

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK ANALISIS SENTIMEN TERHADAP ALUN-ALUN SEBAGAI RUANG TERBUKA PUBLIK DI KABUPATEN SUKABUMI

Farikh Fadhil^{1*}, Asriyanik², Winda Apriandari³

^{1,2,3}Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Sukabumi
Email: ^{1*}farikhfadhil057@ummi.ac.id, ²asriyanik263@ummi.ac.id, ³winda.apriandari@ummi.ac.id

(*: *Corresponding Author*)

(Naskah masuk: 21 Juni 2023, diterima untuk diterbitkan: 4 Juli 2023)

Abstrak

Kabupaten Sukabumi merupakan salah satu Kabupaten di Jawa Barat yang memiliki beberapa fasilitas dan prasarana, salah satunya ruang terbuka publik yang menjadi sebuah tempat yang bermanfaat untuk aktivitas bagi kebutuhan dan kepentingan masyarakat umum sehari-hari seperti contohnya alun-alun yaitu Alun-alun Cisaat, Alun-Alun Palabuhanratu, Alun-Alun Jampang Kulon, Alun-Alun Purabaya dan Alun-Alun Cicurug. Namun di Kabupaten Sukabumi juga masih terdapat alun-alun yang kurang terawat sehingga terdapat artikel atau opini dari pengunjung yang kurang baik mengenai alun-alun tersebut. Oleh karena itu perlu dilakukan pengembangan atau evaluasi oleh pemerintah setempat untuk menjadikan ruang terbuka publik di Kabupaten Sukabumi lebih baik lagi. Sebelum melakukan pengembangan dan evaluasi diperlukan terlebih dahulu informasi data atau analisis terkait opini publik terhadap ruang terbuka publik tersebut salah satunya dengan melakukan analisis sentimen. Penelitian ini melakukan analisis sentimen publik terhadap alun-alun sebagai ruang terbuka publik di Kabupaten Sukabumi menggunakan algoritma *Naïve Bayes* dalam melakukan analisis sentimen dan data yang diambil dari ulasan pengunjung tempat tersebut pada periode 2019-2023 di *website Google Maps* dan didapatkan data sentimen sebanyak 2698 yang terdiri dari data sentimen positif sebanyak 2254 dan data sentimen negatif 444 data dengan hasil akurasi dari algoritma yang digunakan pada model yang dibuat yang memiliki ketepatan akurasi sebesar 92%, nilai *precision* sebesar 90%, nilai *recall* sebesar 53% dan *f1-score* sebesar 67%. Frekuensi kata yang diperoleh dari hasil analisis sentimen mendapati 5 kata yang paling banyak muncul berdasarkan kelas sentimennya yaitu sentimen positif adalah mantap, bagus, bersih, nongkrong dan enak. Sedangkan 5 Frekuensi kata negatif adalah adalah sampah, kotor, kumuh, macet dan tata. Hasil dari penelitian ini terkait analisis sentimen diharapkan dapat membantu pemerintah setempat terutama Dinas Pertanahan dan Tata Ruang Kabupaten Sukabumi untuk dijadikan informasi sebagai bahan acuan atau rekomendasi untuk pengembangan dan evaluasi objek ruang terbuka publik di Kabupaten Sukabumi kedepannya.

Kata kunci: *analisis sentimen, ruang terbuka publik, alun-alun, naïve bayes*

IMPLEMENTATION OF THE NAIVE BAYES ALGORITHM FOR ANALYSIS SENTIMENT OF ALUN-ALUN AS A PUBLIC OPEN SPACE IN SUKABUMI REGENCY

Abstract

Sukabumi Regency is one of the regencies in West Java that has various facilities and infrastructure, including public open spaces, which serve as beneficial places for the daily activities and needs of the general public. One of these spaces is the town squares, namely Alun-Alun Cisaat, Alun-Alun Palabuhanratu, Alun-Alun Jampang Kulon, Alun-Alun Purabaya, and Alun-Alun Cicurug. However, there are still town squares in Sukabumi Regency that are poorly maintained, resulting in negative articles or opinions from visitors regarding these town squares. Therefore, it is necessary for the local government to undertake development or evaluation to further improve the public open spaces in Sukabumi Regency. Before carrying out the development and evaluation, it is essential to gather information, data, or conduct analysis related to public opinions on these public open spaces. One method is by performing sentiment analysis. This research conducts a public sentiment analysis on the town squares as public open spaces in Sukabumi Regency, utilizing the *Naïve Bayes* algorithm for sentiment analysis. The data used for analysis were obtained from visitor reviews of these places from the period of 2019 to 2023 on the *Google Maps website*. A total of 2698 sentiment data points were collected, consisting of 2254 positive sentiment data and

444 negative sentiment data. The algorithm used in the created model achieved an accuracy rate of 92%, precision value of 90%, recall value of 53%, and an F1-score of 67%. The frequency of words obtained from the sentiment analysis revealed the top 5 most frequently mentioned words based on their sentiment class. The top 5 positive sentiment words are excellent, good, clean, hang out, and pleasant. On the other hand, the top 5 negative sentiment words are garbage, dirty, slum, traffic jam, and disorganized. The results of this research, specifically related to sentiment analysis, are expected to assist the local government, especially Regional Office of Land and Spatial Planning in Sukabumi Regency, by providing valuable information as a reference or recommendation for the future development and evaluation of public open spaces in Sukabumi Regency..

Keywords: sentiment analysis, public open space, alun-alun, naïve bayes.

1. PENDAHULUAN

Ruang terbuka publik merupakan sebuah tempat atau area yang sering digunakan oleh masyarakat untuk melakukan sebuah aktivitas sebagaimana keperluan atau kebutuhannya. Menurut [1] Ruang terbuka publik merupakan sebuah ruang yang berfungsi sebagai tempat untuk melakukan kegiatan sosial dalam kehidupan masyarakat umum yang menjadikan tempat bertemunya individu ataupun kelompok masyarakat dalam melakukan kegiatan maupun rutinitas dalam sehari-hari ataupun berulang dan salah satunya adalah alun-alun yang merupakan salah satu tempat atau area ruang terbuka publik yang digunakan untuk masyarakat.

Kabupaten Sukabumi merupakan salah satu wilayah di Jawa Barat yang mempunyai beberapa tempat ruang terbuka publik salah satunya alun-alun. Terdapat alun-alun yang berada di Kabupaten Sukabumi yaitu Alun-Alun Cisaat Alun-Alun Palabuhanratu, Alun-Alun Purabaya, Alun-Alun Cicurug dan Alun-Alun Jampang Kulon yang semuanya berada di Kabupaten Sukabumi. Namun terkait alun-alun di Kabupaten Sukabumi masih terdapat opini atau berita negatif dari publik mengenai alun-alun tersebut seperti dalam artikel berita online dalam *website* [2] yang menyebutkan terkait kebersihan karena masih banyaknya sampah di Alun-Alun Cisaat Kabupaten Sukabumi dan terdapat lagi artikel berita tentang kurangnya kebersihan di alun-alun Kabupaten Sukabumi yaitu Alun-Alun Palabuhanratu yang disebutkan dalam artikel berita online pada *website* [3]. Selain kebersihan terdapat juga artikel berita tentang rusaknya fasilitas di Alun-Alun Palabuhanratu di Kabupaten Sukabumi seperti yang disebutkan dalam artikel berita *online website* berikut [4]. Terdapatnya alun-alun yang kurang terawat menjadikan perlunya pemantauan dari pemerintah setempat agar sarana dan prasarana yang terdapat di alun-alun bisa dievaluasi dan diperbaiki. Oleh karena itu peneliti melakukan wawancara dan diskusi bersama pihak Dinas Pertanahan dan Tata Ruang Kabupaten Sukabumi sebagai pihak yang bertanggungjawab dalam mengelola tempat alun-alun tersebut. Dari diskusi tersebut peneliti mendapatkan hasil bahwa diperlukan penelitian dan analisis lebih lanjut karena belum pernah ada peneliti lain yang melakukan analisis terkait permasalahan ini. Dalam proses pelaksanaan analisis dan penelitian yang akan dilakukan untuk pengembangan atau evaluasi

mengenai tempat alun-alun diperlukan beberapa data dan informasi mengenai apa saja bagian yang perlu diperbaiki atau dipertahankan terhadap sarana dan prasarana yang telah disediakan di tempat alun-alun agar pengembangan dan evaluasi yang dilakukan bisa dilaksanakan secara tepat.

Salah satu cara untuk melakukan analisis terkait tempat tersebut adalah dengan mengetahui informasi opini publik atau masyarakat yang telah berkunjung ke tempat tersebut salah satunya dengan melihat ulasan atau komentar dari pengunjung yang telah mengunjungi tempat tersebut seperti dari situs internet Google Maps karena di dalamnya terdapat ulasan atau komentar pengunjung mengenai tempat yang telah dikunjungi. Setelah mengetahui tempat yang akan diambilnya dataset maka perlu dilakukannya pengumpulan informasi atau analisis dan salah satunya dengan melakukan analisis sentimen menurut [5] analisis sentimen merupakan sebuah proses atau teknik penambahan pada sebuah dokumen teks dengan mengekstrak atau menganalisis yang hasilnya dapat berguna untuk pihak yang membutuhkan hasil dari analisis.

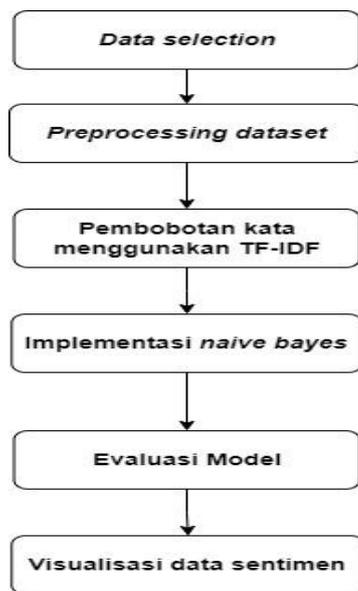
Pada proses analisis sentimen terdapat cara untuk melakukan prosesnya yaitu dengan menggunakan algoritma *machine learning* seperti *Support vector machine* yang telah banyak digunakan untuk analisis sentimen seperti penelitian sebelumnya seperti [6] dalam penelitiannya tersebut menggunakan *Support vector machine* menghasilkan akurasi sebesar 62% dan untuk penelitian sebelumnya yang menggunakan algoritma *K-Nearest neighbor* [7] menghasilkan akurasi dari algoritma tersebut yaitu sebesar 85,14 % dan untuk penelitian sebelumnya yang menggunakan algoritma *Naïve bayes* [8] menghasilkan akurasi sebesar 87%. Proses analisis sentimen dilakukan dengan menggunakan algoritma *machine learning* tersebut bertujuan agar proses analisis sentimen dengan skala data yang besar dapat dilakukan dengan cepat dan akurat dan juga menghasilkan sebuah model yang nantinya dapat digunakan ketika terdapat data baru dengan skala yang besar.

Pada penelitian ini proses analisis sentimen menggunakan algoritma *Naïve bayes* karena menurut dari penelitian terdahulu [8] algoritma ini memiliki akurasi yang bagus dan menurut [9] Algoritma *Naïve bayes* bisa melakukan komputasi dalam melakukan klasifikasi dokumen teks dengan sangat cepat dalam melakukan prosesnya, sehingga sangat cocok untuk

diterapkan dalam penelitian ini karena dapat menghasilkan model dan akurasi yang baik. Hasil dari penelitian ini menghasilkan sebuah visualisasi data mengenai analisis sentimen pada alun-alun di Kabupaten Sukabumi sehingga hasil akhirnya dapat menjadi sebuah informasi dan bahan acuan untuk pemerintah setempat dalam melakukan peningkatan serta pengembangan pada alun-alun sebagai ruang terbuka publik tersebut sehingga kedepannya akan lebih baik lagi.

2. METODE PENELITIAN

Dalam penelitian ini terdapat beberapa tahapan metode yang dilakukan dalam melakukan prosesnya yaitu seperti pada **Gambar 1**.



Gambar 1. Tahapan metode penelitian

2.1 Data Selection

Data selection adalah tahapan pertama dari penelitian ini yang akan memilih dan mengambil dari ulasan komentar pengunjung yang datang ke tempat alun-alun yang berada di Kabupaten Sukabumi yang dikumpulkan menggunakan *web scrapping* dengan ekstensi *browser* Google Chrome menggunakan *software Data miner* dan data yang diambil adalah data yang sesuai dengan kebutuhan untuk selanjutnya dilakukan proses *labelling* data secara manual untuk menentukan data yang termasuk negatif dan positif setelah itu data akan diproses lagi sesuai dengan tahapan analisis sentimen.

2.2 Preprocessing dataset

Tahapan *preprocessing dataset* adalah tahapan mengolah dataset yang asalnya merupakan data mentah atau data kotor yang akan diproses menjadi data yang sesuai dengan kebutuhan untuk analisis berikutnya. *Text processing* adalah sebuah proses yang berfungsi mengubah suatu data yang belum terstruktur menjadi terstruktur [10] Karena pada *text preprocessing* ini terdapat banyak tahapan yang akan dilakukan untuk mengolah data tersebut yaitu

cleaning, case folding, tokenizing, filtering dan *stemming* tujuan dari tahapan ini juga akan menghasilkan data yang berkualitas serta relevan untuk melakukan proses analisis sentimen yang diinginkan.

2.3 Pembobotan kata menggunakan TF-IDF

Setelah data di proses pada tahapan *preprocessing* data tersebut masih dalam berbentuk huruf oleh karena itu perlu diubah kedalam bentuk angka atau numerik dan selanjutnya dilakukan pembobotan pada kata yang ada di dalam tersebut untuk memberikan nilai pada *frequency* sebuah kata yang ada muncul di dalam dokumen teks tersebut dan pada proses pembobotan ini dilakukan menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) dalam prosesnya.

Menurut [11] metode pembobotan kata menggunakan TF-IDF ini adalah suatu cara memberikan bobot hubungan pada suatu kata (*term*) terhadap dokumen dengan melakukan perhitungan bobot dengan dua konsep yaitu *Term Frequency* dan *Inverse Document Frequency* dan dapat dilakukan dengan perhitungan berikut.

1. *Term Frequency* didapatkan dari nilai *Frequency* kemunculan kata atau fitur t pada sebuah d atau dokumen.

$$TF_t = (t, d) \quad (1)$$

2. *Inverse Document Frequency* diperoleh dari perhitungan logaritma d atau dokumen dibagi dengan df atau *document frequency*.

$$IDF_t = \log \frac{n}{df(t)} \quad (2)$$

3. *Term Frequency* (TF) dikalikan dengan *Inverse Document Frequency* (IDF) untuk mendapatkan hasil *Term Frequency Inverse Document Frequency* atau yang disimbolkan sebagai W_t .

$$W_t = TF_t \times IDF_t \quad (3)$$

2.4 Implementasi Naïve Bayes

Pada penelitian ini algoritma *naïve bayes* akan digunakan untuk melakukan proses klasifikasi pada dataset dan untuk datasetnya nanti akan dibagi menjadi sebuah data *training* dan data *testing*. Rumus dari *Naïve Bayes* adalah salah satu algoritma dalam melakukan klasifikasi dengan melakukan sebuah prediksi peluang masa depan berdasarkan data atau pengalaman masa lalu dalam proses klasifikasi *naïve bayes* diasumsikan bahwa ada atau tidak adanya ciri tertentu pada sebuah kelas tidak berhubungan dengan ciri pada kelas lain [10] dan pada penelitian ini akan digunakan untuk menentukan sebuah kelas menjadi kelas positif dan kelas negatif. Menurut [12] algoritma *naïve bayes* dapat dinotasikan seperti berikut 4.

$$P(c|x) = \frac{P(c).P(C)}{P(X)} \quad (4)$$

Keterangan:

P(c|x) : Adalah nilai dari hasil probabilitas *posterior*.

P(x|c) : Adalah nilai probabilitas *conditional*.

P(c) : Adalah nilai probabilitas *prior*.

P(x) : Adalah kemungkinan c.

Untuk menghitung probabilitas *prior* menurut [12] bisa menggunakan rumus berikut 5.

$$P(V_j) = \frac{D_j}{D} \quad (5)$$

Keterangan :

P(V_j) : Adalah probabilitas *prior* pada sebuah kelas.

D_j : Adalah dokumen pada suatu kelas j.

D : Adalah jumlah atau total dokumen.

Untuk menghitung probabilitas *conditional* Menurut [12] dapat dihitung menggunakan rumus 6.

$$P(X_i|V_j) = \frac{W_{i,j}+1}{N+N_j} \quad (6)$$

P(X_i|V_j) : Adalah probabilitas *conditional* pada kata ke-n yang dengan diketahui kelas j.

W_{i,j} : Hasil dari nilai pembobotan kata *Term frequency inverse document frequency* atau W dari sebuah *term* I didalam kelas j.

N : Adalah jumlah dari kata unik dalam semua dokumen yang tersedia.

N_j : Merupakan jumlah total N dari seluruh *term* yang berada didalam kelas j.

Untuk menghitung hasil prediksi menurut [12] dapat dinotasikan sebagai 7.

$$Vmap = \prod_{i=1}^n P(X_i|V_j) P(V_j) \quad (7)$$

Dalam prosesnya *dataset* yang diambil dan dikumpulkan akan dibagi menjadi dua yaitu data latih dan data *testing*. Data latih untuk digunakan dalam melatih algoritma klasifikasi dan sebelumnya data tersebut harus sudah dilabeli dan harus sudah melalui proses dari *text preprocessing*. Data *testing* akan digunakan untuk menguji algoritma maupun model yang sudah dilatih sehingga akan memberikan hasil baik pada proses klasifikasinya terhadap data yang termasuk data positif atau data negatif.

2.5 Evaluasi Model

Pada proses evaluasi model ini dilakukan setelah melakukan proses pengujian model yaitu meninjau nilai hasil akurasi pada model algoritma klasifikasi yang diterapkan dengan menggunakan *confusion matrix* untuk mengukur proses menggunakan perhitungan *precision*, *recall*, *accuracy* dan *f1-score*. Proses evaluasi dilakukan untuk mengetahui akurasi, kinerja dari algoritma yang digunakan [13] Tabel

untuk perhitungan *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. *Confusion Matrix*

		Class prediction	
		True	False
True	True	TP	TF
	False	FP	TN

Menurut [13] rumus untuk menghitung *precision*, *recall* dan *accuracy* adalah seperti berikut.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

2.6 Visualisasi data sentimen

Tahapan terakhir pada analisis sentimen yaitu visualisasi data sentimen yang bertujuan agar hasil dari data analisis sentimen dapat dengan mudah dibaca dan dipahami oleh pihak terkait dengan menampilkan diagram maupun grafik yang terdiri dari *barchart*, *pie chart* dan *wordcloud* sehingga data berbentuk visual terkait jumlah data, frekuensi kata, besaran data lebih mudah dipahami.

3. HASIL DAN PEMBAHASAN

3.1 Data Selection

Pengumpulan dataset merupakan tahapan pertama yang dilakukan pada penelitian ini dengan mengambil dataset dari ulasan atau komentar pengunjung terhadap alun-alun yang ada di Kabupaten Sukabumi yang terdiri dari Alun-Alun Cisaat, Alun-Alun Palabuhanratu, Alun-Alun Cicurug, Alun-Alun Jampang Kulon dan Alun-Alun Purabaya yang diambil dari situs *website* Google Maps dengan menggunakan teknik *web scrapping* dengan menggunakan *software* yang di ekstensi pada sebuah *browser* Google Chrome yang bernama Data Miner.

Tabel 2. Data *scrapping* yang sudah melalui proses *labelling*

No	Ulasan	Label
D1	Nyaman,adem dan sejuk.... 🍃 🍃 🍃 ...	Positif
D2	Tempatnya bersih dan rapih	Positif
D3	Kumuh, kotor tidak keurus	Negatif
D4	Kotor banyak sampahnya	Negatif

Setelah dilakukan proses *scrapping* selanjutnya akan dilakukan proses *labelling* manual dengan memberikan label yang akan membedakan antara ulasan positif dan negatif agar program dapat melakukan klasifikasi dengan memberikan label negatif untuk ulasan yang negatif dan memberikan label positif untuk data yang positif secara manual dan Total data yang diperoleh dari hasil *scrapping*

pada Google Maps adalah 2698 data yang terdiri dari 2254 data positif dan negatif sebesar 444 yang diambil pada periode 4 tahun kebelakang yaitu dari tahun 2019 sampai 2023 agar data yang diambil lebih relevan. Beberapa dataset hasil *scrapping* yang telah melalui proses *labelling* seperti pada Tabel 2.

3.2 Preprocessing Dataset

Tahap *Preprocessing* ini dilakukan pada dataset karena data tersebut masih banyak data yang tidak diperlukan atau tidak sesuai yang harus dibenarkan dan harus dihapus dengan tujuan agar dapat mempermudah dalam proses berikutnya dan dalam melakukan prosesnya menggunakan bantuan *library* dalam bahasa pemrograman seperti *Sklearn*, *Sastrawi* dan *NLTK*. *Text preprocessing* dalam penelitian ini terdiri dari beberapa tahapan yaitu *Cleaning*, *Case folding*, *Tokenizing*, *Filtering* dan *Stemming*.

3.2.1 Cleaning

Pada tahapan pertama dilakukan proses *cleaning* pada *dataset* untuk menghilangkan simbol, *emoticon* atau data *noise* pada data ulasan pengunjung. Pada tahap *cleaning*, implementasi kode program pada bahasa *python* menggunakan bantuan *regex* atau *regular expression* proses hasilnya bisa dilihat pada Tabel 3.

Tabel 3. Hasil *Cleaning*

No	Input process	Output process
D1	Nyaman,adem dan sejuk.... 🤔 🤔 🤔 ...	Nyaman adem dan sejuk
D2	Tempatnya bersih dan rapih	Tempatnya bersih dan rapih
D3	Kumuh, kotor tidak keurus	Kumuh kotor tidak keurus
D4	Kotor banyak sampahnya	Kotor banyak sampahnya

3.2.2 Case Folding

Tahapan *case folding* ini digunakan untuk merubah atau mengganti seluruh kata dalam teks menjadi huruf kecil karena dalam dataset data ulasan tersebut terdapat huruf kapital dan kecil oleh karena itu diperlukan tahapan *case folding* agar semua kata sama menggunakan huruf kecil. Pada tahap implementasi *case folding* menggunakan kelas *lower* untuk mengubah huruf kapital menjadi kecil. Hasilnya dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Case Folding*

No	Input process	Output process
D1	Nyaman adem dan sejuk	nyaman adem dan sejuk
D2	Tempatnya bersih dan rapih	tempatnya bersih dan rapih
D3	Kumuh, kotor tidak keurus	kumuh kotor tidak keurus
D4	Kotor banyak sampahnya	kotor banyak sampahnya

3.2.3 Tokenizing

Pada tahapan ini *tokenizing* berfungsi untuk melakukan sebuah proses pemecah pada kalimat yang berdasarkan dari spasi menjadi sebuah kata per kata menggunakan *library python* dalam prosesnya yaitu *library NLTK* dan hasilnya seperti pada Tabel 5.

Tabel 5. Hasil *Tokenizing*

No	Input process	Output process
D1	nyaman adem dan sejuk	['nyaman','adem','dan','sejuk']
D2	tempatnya bersih dan rapih	['tempatnya','bersih','dan','rapih']
D3	kumuh kotor tidak keurus	['kumuh','kotor','tidak','keurus']
D4	kotor banyak sampahnya	['kotor','banyak','sampahnya']

3.2.4 Filtering

Proses *filtering* atau *Stopwords* pada tahapan *preprocessing* ini adalah untuk menghapus kata kata yang tidak relevan atau kata kata yang tidak akan digunakan untuk proses selanjutnya pada *dataset* tersebut pada prosesnya menggunakan *library python* dalam prosesnya yaitu *library NLTK*. Hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Filtering*

No	Input process	Output process
D1	['nyaman','adem','dan','sejuk']	['nyaman','adem','sejuk']
D2	['tempatnya','bersih','dan','rapih']	['bersih','rapih']
D3	['kumuh','kotor','tidak','keurus']	['kumuh','kotor']
D4	['kotor','banyak','smpahnya']	['kotor','sampahnya']

3.2.5 Stemming

Stemming pada tahapan *preprocessing* ini dilakukan untuk mengubah kata kata menjadi kata dasarnya dengan menghilangkan imbuhan sesuai dengan kata baku standar kamus pada proses ini menggunakan *library sastrawi* dalam bahasa pemrograman *python*. Hasilnya seperti pada Tabel 7.

Tabel 7. Hasil *Stemming*

No	Input process	Output process
D1	['nyaman','adem','sejuk']	['nyaman','adem','sejuk']
D2	['bersih','rapih']	['bersih','rapih']
D3	['kumuh','kotor']	['kumuh','kotor']
D4	['kotor','sampahnya']	['kotor','sambah']

3.3 Pembobotan kata menggunakan TF-IDF

Pada penelitian ini metode TF-IDF atau pembobotan *term* digunakan untuk proses memberikan nilai atau bobot pada setiap kata dalam dokumen ulasan pengunjung agar hasilnya dapat digunakan pada saat melakukan proses klasifikasi. Pada tahapan pertama dilakukan penghitungan *Term Frequency* (TF) kata pada suatu dokumen lalu

menghitung *inverse document frequency* menggunakan perhitungan (1) dan (2). Hasil dari perhitungan *Term Frequency* dan *Inverse Document Frequency* bisa dilihat Pada Tabel 8 berikut.

Tabel 8 Hasil dari menghitung TF dan menghitung IDF

Term	DF	Idf
nyaman	1	0,602
adem	1	0,602
sejuk	1	0,602
bersih	1	0,602
rapih	1	0,602
kumuh	1	0,602
kotor	2	0,301
sampah	1	0,602

Setelah ditemukan hasil dari *Term Frequency* dan *Inverse Document Frequency* yang bisa dilihat pada Tabel 8 maka selanjutnya melakukan proses perhitungan dari *Term Frequency* dikalikan dengan hasil *Inverse Document Frequency* menggunakan perhitungan (3) dan untuk hasilnya bisa dilihat pada Tabel 9.

Tabel 9. Hasil perkalian TF dengan IDF

Term	Tf-Idf (D1)	Tf-Idf (D2)	Tf-Idf (D3)	Tf-Idf (D4)
nyaman	0,602	0	0	0
adem	0,602	0	0	0
sejuk	0,602	0	0	0
bersih	0	0,602	0	0
rapih	0	0,602	0	0
kumuh	0	0	0,602	0
kotor	0	0	0,301	0,301
sampah	0	0	0	0,602

3.4 Implementasi Naïve Bayes

Proses Implementasi klasifikasi menggunakan *naïve bayes* dilakukan dengan menggunakan data yang telah melalui *text preprocessing* dan pembobotan menggunakan TF-IDF. Pertama tama dengan melakukan perhitungan *probability prior* menggunakan rumus (5) dan hasilnya bisa dilihat pada Tabel 10.

Tabel 10. Hasil probabilitas *prior*

Sentimen	$P(V_j) = D_j/D$
Positif	0,5
Negatif	0,5

kemudian melakukan perhitungan *probability conditional* untuk menentukan kelas negatif dan kelas positifnya yang dihitung menggunakan *naïve bayes* dengan rumus (6) dan untuk hasilnya bisa dilihat pada Tabel 11.

Tabel 11. Hasil probabilitas *conditional*

Probabilitas Term	Positif	Negatif
nyaman	0.1232	0.0769
adem	0.1232	0.0769
sejuk	0.1232	0.0769
bersih	0.1232	0.0769
rapih	0.1232	0.0769

kumuh	0.1232	0.0769
kotor	0.0769	0.1183
sampah	0.1232	0.0769

Dalam menentukan hasil prediksi menggunakan *algoritma naïve bayes* berdasarkan hasil probabilitas *conditional* Tabel 11 maka untuk contohnya kita akan menentukan pengujian data baru seperti pada Tabel 12.

Tabel 12. Menentukan label sentimen

Data	Label sentimen
Tamannya nyaman dan bersih	??

Untuk menemukan hasil prediksi dapat menggunakan rumus (7) dan hasilnya bisa dilihat pada Tabel 13.

Tabel 13. Hasil pengujian atau prediksi

Ulasan	Uji Positif	Uji Negatif	Hasil prediksi
Tamannya nyaman dan bersih	0.0076	0.0029	Positif

Karena hasil atau nilai probabilitas tertinggi terdapat pada uji positif maka dapat diputuskan bahwa Ulasan tersebut termasuk ke dalam label sentimen positif.

Dalam pembentukan model menggunakan *algoritma naïve bayes* tahapan pertama yang dilakukan yaitu dengan membuat dulu model *naïve bayes* menggunakan *library sklearn* dalam bahasa pemrograman python yaitu *multinomial naïve bayes* lalu dibentuklah model klasifikasinya menggunakan fungsi *fit* untuk melatih model menggunakan data *training* sebanyak 80% dan untuk dataset yang telah dibagi sebelumnya sebesar 20% akan digunakan untuk melakukan proses pengujian data dengan menggunakan model *naïve bayes* yang telah dibuat.

3.5 Evaluasi Model

Evaluasi model program yang digunakan pada penelitian ini menggunakan *confusion matrix* dalam proses. Pada proses evaluasi ini bertujuan untuk mengetahui nilai dari performa pada algoritma yang digunakan dengan menghitung untuk mendapatkan nilai *accuracy*, *recall*, *precision* dan *f1-score*. Implementasi tahap evaluasi data uji untuk mengukur nilai performa dari model algoritma *naïve bayes* yang telah dibuat diukur performanya menggunakan *library sklearn metrics* dan *seaborn*. Data hasil untuk proses klasifikasi sentimen menggunakan *confusion matrix* bisa dilihat pada Tabel 14.

Tabel 14. *Confusion Matrix Sentimen*

	Positif	Negatif
Positif	40	4
Negatif	35	427

Setelah data *confusion matrix* didapatkan maka langsung dihitung *precision*, *recall*, *accuracy* dan *f1-score* dari data tabel 14 menggunakan perhitungan rumus (8), (9) dan (10), setelah dilakukan proses perhitungannya menghasilkan sebuah nilai performa kinerja dari model sentimen yang digunakan dan untuk hasil pengukuran modelnya bisa dilihat tabel 15.

Tabel 15. Nilai performa model sentimen

Nama performa	Nilai
Presisi	0.90
Recall	0.53
F1-score	0.67
Akurasi	0.92

Setelah mengetahui nilai performa model klasifikasi yang digunakan, maka dapat diketahui bahwa persentase nilai dari performa akurasi bahwa model yang dibuat memiliki kemampuan ketepatan akurasi sebesar 92%, nilai *precision* sebesar 90%, nilai *recall* sebesar 53% dan *fi-score* sebesar 67%.

3.6 Visualisasi Sentimen

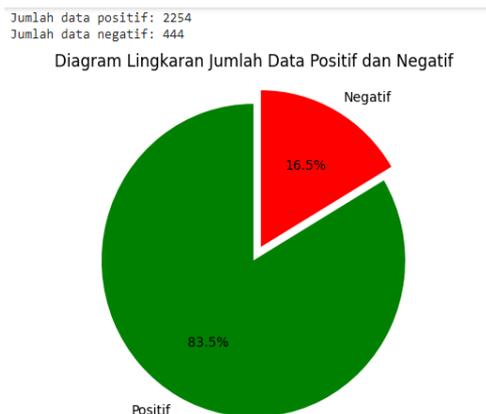
Visualisasi sentimen pada penelitian ini digunakan untuk mempermudah melihat hasil dari visual data sentimen mulai dari total jumlah data sentimen yang diperoleh dengan menampilkan grafiknya sesuai dengan jumlahnya dan juga menampilkan data yang diperoleh sesuai kategorinya yaitu sentimen negatif dan positif sehingga lebih mudah dipahami untuk melihat grafik data yang telah dibuat. Visualisasi sentimen juga menampilkan dari sebuah grafik frekuensi kata yang sering muncul dari label sentimen positif maupun label sentimen negatif dengan menampilkan kata kuncinya lebih besar sesuai dengan banyaknya kata. Selain itu visualisasi sentimen juga menampilkan kata kunci yang sering muncul dan paling banyak menggunakan yang prosesnya dilakukan menggunakan atau memanfaatkan *library* dalam bahasa pemrograman *python* seperti *library wordcloud* dan *matplotlib*.

3.6.1 Visualisasi data total sentimen

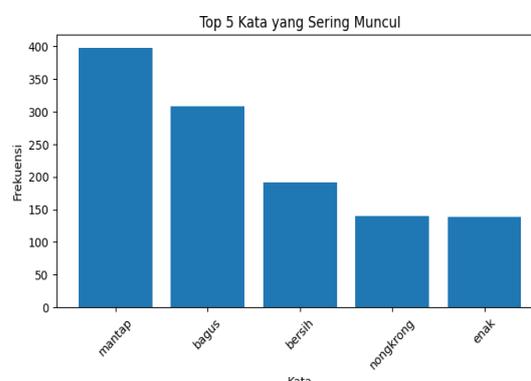
Visualisasi sentimen positif dan negatif yang diperoleh menghasilkan data positif sebesar 83,5% dengan total jumlah data sebanyak 2254 dan data sentimen negatif sebesar 16,5 % dengan total jumlah data sebanyak 444 data yang ditotalkan dari keduanya berjumlah 2698 data dan untuk grafik lingkaran ini dibuat dengan bantuan *library matplotlib pyplot* dalam bahasa pemrograman *python* dalam proses pembuatannya untuk hasilnya bisa dilihat pada **Gambar 2**.

3.6.2 Visualisasi sentimen positif

Visualisasi dari sentimen positif menggunakan *library wordcloud* dan visualisasi grafik frekuensi kata yang sering muncul menggunakan *library matplotlib* hasilnya dapat dilihat pada **Gambar 3**.



Gambar 2. Total data sentimen



Gambar 3. Visualisasi barchart sentimen positif

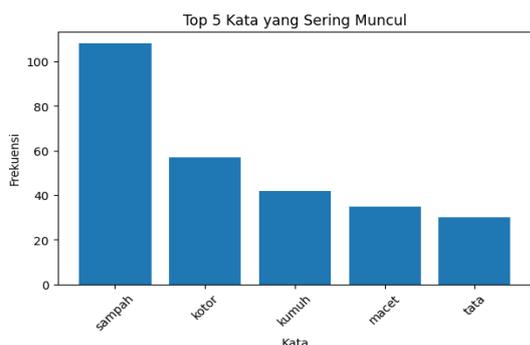
Berdasarkan hasil visualisasi sentimen positif menggunakan *barchart* pada **Gambar 3** untuk 5 kata paling banyak atau paling sering muncul dalam sentimen positif adalah mantap, bagus, bersih, nongkrong dan enak. Terdapat juga visualisasi sentimen positif menggunakan *library wordcloud* yang menunjukkan kata yang paling besar ukurannya merupakan kata yang sering muncul sedangkan kata yang ukurannya paling kecil merupakan kata yang paling jarang muncul yang bisa dilihat pada **Gambar 4**.



Gambar 4. Visualisasi sentiment wordcloud positif

3.6.3 Visualisasi sentimen negatif

Visualisasi dari sentimen negatif menggunakan *library wordcloud* dan visualisasi grafik frekuensi kata yang sering muncul menggunakan *library matplotlib* hasilnya dapat dilihat pada **Gambar 5**.



Gambar 5. Visualisasi barchart sentimen negatif

Berdasarkan hasil visualisasi sentimen negatif menggunakan *barchart* pada **Gambar 5** untuk 5 kata paling banyak atau paling sering muncul dalam sentimen negatif adalah sampah, kotor, kumuh, macet dan tata. Terdapat juga visualisasi sentimen positif menggunakan *library wordcloud* yang menunjukkan kata yang paling besar ukurannya merupakan kata yang sering muncul sedangkan kata yang ukurannya paling kecil merupakan kata yang paling jarang muncul yang bisa dilihat pada **Gambar 6**.



Gambar 6. Visualisasi sentiment wordcloud negatif

4. KESIMPULAN

Analisis sentimen terhadap alun-alun di Kabupaten Sukabumi mendapatkan hasil 2698 data yang diambil dari periode waktu 2019 sampai 2023 dengan total jumlah sentimen positif sebesar 83,5% yaitu 2254 data dan sentimen negatif sebesar 165 % dengan total 444 data. Frekuensi kata yang diperoleh dari hasil analisis sentimen mendapati 5 kata yang paling banyak muncul berdasarkan kelas sentimennya yaitu sentimen positif adalah mantap, bagus, bersih, nongkrong dan enak. dan sentimen negatif adalah sampah, kotor, kumuh, macet dan tata. Hasil tersebut menandakan untuk sentimen positif lebih banyak dibanding sentimen negatif yang menandakan lebih banyak pengunjung yang sudah puas karena lebih dominan memberi ulasan baik dibanding ulasan tidak baik terutama sentimen positif mengenai kondisi tempat tersebut dinilai bagus dan bersih namun perlu diperhatikan juga untuk sentimen negatif terkait sampah yang menjadikan tempat tersebut kotor dan kumuh agar kedepannya dapat lebih di perhatikan dan di evaluasi lagi. Untuk nilai dari performa akurasi bahwa model yang dibuat memiliki ketepatan akurasi sebesar 92%, nilai *precision* sebesar 90%, nilai *recall* sebesar 53% dan *f1-score* sebesar 67% yang menandakan akurasi dari

model yang dibuat sudah cukup baik. Berdasarkan hasil data dan informasi analisis sentimen yang diperoleh dapat membantu pemerintah setempat yaitu Dinas Tata Ruang dan Kabupaten Sukabumi agar dapat mengevaluasi dan melakukan pengembangan terhadap tempat tersebut secara tepat.

DAFTAR PUSTAKA

- [1] A. N. Farestianto and M. S. P. Nugroho, "Identifikasi Alun-Alun Brebes sebagai Ruang Terbuka Publik dengan Pendekatan Fungsional," *Pros. Semin. ...*, 2021, [Online]. Available: <https://proceedings.ums.ac.id/index.php/siar/article/view/1008%0Ahttps://proceedings.ums.ac.id/index.php/siar/article/download/1008/983>.
- [2] M. Jaenudin, "Sampah Berserakan Di Alun Alun Cisaat," *atmago.com*, 2020. https://www.atmago.com/berita-warga/sampah-berserakan-di-alun-alun-cisaat_b52ca246-9e6a-4d6f-8b91-752923668151 (accessed Jun. 07, 2023).
- [3] S. Alamsyah, "Miris, Sampah Berserakan di Alun-alun Palabuhanratu," *detik.com*, 2022. <https://www.detik.com/jabar/berita/d-6016736/miris-sampah-berserakan-di-alun-alun-palabuhanratu> (accessed Jun. 07, 2023).
- [4] S. Alamsyah, "Baru Selesai Dibangun, Plafon Podium Alun-alun Palabuhanratu Berjatuh," *detiknews*, 2022. <https://news.detik.com/berita-jawa-barat/d-5938536/baru-selesai-dibangun-plafon-podium-alun-alun-palabuhanratu-berjatuh> (accessed Jun. 08, 2023).
- [5] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 5, no. 2, pp. 697–711, 2021.
- [6] D. S. Utami and A. Erfina, "Analisis Sentimen Pinjaman Online di Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, vol. 1, no. 1, pp. 299–305, 2021.
- [7] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, pp. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [8] A. I. Tanggraeni and M. N. N. Sitokdana, "Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.
- [9] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 820–826, 2021, doi: 10.29207/resti.v5i4.3146.
- [10] D. Alita, I. Sari, and A. Rahman Isnain, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *Jdmsi*, vol. 2, no. 1, p. 702022, 2021.
- [11] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) Dengan Tf-Idf N-Gram Untuk Text Classification," vol. 6, no. 2, 2021.
- [12] P. Pandunata, K. T. Winarno N, Y. Nurdiansyah, and

N. El Maidah, “Analisis Sentimen Opini Publik Terhadap Program Vaksinasi Covid-19 Di Indonesia Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *INFORMAL Informatics J.*, vol. 7, no. 3, p. 246, 2022, doi: 10.19184/isj.v7i3.34930.

- [13] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,” *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.