MEANS (Media Informasi Analisa dan Sistem)
p-ISSN : 2548-6985, e-ISSN : 2599-3089

Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Analisis Sentimen Komentar Kebijakan Full Day School

[1)] Yarma Agustya Dewi Utami
Universitas Labuhan Batu, Rantau Prapat, Indonesia
E-Mail: yarmadewi@yahoo.com

[2)] Volvo Sihombing
Universitas Labuhan Batu, Rantau Prapat, Indonesia
E-Mail: volvolumbantoruan@gmail.com

[3)] Muhammad Halmi Dar
Universitas Labuhan Batu, Rantau Prapat, Indonesia
E-Mail: mhd.halmidar@gmail.com

ABSTRAK

Sentiment analysis is an important research topic and is currently being developed. Sentiment analysis is carried out to see the opinion or tendency of a person's opinion on a problem or object, whether it tends to have a negative or positive view. The main purpose of this research is to find out public sentiment towards the Full Day school policy comments from the Facebook Page of the Ministry of Education and Culture of the Republic of Indonesia and to determine the performance of the Na-ïve Bayes Classifier Algorithm. The results of this study indicate that the public's negative sentiment towards the Full Day School policy is higher than positive or neutral sentiment. The highest accuracy value is the Naïve Bayes Classifier algorithm with the trigram feature selection of the 300 data training model with a value of 80%. This simulation has proven that the larger the training data and the selection of features used in the NBC Algorithm affect the accuracy of the results. Meanwhile, the simulation results from 10 test data with 5 different NBC and Lexicon algorithms also show that the Full Day School Policy proposed by the Indonesian Minister of Education and Culture has a higher negative sentiment than positive or neutral by most Facebook users who express opinions through comments. The highest accuracy value is the Naïve Bayes Classifier algorithm with the trigram feature selection of the 300 data training model with a value of 80%. This simulation has proven that the larger the training data and the selection of features used in the NBC Algorithm affect the accuracy of the results. Meanwhile, the simulation results from 10 test data with 5 different NBC and Lexicon algorithms also show that the Full Day School Policy proposed by the Indonesian Minister of Education and Culture has a higher negative sentiment than positive or neutral by most users. Facebook that expresses opinions through comments. The highest accuracy value is the Naïve Bayes Classifier algorithm with the tri-gram feature selection of the 300 data training model with a value of 80%. This simulation has proven that the larger the training data and the selection of features used in the NBC Algorithm affect the accuracy results.

Kata kunci: Sentiment Analysis; Naïve Bayes classifier; Lexicon Based Method

PENDAHULUAN

Pesatnya perkembangan teknologi digital dalam beberapa tahun terakhir telah menjadi pemicu di masyarakat untuk menggunakan media sosial sebagai alat utama untuk terhubung dan berinteraksi dengan keluarga, teman dan kolega. Ada platform media sosial teratas yang paling sering digunakan oleh masyarakat Indonesia seperti Facebook. Instagram dan Twitter. Menurut data dari Asosiasi Penyelenggara Jasa Internet Indonesia atau APJII dalam surveinya menyebutkan bahwa Facebook memiliki pengunjung terbesar dengan 71,6 juta pengunjung atau 54% dari populasi Indonesia diikuti oleh Instagram dengan 19,9 juta (15%) dan Youtube dengan 14,5 juta (11%) pada tahun 2016.

Di sisi lain, Menteri Pendidikan dan Kebudayaan RI Muhadjir Effendy menerbitkan peraturan baru tentang Kebijakan Full Day School melalui Permendikbud Nomor 23 Tahun 2017. Permendikbud ini sempat viral dan menjadi pro kontra di kalangan warga setelah peraturan ini dikeluarkan. diterbitkan. Menyadari pertumbuhan media sosial yang selalu meningkat terutama dengan Facebook, Kementerian Pendidikan dan Kebudayaan mulai mensosialisasikan Kebijakan Permendikbud Full Day School melalui Facebook Page resmi mereka yang disebut “Kemdikbud.RI”. Banyak pendapat yang disampaikan melalui komentar

dari pengguna Facebook baik yang positif maupun yang negatif dalam menanggapi Kebijakan Full Day School.

Analisis sentimen adalah proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam sebuah kalimat opini. Analisis sentimen dilakukan untuk melihat opini atau kecenderungan opini seseorang terhadap suatu masalah atau objek, baik yang cenderung berpandangan negatif maupun positif [1] Analisis sentimen dapat menggunakan algoritma klasifikasi dalam bidang text mining seperti Naive Bayes, Decision Tree, C.45, k-NN dan lain sebagainya. Algoritma Naive Bayes dapat digunakan dan memiliki hasil yang cukup baik dalam mengklasifikasikan sentimen dibandingkan dengan algoritma lainnya [2]

Penelitian ini membahas tentang analisis sentimen kebijakan Full Day School dari Facebook Page Kemendikbud RI menggunakan Algoritma Naïve Bayes Classifier. Algoritma Naïve Bayes dalam penelitian penulis menggunakan seleksi fitur karakter tri-gram dan quad-gram serta menggunakan dua model data latih yang berbeda, yaitu 200 dan 300 data latih. Tujuan dari penelitian ini adalah untuk mengetahui sentimen publik dalam komentar kebijakan Full Day school dari Facebook Kemendikbud RI menggunakan Algoritma Naïve Bayes Classifier (NBC) dan untuk mengetahui implementasi dan kinerja Algoritma Naïve Bayes Classifier dengan Tri-gram dan Quad- gram Pemilihan fitur karakter dengan model data latih yang berbeda.

METODE

Dalam penelitian ini ada dua metode yang digunakan yaitu metode pengumpulan data dan metode simulasi.

2.1. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan adalah studi kepustakaan dan studi lapangan menggunakan metode observasi. Dalam studi kepustakaan penulis mengumpulkan data dari buku, jurnal atau literatur sejenis yang berhubungan dengan penelitian penulis sebagai referensi sehingga dapat membantu penulis dalam melakukan penelitian ini. Dalam studi lapangan, penulis mengamati dan mengambil data dari Facebook API tentang komentar Netizen pada postingan Full Day School Policy (FDS) dari Facebook Page resmi Kemendikbud RI. Peneliti mengambil data pada tanggal 25 September 2017.

2.2. Metode Simulasi

Ada berbagai jenis siklus hidup yang dapat digunakan untuk studi tentang pemodelan dan simulasi. Tahap simulasi dalam penelitian ini adalah Perumusan Masalah, Model Konseptual, Pengumpulan Data Input/Output, Tahap Modeling, Tahap Simulasi, Verifikasi, Validasi, dan Eksperimen, Tahap Analisis Output [3].

2.3. Simulasi Masalah

Berdasarkan ketiga penelitian sebelumnya pada bab kedua, penulis menemukan hasil analisis bahwa Algoritma Naïve Bayes Classifier pada penelitian sebelumnya hanya menggunakan satu seleksi fitur dan klasifikasi sentimen data latih dilakukan secara manual.

Dalam penelitian ini penulis menggunakan Algoritma Naïve Bayes Classifer dengan pemilihan dua fitur yaitu karakter Trigram dan Quadgram dengan klasifikasi sentimen data latih menggunakan metode Lexicon Based.

Setiap pemilihan fitur menggunakan dua model data latih yang berbeda, yaitu model latih 200 dan 300 data.

2.4. Model konseptual

Dalam penelitian ini, model konseptual membahas keseluruhan penelitian ini. Pertama dengan mengidentifikasi masukan dalam penelitian ini, adalah komentar netizen terkait kebijakan Full Day School. Kedua, komentar yang telah terkumpul selanjutnya diolah dengan Algoritma Naïve Bayes Classifier dengan seleksi fitur Tri-gram dan seleksi fitur Quad-gram secara manual.

Setelah diolah secara manual, kemudian dilakukan perbandingan antara Algoritma Naïve Bayes dengan Trigram feature selection model 200 Data Training dan model 300 Data Training dan Quad-gram feature selection 200 Data Training dan 300 Data Training Model dengan klasifikasi atau pelabelan sentimen training data dengan metode Lexicon Based. Jadi ada empat hasil tingkat akurasi yang bisa dibandingkan.

2.5. Pengumpulan Data Input/Output

Data diperoleh dari komentar netizen tentang kebijakan full day school di postingan Halaman Resmi Kemmikbud RI. Total ada 310 komentar yang terbagi menjadi 300 data latih dan 10 data uji. Selain data dari Facebook API, terdapat leksikon positif dan leksikon negatif untuk pelabelan data latih dan data uji (kelas aktual). Hasil atau keluaran yang diperoleh dari penelitian ini adalah hasil seleksi fitur N gram dan data pengujian yang telah diklasifikasikan.

Setelah diolah secara manual, kemudian dilakukan perbandingan antara Algoritma Naïve Bayes dengan Trigram feature selection model 200 Data Training dan model 300 Data Training dan Quad-gram feature selection 200 Data Training dan 300 Data Training Model dengan klasifikasi atau pelabelan sentimen training data dengan metode Lexicon Based. Jadi ada empat hasil tingkat akurasi yang bisa dibandingkan.

MEANS (Media Informasi Analisa dan Sistem)
p-ISSN : 2548-6985, e-ISSN : 2599-3089

2.6. Fase Pemodelan

Pada konstruksi algoritma Naïve Bayes Classifier terdiri dari dua tahap yaitu tahap pelatihan dan tahap pengujian.

1.Tahap Pelatihan

a. Ambil data pelatihan dari setiap sentimen (positif, negatif, dan netral)
b. Data latih kemudian diolah menggunakan pengolahan teks (case folding, cleansing, stopword removal)
c. Setelah diolah kemudian data latih di-stem. Untuk prosedur stemming penulis menggunakan Algoritma Nazief & Andriani. Prosedur algoritma stemming Nazief dan Andriani dijelaskan pada Gambar 1 [4][5]
d. Pilih pemilihan fitur Tri-gram atau Quad-gram.
e. Pemilihan fitur dengan n gram dipilih
f. Hitung frekuensi setiap n gram
g. Hitung total n gram kata
h. Hitung peluang setiap diagram[6]

$$P\langle Xi|Vj\rangle = \frac{nk+1}{n+|word|} \qquad (1)$$

i. Hitung probabilitas setiap dokumen sentimen [7]

$$P(Vj) = \frac{|docsj|}{|sample|} \qquad (2)$$

2.7. Fase Pengujian

a. Ambil data pengujian
b. Data pengujian kemudian diolah menggunakan pemrosesan teks (case folding, cleansing, stopword removal)
c. Setelah diolah kemudian data pengujian di-stem. Untuk prosedur stemming penulis menggunakan Algoritma [8][9][10]. Prosedur algoritma stemming Nazief dan Andriani dijelaskan pada Gambar 1 [11][12].
d. Pilih pemilihan fitur Tri-gram atau Quad-gram
e. Pilihan fitur N gram
f. Hitung frekuensi kemunculan setiap n gram
g. Membandingkan n gram data pengujian dengan n gram data pelatihan setiap sentimen.
h. Jika n gram data uji sama dengan n gram data latih, maka probabilitas n gram data latih menjadi n gram probabilitas data uji.
i. jika tidak, maka frekuensi n gram data pengujian adalah 0, lalu hitung peluang data pengujian n gram
j. Hitung nilai Vmap masing-masing kategori sentimen [13]

$$Vmap = \prod_{VjeV}^{ar\max} \prod_{i=1}^{n} P(Xi|Vj)P(Vj) \qquad (3)$$

k. Ambil Nilai Vmap tertinggi dari setiap kategori sentimen
l. Data pengujian diklasifikasikan

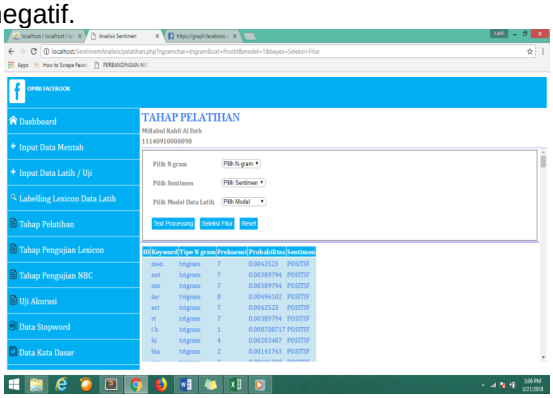
2.8. Fase Simulasi

Pada tahap simulasi ini akan disimulasikan Analisis Sentimen Kebijakan Full Day School Menggunakan Algoritma Naïve Bayes Classifier sesuai tabel 1.

Tabel 1. Fase Simulasi Algoritma NBC

Variabel/Simulasi Parameter	Fase Simulasi
Faktor 1	Fase pelatihan data dengan Naïve Bayes Classifier Algoritma berdasarkan sentimen, pemilihan fitur, dan data
Faktor 2	Tahap pengujian data dengan Algoritma Naïve Bayes Classifier berdasarkan pemilihan fitur dan model pelatihan data
Faktor 3	Fase pengujian akurasi setiap Algoritma NBC

Hasil simulasi dari tabel I sesuai dengan Gambar 1, Gambar 2 dan Gambar 3. Gambar 1 merupakan simulasi tahap pelatihan sentimen positif dengan pemilihan fitur trigram dari 200 model data pelatihan yang telah dihitung probabilitas setiap kata kunci. Gambar 3 menggambarkan hasil klasifikasi algoritma NBC dengan pemilihan fitur trigram dari 200 model data latih. Proses klasifikasi dilakukan dengan menghitung nilai probabilitas dari setiap sentimen. Sebelum menghitung probabilitas setiap sentimen, kata kunci dari data n gram yang telah disimpan dalam database dibandingkan dengan kata kunci data latih. Jika kata kunci pada data uji sama dengan kata kunci pada data latih, maka peluang kata kunci pada data latih menjadi peluang kata kunci dalam data uji. Sebaliknya jika tidak sama maka sistem akan menghitung probabilitas kata kunci. Setelah probabilitas semua kata kunci di setiap sentimen diperoleh, maka nilai Vmap dapat dihitung. Untuk menentukan sentimen diambil nilai Vmap tertinggi dari semua sentimen. Pada Gbr.3 didapatkan nilai Vmap tertinggi 1.9637096024859E-255 dengan sentimen negatif.



Gambar 1. Simulasi Tahap Pelatihan

MEANS (Media Informasi Analisa dan Sistem)
p-ISSN : 2548-6985, e-ISSN : 2599-3089

Gambar 2. Simulasi Algoritma NBC Tahap Pengujian

G**ambar 3.** Hasil Vmap

Pada simulasi pengujian akurasi diperoleh hasil bahwa akurasi algoritma NBC dengan pemilihan fitur trigram data latih model 300 lebih tinggi dibandingkan dengan nilai akurasi algoritma NBC lainnya yaitu sebesar 80%. Untuk menghitung nilai akurasi dari setiap pemilihan fitur, penulis menggunakan rumus sebagai berikut (4).

$$Accuracy = \frac{TruePositive + TrueNegative + TrueNeutral}{TotalofTestingData} \quad (4)$$

HASIL DAN PEMBAHASAN

Berikut adalah hasil penelitian kami. Verifikasi, Validasi, dan Eksperimen, Peneliti melakukan verifikasi untuk memastikan setiap tahapan yang telah dilakukan sebelumnya memiliki keterkaitan dan menghasilkan output sesuai dengan yang diharapkan. Selanjutnya pada proses validasi dilakukan pengujian kebenaran algoritma NBC dengan membandingkan performansi masing-masing algoritma NBC dari segi akurasi dan eksperimen dengan membandingkan hasil skenario yang merupakan hasil klasifikasi sentimen dari data pengujian pada masing-masing Algoritma NBC. Dari percobaan, analisis keluaran akan dibahas pada Tahap Analisis Keluaran.

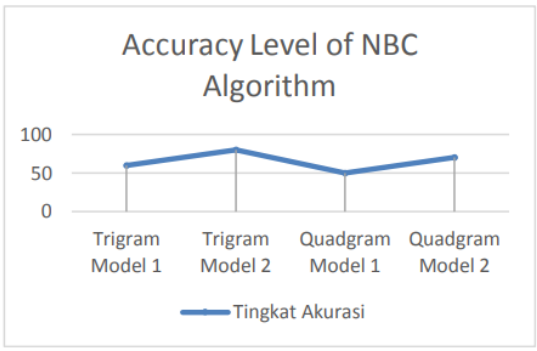
Fase Analisis Output

Pada fase ini dijelaskan output dari klasifikasi 10 data pengujian yang diambil dari komentar netizen atas kebijakan full day school Kementerian Pendidikan dan Kebudayaan RI. Karena pengklasifikasian menggunakan algoritma NBC dengan pemilihan dua fitur yang berbeda dan dua model data latih, serta menggunakan metode Lexicon Based untuk mengetahui sentimen aktual dari 10 data pengujian, maka dalam penelitian ini terdapat lima keluaran klasifikasi sentimen yang digambarkan pada Tabel 2.

Tabel 2. Kalisifikasi Hasil Sentimen NBC

Data ke n	KLASIFIKASI HASIL SENTIMEN NBC				
	Trigram Model 1*	Trigram Model 2*	Segiempat Model 1	Segiempat Model 2	Segiempat Model 3
1	Netral	Negatif	Netral	Negatif	Negatif
2	Negatif	Negatif	Negatif	Negatif	Negatif
3	Negatif	Negatif	Netral	Negatif	Negatif
4	Negatif	Negatif	Negatif	Negatif	Negatif
5	Negatif	Negatif	Negatif	Negatif	Negatif
6	Negatif	Negatif	Negatif	Negatif	Negatif
7	Negatif	Negatif	Negatif	Negatif	Negatif
8	Negatif	Negatif	Netral	Negatif	Positif
9	Netral	Negatif	Netral	Negatif	Positif
10	Netral	Negatif	Netral	Negatif	Negatif

Pada bab ini disajikan analisis akurasi algoritma Naïve Bayes Classifier (NBC) dengan pemilihan fitur tri-gram dan quad-gram dengan dua model data latih yang berbeda. Pada simulasi yang telah dilakukan didapatkan hasil grafik sebagai berikut, gambarkan pada gambar 4.



Gambar 4. Grafik Tingkat Akurasi ALgoritma NBC

Tabel 3. Tingkat Akurasi Algoritma NBC

No	Algoritma NBC	Tingkat Keakuratan
1	Trigram Model 1*	60%
2	Trigram Model 2**	80%
3	Quadram Model 1 *	50%
4	Quadram Model 2 **	70%

*Model 1 mewakili 200 data pelatihan **Model 2 mewakili 300 data pelatihan.

Berdasarkan grafik pada gambar 5 dan tabel 3, diperoleh analisis sebagai berikut :

Pemilihan fitur Trigram Algoritma NBC memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma NBC pemilihan fitur quad gram baik menggunakan 200 data latih maupun 300 data latih. Pada penelitian "Implementasi Karakter N-Gram Untuk Analisis Sentimen Hotel Review Menggunakan Algoritma Naive Bayes" nilai akurasi pemilihan fitur quad gram lebih tinggi dari akurasi pemilihan fitur trigram yaitu masing-masing sebesar 83,67% dan 84,67% (Indrayuni & Wahyudi, 2015). Hasil akurasi seleksi fitur trigram dan quad gram dari penelitian sebelumnya berbeda dengan penelitian penulis. Perbedaan ini terjadi karena pada penelitian sebelumnya klasifikasi data latih masih manual sedangkan pada penelitian penulis klasifikasi data latih sudah menggunakan metode berbasis Lexicon. para penulis analisis didukung oleh penelitian lain menggunakan metode Lexicon Based "Improved Twitter Sentiment Analysis Using N Gram Feature Selection and Combinations" akurasi pemilihan fitur trigram lebih rendah dari dua seleksi fitur lainnya (unigram dan bigram) sebesar 83,652% untuk unigram, 84,064% untuk bigram dan 70,532% untuk trigram (Awachate & Kshirsagar, 2007). Berdasarkan analisis penulis, ditemukan bahwa penambahan n gram feature selection pada Algoritma NBC tidak berpengaruh dalam meningkatkan hasil akurasi jika klasifikasi data latih menggunakan Lexicon Based. 652% untuk unigram, 84,064% untuk bigram dan 70,532% untuk trigram (Awachate & Kshirsagar, 2007). Berdasarkan analisis penulis, ditemukan bahwa penambahan n gram feature selection pada Algoritma NBC tidak berpengaruh dalam meningkatkan hasil akurasi jika klasifikasi data latih menggunakan Lexicon Based. 652% untuk unigram, 84,064% untuk bigram dan 70,532% untuk trigram (Awachate & Kshirsagar, 2007). Berdasarkan analisis penulis, ditemukan bahwa penambahan n gram feature selection pada Algoritma NBC tidak berpengaruh dalam meningkatkan hasil akurasi jika klasifikasi data latih menggunakan Lexicon Based.

Algoritma NBC dengan model 300 data latih memiliki tingkat akurasi yang lebih tinggi dari algoritma NBC dengan model latih 200 data baik menggunakan seleksi fitur trigram atau seleksi fitur quad gram. Hasil tersebut menunjukkan bahwa semakin banyak data latih yang digunakan maka semakin tinggi akurasi Algoritma NBC, karena algoritma NBC merupakan metode pembelajaran terawasi yang sangat mengandalkan data latih

KESIMPULAN

Berdasarkan hasil klasifikasi lima sentimen menggunakan Algoritma Naïve Bayes dengan dua seleksi fitur yang berbeda dan dua model data latih, ditemukan sentimen Negatif publik dalam hal ini netizen yang mengomentari Kebijakan Full Day School (FDS) di Halaman Facebook Kemendikbud RI lebih besar dari sentimen positif atau netral. Berdasarkan simulasi yang dilakukan dengan menggunakan aplikasi Analisis Sentimen yang dibangun oleh penulis menggunakan Bahasa Pemrograman PHP dan database Mysql diperoleh hasil bahwa akurasi Algoritma Naïve Bayes Classifier (NBC) dengan pemilihan fitur trigram model 300 data latih lebih tinggi daripada algoritma NBC dengan trigram 200 data pelatihan dan algoritma NBC dengan quadgram 200 dan 300 data pelatihan

DAFTAR PUSTAKA

[1] F. A. Sianturi, B. Sinaga, P. M. Hasugian, T. Informatika, and S. Utara, "Fuzzy Multiple Attribute Decisison Macking Dengan Metode Oreste Untuk Menentukan Lokasi Promosi," *J. Inform. Pelita Nusant.*, vol. 3, no. 1, pp. 63–68, 2018, [Online]. Available: http://e-jurnal.pelitanusantara.ac.id/index.php/JIPN/article/view/289.

[2] Fricles Ariwisanto Sianturi, "Analisa metode teorema bayes dalam mendiagnosa keguguran pada ibu hamil berdasarkan jenis makanan," *Tek. Inf. dan Komput.*, vol. 2, no. 1, pp. 87–92, 2019, [Online]. Available: http://jurnal.murnisadar.ac.id/index.php/Tekinkom/article/view/78.

[3] T. Juninda, E. Andri, U. Kahirunnisa, N. Kurniawati, and M. Mustakim, "Penerapan Metode Promethee Untuk Pendukung Keputusan Pemilihan Smartphone Terbaik," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 2, p. 224, 2019, doi: 10.24014/rmsi.v5i2.7677.

[4] S. P. Tamba, D. R. Hia, D. Prayitna, and A. Tryvaldy, "Pemanfaatan Teknologi Berbasis Mobile Untuk Manajemen Kontrol Nilai Dan Absensi Siswa Pada Mts Al-Ittihadiyah Medan," vol. 2, no. 1, pp. 18–22, 2020.

[5] D. A. Butar-butar, D. Amalia, K. Mayra, A. Nst, and Y. Naibaho, "Pemanfaatan Teknologi Informasi Dalam Pengambilan Keputusan Penilaian Karyawan Terbaik," vol. 2, no. 1, pp. 43–46, 2020.

[6] J. Banjarnahor and A. X. Lim, "Aplikasi Pembayaran Uang Kuliah Pada Universitas Prima Indonesia Menggunakan Metode Fuzzy Logic Berbasis Android," vol. 2, no. 1, pp. 7–13,

2020.

[7] O. Sihombing, N. S. Nainggolan, B. L. Gaol, and N. Kesuma, “Rancang Bangun Aplikasi Objek Wisata Kabupaten Tapanuli Tengah Berbasis Android,” vol. 2, no. 1, pp. 14–17, 2020.

[8] J. Wijaya, V. Frans, and F. Azmi, “Aplikasi Traveling Salesman Problem Dengan GPS dan Metode Backtracking,” vol. 3, no. 2, pp. 81–90, 2020.

[9] B. Krismoyo and J. R. Sagala, “PENERAPAN METODE WEIGHTED PRODUCT (WP) MENENTUKAN SISWA DROP OUT PADA,” vol. 3, no. 2, pp. 8–14, 2020.

[10] W. Purba, D. Ujung, T. Wahyuni, L. Sihaloho, and J. Damanik, “PERANCANGAN SISTEM INFORMASI PEMESANAN TIKET ONLINE PADA KMP . IHAN BATAK BERBASIS,” vol. 3, no. 2, pp. 65–75, 2020.

[11] A. Firman, H. F. Wowor, X. Najoan, J. Teknik, E. Fakultas, and T. Unsrat, “Sistem Informasi Perpustakaan Online Berbasis Web,” *E-Journal Tek. Elektro Dan Komput.*, 2016.

[12] D. Sitanggang, S. Simangunsong, R. U. Sipayung, and A. S. Nababan, “Perancangan Aplikasi Penyeleksian Penerimaan Siswa Untuk Mengikuti Oliampiade Sains Berbasis Android,” vol. 3, no. 2, pp. 34–43, 2020.

[13] B. Kurniawan, S. Effendi, and O. S. Sitompul, “Klasifikasi Konten Berita Dengan Metode Text Mining,” *J. Dunia Teknol. Inf.*, 2012.