



SENTIMENT ANALYSIS KINERJA KARYAWAN DI OKE GARDEN MENGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Alif Zidan Mubarak¹, Hidayatus Sibyan², Nur Hasanah³

Program Studi Teknik Informatika, Universitas Sains Al-Qur'an^{1,2,3}

alifzidanmubarak@gmail.com¹, hsibyan@unsiq.ac.id², nurhasanah@unsiq.ac.id³

ABSTRACT

This research was conducted to determine the accuracy of the Naïve Bayes Classifier algorithm in determining sentiment analysis based on feedback data from employees at OKE garden. Additionally, this research also presented data visualization of the feedback using the Microsoft Power BI application. The results of the Naïve Bayes algorithm calculations in determining the type of sentiment obtained a good accuracy score of 80% using a training data set of 10%. In terms of data visualization, the utilization of the Microsoft Power BI application in this analysis serves as a tool for visualizing data using the available visualization attributes in the application. Some of the visualizations displayed in this data visualization include the Donut Chart, Stacked Column Chart, LinearGauge, Word Cloud, Slicer, and Card. The results of the calculations using the LinearGauge to determine the average sentiment score from all visualizations are below the number 0.70.

Keywords : Sentiment Analysis; Feedback; Human Resource; Naïve Bayes Classifier; OKE Garden;

ABSTRAK

Penelitian ini dilakukan untuk mengetahui seberapa akurat algoritma Naïve Bayes Classifier dalam menentukan sentiment analysis berdasarkan data feedback dari para karyawan di OKE garden. Selain itu, ada penelitian ini juga disajikan visualisasi data dari feedback menggunakan aplikasi Microsoft Power BI. Hasil dari perhitungan algoritma Naïve Bayes dalam menentukan jenis sentiment, mendapatkan nilai akurasi yang baik yaitu sebesar 80% dengan menggunakan data latih sebesar 10%. Pada visualisasi data, Pemanfaatan aplikasi Microsoft Power BI pada analisis ini, sebagai alat visualisasi data dengan menggunakan atribut visualisasi yang tersedia pada aplikasi ini. Adapun visualisasi yang ditampilkan pada visualisasi data ini antara lain Donut Chart, Stacked Column Chart, LinearGauge, Word Cloud, Slicer, dan Card. Hasil dari perhitungan menggunakan Lineargauge untuk mengetahui hasil sentiment average dari seluruh visualisasi berada dibawah angka 0.70.

Kata Kunci : Analisis Sentimen; Feedback; Human Resource; Naïve Bayes Classifier; OKE Garden;

1. PENDAHULUAN

PT. RISE LANSKAP Holtikultura atau yang lebih dikenal dengan OKEGarden adalah sebuah platform digital yang dibuat untuk mendukung ekosistem pertamanan menjadi lebih baik khususnya di Indonesia. Dibantu dengan

teknologi, OKE Garden mampu menghubungkan pelanggan dengan tukang taman yang berkompeten serta produk-produk dari petani tanaman hias. Berdiri sejak 2017, OKEGarden merupakan salah satu startup yang berfokus pada bidang pertamanan, yang merupakan

turunan dari program Gerakan Nasional 1000 Startup Digital di tahun 2017.

Seperti perusahaan pada umumnya, pada OKEGarden juga memiliki divisi HR (*Human Resource*). Divisi ini bertanggung jawab pada proses recruitment, setting meet untuk setiap divisi, hingga monitoring kinerja karyawan. Maka dari itu, peran dari HR begitu vital bagi perusahaan karena HR bertanggung jawab secara langsung dalam menangani pengelolaan karyawan. Lebih dari itu, HR juga turut serta sebagai penunjang aktivitas perusahaan dalam mencapai target yang sudah ditentukan.

Dari peran penting HR tersebut, divisi ini belum memiliki sebuah report yang propper dalam proses pengambilan strategi dan keputusan yang akan dilakukan di kemudian hari. Laporan yang dimaksud adalah laporan mengenai kepuasan karyawan OKEGarden, dan juga feedback satu bulan kebelakang mengenai sistem kerja, ekspektasi terkait job desc karyawan, dan apa masukan untuk sesuatu yang bisa dikembangkan dikemudian hari. Selain itu, mengetahui faktor apa saja memengaruhi kinerja karyawan juga dibutuhkan. Maka akan dilakukan analisis faktor – faktor itu yang nantinya akan dicari untuk memenuhi kebutuhan dari HR untuk menganalisa kinerja karyawan dalam melaksanakan pekerjaannya, dan perasaan seperti apa yang para karyawan rasakan dalam satu bulan terakhir. Faktor – faktor ini juga berfungsi sebagai acuan dalam penentuan keputusan.

Tracking and Monitoring sangat diperlukan dalam memahami kinerja agar kinerja para karyawan terpantau dengan baik. Namun, selama ini HR hanya memiliki sebuah lembar spreadsheet yang berisi feedback tanpa diolah, sehingga data yang dihasilkan berantakan, dan membuat HR kesulitan dalam melakukan pengambilan keputusan strategi yang tepat untuk memenuhi ekspektasi karyawan OKEGarden.

Maka dibutuhkan sebuah sentiment analysis report untuk dapat melihat dan memprediksi kepusanan karyawan terhadap perusahaan, sehingga memudahkan HR dalam menentukan sebuah keputusan. Melalui laporan ini nantinya HR diharapkan dapat melakukan pengambilan keputusan secara tepat, sehingga berdampak pada performa karyawan yang meningkat.

Sehingga mendongkrak performa perusahaan secara menyeluruh.

Sentiment analisis merupakan sebuah teknik pengolahan data yang digunakan untuk membantu memahami opini, emosi, dan perasaan pelanggan atau masyarakat umum terhadap suatu topik atau produk. Sentiment analisis memiliki sejarah yang panjang dan telah berkembang selama beberapa decade (Rakhmah, 2019).

Pada awalnya, teknik ini digunakan untuk menganalisis sentiment di media cetak, termasuk surat kabar dan majalah. Namun, seiring dengan perkembangan teknologi, sentiment analisis sekarang dapat diterapkan pada berbagai platform media sosial, situs web, blog, dan forum online lainnya. Teknik ini sangat penting bagi perusahaan untuk memahami kepuasan pelanggan, melakukan perubahan produk yang tepat, serta menjaga reputasi merek dan meningkatkan penjualan (Haryani, 2018). Sentiment analisis juga digunakan dalam pemasaran digital untuk mendapatkan wawasan tentang preferensi dan perilaku konsumen (Singgalen, 2021). Pada pembuatan sentiment analysis report ini, menggunakan Naïve Bayes sebagai metode pengklasifikasian data. Karena pada klasifikasi Naïve Bayes membutuhkan jumlah data pelatih (*Training data*) yang relatif kecil, untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Romadhon, 2022).

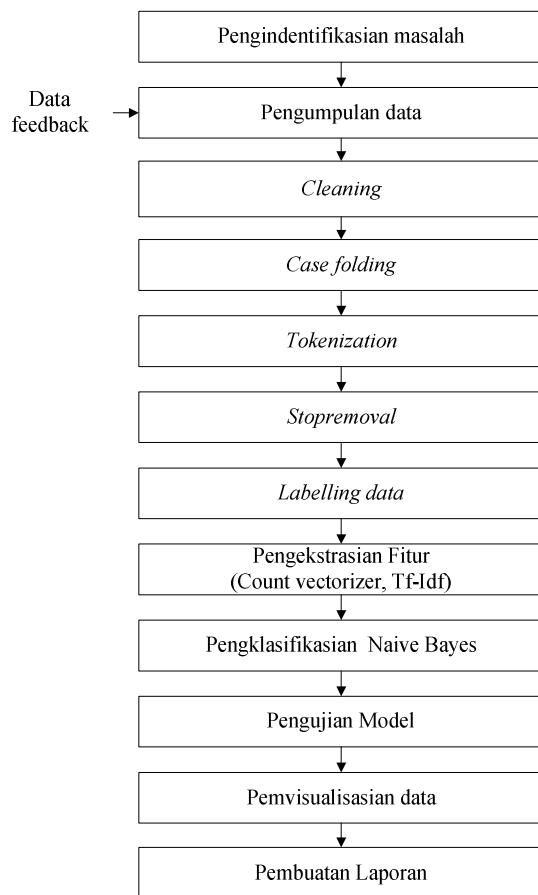
Selain itu, Naïve Bayes juga merupakan metode klasifikasi algoritma yang dinilai efektif. Setelah selesai dengan pemodelan data dengan Naïve Bayes, selanjutnya adalah pembuatan Report menggunakan aplikasi Microsoft Power BI. Microsoft Power BI ini digunakan sebagai alat dalam memvisualisasikan data hasil klasifikasi. Microsoft Power BI digunakan karena kemudahan dalam pengoperasiannya. Aplikasi ini berbasis desktop sehingga memberikan kemudahan dalam aksesnya, karena tidak membutuhkan koneksi internet apabila data yang diolah berasal dari komputer. Selain itu, Microsoft Power BI ini *compatible* dengan berbagai ekstensi data, dan data bisa diambil melalui daring maupun luring (Bororing, 2022).

Oleh karena itu, berdasarkan uraian latar belakang yang ada, pada penelitian ini, peneliti akan memberikan solusi bagi divisi Human

Resource di OKEGarden berdasarkan Sentiment Analysis Report yang telah dibuat. Peneliti berharap dapat laporan analisis sentiment ini dapat mempermudah divisi HR untuk menemukan insight dalam menentukan strategi dan kebijakan untuk kemajuan perusahaan.

2. METODE

Alur penelitian ini meliputi beberapa tahap yang dilakukan, antara lain :



Gambar 1. Alur Penelitian

Alur pengerjaan penelitian ini dimulai dari proses identifikasi masalah. Identifikasi masalah dimulai dengan mengumpulkan segala kesulitan atau kendala yang dihadapi oleh divisi *Human Resource* untuk kemudian disimpulkan menjadi satu permasalahan dan dihimpun menjadi sebuah data, untuk kemudian data ini dikumpulkan menjadi satu. Proses pengidentifikasi masalah ini dilakukan untuk mengetahui apa saja kendala yang dihadapi oleh HR dalam melakukan *monitoring*, *tracking*, serta kepuasan karyawan dengan sistem yang sudah atau sedang berjalan. Proses ini, juga menjadi

acuan dalam pembuatan visualisasi data agar visualisasi yang dibuat sesuai dengan identifikasi. Pada proses pengumpulan data ini, data didapatkan dengan berkomunikasi langsung kepada koordinator divisi *Human Resource* dari OKE Garden. Selain itu, data juga didapatkan dari kuisisioner yang peneliti buat, dengan mengacu pada *form feedback* yang peneliti dapatkan dari divisi HR.

Pengumpulan data menggunakan metode kuisisioner, dan disebarluaskan dalam sebulan sekali. Data yang digunakan adalah data yang bersumber dari *form feedback* yang dibuat oleh divisi *Human Resource* sebagai media evaluasi bagi HR dalam melakukan pengawasan kepuasan dan kinerja karyawan di OKE Garden. Data yang diberikan merupakan hasil pengumpulan timbal balik dari seluruh divisi di OKE Garden, yang dikumpulkan menjadi satu. Selain menggunakan data kuisisioner dari divisi HR, data juga didapatkan dengan menyebarkan kuisisioner kembali, dengan pertanyaan yang telah disesuaikan dengan hasil kuisisioner dari HR.

Data yang telah didapatkan, merupakan data dengan format *xlsx*. Namun, kumpulan *feedback* ini masih berada pada *cloud*, yaitu *Google Spreadsheet*. Data dari *Google Spreadsheet* ini, kemudian diunduh dan diolah dengan *Microsoft Excel* untuk dilakukan penyesuaian dibagian nama tabel.

Setelah terkumpul, maka peneliti akan melakukan *preprocessing data* dari data yang sudah digabungkan dan mempersiapkan data sebelum dianalisis. Pada *pre-processing data*, tahapan yang dilakukan antara lain *Cleaning Data*, *Case Folding*, *Tokenization*, dan *Stop Removal* (Sibyan, 2022).

Tahap selanjutnya yang dilakukan pada penelitian ini adalah melakukan *Labelling* pada data yang sebelumnya sudah melewati proses *preprocessing data*. *Labelling data* ini berfungsi untuk memilih dan memilah sentimen yang tepat dari *feedback* yang terkumpul.

Setelah *Labelling* selesai dilakukan, selanjutnya dilakukan ekstraksi fitur untuk mempermudah pengklasifikasian dengan *Naïve Bayes*. Ekstraksi fitur menggunakan *Count Vectorizer* dan *Tf-Idf* yang kemudian hasil ekstraksi ini dapat masuk tahap keempat .

Pada tahap keempat adalah pengklasifikasian menggunakan *Naïve Bayes*

Classifier. Pengklasifikasian ini dilakukan berdasarkan data yang sudah melewati tahap – tahap sebelumnya. Dilakukan klasifikasi untuk menghasilkan klasifikasi sentimen berdasarkan data.

Selanjutnya uji model dilakukan untuk memeriksa akurasi dan ketepatan *Naïve Bayes* dalam melakukan klasifikasi. Tingkat akurasi dari klasifikasi yang sudah dilakukan. Dari klasifikasi tersebut, kemudian akan disajikan nilai *Precision*, *recall*, *f1-Score*, dan *accuracy* dari analisis yang telah dilakukan.

Lalu visualisasi dilakukan, untuk mempermudah dalam menentukan *insight* dan strategi yang dapat diterapkan untuk kemajuan perusahaan. Visualisasi ini menggunakan data yang telah selesai diolah, kemudian data tersebut di *import* dan divisualisasikan menggunakan aplikasi *Microsoft Power BI*.

3. HASIL DAN PEMBAHASAN

3.1. Preprocessing Data

Pada tahap *preprocessing data*, data akan melewati beberapa proses. Antara lain *Remove Punctuation*, *Case Folding*, *Tokenization*, dan *stop removal*.

Remove Punctuation ini berfungsi membersihkan data dari tanda baca, atau angka, dan elemen lain yang mempengaruhi akurasi. *Case Folding* berfungsi untuk membuat data yang memiliki huruf kapital, berubah menjadi huruf kecil. *Tokenization* berguna untuk memilah kalimat menjadi kata, dan *stop removal* digunakan untuk menghilangkan kata yang tidak diperlukan dalam data.

Contoh akhir dari tahap *preprocessing data* adalah sebagai berikut.

Tabel 1. Hasil *preprocessing data*

Sebelum	Sesudah
mendapat relasi belajar mengelola social media bekerja bersama team yang seru	relasi belajar mengelola social media team seru
sistematika bekerja file yang tercatat lengkap	sistematika file tercatat lengkap
dapat pengetahuan mengenai cara	pengetahuan kerja produk

kerja membuat produk	
tugas dan peran dalam bidang kol	tugas peran bidang kol

3.2. Labelling Data

Labeling pada intinya adalah sebuah proses melabeli atau menandai sebuah informasi yang relevan pada sebuah *dataset* untuk kemudian dipelajari mengenai makna yang terkandung didalamnya. *Dataset* bisa didapatkan dari berbagai macam sumber, bisa berasal dari sebuah gambar, berkas audio, potongan video, maupun *text*. Dalam konteks penelitian ini, *labeling* dilakukan pada *feedback* yang telah melewati tahap *preprocessing*.

Tabel 2. *Labelling data*

<i>Feedback</i>	<i>Label</i>
push anggota divisi kadang ngerasa gabut habisnya karna task	Negatif
kendala peserta daily huddle	Netral
daily huddle malem kali	Positif

3.3. klasifikasi *Naïve Bayes*

Klasifikasi *naïve bayes* dilakukan untuk mengitung nilai akurasi dari analisis yang telah dibuat. Perhitungan ini akan melewati 2 proses, yaitu klasifikasi data latih, dan klasifikasi data uji.

Pada penelitian ini, rumus yang digunakan adalah rumus

$$P\left(X \left| \begin{matrix} Pos \\ Neg \\ Net \end{matrix} \right. \right) = \frac{Nk \left(\frac{pos}{net} \right) + 1}{n \left(\frac{pos}{net} \right) + |T|} \quad (1)$$

$$P = p \left(\frac{Pos}{Net} \right) \times X^1 \times X^2 \dots \dots \times X^n \quad (2)$$

Dengan keterangan :

P : Probabilitas

X : nilai data yang akan dihitung probabilitasnya

Nk : Jumlah frekuensi kata, pada masing – masing data

N : Total data, pada masing – masing *sentiment*

T : Total keseluruhan data

Kemudian sebagai contoh data dilatih menggunakan rumus yang telah ditentukan diatas. Angka probabilitas yang didapatkan pada

kata latih “daily” dari masing – masing sentimen menggunakan rumus 1.

$$P(\text{daily} | \text{positif}) = \frac{1 + 1}{4 + 17} = 0.095$$

$$P(\text{daily} | \text{negatif}) = \frac{0 + 1}{9 + 17} = 0.038$$

$$P(\text{daily} | \text{netral}) = \frac{1 + 1}{4 + 17} = 0.095$$

Kemudian hasil perhitungan probabilitas tersebut, diuji pada data uji, untuk menentukan hasil

klasifikasi berdasarkan kata latih, menggunakan rumus 2.

Sebagai contoh, perhitungan probabilitas data uji adalah data “proses daily huddle oke membantu kendala” kata yang masuk kedalam data uji adalah “daily”, “huddle”, “kendala”.

$$P = \text{Uji} | \text{positif}$$

$$P = p(\text{Positif}) \times p(\text{daily}) \times p(\text{huddle}) \times p(\text{kendala})$$

$$P = 0.3333 \times 0.095 \times 0.095 \times 0.047$$

$$P = 0.00014137752$$

$$P = \text{Uji} | \text{netral}$$

$$P = p(\text{netral}) \times p(\text{daily}) \times p(\text{huddle}) \times p(\text{kendala})$$

$$P = 0.3333 \times 0.095 \times 0.095 \times 0.047$$

$$P = 0.00014137752$$

$$P = \text{Uji} | \text{negatif}$$

$$P = p(\text{negatif}) \times p(\text{daily}) \times p(\text{huddle}) \times p(\text{kendala})$$

$$P = 0.3333 \times 0.038 \times 0.038 \times 0.076$$

$$P = 0.00003657767$$

Berdasarkan perhitungan Probabilitas tersebut, maka data uji merupakan data dengan muatan sentimen Positif.

3.4. Uji Model

Perhitungan dari analisis ini dilakukan pada dataset dengan jumlah 50 data, dengan data test sebesar 10% dari jumlah data yang ada. Menggunakan Confusion Matrix, didapatkan akurasi dari perhitungan adalah sebesar 80%.

Tabel 3. Confusion Matrix

fij		Kelas hasil Prediksi (j)		
		Positif	Negatif	Netral
	Positif	4	0	1

Kelas asli (i)	Negratif	0	0	0
	Netral	0	0	0

Dari total 50 data training, dan data testing 10 terdapat 4 data yang berhasil diklasifikasikan dengan benar dalam kategori Positif, sedangkan 1 data diklasifikasikan Netral dan kategory negatif data 0. Sehingga presentase hasil correct Classified instance (klasifikasi yang benar) adalah sebesar 80.00 % dan untuk incorrect classifield instance (klasifikasi tdk benar) adalah 20.00%. hasil dari rumus perhitungan akurasi dapat dilihat dibawah.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP} \times 100 \%$$

$$= \frac{4 + 0 + 0}{4 + 0 + 0 + 0 + 1} \times 100 \%$$

$$= \frac{4}{5} \times 100 \%$$

$$= 0,8 \times 100\% = 80,00\%$$

Selanjutnya data diuji secara otomatis menggunakan library MultinomialNB. Tahap klasifikasi dilakukan dengan mengambil data test secara acak 10% data training. Data test ini berfungsi untuk menguji akurasi dari klasifikasi yang dilakukan, dan merupakan keberhasilan penelitian. Proses klasifikasi dimulai dengan proses penentuan X dan Y dari data test dan data training.

	precision	recall	f1-score	support
0	0.00	0.00	0.00	1
1	0.00	0.00	0.00	1
2	0.00	0.00	0.00	1
accuracy			0.80	3
macro avg	0.00	0.00	0.00	3
weighted avg	0.00	0.00	0.00	3

Gambar 2. Hasil Uji Model

3.5. Visualisasi Data

Setelah uji model selesai, dataset kemudian diarahkan ke aplikasi Microsoft Power BI untuk dilakukan visualisasi data. Adapun visualisasi yang digunakan antara lain :

a. Donut Chart

Donut Chart pada visualisasi ini berfungsi untuk menunjukkan prosentase perbedaan antara sentimen positif, netral, dan negative.

b. Stacked Column Chart

Pada visualisasi ini digunakan untuk menampilkan jumlah sentimen positif, netral

dan negatif, serta menampilkan grafik perbedaannya.

c. *LinearGauge*

LinearGauge berfungsi untuk menunjukkan nilai sentimen rata-rata pada dataset.

d. *Word Cloud*

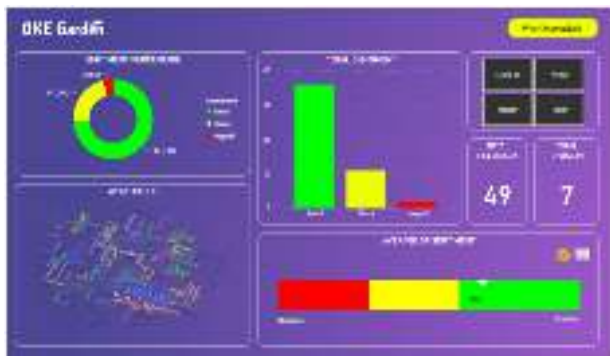
Word Cloud berfungsi untuk menampilkan frekuensi kata terbanyak yang disebutkan dalam dataset.

e. *Slicer*

Slicer digunakan untuk memilih visualisasi mana yang akan ditampilkan secara spesifik.

f. *Card*

Card digunakan untuk menampilkan jumlah *Feedback* yang masuk dan jumlah divisi yang ada di OKE Garden.



Gambar 3. Visualisasi Data

4. PENUTUP

4.1. Kesimpulan

- Nilai akurasi yang dihasilkan pada *Sentiment Analysis* menggunakan metode klasifikasi *naive bayes* adalah sebesar 80% atau 0.80.
- Pemanfaatan aplikasi *Microsoft Power BI* pada analisis ini, sebagai alat visualisasi data dengan menggunakan atribut visualisasi yang tersedia pada aplikasi ini. Adapula visualisasi yang ditampilkan pada visualisasi data ini antara lain *Donut Chart*, *Stacked Column Chart*, *LinearGauge*, *Word Cloud*, *Slicer*, dan *Card*.
- Hasil dari perhitungan menggunakan *Lineargauge* untuk mengetahui hasil *sentiment average* dari seluruh visualisasi berada dibawah angka 0.70. ini menandakan bahwa kepuasan karyawan dengan sistem yang ada di OKE garden dapat dikatakan baik, namun belum mencapai tahap puas.
- Kelebihan dalam analisis ini adalah visualisasi data yang dibuat dengan rapi dan

jasas, sehingga mempermudah HR dalam memahami, dan menentukan *Insight* dari data.

4.2. Saran

- Perlu dilakukan analisis dengan metode klasifikasi yang berbeda, sehingga dapat membandingkan hasil uji model untuk menentukan metode mana yang tepat.
- Sesuai dengan kekurangan dari analisis ini, peneliti menyarankan untuk jumlah data latih agar ditambah kuantitasnya, dan tetap berusaha untuk mempertahankan kualitas data pada standar yang baik.
- Kematangan data perlu ditingkatkan, dengan cara melakukan penyesuaian pertanyaan dengan analisis, sehingga akurasi dari data meningkat.

5. DAFTAR PUSTAKA

- Bororing, J. E. (2022). IMPLEMENTASI DASHBOARD MICROSOFT POWER BI UNTUK VISUALISASI DATA COVID 19 INDONESIA. *Informasi Interaktif*, 7(1), 21-29.
- Haryani, C. A., Tohari, H., Marhamah, M., & Nurrahman, Y. A. (2018). Sentimen Analisis Kepuasan Pelanggan E-commerce Menggunakan Lexicon Classification dengan R. Konferensi Nasional Sistem Informasi (KNSI) 2018.
- Rakhmah, A. H., & Putri, T. A. (2022). Analisis Sentimen Terhadap Pasangan Calon Presiden 2019 Pada Media Sosial Twitter. *JURNAL LENTERA ICT*, 5(1), 1-11.
- Romadhon, F., Suwondo, A., & Sibyan, H. (2022). IMPLEMENTASI ALGORITMA NAÏVE BAYES PADA APLIKASI PREDIKSI PERTUMBUHAN PENDUDUK TAHUNAN DI KABUPATEN WONOSOBO. *Journal of Economic, Business and Engineering (JEBE)*, 3(2), 301-310.
- Sibyan, H., & Hasanah, N. (2022). Analisis Sentimen Ulasan Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn). *Jurnal Penelitian dan Pengabdian Kepada Masyarakat UNSIQ*, 9(1), 38-47.
- Singgalen, Y. A. (2021). Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi

Pemasaran Destinasi Pariwisata Prioritas di Indonesia. Journal of Information Systems and Informatics, 3(3), 459-470.