data ulasan kelas positif dan negatif. Berikut merupakan hasil dalam proses pelabelan kelas untuk kelas positif dan kelas negatif yang akan ditunjukkan pada Gambar 3.

	userName	Ulasan_clean	label
1	pranoto hadi	server shopee lambat mohon baik transaksi lancar lambat respon server mohon baik temu review produk tanda muat gambar coba cek hp teman	0
2	fahmi azis	kecewa update baru karuan masuk halaman flash sale kategori jam jam jam masuk otomatis halaman jam kota coba flashsale jam masuk sub kategori pakai pria salah item tawar situ cocok item klik tombol halaman flashsale jam	0
3	Dianta N	filter jual fungsi ya sedia opsi jual premium pilih produk muncul sulit cari bantu kembali star mall dst tolong baik	0
4	noppp	bagus bagus barang banget pengirimannya minggu tunggu cuman jakarta barat selatan pakai simpan gede suka lag kalo scroll scroll cari cari barang	0
5	Supangkat Dema	beli shopee lancar aman kendala gratis ongkir mudah mudah gratis ongkir langgan shopee	1

Gambar 3. Hasil Pelabelan data

Berdasarkan dari jumlah data pengguna aplikasi shopee yang terdapat dalam penelitian ini sebanyak 1528 data akan dibagi untuk data yang diklasifikasikan kelas positif dan kelas negatif. Sehingga akan didapatkan hasil pembagian data untuk kelas positif (1)

yaitu sebanyak 764 data dan untuk kelas negatif (0) yaitu sebanyak 764 data dari proses pembagian label kelas tersebut. berikut merupakan hasil pembagian label kelas untuk kelas positif dan kelas negatif yang ditunjukkan pada Gambar 4.



Gambar 4. Grafik Hasil Pembagian Label Kelas Positif (1) dan Negatif (0)

Tahapan selanjutnya dalam penelitian ini, penulis akan menampilkan dalam bentuk visualisasi *wordcloud* untuk mengetahui data ulasan untuk kelas positif

dan kelas negatif terhadap pengguna aplikasi shopee di play store, dapat ditunjukkan pada Gambar 5.



Gambar 5. Visualisasi *wordcloud* untuk Ulasan Kelas Positif (a) dan Kelas Negatif (b)

Hasil visualisasi ulasan positif dan negatif merupakan hasil dari ekstraksi fitur yaitu pengenalan pola, terdapat pengenalan pola dalam penelitian ini menggunakan *wordcloud* untuk mengetahui kata yang sering dibahas oleh pengguna aplikasi shopee di play store.

### 3.4 Evaluasi Model

Tahapan dalam melakukan pelabelan data ulasan pengguna aplikasi shopee menggunakan metode klasifikasi *naïve bayes*. Sebelum melakukan evaluasi model hasil klasifikasi data ulasan, dalam penelitian ini akan dilakukan pembagian data uji dan data latih dengan perbandingan 80%:20%, sehingga

digunakannya sebanyak 306 data uji dan 1222 data latih.

Evaluasi model digunakan untuk mengetahui dan membantu dalam

mengevaluasi hasil akurasi dari metode yang digunakan. Berikut merupakan hasil akurasi *confusion matrix* dengan *python*, seperti yang ditunjukkan Gambar 6.

```

----- confusion matrix -----
[[116 37]
 [ 4 149]]
----- classification report -----
              precision    recall  f1-score   support

     0           0.97       0.76       0.85        153
     1           0.80       0.97       0.88        153

 accuracy          0.87        306
 macro avg         0.88        0.87       0.86        306
 weighted avg     0.88        0.87       0.86        306

```

Gambar 6. Hasil Evaluasi Model dengan *Confusion Matrix*

Berdasarkan hasil yang telah ditampilkan dengan menggunakan metode *confusion matrix* pada Gambar 6. Diketahui bahwa dengan jumlah data uji sebanyak 306 data, didapatkan *true positive* 116, *false negative* 37, *true negative* 149, *false positive* 4. Sehingga dapat menghitung akurasi dengan persamaan pada rumus akurasi *confusion matrix*, yaitu.

$$\begin{aligned}
 \text{Akurasi} &= \frac{(116+149)}{(116+149+4+37)} \\
 &= \frac{265}{306} \\
 &= 0,866 \approx 0,87
 \end{aligned}$$

Berdasarkan hasil tersebut didapatkan hasil akurasi evaluasi model dengan *confusion matrix* yaitu sebesar 87%.

#### 4. UCAPAN TERIMA KASIH

Penulis ingin mengucapkan banyak rasa puji serta syukur yang telah diberikan oleh Tuhan, karena berkat ridha dan kehendaknya penulis telah diberikan kelancaran dalam melakukan pelaksanaan dan penulisan penelitian ini dapat di selesaikan.

Penulis ingin mengucapkan terima kasih kepada JIMI atas selaku penyelenggara jurnal ini yang telah memberikan tempat untuk penelitian ini.

#### 5. KESIMPULAN

Berdasarkan hasil pengolahan data penelitian ini diperoleh melalui metode klasifikasi *naïve bayes* dan *confusion matrix* yaitu didapatkan hasil akurasi sebesar 87%, sehingga dapat disimpulkan dari hasil tersebut, bahwa hasil dalam proses klasifikasi data pengguna aplikasi shopee di play store dapat diklasifikasikan dengan baik menggunakan metode klasifikasi *naïve bayes*.

Penelitian selanjutnya, perlu adanya untuk menggunakan metode klasifikasi lain, sehingga dapat membandingkan dan mencari hasil klasifikasi dengan metode terbaik.

#### 6. REFERENSI

- [1] K. C. Laudon and J. P. Laudon, *Management Information Systems Managing the Digital Firm*, 13<sup>th</sup> ed. Prentice Hall, 2014.
- [2] W. Social, Are, "Penggunaan E-Commerce Indonesia Tertinggi di Dunia," *KataData*, no. April, p. 2021, 2021, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2021/06/04/penggunaan-e-commerce-indonesia-tertinggi-di-dunia>.

- [3] Romindo *et al.*, *E-Commerce: Implementasi, Strategi dan Inovasinya*. 2019.
- [4] IPrice, "10 E-Commerce dengan Pengunjung Terbanyak Kuartal I 2022," *Katadata*, p. 2022, 2022, [Online]. Available: <https://databoks.katadata.co.id/datapublish/2022/07/19/10-e-commerce-dengan-pengunjung-terbanyak-kuartal-i-2022>.
- [5] D. Yani, Deviacita, S. Pratiwi, Helen, and H. Muhandi, "Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [6] K. Suryadewiansyah, Muhammad and T. Tju, Eng, Endra, "Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," vol. 08, no. 02, pp. 81–88, 2022.
- [7] D. Normawati and S. Prayogi, Aliit, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," vol. 5, no. November 2019, pp. 697–711, 2021.
- [8] Riskania and F. Thalib, "Jurnal Teknologi," *Jurnal Teknologi*, vol. 8, no. 1, pp. 31–39, 2020, [Online]. Available: [www.jurnalteknologi.utm.my](http://www.jurnalteknologi.utm.my).
- [9] A. Kusuma and H. N. Irmanda, "Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine," vol. 3, no. 2, pp. 563–574, 2022.
- [10] R. Prasetya, "Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbor Algorithm With Model Testing of Supervised Learning Probabilistic Brier Score, Confusion Matrix and ROC," *Jaict*, vol. 4, no. 2, p. 25, 2020, doi: 10.32497/jaict.v4i2.1690.
- [11] I. Cholissodin and A. A. Soebroto, "AI , Machine Learning & Deep Learning ( Teori & Implementasi )," 2021.