

Analysis of Customer Sentiment at Paga Lewu Store Using the Naive Bayes Classifier Method

Gergorius Kopong Pati^{1*}, Apliana Mata², Fiandro Markus Laki Riti³, Apliana Umbu Lele⁴, Kristofel Bili⁵, Ferianus Dappa Ole⁶, Samuel Weri⁷

¹⁻⁷ Teknik Informatika, Univeristas Stella Maris Sumba, Indonesia

^{1*} gregkopong80@gmail.com, ³ yandromarkus@gmail.com, ⁴ aplianaumbu5@gmail.com,

⁵ ristobili07@gmail.com, ⁶ ferianusdappaole@gmail.com

Alamat: Jl. Pastoran No. 05 Tambolaka Kab. Sumba Barat Daya

Korespondensi penulis: gregkopong80@gmail.com

Abstract: Sentiment Analysis is a technique for extracting text data to obtain information about positive, neutral or negative sentiments. The purpose of sentiment analysis is given by internet users on social media to provide a personal assessment or opinion. Paga Lewu Shop that often gets user sentiment through social media is Paga Lewu Shop. The existence of consumer opinion sentiments about Paga Lewu Shop can be analyzed and utilized to obtain useful information for other customers and the Paga Lewu Shop. By using the Text Mining technique classification method, a sentiment will be known as positive, neutral or negative. One of the algorithms widely used in sentiment analysis is the Naive Bayes classification method. This study uses the Naive Bayes Classifier (NBC) method with tf-idf weighting accompanied by the addition of an emotion icon conversion feature (emoticon) to determine the existing sentiment class from tweets about the Paga Lewu Shop. The results of the study show that the Naive Bayes method without additional features is able to classify sentiment with an accuracy value of 96.44%, while if the tf-idf weighting feature is added along with the conversion of emotion icons, the accuracy value can be increased to 98%.

Keywords: Navie, Bayes, Classifier

Abstrak: Analisis Sentimen adalah suatu teknik mengekstrak data teks untuk mendapatkan informasi tentang sentimen bernilai positif, netral maupun negatif. Tujuan analisis sentimen diberikan oleh pengguna internet pada media sosial untuk memberikan suatu penilaian atau opini pribadi. Toko Paga Lewu yang sering mendapatkan sentimen pengguna melalui media sosial adalah Toko Paga Lewu. Adanya sentimen opini dari konsumen tentang Toko Paga Lewu dapat dianalisis dan dimanfaatkan untuk mendapatkan informasi yang berguna bagi pelanggan lain maupun pihak toko Toko Paga Lewu. Dengan menggunakan teknik Text Mining metode klasifikasi, akan diketahui suatu sentimen bernilai positif, netral atau negatif. Salah satu algoritme yang banyak digunakan dalam analisis sentimen adalah metode klasifikasi Naive Bayes. Penelitian ini menggunakan metode Naive Bayes Classifier (NBC) dengan pembobotan tf-idf disertai penambahan fitur konversi ikon emosi (emoticon) untuk mengetahui kelas sentimen yang ada dari tweet tentang toko Toko Paga Lewu. Hasil penelitian menunjukkan bahwa metode Naive Bayes tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 96,44%, sementara jika ditambahkan fitur pembobotan tf-idf disertai konversi ikon emosi mampu meningkatkan nilai akurasi menjadi 98%.

Kata Kunci : Navie, Bayes, Classifier

1. PENDAHULUAN

Kepuasan mahasiswa ditentukan oleh kualitas yang dikehendaki mahasiswa, sehingga jaminan kualitas menjadi prioritas utama bagi setiap perguruan tinggi, yang saat ini dijadikan sebagai salah satu tolak ukur keunggulan daya saing perguruan tinggi. Banyaknya opini masyarakat pada facebook yang membahas mengenai toko Toko Paga Lewu tersebut bersifat subyektif. Opini tersebut merupakan textual data yang dapat dianalisa dan dimanfaatkan untuk mendapatkan informasi guna memantau sentimen masyarakat terhadap kinerja Toko Paga Lewu menggunakan menggunakan teknik text mining yaitu sentimen analisis. Sentimen

tersebut dikategorikan menjadi sentimen positif, netral dan negatif. Penelitian ini memodelkan proses klasifikasi sentimen pengguna facebook terhadap toko Toko Paga Lewu untuk mengetahui layanan maupun produk dari marketplace yang bermanfaat bagi pelanggan maupun penjual menggunakan metode Naïve Bayes Classifier (NBC) yang dikembangkan pembobotan tf-idf dan fitur konversi ikon emosi.

Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Sebuah pengklasifikasian yang sangat diperlukan untuk mengatasi masalah yang terjadi dalam hal perekapan angket mahasiswa.

Naïve bayes classifier adalah metode klasifikasi yang berdasarkan probabilitas dan Teorema Bayesian dengan asumsi bahwa setiap variabel X bersifat bebas atau berdiri sendiri dan tidak ada kaitannya dengan variabel lainnya. Klasifikasi Naive Bayes juga memperlihatkan tingginya akurasi dan cepat ketika digunakan untuk dataset dengan jumlah besar. Berdasarkan latar belakang yang telah dijelaskan, penulis akan membuat Analisis Sentimen Pelanggan Toko Paga Lewu Menggunakan Metode Naive Bayes Clasifer ditampilkan.

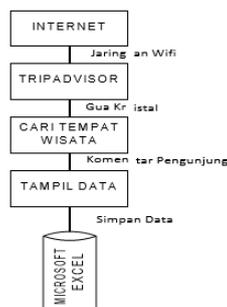
2. METODOLOGI PENELITIAN

Penentuan Tempat Wisata

Penelitian ini dimulai dengan menentukan tempat wisata, selanjutnya mengumpulkan data yang berkaitan dengan penelitian sehingga penelitian akan lebih terarah.

Pengumpulan Data

Pengumpulan data dilakukan berdasarkan komentar pengunjung. Data didapat dengan memanfaatkan web Trip Advisor dengan tools yang digunakan yaitu Rapidminer Studio 9.6.0. Selanjutnya data akan disimpan ke Microsoft Excel 2007. Data yang diambil untuk melakukan analisis sentimen sebanyak 100 data. Gambar 1 menunjukkan proses pengumpulan data.



Gambar 1. Proses Pengumpulan Data

☐ ☐	Sedih
:-) :\$ >:[☐	Kecewa
;-) ;)	Kedip
:-P :P	Ejek
:-:/	Ragu

c. Menghapus Simbol

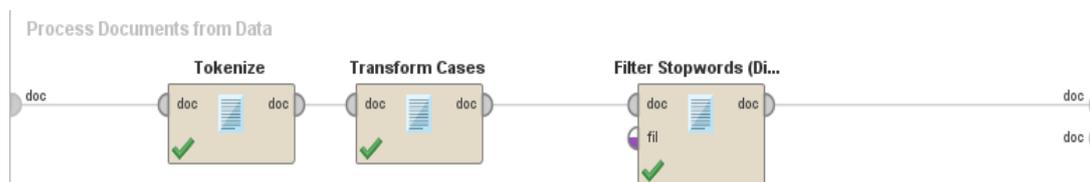
Tahapan ini dilakukan proses untuk membuang simbol dan tanda baca dalam komentar.

d. Normalisasi Kalimat

Normalisasi kalimat adalah proses untuk mengubah data yang tidak baku menjadi kalimat baku. Hal ini dilakukan karena banyak komentar yang menggunakan kalimat tidak baku sehingga akan sulit dilakukan dalam pengujian data. Contoh kalimat tidak baku yang ditemukan “bangeetttt”, “engga”, “tdk”, “gak”, “tak”, “keren”.

e. *Tokenizing*

Tokenizing adalah Proses membagi teks yang dapat berupa kalimat atau paragraf untuk menjadi bagian-bagian tertentu. Proses ini merupakan proses akhir untuk diolah datanya menjadi suatu sentimen analisis yang setiap kalimat dibagi menjadi masing-masing token biasa dilakukan oleh peneliti. Berikut contoh penerapan dari *tokenizing*.



Gambar 3. *Tokenizing* pada Tolls RapidMiner

Tabel 2. Penerapan *Tokenizing*

Data Latih	
Input	Output
Tempat ini begitu unik, gua dengan kolam. Airnya seperti kristal dan bersih. Tetapi ketika saya mengunjungi begitu banyak orang local yang berenang di sini	[Tempat, ini, begitu, unik, gua, dengan, kolam, airnya, seperti, kristal, dan, bersih, tetapi, ketika, saya, mengunjungi, begitu, banyak, orang, local, yang, berenang, di, sini]

Naive Bayes Classifier

Naive Bayes merupakan sebuah metoda klasifikasi menggunakan metode probabilitas dan statistik. Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Rumus Naive Bayes :

$$P(C_j/W_i) = P(c_j) \times P(w_i/c_j)$$

$P(w_i)$

Keterangan :

x : Data dengan class yang belum diketahui

c : Hipotesis data merupakan suatu class spesifik

$P(c/x)$: Probabilitas hipotesis berdasar kondisi (posteriori probability)

$P(c)$: Probabilitas hipotesis (prior probability)

$P(x/c)$: Probabilitas berdasarkan kondisi pada hipotesis

$P(x)$: Probabilitas c

Algoritma K-Nearest Neighbor (K-NN)

K-NN adalah sebuah metode klasifikasi terhadap sekumpulan data berdasarkan data *training* yang sudah terklasifikasikan sebelumnya. Rumus k-NN ditunjukkan pada persamaan dibawah ini:

$$d_i = \sum_{i=1}^p (x_{2i} - x_{1i})^2$$

Keterangan:

x_1 : data latih

x_2 : data uji

i : variabel data

d : jarak

p : dimensi data

3. HASIL DAN PEMBAHASAN

Sumber Data

Proses data pada penelitian ini membagi data-data yang telah dilakukan proses preprocessing menjadi dua bagian yaitu data latih dan data uji. Pembuatan data latih ini bertujuan sebagai pedoman dalam menentukan hasil data uji. Data komentar yang disimpan dalam excel bisa dilihat pada gambar dibawah ini:

1	TEXT	SENTIMEN
2	saya puas dengan pekerjaan sebagai guru	positif
3	saya puas dengan pilihan pekerjaan saya karena menjadi guru itu sangat mulia	positif
4	saya puas dengan apa yang saya kerjakan karena ada gambaran tugas yang jelas untuk saya	positif
5	saya puas dengan gaji yang saya terima karena cukup untuk memenuhi kebutuhan keluarga	positif
6	saya puas karena diberi kesempatan pelatihan untuk mengembangkan diri	positif
7	saya puas dengan relasi kerja yang terbangun antara saya dengan kepala sekolah	positif
8	saya puas dengan relasi yang terbangun antara sesama guru di sekolah	positif
9	saya puas karena bisa dengan leluasa berkreasi dan berinovasi untuk mendorong kesuksesan	positif
10	saya puas dengan relasi yang terbangun antara sesama guru di sekolah	positif
11	saya puas dengan relasi yang terbangun antara sesama guru di sekolah	positif
12	saya puas dengan pilihan pekerjaan saya karena menjadi guru itu sangat mulia	positif
13	saya puas dengan gaji yang saya terima karena cukup untuk memenuhi kebutuhan keluarga	positif
14	saya puas dengan pilihan pekerjaan saya karena menjadi guru itu sangat mulia	positif
15	saya puas dengan gaji yang saya terima karena cukup untuk memenuhi kebutuhan keluarga	positif
16	saya puas dengan relasi yang terbangun antara sesama guru di sekolah	positif
17	kepala sekolah membuat guru-guru merasa nyaman berada di sekitarnya	positif
18	kepala sekolah Tidak memberikan gambaran tentang apa yang harus guru lakukan	negatif
19	kepala sekolah mendorong para guru untuk berpikir tentang sesuatu yang lama dengan cara	positif
20	kepala sekolah membantu guru-guru untuk menyelesaikan permasalahannya sendiri	positif

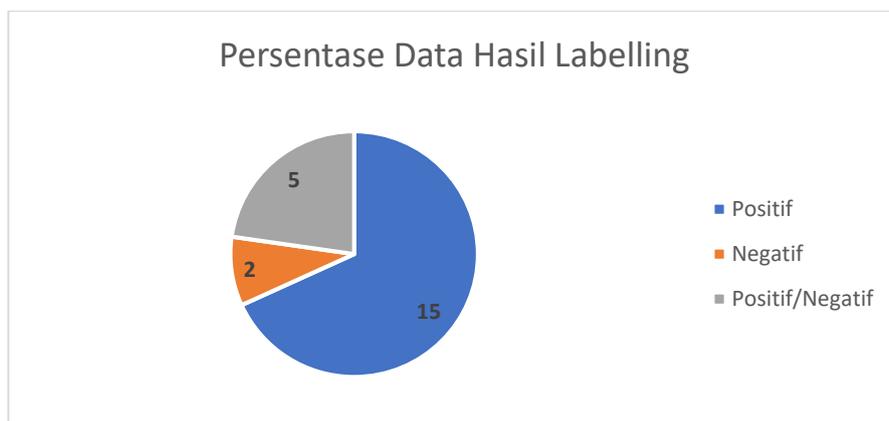
Gambar 4 Contoh Data Latih

Persentase Data Hasil Labelling

Pada proses pengambilan data yang dilakukan berdasarkan koisioner yang dibagikan guru-guru dengan menghasilkan data dengan jumlah 21 komentar.

Setelah seluruh komentar berhasil disimpan selanjutnya dilakukan proses pelabelan. Dalam proses pelabelan yang dilakukan seharusnya untuk menentukan komentar yang didapatkan apakah bernilai positif, negative. amun pada proses pelabelan pada penelitian ini terdapat beberapa komentar yangng tidak sesuai dengan sentimen komentar yang sebenarnya.

Dan hasil akhir dari proses pelabelan data menghasilkan data sebanyak 15 komentr bernilai positif, 2 komentar bernilai negatif dan 5 komentar bernilai positif negatif. Persentase data bernilai positif, negatif dan positif negatif dan dapat dilihat pada gambar 4.8

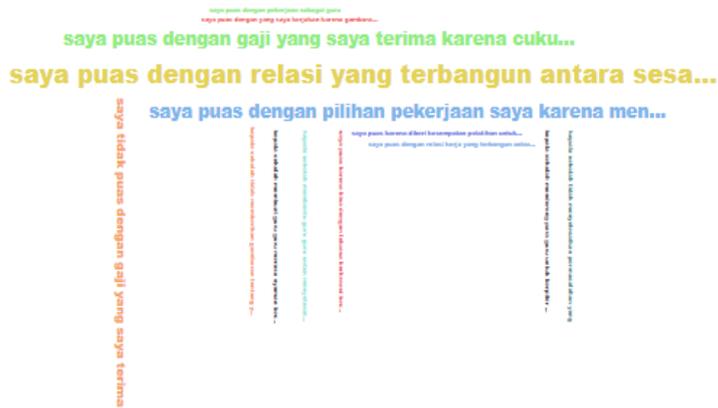


Gambar 5 Persentase Data Hasil Labelling

Ekstraksi Fitur

Setelah terbentuknya file yang akan dijadikan dataset, maka selanjutnya data tersebut akan dibentuk menjadi sebuah model klasifikasi. Namun sebelum membentuk model, ada beberapa tahapan yang harus dilakukan agar terbentuknya suatu model yang baik. Yang

pertama dilakukan adalah membaca file xlsx dan kemudian dilakukan tokenisasi terhadap seluruh dokumen dalam file tersebut. Berdasarkan hasil tokenisasi yang dilakukan, maka penulis juga ingin mengetahui frekuensi kata yang banyak diperbincangkan oleh pengunjung wisata, untuk itu penulis memvisualisasikannya dalam bentuk wordcloud pada gambar 4.18.



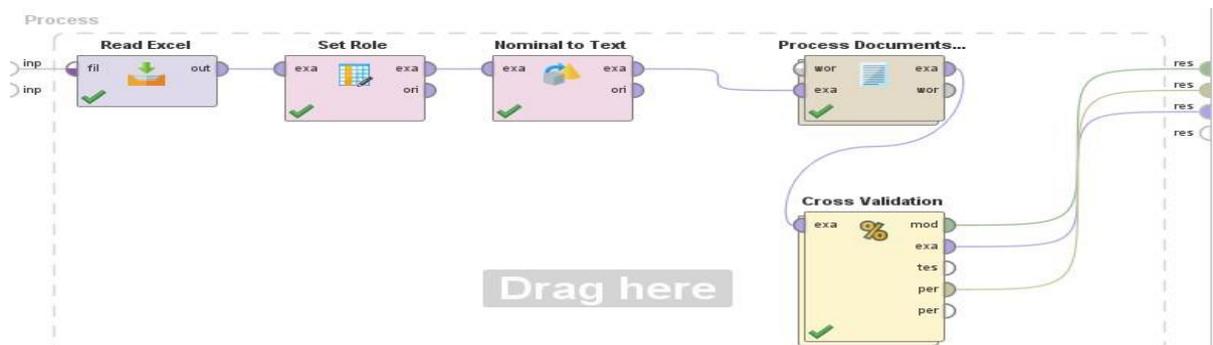
Gambar 6 Proses Document From Data

Perhitungan Akurasi Dengan RapidMiner

Penelitian ini menggunakan tools RapidMiner versi 10.1. perhitungan akurasinya akan proses dengan menggunakan metode Naive Bayes Classifier dan Support Vektor Machine. Untuk mengetahui tingkat akurasi dari kedua metode tersebut.

Proses Dokument dengan metode Naive Bayes Classifier

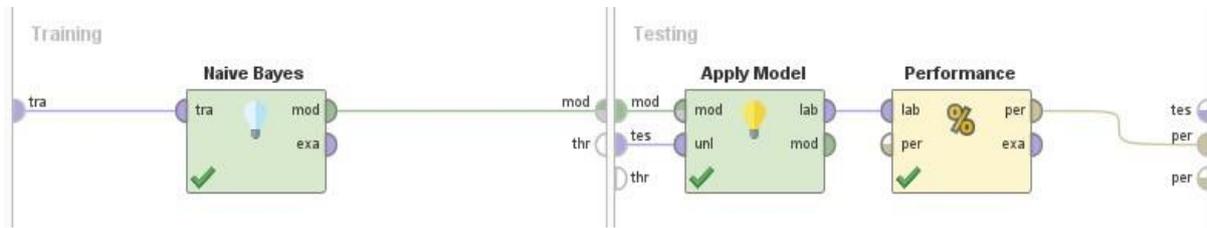
Pada proses ini diuji dengan data latih dari excel. Seperti yang terlihat pada gambardibawah ini



Gambar 7. Proses Dokumen Data Latih

Perhitungan Akurasi dengan Naive Bayes Classifier pada RapidMiner

Proses *Training* dan *Testing* dengan metode Naive Bayes Classifier pada RapidMiner ditampilkan pada gambar dibawah ini:



Gambar 8. Proses *Training* dan *Testing* dengan Naive Bayes Classifier pada RapidMiner

Hasil perhitungan klasifikasi menggunakan metode Metode Naive Bayes Classifier diperoleh sebesar 73.47% seperti yang terlihat pada gambar dibawah.

precision	accuracy: 77.27% +/- 42.89% (micro average: 77.27%)		
recall			
AUC (optimistic)			
AUC			
AUC (pessimistic)			
	true positif	true negatif	class precision
pred. positif	15	2	88.24%
pred. negatif	3	2	40.00%
class recall	83.33%	50.00%	

Gambar 9. Akurasi dengan Metode Naive Bayes Classifier pada RapidMiner

Hasil Pengujian Naive Bayes

Akurasi yang diuji pada tools RapidMiner dengan menggunakan metode Naive Bayes Classifier. Hasil pengujian akurasi menggunakan metode Naive Bayes Classifier sebesar 95,38%. oleh karena itu tingkat akurasi yang diperoleh dengan menggunakan metode yang diuji pada tools RapidMiner.

Tabel 3. Hasil Perhitungan Akurasi

Metode	Tingkat Predikat	True Positif	True Negatif
Naive Bayes Classifier	Pred. Positif	15	2
	Pred. Negatif	3	2
	Akurasi pada tools RapidMiner		
Naive Bayes Classifier		80.95%	

Text Mining

Teks merupakan sesuatu yang umum dalam melakukan pertukaran informasi. Syarat umum data dan teks mining adalah informasi yang diambil dan dapat menjadi data yang berguna.

Teks mining merupakan proses menganalisa teks untuk menjadi informasi yang berguna untuk tujuan tertentu. Informasi yang diambil harus jelas dan eksplisit, karena teks mining merubah menjadi bentuk yang dapat digunakan oleh komputer atau orang yang tidak memiliki waktu untuk membaca full teks (Bramer, 2007).

Analisis Sentimen

Analisis sentimen menentukan terhadap komentar mengenai permasalahan, apakah masalah tersebut memiliki kecenderungan positif atau negatif dan dapat dijadikan acuan untuk meningkatkan pelayanan atau kualitas produk (Nurjanah et al., 2017). Sentimen analisis adalah riset komputasional dari opini, sentimen dan emosi yang di ekspresikan secara tekstual (Zulfa & Winarko, 2017).

K- Nearest Neighbor

Metode K-Nearest Neighbor (k-NN) pertama kali di perkenalkan awal tahun 1960. Metode ini bekerja intensif ketika diberikan data training yang besar dan tidak populer sampai dengan tahun 1960. Ketika tekin komputasi sedang berkembang (Han & Kamber, 2006). k-NN memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih. Sedangkan kekurangan dari k-NN yaitu computational cost dan sensitif terhadap derau dan input yang tidak sesuai (Budianto et al., 2018).

Naive Bayes

Naive Bayes adalah metode klasifikasi data berdasarkan probabilitas yang bisa saja terjadi di masa depan, metode ini digunakan untuk klasifikasi biner dan *multiclass*. Metode yang juga dikenal sebagai Naive Bayes Classifier ini menerapkan teknik *supervised* klasifikasi objek di masa depan dengan menetapkan label kelas ke *instance*/catatan menggunakan probabilitas bersyarat. Probabilitas bersyarat adalah ukuran peluang suatu peristiwa yang terjadi berdasarkan peristiwa lain yang telah (dengan asumsi, praduga, pernyataan, atau terbukti) terjadi.

Teks Preprocessing

Teks preprocessing merupakan tahap awal dari text mining, proses ini untuk mempersiapkan data (Rustiana & Rahayu, 2017).

Methods

Metode yang digunakan pada penelitian ini merupakan metode penelitian eksperimen, dengan tahapan sebagai berikut:

- a. Pengumpulan Data Penulis mengambil data dari situs review www.tripadvisor.com. Data yang diambil terdiri dari 50 review positif dan 50 review negative.
- b. Pengolahan Awal Data Preprocessing pada dataset yang dilakukan yaitu:
 - 1) Tokenize
Tokenize merupakan tahap preprocessing dengan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.
 - 2) Filter Stopword
Pada tahap filter stopwords membuang kata-kata berupa kata sambung, imbuhan, akhiran. Seperti kata “dari”, “yang”, “di”, dan “ke”.
 - 3) N-gram
Pada tahap ini dilakukan pemotongan karakter dengan bi-gram yaitu pemotongan string sebanyak 2 bagian.
- c. Metode yang digunakan Pada penelitian ini menggunakan metode K-Nearest Neighbor (k-NN) dan Naive Bayes untuk mengetahui akurasi yang dihasilkan.
- d. Eksperimen dan Pengujian Metode Eksperimen pada penelitian ini menggunakan Rapid Miner 5 dalam pengolahan data

Evaluasi dan Validasi Hasil Validasi dilakukan menggunakan fold cross validation, untuk melakukan pengukuran akurasi menggunakan confusion matrix dan kurva ROC untuk mengukur nilai AU

4. KESIMPULAN

Berdasarkan Hasil pengujian akurasi menggunakan metode Naive Bayes Classifier adalah: tingkat akurasi dengan metode Naive Bayes Classifier sebesar 80.95%. memiliki hasil akurasi yang cukup baik untuk menganalisis sentimen komentar atau tweet dari guru-guru. Hasil tingkat akurasi juga menunjukkan adanya kepuasan tentang kinerja guru.

REFERENSI

- Ernawati, Siti, and Risa Wati. 2018. “Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel.” *Jurnal Khatulistiwa Informatika* 6(1): 64–69.
- Hendri, Rifwan. 2018. “No Title.” <https://travel.tempo.co/read/1149739/enam-manfaat-traveling-di-kehidupan-yang-semakin-sibuk>.

- Mentari, Nurul Dyah, M Ali Fauzi, and Lailil Muflikhah. 2018. "Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor Dan Feature Selection Query Expansion Ranking." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya* 2(8): 2739–43.
- Salam, Abu, Junta Zeniarja, and Rima Septiyan Uswatun Khasanah. 2018. "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia)." *Prosiding SINTAK*: 480–86.
- Wilianto, L., Pudjiantoro, T. H., & Umbara, F. R. 2017. Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naive Bayes Classifier Studi Kasus Jawa Barat. *Jurnal Prosiding Sntatif* Vol. 4.
- Liu, B. 2012. *Sentiment analysis and opinion mining. Synthesis lectures on human language technologies*. California: Morgan & Claypool Publishers.
- Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H. 2006. "Some effective techniques for naive bayes text classification." *IEEE transactions on knowledge and data engineering*, 18(11), 1457- 1466.