

Klasifikasi Sentimen Ulasan Pelanggan Layanan Ekspedisi Menggunakan Algoritma *Naïve Bayes*

Amada Uzlyva Almie¹, Agung Nilogiri², Ari Eko Wardoyo³

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember^{1,2,3}

Email: amadauzlyva@outlook.com^{1*}, agungnilogiri@unmuhjember.ac.id², arieko@unmuhjember.ac.id³

ABSTRAK

Pemanfaatan media sosial *twitter* memiliki potensi besar dalam menghasilkan informasi, dari banyaknya jumlah opini dari masyarakat yang didapatkan, diperlukan analisis sentimen untuk mengklasifikasikan opini tersebut agar memudahkan dalam menemukan kecenderungan opini dari ulasan pengguna ekspedisi guna mengetahui seberapa besar persentase sentimen positif, dan negatif. Penelitian ini akan membandingkan antara model klasifikasi *multinomial naïve bayes* dan *gaussian naïve bayes* tanpa dan dengan menerapkan teknik *balancing* untuk mengatasi *imbalance* data. Hasil akurasi yang didapatkan pada metode *multinomial naïve bayes* tanpa menggunakan teknik *balancing* hasil akurasi sebesar 95%, sedangkan dengan *random undersampling* sebesar 95%, dan untuk *random oversampling* sebesar 97%. Pada metode *gaussian naïve bayes* tanpa menggunakan teknik *balancing* hasil akurasi sebesar 85%, sedangkan dengan *random undersampling* sebesar 80%, dan untuk *random oversampling* didapatkan hasil akurasi sebesar 93%.

Kata Kunci: Analisis Sentimen, *Twitter*, *Multinomial Naïve Bayes*, *Gaussian Naïve Bayes*, *Random Oversampling*, *Random Undersampling*

ABSTRACT

The use of twitter social media has great potential in generating information, from the large number of opinions from the public that are obtained, sentiment analysis is needed to classify these opinions so that it makes it easier to find opinion trends from expedition user reviews to find out how much the percentage of positive and negative sentiments is. This research will compare the multinomial naïve bayes and gaussian naïve bayes classification models without and by applying balancing techniques to overcome imbalance data. The accuracy results obtained for the multinomial naïve bayes method without using a balancing technique yields an accuracy of 95%, while for random undersampling it is 95%, and for random oversampling it is 97%. In the gaussian naïve bayes method without using a balancing technique the accuracy is 85%, while with random undersampling it is 80%, and for random oversampling the accuracy is 93%.

Keywords: *Sentiment Analysis, Twitter, Multinomial Naïve Bayes, Gaussian Naïve Bayes, Random Oversampling, Random Undersampling*

1. PENDAHULUAN

Penggunaan jasa ekspedisi meningkat karena semakin dibutuhkan masyarakat, namun terdapat permasalahan dalam proses pengiriman, yang salah satunya dapat diatasi dengan menganalisis ulasan pelanggan pada media sosial *twitter*. Banyaknya jumlah opini dari masyarakat di *twitter*, maka diperlukan analisis sentimen untuk menemukan kecenderungan opini dari ulasan pelanggan apakah bersifat positif, negatif atau netral terkait topik permasalahan (Liu, 2012). Klasifikasi sentimen memiliki berbagai jenis metode salah satunya metode *naïve bayes classifier*.

Metode *naïve bayes* dibagi atas beberapa tipe yaitu *multinomial naïve bayes* dan *gaussian naïve bayes*. *Multinomial naïve bayes* merupakan *naïve bayes* yang mengasumsikan independensi di antara kemunculan kata dalam dokumen tanpa menghitung urutan kata (Siniwi dkk., 2021). *Gaussian naïve bayes* adalah *naïve bayes* dengan metode distribusi *gaussian* yang biasanya digunakan pada data numerik (Lestari dkk., 2017).

Penelitian sebelumnya menggunakan metode *naïve bayes* dalam klasifikasi sentimen. Pada penelitian tentang klasifikasi sentimen pada *movie review* menggunakan metode *multinomial naïve bayes* dengan kombinasi *negation handling*, *preprocessing*, dan TF-IDF menghasilkan nilai akurasi terbesar yaitu 85,16% (Wisudawati dkk., 2017). Penelitian lainnya terkait perbandingan ekstraksi fitur TF-IDF dan *Word2Vec* menggunakan *gaussian naïve bayes* terhadap data *review* analisis sentimen

pelayanan Indihome menghasilkan akurasi terbaik pada ekstraksi fitur TF-IDF sebesar 96% (Tyas dkk., 2022).

Imbalance data dan dapat mempengaruhi hasil akurasi klasifikasi, maka diperlukan suatu cara untuk menyeimbangkan jumlah distribusi data dengan teknik *balancing* yaitu *random oversampling* dan *random undersampling* untuk menyeimbangkan jumlah sentimen negatif dan positif pada penelitian ini. *Random oversampling* adalah metode re-sampling penggantian kelas minor secara acak sampai menjadi kelas major, sedangkan *random undersampling* adalah metode re-sampling eliminasi kelas major secara acak sampai menjadi kelas minor (Arifiyanti & Wahyuni, 2020).

Penelitian lainya terkait membandingkan kinerja random forest dengan 11 model algoritma menggunakan metode re-sampling random over-under sampling pada klasifikasi penilaian kredit, dengan hasil akurasi random forest tanpa metode re-sampling sebesar 76%, sedangkan penerapan random over-under sampling dan random forest didapatkan yaitu 90,1% (Syukron & Subekti, 2018).

Perbedaan dari penelitian sebelumnya adalah penelitian ini akan berfokus pada membandingkan performa akurasi model klasifikasi multinomial *naïve bayes* dan gaussian *naïve bayes* tanpa teknik *balancing* dan dengan teknik *balancing* (*random oversampling* dan *random undersampling*) pada analisis sentimen ulasan pelanggan pelayanan jasa ekspedisi pada media sosial twitter.

2. KAJIAN PUSTAKA

A. Analisis Sentimen

Analisis sentimen untuk menemukan kecenderungan opini dari ulasan pelanggan apakah bersifat positif, negatif atau netral terkait topik permasalahan (Liu, 2012).

B. Text Mining

Text mining juga dapat didefinisikan sebagai usaha menambang data yang berupa teks dimana sumber data biasanya didapat dari dokumen untuk mencari kata-kata yang dapat mewakili isi dari dokumen (Pudjiantoro, 2013).

C. Jasa Ekspedisi

Jasa ekspedisi adalah jasa pengiriman paket atau barang dengan setiap pelanggan diharuskan membayar ongkos kirim sesuai daerah tujuan.

D. Twitter

Twitter ialah media sosial yang sudah ada pada tahun 2006 dibuat oleh Jack Dorsey yang telah berubah menjadi situs berbagi informasi dan berkomunikasi secara cepat (Kumar & Pamdey, 2013).

E. Multinomial Naïve Bayes

Metode *naïve bayes* yang mengasumsikan independensi di antara kemunculan kata dalam dokumen tanpa menghitung urutan kata (Siniwi, 2021) dengan rumus sebagai berikut (Manning dkk., 2013):

$$P(c|d) \propto P(c) \prod_{k=1}^n P(t_k|c) \quad (1)$$

Dimana $P(t_k|c)$ adalah *conditional probability* terhadap dokumen yang dimiliki kategori c , dan $P(c)$ adalah *prior probability* dokumen pada kategori c .

F. Gaussian Naïve Bayes

Metode *naïve bayes* dengan metode distribusi *gaussian* yang biasanya digunakan pada data numerik (Lestari dkk., 2017).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

G. Balancing Data

Teknik *balancing* yaitu *random oversampling* dan *random undersampling* untuk menyeimbangkan jumlah sentimen negatif dan positif pada penelitian ini. *Random oversampling* dimana untuk menyeimbangkan data dengan menjadikan jumlah data terbesar sebagai patokan, sedangkan *random undersampling* untuk menyeimbangkan data dengan menjadikan jumlah data terkecil sebagai patokan (Arifiyanti & Wahyuni, 2020).

H. Evaluasi Kinerja Model

Evaluasi kinerja sistem dilakukan untuk mengevaluasi hasil percobaan dari klasifikasi dan menganalisis kinerja klasifikasi teks dengan mengukur nilai performansi dari sistem yang telah dibuat (Novantirani dkk., 2015).

3. METODE PENELITIAN

Adapun langkah-langkah dalam penelitian klasifikasi sentimen ulasan pelanggan layanan ekspedisi menggunakan algoritma naïve bayes dapat ditunjukkan pada Gambar 1



Gambar 1. Tahapan Penelitian

Pada gambar diatas dimulai dengan identifikasi masalah yang telah dibahas pada pendahuluan, lalu studi literatur untuk mempermudah dalam pemahaman dalam penelitian dari jurnal dan artikel terkait penelitian, lalu pengumpulan data dengan teknik *crawling*, setelah itu implementasi dengan membandingkan dua model klasifikasi yang digunakan, kemudian melakukan pengujian dan analisis hasil.

A. Pengumpulan Data

Teknik pengumpulan data yang digunakan yaitu *crawling* dengan API *twitter* untuk memudahkan pengambilan data pada ulasan-ulasan ekspedisi di *twitter* pada bulan Juli 2022 dengan total 5600 data *crawling* menggunakan kata kunci *jne id*, *jnt express*, *id tiki*, *pos Indonesia*, *sicepat ekspres*, *anteraja*, *shopee express*.

B. Labelisasi

Pelabelan dilakukan secara manual oleh ahli yang berkompeten di bidangnya. Hasil labelisasi didapatkan 1000 data yang terdiri dari 700 data negatif dan 300 data positif.

C. Balancing Data

Dikarenakan terjadi *imbalance* data pada data sampel, maka digunakan teknik *balancing* yaitu *random oversampling* dan *random undersampling*. *Random oversampling* dimana untuk menyeimbangkan data dengan menjadikan jumlah data terbesar sebagai patokan, sedangkan *random undersampling* untuk menyeimbangkan data dengan menjadikan jumlah data terkecil sebagai patokan (Arifiyanti & Wahyuni, 2020).

Tabel 1. Hasil Balancing Data

Label		Random oversampling	Random Undersampling
Positif	700	700	300
Negatif	300	700	300
Total	1000 data	1400 data	600 data

Dapat di lihat pada tabel 1 bahwa total data *random oversampling* yaitu 1400 data, dan total data *random undersampling* yaitu 600 data.

D. Preprocessing Text

Preprocessing text untuk meningkatkan akurasi klasifikasi data dengan mengubah bentuk dokumen menjadi data yang terstruktur (Manning dkk., 2009). Tahapan-tahapan yang dilakukan seperti *cleansing*, *case folding*, *stopword removal*, *tokenizing*, dan *stemming*. *Cleansing* untuk menghilangkan url, *hashtag*, *username*, simbol dan tanda baca lainnya. *Case folding* dengan merubah kata pada dokumen menjadi huruf kecil. *Stopword removal* untuk menghilangkan kata-kata yang dapat mempengaruhi hasil akurasi. *Tokenizing* adalah pemisahan tiap kata menjadi token. *Stemming* untuk membuat kata berimbuhan menjadi kata dasar.

E. Pembobotan TF-IDF

Rumus perhitungan tf-idf sebagai berikut:

$$tf_{t,d} = \frac{n_{t,d}}{r_{t,d}} \quad (1)$$

$$idf = \log \frac{D}{df} \quad (2)$$

$$W_{t,d} = tf_{t,d} \times idf \quad (3)$$

$$W_{t,d} = tf_{t,d} \times \log \frac{D}{df} \quad (4)$$

Dimana $tf_{t,d}$ adalah frekuensi kemunculan *term* pada dokumen, *idf* adalah hasil *inverse document frequency*, *df* adalah frekuensi dokumen yang mengandung *term*, *D* adalah total dokumen, dan $W_{t,d}$ yaitu hasil bobot *term* pada dokumen (Yoren, 2018).

F. Pembagian Data

Proses dimana data dibagi menjadi k bagian ama banyak, untuk menentukan data testing pada data yang digunakan (Torres dkk., 2012). Penelitian ini menggunakan 3 jenis *k-fold cross validation* dengan rumus perhitungan sebagai berikut:

$$D = \frac{n}{k} \quad (5)$$

Dimana *D* adalah hasil pembagian data dalam file dokumen, *n* yaitu jumlah data, *k* yaitu nilai *k-fold cross validation*.

1) Normal Dataset:

- a. *2-fold cross validation*, data akan dibagi menjadi 2 bagian sama banyak, masing-masing bagian 500 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1000}{2} = 500 \quad (6)$$

Maka 500 data testing dan 500 data sisanya data *training*.

- b. *5-fold cross validation*, data akan dibagi menjadi 5 bagian sama banyak, masing-masing bagian 200 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1000}{5} = 200 \quad (7)$$

Maka 200 data testing dan 800 data sisanya data *training*.

- c. *10-fold cross validation*, data akan dibagi menjadi 10 bagian sama banyak, masing-masing bagian 100 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1000}{10} = 100 \quad (8)$$

Maka 100 data testing dan 900 data sisanya data *training*.

2) *Random undersampling* Dataset:

- a. *2-fold cross validation*, data akan dibagi menjadi 2 bagian sama banyak, masing-masing bagian 300 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{600}{2} = 300 \quad (9)$$

Maka 300 data testing dan 300 data sisanya data *training*.

- b. *5-fold cross validation*, data akan dibagi menjadi 5 bagian sama banyak, masing-masing bagian 120 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{600}{5} = 120 \quad (10)$$

Maka 120 data testing dan 480 data sisanya data *training*.

- c. *10-fold cross validation*, data akan dibagi menjadi 10 bagian sama banyak, masing-masing bagian 300 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{600}{10} = 60 \quad (11)$$

Maka 60 data testing dan 540 data sisanya data *training*.

3) *Random oversampling* Dataset:

- a. *2-fold cross validation*, data akan dibagi menjadi 2 bagian sama banyak, masing-masing bagian 700 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1400}{2} = 700 \quad (12)$$

Maka 700 data testing dan 700 data sisanya data *training*.

- b. *5-fold cross validation*, data akan dibagi menjadi 5 bagian sama banyak, masing-masing bagian 1400 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1400}{5} = 280 \quad (13)$$

Maka 280 data testing dan 1120 data sisanya data *training*.

- c. *10-fold cross validation*, data akan dibagi menjadi 10 bagian sama banyak, masing-masing bagian 1400 data.

$$D = \frac{n}{k} = \frac{\text{banyak data}}{k \text{ fold}} = \frac{1400}{10} = 140 \quad (14)$$

Maka 140 data testing dan 1260 data sisanya data *training*.

G. Implementasi Model Klasifikasi

Naïve bayes termasuk dalam model algoritma *supervised learning* yang memiliki label sebagai acuan. Model *naïve bayes* yang digunakan pada penelitian ini adalah *multinomial naïve bayes* dan *gaussian naïve bayes*.

1) Multinomial *naïve bayes*:

Metode *naïve bayes* yang mengasumsikan independensi di antara kemunculan kata dalam dokumen tanpa menghitung urutan kata (Siniwi dkk., 2021) dengan rumus sebagai berikut (Manning dkk., 2009).

$$P(c|d) \propto P(c) \prod_{k=1}^n P(t_k|c) \tag{6}$$

Dimana $P(t_k|c)$ adalah *conditional probability* t_k pada dokumen yang dimiliki kategori c , dan $P(c)$ adalah *prior probability* dokumen pada kategori c .

2) Gaussian *naïve bayes*:

Metode *naïve bayes* dengan metode distribusi *gaussian* yang biasanya digunakan pada data numerik (Lestari dkk., 2017).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{7}$$

Dimana $f(x)$ adalah nilai gaussian, x adalah nilai data, μ yaitu mean, dan σ adalah standar deviasi.

H. Evaluasi Model

Evaluasi kinerja model dengan menggunakan metode *k-fold cross validation*, kemudian *confusion matrix* sebagai alat visualisasi menentukan perfoma dari suatu model (Novantirani dkk., 2015). Penelitian ini menguji dan membandingkan akurasi dari kedua model klasifikasi tersebut.

4. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan pengujian dengan membandingkan hasil akurasi multonomial *naïve bayes* dan *gaussian naïve bayes*, antara dataset tanpa dilakukan *balancing* data dan dataset yang dilakukan *balancing* data, dengan menggunakan *2-fold cross validation*, *5-fold cross validation*, dan *10-fold cross validation*.

A. *Multinomial naïve bayes*

Pengujian menggunakan *multinomial naïve bayes* menghasilkan nilai akurasi ditunjukkan pada Tabel 2.

Tabel 2. Hasil Akurasi *Multinomial Naive Bayes*

k-fold cross	Langkah Uji	<i>Multinomial naïve bayes</i>		
		Normal	<i>Undersampling</i>	<i>Oversampling</i>
2-fold	Langkah Uji 1	78%	86%	91%
	Langkah Uji 2	86%	85%	91%
5-fold	Langkah Uji 1	84%	86%	92%
	Langkah Uji 2	80%	85%	92%
	Langkah Uji 3	81%	85%	92%
	Langkah Uji 4	90%	92%	94%
	Langkah Uji 5	86%	85%	91%
10-fold	Langkah Uji 1	90%	83%	93%
	Langkah Uji 2	80%	93%	92%
	Langkah Uji 3	84%	90%	91%
	Langkah Uji 4	82%	86%	90%
	Langkah Uji 5	80%	86%	92%
	Langkah Uji 6	84%	86%	92%
	Langkah Uji 7	95%	93%	97%
	Langkah Uji 8	87%	95%	92%
	Langkah Uji 9	84%	86%	92%
	Langkah Uji 10	91%	85%	92%
Rata-rata akurasi		85%	87%	92%

Proses pengujian menggunakan *k-fold cross validation* dengan nilai akurasi yang dihasilkan metode *multinomial naïve bayes* pada normal data set dengan *10-fold cross validation* di langkah uji ketujuh diperoleh nilai akurasi sebesar 95%, sedangkan pada *random undersampling* data set dengan *10-fold cross validation* di langkah uji kedelapan diperoleh nilai akurasi sebesar 95%, dan pada *random oversampling* data set dengan *10-fold cross validation* di langkah uji ketujuh diperoleh nilai akurasi sebesar 97%.

B. Gaussian naïve bayes

Pengujian menggunakan *gaussian naïve bayes* menghasilkan nilai akurasi ditunjukkan pada Tabel 3.

Tabel 3. Hasil Akurasi *Gaussian Naive Bayes*

k-fold cross	Langkah Uji	<i>Gaussian naïve bayes</i>		
		Normal	<i>Undersampling</i>	<i>Oversampling</i>
2-fold	Langkah Uji 1	80%	75%	82%
	Langkah Uji 2	75%	73%	85%
5-fold	Langkah Uji 1	82%	74%	83%
	Langkah Uji 2	80%	72%	86%
	Langkah Uji 3	76%	74%	86%
	Langkah Uji 4	80%	75%	90%
	Langkah Uji 5	72%	68%	87%
	Langkah Uji 1	85%	76%	88%
10-fold	Langkah Uji 2	82%	70%	85%
	Langkah Uji 3	81%	68%	84%
	Langkah Uji 4	82%	75%	87%
	Langkah Uji 5	82%	78%	87%
	Langkah Uji 6	74%	70%	87%
	Langkah Uji 7	81%	71%	93%
	Langkah Uji 8	75%	80%	87%
	Langkah Uji 9	77%	73%	88%
	Langkah Uji 10	72%	66%	90%
	Rata-rata akurasi		79%	73%

Kemudian pengujian menggunakan *k-fold cross validation* metode *gaussian naïve bayes* dihasilkan nilai akurasi pada normal data set dengan *10-fold cross validation* di langkah uji pertama diperoleh nilai akurasi sebesar 85%, sedangkan pada *random undersampling* data set dengan *10-fold cross validation* di langkah uji kedelapan diperoleh nilai akurasi sebesar 80%, dan pada *random oversampling* data set dengan *10-fold cross validation* di langkah uji ketujuh diperoleh nilai akurasi sebesar 93%.

5. KESIMPULAN

Pada penelitian klasifikasi sentimen ulasan pelanggan layanan ekspedisi perolehan nilai akurasi tertinggi untuk metode *multinomial naïve bayes* terdapat pada *random oversampling* data set yaitu sebesar 97% dengan *10-fold cross validation* pada langkah uji setujuh, sedangkan untuk nilai akurasi tertinggi metode *gaussian naïve bayes* terdapat pada *random oversampling* data set yaitu sebesar 93% dengan *10-fold cross validation* pada langkah uji ketujuh. Hal ini menunjukkan bahwa tingkat akurasi terbaik pada penelitian ini dihasilkan oleh *random oversampling* data set yang lebih besar daripada tingkat akurasi yang dihasilkan oleh normal data set, dan *random undersampling* data set.

Saran untuk penelitian selanjutnya untuk mendapatkan akurasi yang lebih baik dapat menambahkan jumlah data dan mencoba menggunakan teknik *balancing* yang lain.

6. DAFTAR PUSTAKA

- Arifiyanti, A. A., & Wahyuni, E. D. (2020). Smote: Metode Penyeimbang Kelas Pada Klasifikasi Data Mining. *SCAN - Jurnal Teknologi Informasi dan Komunikasi*, 15(1), 34–39.
- Kumar, S., & Pamdey, A. K. (2013). Chemistry and Biological Activities of Flavonoids: An Overview. *Chemistry and Biological Activities of Flavonoids: An Overview*, 1–16.
- Lestari, A. R. T., Setya Perdana, R., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji. *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(12), 1718–1724.
- Liu, B. (2012). *Sentiment analysis and opinion mining. Synthesis lectures on human language technologies*. 5(1), 1–167.
- Manning, C. ., Raghavan, P., & Schütze, H. (2013). Probabilistic information retrieval. *Introduction to Information Retrieval*, 220–235.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). Probabilistic information retrieval. *Introduction to Information Retrieval*, 220–235.
- Novantirani, A., Sabariah, M. K., & Effendy, V. (2015). Analisis Sentimen pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine. *e-Proceeding of Engineering*, 2(1), 1–7.
- Pudjiantoro, T. . (2013). Analisa Kompetensi Calon Pegawai Menggunakan Metode TF-IDF. *Aristoteles*, 10(2).
- Siniwi, L. M., Prahutama, A., & Hakim, A. R. (2021). Query Expansion Ranking Pada Analisis Sentimen Menggunakan Klasifikasi Multinomial Naïve Bayes (Studi Kasus : Ulasan Aplikasi Shopee pada Hari Belanja Online Nasional 2020). *Jurnal Gaussian*, 10(3), 377–387.
- Syukron, A., & Subekti, A. (2018). Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit. *Jurnal Informatika*, 5(2), 175–185.
- Torres, T. J. G. M., Saez, J. ., & Herrera, F. (2012). Study on the impact of partition- induced dataset shift on k-fold cross- validation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8), 1304– 1312.
- Tyas, S. M. P., Rintyarna, B. S., & Suharso, W. (2022). The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 1–10.
- Wisudawati, J., Adiwijaya, & Faraby, S. Al. (2017). Klasifikasi Sentimen pada Movie Review dengan Metode Multinomial Naïve Bayes Sentiment Classification Movie Review Using Multinomial Naïve Bayes Method. *e-Proceeding of Engineering*, 4(2), 2978–2988.
- Yoren. (2018). *Perbandingan Raw TF dan Binary TF pada pencarian di situs Museum Wayang Kekayon Yogyakarta*. Sanata Dharma University.