

Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata di Kabupaten Sukabumi Dengan Naïve Bayes

Boby Rizki Atmadja

Universitas Muhammadiyah Sukabumi

Jalan R. Syamsudin, S.H. No. 50, Cikole, Kec. Cikole, Kota Sukabumi, Jawa Barat 43113,

bobyatmadja1999@gmail.com

ARTICLE INFO

Article history:

Received 30 Juli 2022

Received in revised form 12 Agustus 2022

Accepted 12 September 2022

Available online 2 Desember 2022

ABSTRACT

Sentiment analysis of comments from visitors to tourist attractions and the public on tourist attractions in Sukabumi Regency which is one of the areas with various categories of tourist objects and is a sector of economic income for the surrounding community or for related parties such as the government and managers, in sentiment analysis research This includes using the Naive Bayes classification algorithm to examine the sentiment of tourist visitors and the performance of the classification model used. The data used in this research was taken from the website from Tripadvisor and Google Maps using a crawling technique, which then processed the data by a pre-processing process and then applied a classification to the data and got a sentiment visualization by processing word frequency on tourist visitor sentiment data. The results of the accuracy of the model used were re-tested with the k-fold cross validation method and the results of sentiment visualization got the frequency of words that most often appear on negative sentiment labels are garbage, beaches, lacking, places, roads, parking, dirty, entering, caring, clean , expensive, pay, manage, good and water.

Keywords: *Analysis, Sentiment, Bayes, Tourism*

1. Pendahuluan

Objek wisata merupakan salah satu sektor penunjang sumber perekonomian masyarakat dan juga sebagai salah satu sumber pendapatan penghasilan pemerintah daerah. Kabupaten Sukabumi merupakan daerah yang memiliki keunggulan dalam segi letak geografis, kekayaan alam, dan budaya yang tinggi sebagai modal daya tarik untuk pengembangan tempat objek wisata sehingga dapat mendatangkan banyak parawisatawan.

Pengertian wisatawan menurut [1] menyebutkan pada penelitiannya wisatawan merupakan pelaku (konsumen) yang melakukan kegiatan di luar aktivitas sehari hari untuk melakukan kunjungan wisata ataupun suatu perjalanan yang direncanakan untuk mendapatkan kepuasan bagi diri sendiri.

Terdapat banyak destinasi wisata di Kabupaten Sukabumi dan pada situs [2] menyebutkan 10 tempat wisata populer di Kabupaten Sukabumi adalah:

1. Jembatan Situ Gunung
2. Danau Situ Gunung
3. Geopark Ciletuh
4. Curug Cikaso
5. Curug Sawyer
6. Goa Buni Ayu
7. Pantai Karang Hawu
8. Pantai Pelabuhan Ratu
9. Pantai Penyu Pangumbahan
10. Pantai Ujung Genteng

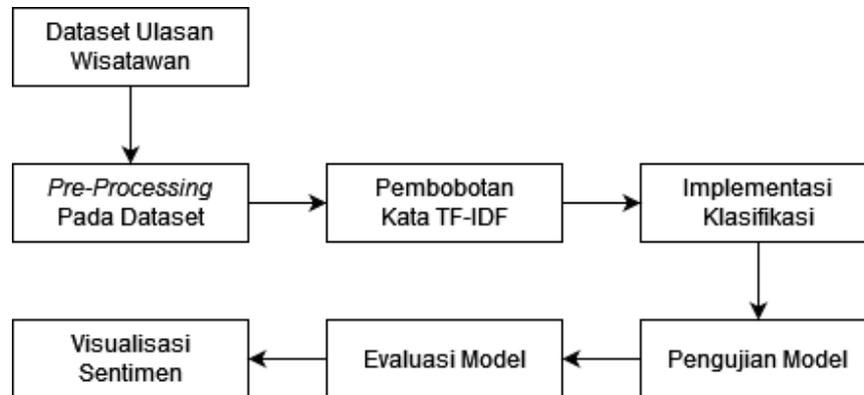
Dan dengan kemajuan teknologi saat ini, pengunjung dapat berkomentar dan mengulas tempat yang mereka kunjungi, termasuk pada objek wisata. Salah satu dari banyak situs daring yang berisikan ulasan dan komentar wisatawan adalah situs Tripadvisor dan Google Maps. Situs daring tersebut memiliki banyak informasi komentar dan ulasan wisatawan yang membantu menemukan informasi tentang tempat atau objek wisata yang dikunjungi. Informasi dari wisatawan ini sangat berguna untuk wisatawan yang akan berkunjung maupun untuk pemerintah dan pengelola tempat wisata karena dapat menjadi salah satu acuan untuk program pemerintah dalam pengembangan hal-hal terkait pada pariwisata yang dapat menjadi lebih baik lagi.

Analisis sentimen adalah proses memahami, mengekstraksi, dan memproses data tekstual otomatis untuk mendapatkan informasi sentimen yang terdapat dalam kalimat opini. Kemudian mengklasifikasikan berbagai polaritas teks dalam kalimat atau dokumen dalam dua kelas kemungkinan, baik positif maupun negatif [3]. Salah satu cara untuk mengetahui sentimen wisatawan berdasarkan ulasan dan komentar tentang tempat wisata di Kabupaten Sukabumi adalah dengan analisis sentimen yang diambil dari ulasan dan komentar dari situs daring seperti Tripadvisor dan Google Maps. Mengetahui sentimen pengunjung wisata dapat membantu untuk memverifikasi pengalaman wisatawan dalam berwisata dan juga sebagai peninjauan untuk pengelola dan pemerintah mengenai tempat wisata di Kabupaten Sukabumi.

Algoritma *Naïve Bayes* termasuk pada sepuluh klasifikasi metode *data mining* terpopuler diantara algoritma-algoritma lainnya. Algoritma *Naïve Bayes* pun tergolong memiliki nilai yang berpotensi baik dalam klasifikasi teks dokumen dalam hal akurasi maupun efisiensi komputasi diantara algoritma lain [4]. Berdasarkan hal tersebut, penulis akan menggunakan algoritma naïve bayes pada proses klasifikasi dalam penelitian analisis sentimen wisatawan mengenai tempat wisata di Kabupaten Sukabumi. Dengan landasan tersebut, penelitian ini meneliti tentang analisis sentimen pada ulasan dan komentar wisatawan berbahasa indonesia di Kabupaten Sukabumi dari situs daring Tripadvisor dan Google Maps dengan menggunakan algoritma klasifikasi naïve bayes.

2. Metode Penelitian

Tahapan-tahapan metode yang digunakan pada penelitian ini yang dilakukan dari awal hingga akhir terdapat pada gambar berikut:



Gambar 1. Tahapan Metode Penelitian

2.1 Dataset Ulasan Wisatawan

Dataset adalah suatu database didalam memori (in-memory). Dataset memiliki semua karakteristik, fitur dan fungsi dari basis data biasa. Dataset dapat memiliki banyak tabel, dan tabel-tabel dapat memiliki hubungan (*relationship*). Tabel-tabel pada suatu dataset dapat memiliki *foreign key* dan integritas referensial [5].

Dataset dalam penelitian ini diambil dengan menggunakan teknik *crawling* dengan ekstensi browser Google Chrome pada komentar dan ulasan wisatawan tempat wisata di Kabupaten Sukabumi dalam situs daring Tripadvisor dan Google Maps. Pengertian dari *crawling* adalah suatu teknik yang digunakan untuk mengumpulkan informasi yang ada dalam *web* [6]. Setelah data ulasan dan komentar wisatawan didapatkan maka data tersebut diberi label positif, negatif, atau netral berdasarkan kategori sentimennya secara manual.

2.2 Pre-Processing Pada Dataset

Text Preprocessing merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Suatu teks tidak dapat diproses langsung oleh algoritma pencarian, oleh karena itu dibutuhkan *preprocessing text* untuk mengubah teks menjadi data numerik [7].

Dalam tahap *pre-processing* data pada dataset akan diolah agar data dapat menjadi lebih baik atau maksimal. Terdapat empat tahap dalam tahap *pre-processing*, yaitu *tokenizing*, *cleaning*, *stemming*, dan *filtering*.

2.3 Pembobotan Kata TF-IDF

Pada proses pembobotan kata ini menggunakan metode proses pembobotan kata TF-IDF (*Term Frequency-Inverse Document Frequency*). *Frequency Invers Document Frequency* (TF-IDF) merupakan metode yang digunakan menentukan seberapa jauh keterhubungan kata (*term*) Terhadap dokumen dengan memberikan bobot setiap kata [8].

Perhitungan pembobotan kata dengan metode TF-IDF terdapat menggunakan persamaan menurut [9] penelitian terdapat sebagai berikut.

$$W_{j,i} = \frac{n_{j,i}}{\sum_k n_{k,i}} \cdot \log_2 \frac{D}{d_j} \dots \dots \dots (1)$$

Keterangan:

- $W_{j,i}$ = Pembobotan TF-IDF untuk term ke j pada dokumen ke i.
 $n_{j,i}$ = Jumlah kemunculan term ke j pada dokumen ke i.
 $\sum_k n_{k,i}$ = Jumlah kemunculan seluruh term pada dokumen ke i.
 D = Banyaknya dokumen yang dibangkitkan.
 d_j = Banyaknya dokumen yang mengandung term ke j.

2.4 Implementasi Klasifikasi

Pada proses implementasi klasifikasi dalam penelitian ini terdapat menggunakan algoritma klasifikasi naïve bayes. Metode Bayes adalah pendekatan secara statistic untuk menghitung diantara keputusan yang berbeda beda, dengan menggunakan probabilitas dan costs yang menyertai suatu pengambilan keputusan [10].

Pada penjelasan pada penelitian [11] , pada penelitian tersebut menjelaskan Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots(2)$$

Keterangan:

- X = Data dengan *class* yang belum diketahui.
 H = Hipotesis data X merupakan suatu class spesifik.
 $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probabilitas*).
 $P(H)$ = Probabilitas hipotesis H (*prior probabilitas*).
 $P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H.
 $P(X)$ = Probabilitas X.

Dalam klasifikasi dokumen teks, dataset dibagi dari total dataset komentar dan rating wisatawan dari survei ini menjadi 85% untuk data latih dan 15% untuk data uji. Kemudian, saat melakukan analisis sentimen, algoritma *Naive Bayes* digunakan untuk mengklasifikasikan dataset untuk mendapatkan nilai probabilitas tertinggi dari *insight* yang diperoleh dari pengujian sebelumnya, dari kelas sentimen positif, negatif, dan netral.

2.5 Pengujian Model

Kemudian setelah melakukan proses tahap implemetasi klasifikasi, maka selanjutnya perlu untuk melakukan pengujian untuk mengetahui informasi nilai performa model yang telah diterapkan.

2.6 Evaluasi Model

Pada tahap ini dilakukan untuk mengevaluasi mengenai hasil dan nilai performa model dalam penggunaan metode klasifikasi yang diterapkan, adapun evaluasi model yang digunakan yaitu dengan melakukan proses perhitungan pada nilai akurasi, presisi, *f1-score*, dan *recall* pada hasil model yang digunakan.

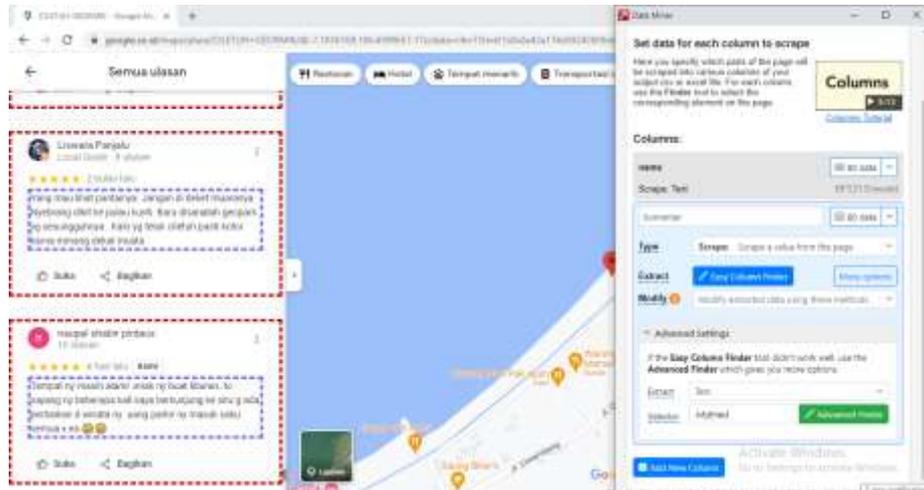
2.7 Visualisasi Sentimen

Pada tahap ini informasi data pada dataset akan diekstrak kedalam bentuk frekuensi topik kata atau data yang paling sering muncul pada komentar atau ulasan dari wisatawan, yang kemudian hasil informasi data tersebut dapat diolah dan akan mempermudah untuk menampilkan data informasi yang dibutuhkan dalam penelitian ini.

3. Hasil dan Pembahasan

3.1 Dataset Ulasan Wisatawan

Proses pengambilan data menggunakan scraper data dimulai dengan mengaktifkan ekstensi *Instant Data Scraper* di browser Google Chrome, kemudian menentukan kueri untuk mengambil data pada bagian yang ditandai saat kueri ditentukan.



Gambar 2. Proses Menentukan *Query* Saat *Crawling*.

Setelah memutuskan data mana yang akan diambil, selanjutnya untuk memulai proses *crawling* yaitu menekan tombol *run* pada *Instant Data Scraper* untuk memulai mengambil data. Ketika proses *crawling* data selesai, maka semua data komentar wisatawan akan terkumpul dan kemudian data disimpan dalam file dengan ekstensi *xlsx*.

Data ulasan dan komentar wisatawan yang disimpan dari hasil *crawling* kemudian dilakukan proses pelabelan untuk membedakan antara komentar bersentimen positif, negatif, atau netral. Pelabelan sentimen pada data ulasan dan komentar wisatawan berfungsi untuk memudahkan pemrosesan program sistem selama proses klasifikasi. Kategori kelas sentimen dengan label 0 menunjukkan label sentimen negatif, label 1 menunjukkan label sentimen positif, dan label 2 menunjukkan label sentimen netral. Setelah tahap *crawling* dan pelabelan sentimen pada dataset telah selesai dilakukan, terdapat total jumlah data ulasan dan komentar wisatawan pada dataset yaitu sebanyak 3.194, dengan komentar berlabel positif 1527, komentar berlabel negatif 1048, dan komentar berlabel netral dengan total ada 618.

3.2 Pre-Processing Pada Dataset

Dalam tahap *pre-processing*, terdapat menggunakan *library* dalam bahasa pemrograman *Python3*. Proses *pre-processing* data dalam penelitian ini dilakukan dalam empat tahap, yaitu:

a. *Tokenizing*

Dalam tahap *tokenizing*, data ulasan wisatawan yang berisi kalimat pada dataset dipisahkan menjadi kata-perkata (*token*). Proses *tokenizing* dalam penelitian ini menggunakan *library nltk*.

Tabel 1. Proses Tahap *Tokenizing*

No.	Input proses	Output Proses
1	mantap lah g bakal nyesel kalo main ke pantai pelabuanratu mah asli dan pasti mau dei ...	'mantap' 'lah' 'g' 'bakal' 'nyesel' 'kalo' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'dei ...'

2	ramai, banyak sampah. air kotor.	'ramai,' 'banyak' 'sampah.' 'air' 'kotor.'
3	pantai indah tapi sayang kurang perhatian dan pengelolaan	'pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'

b. *Cleaning*

Tahap ini membersihkan data pada dataset dalam hal-hal yang tidak dibutuhkan dalam proses penelitian. Seperti mengilangkan tanda baca, menghapus angka, ubah kata menjadi huruf kecil, dan lainnya yang disesuaikan penelitian.

Tabel 2. Proses Tahap *Cleaning*

No.	<i>Input</i> proses	<i>Output</i> Proses
1	'mantap' 'lah' 'g' 'bakal' 'nyesel' 'kalo' 'main' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'dei ...'	'mantap' 'lah' 'g' 'bakal' 'nyesel' 'kalo' 'main' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'dei'
2	'Ramai,' 'banyak' 'sampah.' 'Air' 'kotor.'	'ramai' 'banyak' 'sampah' 'air' 'kotor'
3	'Pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'	'pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'

c. *Stemming*

Dalam tahap ini, teks data pada dataset dikonversi ke dalam kata dasarnya dan juga menghapus imbuhan sebelum dan sesudah kata, lalu mengkoreksi kata-kata yang salah eja, memperbaharui kata-kata bahasa Indonesia dengan ejaan yang belum disempurnakan, atau mengubah kata yang disingkat. Dalam proses tahap *stemming ini* menggunakan *library* Sastrawi.

Tabel 3. Proses Tahap *Stemming*

No.	<i>Input</i> proses	<i>Output</i> Proses
1	'mantap' 'lah' 'g' 'bakal' 'nyesel' 'kalo' 'main' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'dei ...'	'mantap' 'lah' 'tidak' 'bakal' 'sesal' 'kalau' 'main' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'lagi'
2	'ramai,' 'banyak' 'sampah.' 'air' 'kotor.'	'ramai' 'banyak sampah' 'air' 'kotor'
3	'pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'	'pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'

d. *Filtering*

Filtering merupakan tahapan proses *pre-processing* teks data pada dataset yang bertujuan untuk menghilangkan kata-kata yang tidak gunakan atau berarti dalam penelitian. Pada penelitian dibuat daftar kata-kata yang masuk pada daftar stopwords dan tahapan ini terdapat menggunakan *library nltk*.

Tabel 4. Proses Tahap *Filtering*

No.	<i>Input</i> proses	<i>Output</i> Proses
1	'mantap' 'lah' 'tidak' 'bakal' 'sesal' 'kalau' 'main' 'ke' 'pantai' 'pelabuanratu' 'mah' 'asli' 'dan' 'pasti' 'mau' 'lagi'	'mantap' 'sesal' 'main' 'pantai' 'pelabuanratu' 'asli' 'pasti' 'mau' 'lagi'
2	'ramai,' 'banyak sampah' 'air' 'kotor'	'ramai,' 'banyak sampah.' 'air' 'kotor'.
3	'pantai' 'indah' 'tapi' 'sayang' 'kurang' 'perhatian' 'dan' 'pengelolaan'	'pantai' 'indah' 'sayang' 'kurang' 'perhatian' 'pengelolaan'

Setelah semua tahapan proses *pre-processing* dilakukan pada teks data dalam dataset, terdapat total 3,181 jumlah data pada dataset dan dataset tersebut disimpan pada file dokumen teks berekstensi (CSV) *comma separated value*, merupakan format data dalam standar file ASCII, di mana setiap *record* dipisahkan dengan tanda koma (,) atau titik koma (;) [12].

3.3 Pembobotan Kata TF-IDF

Dalam pembobotan *term*, dijalankan suatu proses untuk memberikan nilai atau bobot pada setiap dokumen kata komentar dan pada tahap ini digunakan metode TF-IDF. Untuk tujuan memberikan nilai pada *term* yang nilainya akan digunakan pada saat entri dalam proses klasifikasi. Pada proses implementasi pembobotan dengan menggunakan TF-IDF terdapat menggunakan *library* dari *scikit-learn*.

3.4 Implementasi Klasifikasi Naïve Bayes

Pada tahap implementasi klasifikasi *Naïve Bayes*, terdapat menggunakan beberapa *library* seperti *Scikit-learn*, *Comma Separated Value (CSV)*, *Pandas*, *NumPy*, *Seaborn*, *Matplotlib*, *Pickle*, dan *library* untuk memasang *drive* pada *Google Drive*. Pada tahap klasifikasi penelitian ini, data di *split* menjadi 20% untuk data *test* dan 80% untuk data *train*. Setelah sebelumnya telah dilakukan pembobotan kata dengan menggunakan TF-IDF, lalu untuk proses klasifikasi menggunakan *library Multinomial Naïve Bayes*. Proses selanjutnya melatih klasifikasi dengan data latih yang kemudian data modelnya di simpan dengan nama dan format “Model_NBC.pickle” pada *google drive* yang sudah di *mounting*. Lalu *load* kembali data model, yang dilanjut dengan dilakukan test klafikasi dengan menggunakan data *test*.

Setelah melakukan semua proses pada tahap klasifikasi dengan menghitung probabilitas antar kelas sentimen pada kalimatnya, maka selanjutnya dapat mendapatkan hasil prediksi. Maka selanjutnya baru dapat dilakukan tahap untuk menghitung performa dari algoritma klasifikasi.

3.5 Pengujian Model

Untuk menghitung dan mengetahui performa dari algoritma klasifikasi yang digunakan yaitu algoritma *Naïve Bayes*, maka perlu dilakukan pengujian model klasifikasi. Hasil performa dari algoritma ditampilkan dalam bentuk nilai *accuracy* dan bentuk tabel *confusion matrix*, salah satu metode yang dapat digunakan untuk mengukur performansi suatu metode klasifikasi. *Confusion matrix* berisi informasi yang membandingkan hasil dari klasifikasi yang dilakukan oleh sistem dengan *default* hasil klasifikasi [13]. Berikut hasil nilai *accuracy* dan tabel *confusion matrix* terdapat nilai sebagai berikut.

Accuracy: 0.7786499215070644

Confusion Matrix:

```
[[168  36   0]
 [  5 316   0]
 [ 27  73  12]]
```

3.6 Evaluasi Model

Pada tahap evaluasi model ini bertujuan untuk mengetahui nilai performa algoritma yang digunakan pada penelitian ini. Untuk mendapatkan nilai performa algoritma *precision score*, *recall score*, dan *f1-score* sebelumnya harus mendapat nilai tabel *confusion matrix* dahulu seperti yang dilakukan pada tahap sebelumnya.

Tabel 5. *Confusion Matrix*

	Predict Class		
	Positif	Negatif	Netral
True	168	36	0

Analisis Sentimen Bahasa Indonesia Pada Tempat Wisata di Kabupaten Sukabumi Dengan Naïve Bayes (Boby Rizki Atmadja)

<i>Class</i>	Negatif	5	316	0
	Netral	27	73	12

Pada penelitian ini untuk mendapatkan nilai performa model dibantu dengan menggunakan *library scikitlearn metrics*, adapun kode program implementasi untuk mendapatkan performa nilai model terdapat pada gambar berikut. Adapun hasil nilai performa dari penggunaan model sebelumnya terdapat pada tabel berikut.

Tabel 6. Nilai Performa Model Klasifikasi.

Nama Performa	Nilai Performa
Precision	0.78
Recall	0.78
F1-Score	0.78

Setelah mengetahui nilai performa model klasifikasi yang digunakan, maka dapat diketahui bahwa presentase sistem memiliki kemampuan ketepatan dalam mencari informasi yang diberikan pengguna dan kemampuan keberhasilan menemukan kembali informasi yaitu 78%.

Kemudian untuk mengetahui nilai performa model klasifikasi pada setiap kelasnya. Setelah meimplementasi kode program, maka hasil nilai performa model klasifikasi *precision score*, *recall score*, dan *f1-score* terdapat pada tabel sebagai berikut.

Tabel 7. Nilai Performa Setiap Kelas

Kelas	Precision	Recall	F1-Score
Negatif	0.84	0.82	0.83
Positif	0.74	0.98	0.85
Netral	1.00	0.11	0.19

Dengan nilai performa model pada setiap kelasnya diatas, maka hasil tahap evaluasi model pada sistem dalam mencari informasi yang diberikan oleh pengguna terdapat presentase 84% pada kelas negatif, 74% pada kelas positif, dan 100% pada kelas netral atau positif negatif. Lalu kemampuan sistem dalam menemukan kembali informasi pada pengguna terdapat presentase 82% pada kelas negatif, 98% pada kelas positif, dan 11% pada kelas netral atau positif negatif.

Lalu pada tahap evaluasi model ini, maka dilakukan pengujian selanjutnya dengan menggunakan *k-fold cross validation* yang inialisasi dengan 10 jumlah iterasi. Pada pengujian *k-fold cross validation* menggunakan *library scikit k-fold* dan *tabulate*. Berikut adalah nilai hasil performa model setelah dilakukan pengujian dengan menggunakan *k-fold cross validation* yang disajikan dalam bentuk gambar.

Iteration	Accuracy	Precision	Recall	F-Measure
Iteration-0	0.77836	0.84972	0.65900	0.65918
Iteration-1	0.72775	0.84437	0.61843	0.60563
Iteration-2	0.72426	0.84349	0.61958	0.60764
Iteration-3	0.72949	0.85178	0.61639	0.60109
Iteration-4	0.73473	0.85463	0.61690	0.60713
Iteration-5	0.73298	0.85007	0.61176	0.59384
Iteration-6	0.74346	0.85518	0.62287	0.61451
Iteration-7	0.76789	0.86605	0.63009	0.62495
Iteration-8	0.75916	0.86182	0.62596	0.62203
Iteration-9	0.78185	0.87499	0.63542	0.63139

Average Accuracy	: 0.78185 / 78.185
Average Precision	: 0.87499 / 87.49900000000001
Average Recall	: 0.63542 / 63.542
Average F-Measure	: 0.63139 / 63.139

	precision	recall	f1-score	support
0	0.91	0.79	0.85	212
1	0.71	1.00	0.83	271
2	1.00	0.12	0.22	90

	accuracy	macro avg	weighted avg
		0.67	0.63
		0.83	0.74

Gambar 3. Nilai Hasil Pengujian *K-Fold Cross Validation*

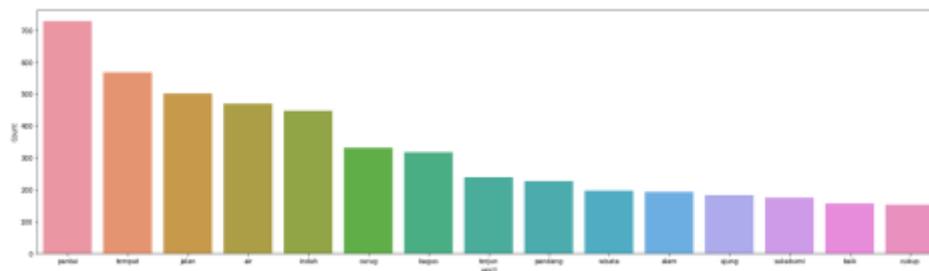
Pada hasil nilai performa model pada pengujian dengan k-fold cross validation pada sistem dalam mencari informasi yang diberikan oleh pengguna terdapat presentase 84% pada kelas negatif, 74% pada kelas positif, dan 100% pada kelas netral atau positif negatif. Lalu kemampuan sistem dalam menemukan kembali informasi pada pengguna terdapat presentase 82% pada kelas negatif, 98% pada kelas positif, dan 11% pada kelas netral atau positif negatif.

3.7 Visualisasi Sentimen

Maka untuk mempermudah melihat hasil visualisasi sentimen sesuai dengan kebutuhan berdasarkan kategori label sentimennya, maka pada tahap visualisasi sentimen ini dibagi menjadi tiga bagian yaitu terdapat sebagai berikut:

a. Visualisasi Sentimen Positif

Sebelum implementasi kode program pada tahap visualisasi sentimen positif dilakukan deklarasi *library string* dan berikut hasil visualisasi sentimen positif berdasarkan pengekstrakan informasi pada dataset ulasan wisatawan.



Gambar 4. Hasil Visualisasi Sentimen Positif.

Berdasarkan hasil visualisasi sentimen positif diatas maka frekuensi kata yang paling sering muncul dari wisatawan adalah pantai, tempat, jalan, air, indah, curug, bagus, terjun, pandang, wisata, alam, ujung, sukabumi, baik, dan cukup. Lalu selanjutnya frekuensi kata-kata

Daftar Pustaka

- [1] S. Wahyulina, S. Darwini, W. Retnowati, and S. Oktaryani, "Persepsi Wisatawan Muslim Terhadap Sarana Penunjang Wisata Halal Dikawasan Desa Sembalun Lawang Lombok Timur," *Jmm Unram - Master Manag. J.*, vol. 7, no. 1, pp. 32–42, 2018.
- [2] Tripadvisor, "Objek Wisata Terpopuler di Sukabumi," 2022. [Online]. Available: https://www.tripadvisor.co.id/Attractions-g297714-Activities-Sukabumi_West_Java_Java.html. [Accessed: 05-Feb-2022].
- [3] N. Hayatin, G. I. Marthasari, and L. Nuraini, "Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization," *J. Online Inform.*, vol. 5, no. 1, pp. 81–88, 2020.
- [4] R. A. Saputra, A. R. Taufik, L. S. Ramdhani, R. Oktapiani, and E. Marsusanti, "Sistem Pendukung Keputusan Dalam Menentukan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes," *Snit 2018*, vol. 1, no. 1, pp. 106–111, 2018.
- [5] Yahya and Mahpuz, "Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat," *Infotek J. Inform. dan Teknol.*, vol. 2, pp. 109–118, 2019.
- [6] N. A. O. Saputri and K. Zuhri, "Analisis Sentimen Masyarakat Terhadap Pilpres 2019 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier," *J. Inform.*, vol. 7, 2021.
- [7] Y. Findawati, *Buku Ajar Text Mining*. Sidoarjo, Jawa Timur: UMSIDA Press, 2020.
- [8] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, pp. 306–312, 2018.
- [9] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, "Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)," *J. Gaussian*, vol. 9, pp. 376–390, 2020.
- [10] E. Sitepu, M. Simanjuntak, and H. Khair, "Sistem Pakar Mendiagnosa Penyakit Kelainan Darah Pada Manusia Menggunakan Metode Bayes Berbasis Website," *J. Tek. Inform. Kaputama*, vol. 6, 2022.
- [11] W. Muslehatin, M. Ibnu, and Mustakim, "Penerapan Naive Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, 2017.
- [12] Eriyanto, *Analisis Jaringan Media Sosial Dasar-Dasar dan Aplikasi Metode Jaringan Sosial untuk Membedah Percakapan di Media Sosial*. 2021.
- [13] F. Rahmad, Y. Suryanto, and K. Ramli, "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 879, no. 1, pp. 0–11, 2020.