Terbit online pada laman web jurnal: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

Journal of Dinda

Kelompok Keahlian Rekayasa Data
Institut Teknologi Telkom Purwokerto

Vol. 2 No. 2 (2022) 64 - 74 ISSN Media Elektronik: 2809-8064

Sentiment Analysis Destinasi Wisata Kabupaten Bekasi Berdasarkan Opini Masyarakat Menggunakan *Naive bayes*

Rizki Alamsyah[1*], Tb Ai Munandar[2*], Fata Nidaul Khasanah[3*], Siti Setiawati[4]
[1*, 2, 3, 4] Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
[1*]rizki.alamsyah18@mhs.ubharajaya.ac.id, [2]tb.aimunandar@dsn.ubharajaya.ac.id, [3]fatanidaul@gmail.com,
[4]siti.setiawati@dsn.ubharajaya.ac.id

Abstract

The topic used in this research is to discuss the problem of public opinion on social media related to tourist destinations in Bekasi Regency by implementing the Naive bayes algorithm to conduct sentiment analysis on existing opinions. This study aims to analyze public opinion on social media towards tourist destinations in Bekasi Regency using the Naive bayes algorithm. The data used in this study are posts or comments from the public on social media facebook as much as 1000 data. The method of data collection is done manually. The data analysis technique in this study are changing non-standard words, labelling, text preprocessing and naive bayes analysis methods. The results of this study indicate that positive opinion dominates compared to negative and neutral opinions with the results obtained at F1 positive score 83.5%, F1 negative score 68.2% and F1 neutral score 59.5% with positive recall 81%, negative 82% and neutral 55% precision positive 85%, negative 58% and neutral 64% with an accuracy rate of 76%.

Keywords: public opinion, social media, sentiment analysis, naive bayes, classification

Abstrak

Topik yang diangkat pada penelitian ini adalah membahas permasalahan tentang opini masyarakat di media sosial terkait destinasi wisata Kabupaten Bekasi dengan mengimplementasikan algoritma *naive bayes* untuk melakukan *sentiment analysis* terhadap opini yang ada. Penelitian ini bertujuan untuk menganalisis opini masyarakat di media sosial terhadap destinasi wisata Kabupaten Bekasi dengan menggunakan algoritma *naive bayes.* Data yang digunakan dalam penelitian ini adalah postingan atau komentar masyarakat di media sosial *facebook* sebanyak 1000 data. Metode pengumpulan data dilakukan secara manual. Teknik analisis data dalam penelitian ini melalui tahap pengubahan kata tidak baku, pelabelan, *text preprocessing* dan metode analisis *naive bayes.* Hasil penelitian ini menunjukkan bahwa opini positif mendominasi dibandingkan dengan opini negatif dan netral dengan hasil yang di dapat pada F1 *score* positif 83,5%, F1 *score* negatif 68,2% dan F1 *score* netral 59,5% dengan *recall* positif 81%, negatif 82% dan netral 55% presisi positif 85%, negatif 58% dan netral 64% dengan tingkat akurasi 76%.

Kata kunci: opini masyarakat, media sosial, s*entiment analysis, naive bayes,* klasifikasi

1. Pendahuluan

Kabupaten Bekasi merupakan salah satu daerah di Provinsi Jawa Barat yang memiliki banyak destinasi wisata, baik itu wisata alami dan wisata buatan. Wisata alami yang ada di Kabupaten Bekasi meliputi hutan dan pantai, sedangkan wisata buatan adalah taman, danau, dan wisata air. Destinasi wisata yang ada menciptakan daya tarik tersendiri bagi wisatawan lokal maupun luar daerah. Pada tahun 2021 Kabupaten Bekasi menduduki peringkat ketiga dengan jumlah kunjungan wisatawan terbanyak se-Kabupaten yang ada di Provinsi Jawa Barat yaitu mencapai 2.043.000 [1]. Banyaknya wisatawan yang berkunjung menyebabkan banyak opini positif, negatif maupun netral yang dilontarkan masyarakat melalui media sosial seperti *facebook.*

Keberadaan opini masyarakat di media sosial memberikan pengaruh terhadap penilaian eksistensi destinasi wisata di Kabupaten Bekasi sehingga dapat

Diterima Redaksi : 22-06-2022 | Selesai Revisi : 07-07-2022 | Diterbitkan Online : 01-08-2022

menimbulkan beberapa keputusan yang akan diambil oleh masyarakat lainnya terhadap destinasi wisata yang ada. Oleh karena itu sangat penting untuk dapat mengelola opini yang berkembang di media sosial untuk kebutuhan alternatif dalam penentuan keputusan para wisatawan untuk menganalisis opini di media sosial ada banyak teknik yang dapat digunakan, salah satunya sentiment analysis. Sentimen analisis adalah pengolahan data testimoni dimulai dari *preprocessing* sampai ke tahap klasifikasi [2] yang bertujuan untuk memperoleh bermacam sumber informasi dari internet serta bermacam-macam platform media sosial dan dapat mengetahui *class* positif, negatif dan netral [3]. Terdapat berbagai metode yang digunakan dalam sentimen analisis, salah satunya yaitu *Naive bayes*. *Naive bayes* merupakan metode yang digunakan untuk kebutuhan pengenalan pola dan klasifikasi sesuatu objek. Dalam teorema ini probabilitas di hitung untuk hipotesis menjadi benar [4].

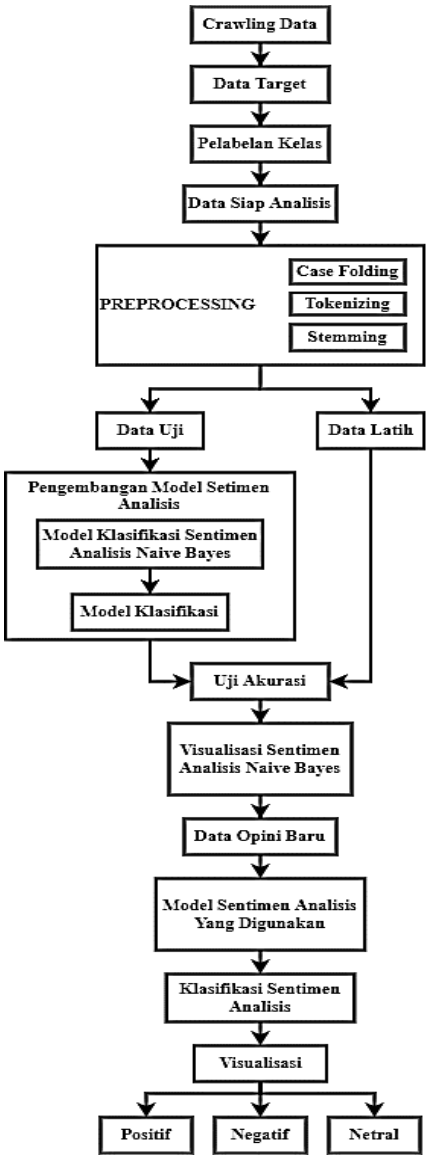
Penelitian ini dilakukan dengan alasan Terdapat banyak destinasi wisata di Kabupaten Bekasi namun masyarakat belum mengetahui opini positif, negatif dan netral terkait destinasi wisata tersebut serta Opini masyarakat yang ada tentang destinasi wisata di Kabupaten Bekasi harus diklasifikasikan menggunakan *sentiment analysis*.

Rumusan masalah yang akan dibahas dalam penelitian ini mencakup bagaimana cara menganalisis opini masyarakat di media sosial terhadap destinasi wisata Kabupaten Bekasi menggunakan pendekatan *naive bayes*? bagaimana mengimplementasikan algoritma *naive bayes* untuk melakukan *sentiment analysis* destinasi wisata Kabupaten Bekasi berdasarkan opini masyarakat di media sosial?

Tujuan dari penelitian ini adalah untuk melakukan analisis opini masyarakat di media sosial terhadap destinasi wisata Kabupaten Bekasi dengan *sentiment analysis*. Tujuan selanjutnya adalah untuk mengimplementasikan algoritma *naive bayes* untuk melakukan *sentiment analysis* destinasi wisata Kabupaten Bekasi berdasarkan opini masyarakat di media sosial.

2. Metode Penelitian

Metode penelitian ini mengggunakan beberapa tahapan untuk menganalisis opini masyarakat tentang destinasi wisata seperti pada Gambar 1.



Gambar 1.Model Analisis Sentimen

2.1. *Tearm Frequency - Inverse Document Frequency (TF-IDF)*

TF-IDF adalah frekuensi dokumen terbalik frekuensi dan dapat digunakan untuk menanyakan korpus dengan menghitung skor yang dinormalisasi yang menyatakan kepentingan relatif dari istilah dalam dokumen [5]. Adapun formulasi untuk menghitung bobot dari masing-

masing dokumen terhadap kata kunci menggunakan persamaan (1).

$$W_{dt} = TF_{dt} * IDF_{ft} \quad (1)$$

Dimana:

W_{dt} = Bobot dokumen ke-d terhadap kata ke-t

TF_{dt} = Banyaknya kata yang dicari pada dokumen

IDF_{ft} = Inversed Document Frequency $\left(\log \frac{N}{df}\right)$

N = Total dokumen

df = Banyak dokumen mengandung kata yang dicari

Frekuensi atau istilah mudah direpresentasikan sebagai berapa kali kemunculan kata dalam teks atau perbedaan pada panjang dokumen, sering terjadi jika tidak melakukan tahap normalisasi pada setiap kata yang ada di dalam dokumen. Oleh karena itu tahap normalisasi sangat dibutuhkan dalam perbaikan teks yang ada di dalam dokumen dengan menggunakan persamaan (2)

$$tf_{td} = \frac{\text{Frekuensi kemunculan term t pada dokumen d}}{\text{Total term pada dokumen d}} \quad (2)$$

2.2. *Naive bayes*

Berdasarkan buku teks [6] Larose, D.T., (2012) mendefinisikan bahwa teori *naive bayes* merupakan metode yang menentukan kebutuhan untuk pengenalan pola dan klasifikasi pada objek. Konsep ini berjalan dengan berawal pada anggapan jika pola klasifikasi bersumber atas nilai-nilai probabilistik yang memiliki suatu objek yang berlandaskan pada pola natural serta karakteristik yang didapat. Adapun formulasi untuk menghitung nilai probabilitas menggunakan persamaan (3).

$$P(X|Y) = \frac{P(Y|X) x P(X)}{P(Y)} \quad (3)$$

Dimana:

Y = Data dengan *class* yang belum diketahui

X = Hipotesis data atau suatu *class* yang spesifik

P(X|Y) = Probabilitas hipotesis berdasarkan kondisi tertentu *(posteriori probability)*

P(X) = Probabilitas hipotesis *(prior probability)*

P(Y|X) = Probabilitas berdasarkan kondisi pada hipotesis

P(Y) = Probabilitas dari Y

Uraian terhadap persamaan *bayes* dilakukan dengan menguraikan kondisi data Y dengan jumlah contoh tertentu dan belum memiliki *class* X. Adapun formulasi untuk menghitung nilai probabilitas menggunakan persamaan (4).

$$P(X_i|Y_1, Y_2, \ldots Y_n) = \frac{P(Y_1, Y_2, \ldots Y_n|X_i)}{\sum_{k=1}^{n} P(Y_1, Y_2, \ldots Y_k|X_k) x P(X_k)} \quad (4)$$

Dimana Xi merupakan *class* spesifik ke-i dan Yn atau Yk merupakan sejumlah data yang belum memiliki *class.*

2.3. *Crawling* Data

Crawling data merupakan tahapan pertama dalam menganalisis data untuk mencari dan mengumpulkan data melalui media sosial, web dan *marketplace*. Data tersebut dapat diperoleh dari masing-masing sumber dengan menggunakan *API* atau secara manual.

2.4. Data Target

Data target merupakan tahap lanjutan dari *crawling* data dengan tujuan untuk menggunakan data tersebut agar dapat dianalisis lebih lanjut.

2.5. Pelabelan Kelas

Pelabelan kelas merupakan proses mengklasifikasikan hasil data target ke dalam tiga kelas, yaitu kelas positif, negatif dan netral.

2.6 Data Siap Analisis

Setelah melalui beberapa tahapan di atas, data yang diperoleh telah siap untuk dianalisis. Proses analisis data dapat dilakukan menggunakan *tools* yang sudah tersedia maupun secara manual.

2.7 *Preprocessing*

Tahap *preprocessing* merupakan langkah penting dalam proses penemuan pengetahuan, sebab keputusan yang berkualitas harus didasarkan pada data yang berkualitas. Mengetahui anomali data, memperbaikinya lebih awal, serta mereduksi data, membuat hasil yang besar untuk pengambilan keputusan [7]. Pengumpulan data biasanya merupakan proses yang tidak ketat dalam pengontrolan, sehingga mendapatkan hasil yang tidak akurat [8]. Tahap *prepocessing* terdiri dari *case folding*, *tokenizing* dan *stemming* [9]. *Case folding* merupakan tahap untuk mengubah semua huruf kapital menjadi huruf kecil (*lowercase*). Hal ini dilakukan untuk menyamaratakan penggunaan huruf agar dapat mempermudah identifikasi. *Case folding* juga berguna untuk menghapus angka, tanda baca dan spasi yang berlebih. *Tokenizing* merupakan tahap untuk memecah kalimat menjadi sebuah kata atau disebut dengan token. Tujuannya yaitu untuk memudahkan dalam proses analisis data. *Stemming* merupakan tahap untuk menghapus atau mengubah kata yang tidak digunakan dalam kalimat, hanya mengambil kata inti saja dari setiap kalimat dan mengubahnya ke dalam bentuk kata dasar.

2.8 Pengembangan Model Sentimen Analisis

Dalam pengembangan model sentimen analisis terdiri dari tahapan model klasifikasi sentimen analisis *naive bayes* yang akan menghasilkan model klasifikasi untuk dianalisis lebih lanjut.

2.9 Uji Akurasi

Uji akurasi digunakan untuk menguji tingkat akurasi hasil pengembangan model sentimen analisis yang telah diperoleh. Sehingga dapat diketahui total akurasi keseluruhan dari *recall* dan presisi pada setiap klasifikasi yang tersedia.

2.10 Visualisasi Sentimen Analisis *Naive bayes*

Hasil sentimen analisis *naive bayes* yang telah diperoleh akan divisualisasikan dalam bentuk jumlah opini positif, negatif dan netral serta kata pada tampilan *wordcloud*. Kata tersebut dapat terbentuk dengan menarik dan informatif. Semakin sering kata digunakan, maka semakin besar ukuran kata yang keluar pada *wordcloud*.

2.11 Data Opini Baru

Data opini baru merupakan tahapan yang dilakukan untuk menguji data baru yang ditambahkan dengan menggunakan model sentimen analisis.

2.12 Model Sentimen Analisis yang Digunakan

Model sentimen analisis yang digunakan yaitu model sentimen analisis *naive bayes.*

2.13 Klasifikasi Sentimen Analisis

Klasifikasi sentimen analisis yang digunakan untuk menguji data opini baru yaitu model klasifikasi sentimen analisis *naive bayes*.

2.14 Visualisasi

Hasil sentimen analisis *naive bayes* yang telah diperoleh akan divisualisasikan dalam bentuk jumlah opini positif, negatif dan netral serta kata pada tampilan *wordcloud*. Kata tersebut dapat terbentuk dengan menarik dan informatif. Semakin sering kata digunakan, maka semakin besar ukuran kata yang keluar pada *wordcloud*.

3. Hasil dan Pembahasan

3.1. Data Hasil Penelitian

Penelitian ini menggunakan 1000 data yang diambil secara manual pada media sosial facebook dengan mencari postingan dan komentar terkait destinasi wisata di Kabupaten Bekasi. Data yang diambil berupa postingan atau komentar, tanggal upload, id pengguna, nama pengguna dan bagikan. Proses pengambilan data dilakukan pada postingan atau komentar *yang tersedia sejak tanggal 07 Januari 2022 sampai 30 Mei 2022. Data yang diperoleh berasal dari grup facebook yaitu "*Explore *Wisata Bekasi" dengan jumlah anggota yang bergabung sebanyak 92.100 anggota. Pada grup tersebut membahas tentang destinasi wisata di Kabupaten Bekasi yang ditunjukkan pada Gambar 2.*

Gambar 2.Grup *Explore* Wisata Bekasi

Beberapa data postingan atau komentar yang diperoleh dari *facebook* dapat dilihat pada Tabel 1.

Tabel 1.Beberapa data postingan atau komentar

Text	Tanggal	Id Profile	Nama Profil	Bagikan
Bagus ya teh	Jumat, 13 Mei 2022	100063530302727	Adelia Lontoh	FALSE
hooh....spot fotony bagus" teh	Jumat, 13 Mei 2022	100053893980198	Incees Yusni	FALSE
jeleek daah gk usah kesanq pook situ maah	Sabtu, 14 Mei 2022	100077310652424	Yudi Setia	FALSE
Alhmdulillah aku sudh ke sana.. Dan tempat nya baguuss bgt adem angin semriwing2 hhheee	Jumat, 13 Mei 2022	100007477666841	Pipiet Irul Zio Aziel	FALSE
bayar berapa masuk nya ?	Jumat, 13 Mei 2022	100003302648325	Faiz Fikrian to	FALSE

Postingan atau komentar yang sudah diambil akan melalui proses pengubahan kata. Pengubahan kata ini dilakukan secara manual karena postingan atau komentar yang sudah diambil masih menggunakan kata tidak baku. Tabel 2 merupakan sebagian kata tidak baku yang diperoleh dari data penelitian keseluruhan.

Tabel 2.Kata Tidak Baku

Sebelum	Sesudah
Bagen	Biarin
Ge	Aja
Olog	Boros
Ilok	Masa
Ora Danta	Engga Jelas
Uantri Pool	Antri Banget
Kongkow	Berkumpul
Nyo	Ayo
Gretong	Gratis
Misquen	Miskin
Now	Sekarang

Sebelum	Sesudah
Awang	Males
Sediain	Menyediakan
Pas	Tepat
Kne	Sini

Setelah melakukan pengubahan kata, selanjutnya memberikan label pada setiap kalimat. Pemberian label dilakukan secara manual untuk menentukan postingan atau komentar tersebut dinyatakan positif, negatif dan netral. Tabel 3 memperlihatkan sebagian data yang sudah diberikan label.

Tabel 3.Sebagian Hasil Pelabelan Data

Text	Label
Bagus ya teh.	Positif
engga ada musholanya doang. Dan kalau hujan susah neduh dan hasilnya basah kuyup	Negatif
Kalo ga salah ini bekas danau samba cuma beda pintu masuknya di rubah	Netral
iya tempat fotonya bagus-bagus mbak	Positif
jelek dah engga usah kesana mbak situ mah	Negatif

Setelah melakukan proses pelabelan, kemudian melalui tahap *text preprocessing*. Tahap ini dilakukan secara manual yang terdiri dari *case folding*, *tokenizing* dan *stemming*. Tahapan pertama yang harus dilakukan yaitu *case folding* dengan tujuan mengubah semua huruf kapital menjadi huruf kecil (*lowercase*). Tabel 4 menunjukkan beberapa hasil *case folding*.

Tabel 4.Beberapa Hasil *Case folding*

Sebelum	Sesudah
Bagus ya teh.	bagus ya teh
engga ada musholanya doang. Dan kalau hujan susah neduh dan hasilnya basah kuyup	engga ada musholanya doang dan kalau hujan susah neduh dan hasilnya basah kuyup
Kalo ga salah ini bekas danau samba cuma beda pintu masuknya di rubah	kalo ga salah ini bekas danau samba cuma beda pintu masuknya di rubah
iya tempat fotonya bagus-bagus mbak	iya tempat fotonya bagus-bagus mbak
jelek dah engga usah kesana mbak situ mah	jelek dah engga usah kesana mbak situ mah

Setelah proses *case folding* selesai selanjutnya masuk ke dalam tahap *tokenizing* dengan tujuan untuk memudahkan dalam proses analisis data. Tabel 5 menunjukkan beberapa hasil *tokenizing*.

Tabel 5.Beberapa Hasil *Tokenizing*

Sebelum	Sesudah
bagus ya teh	['bagus', 'ya', 'teh']
engga ada musholanya doang dan kalau hujan susah neduh dan hasilnya basah kuyup	['engga', 'ada', 'musholanya', 'doang', 'dan', 'kalau', 'hujan', 'susah', 'neduh', 'dan', 'hasilnya', 'basah', 'kuyup']
kalo ga salah ini bekas danau samba cuma beda pintu masuknya di rubah	['kalo', 'ga', 'salah', 'ini', 'bekas', 'danau', 'samba', 'cuma', 'beda', 'pintu', 'masuknya', 'di', 'rubah']
iya tempat fotonya bagus-bagus mbak	['iya', 'tempat', 'fotonya', 'bagus', 'mbak']
jelek dah engga usah kesana mbak situ mah	['jelek', 'dah', 'engga', 'usah', 'kesana', 'mbak', 'situ', 'mah']

Selanjutnya yaitu tahap *stemming*, yaitu tahap untuk menghapus atau mengubah kata yang tidak digunakan dalam kalimat, hanya mengambil kata inti saja dari setiap kalimat dan mengubahnya ke dalam bentuk kata dasar. Tabel 6 menunjukkan beberapa hasil *stemming.*

Tabel 6.Beberapa Hasil *Stemming*

Sebelum	Sesudah
bagus ya teh	bagus
engga ada musholanya doang dan kalau hujan susah neduh dan hasilnya basah kuyup	Enggak mushola hujan susah teduh hasil basah kuyup
kalo ga salah ini bekas danau samba cuma beda pintu masuknya di rubah	Enggak salah bekas danau samba beda pintu masuk ubah
iya tempat fotonya bagus-bagus mbak	iya tempat foto bagus
jelek dah engga usah kesana mbak situ mah	jelek enggak kesana

Setelah melakukan tahap *text preprocessing*, selanjutnya data tersebut dapat digunakan untuk melakukan analisis data di dalam web *kaggle.com* menggunakan bahasa pemrograman R.

3.2. Data Hasil *Tokenizing* dan *Stemming*

Tabel 7 memperlihatkan memperlihatkan data hasil *tokenizing* yang sudah dilakukan secara manual.

Tabel 7.Hasil *Tokenizing*

No	*Text*	*Tokenizing*
1	bayar berapa masuknya	['bayar', 'berapa', 'masuknya']
2	kemarin baru dari sana tempatnya masih bersih banget cuma kurang tempat permainan anak aja kalo buat foto mah oke aja kalo ajak anak bakal cepet bosen	['kemarin', 'baru', 'dari', 'sana', 'tempatnya', 'masih', 'bersih', 'banget', 'cuma', 'kurang', 'tempat', 'permainan', 'anak', 'aja', 'kalo', 'buat', 'foto', 'mah', 'oke', 'aja', 'kalo', 'ajak', 'anak', 'bakal', 'cepet', 'bosen']

No	*Text*	*Tokenizing*
3	kurang berasa minimal ada tempat bermain pasti anak betah ngajak balita paling liat ikan aja sukanya tapi lama bosen naik bebek mah emaknya juga yang goes	['kurang', 'berasa', 'minimal', 'ada', 'tempat', 'bermain', 'pasti', 'anak', 'betah', 'ngajak', 'balita', 'paling', 'liat', 'ikan', 'aja', 'sukanya', 'tapi', 'lama', 'bosen', 'naik', 'bebek', 'mah', 'emaknya', 'juga', 'yang', 'goes']

Tabel 8 memperlihatkan data hasil *stemming* yang sudah dilakukan secara manual.

Tabel 8.Hasil *Stemming*

No	*Text*	*Stemming*
1	kalo engga salah ini bekas danau samba cuma beda pintu masuknya di rubah tapi perasaan lokasinya sama aja dulu dua tahun	enggak salah bekas danau samba beda pintu masuk
2	lalu saya ke danau samba emang kayak gitu lokasi nya hanya beda titik lokasi pintu masuk kalo di google maps	lokasi hanya beda titik pintu masuk kalau google maps
3	air terjunnya kok engga ada airnya	air terjun enggak air

3.3. Perhitungan Pembobotan Dengan *TF-IDF*

Perhitungan pembobotan yang dilakukan pada setiap term atau teks yang ada pada postingan atau komentar *facebook* yang sudah melalui tahap *preprocessing*. Dalam pembobotan ini menggunakan metode *TF-IDF* yang bertujuan untuk memberikan nilai pada setiap term atau text. Tabel 9 memperlihatkan perhitungan *TF-IDF* untuk sebagian data berdasarkan tahap pada Tabel 8.

Tabel 9.Sebagian Dokumen Perhitungan *TF-IDF*

N	*Text*
C1	enggak salah bekas danau samba beda pintu masuk
C2	lokasi hanya beda titik pintu masuk kalau google maps
C3	air terjun enggak ada air

Tabel 9 memuat komponen yang akan diolah pada table selanjutnya berupa *class* yang dinotasikan dalam huruf C. Terdapat tiga jenis *class* dalam tabel, yaitu C1 yang memperlihatkan *class* positif, C2 memperlihatkan *class* negatif, dan C3 memperlihatkan *class* netral. Tahap selanjutnya yaitu memisahkan kalimat menjadi sebuah kata, menghitung panjang dokumen dan menormalisasikan kata yang ada pada dokumen. Tabel 10 menampilkan hasil dari normalisasi kata.

Tabel 10.Hasil Normalisasi Kata

Text	*TF*			*TF* Normalisasi		
	c1	*c2*	*c3*	*c1*	*c2*	*c3*
enggak	1	0	1	0,125	0	0,2
salah	1	0	0	0,125	0	0
bekas	1	0	0	0,125	0	0
danau	1	0	0	0,125	0	0
samba	1	0	0	0,125	0	0
beda	1	1	0	0,125	0,111	0
pintu	1	1	0	0,125	0,111	0
masuk	1	1	0	0,125	0,111	0
lokasi	0	1	0	0	0,111	0
hanya	0	1	0	0	0,111	0
titik	0	1	0	0	0,111	0
kalau	0	1	0	0	0,111	0
google	0	1	0	0	0,111	0
maps	0	1	0	0	0,111	0
air	0	0	2	0	0	0,4
terjun	0	0	1	0	0	0,2
ada	0	0	1	0	0	0,2
panjang *text*	8	9	5			

Setelah melakukan tahap normalisasi, selanjutnya menghitung dokumen frekuensi, yaitu jumlah dokumen yang mengandung term atau teks pada setiap dokumen. Tabel 11 memperlihatkan hasil dari dokumen frekuensi.

Tabel 11.Hasil Dokumen Frekuensi

Text	*TF*			*TF* Normalisasi			*DF*
	c1	*c2*	*c3*	*c1*	*c2*	*c3*	
enggak	1	0	1	0,125	0	0,2	2
salah	1	0	0	0,125	0	0	1
bekas	1	0	0	0,125	0	0	1
danau	1	0	0	0,125	0	0	1
samba	1	0	0	0,125	0	0	1
beda	1	1	0	0,125	0,111	0	2
pintu	1	1	0	0,125	0,111	0	2
masuk	1	1	0	0,125	0,111	0	2
lokasi	0	1	0	0	0,111	0	1
hanya	0	1	0	0	0,111	0	1
titik	0	1	0	0	0,111	0	1
kalau	0	1	0	0	0,111	0	1
google	0	1	0	0	0,111	0	1
maps	0	1	0	0	0,111	0	1
air	0	0	2	0	0	0,4	2
terjun	0	0	1	0	0	0,2	1
ada	0	0	1	0	0	0,2	1
panjang *text*	8	9	5				

Setelah melakukan tahap normalisasi, selanjutnya menghitung *Inverse Document Frequency (IDF)*, yaitu kebalikan dari dokumen frekuensi. Tabel 12 memperlihatkan hasil dari *Inverse Document Frequency (IDF)*.

Tabel 12.Hasil *Inverse Document Frequency (IDF)*

Text	*TF* Normalisasi			*DF*	*IDF*
	c1	*c2*	*c3*		
enggak	0,125	0	0,2	2	0,176
salah	0,125	0	0	1	0,477
bekas	0,125	0	0	1	0,477
danau	0,125	0	0	1	0,477
samba	0,125	0	0	1	0,477
beda	0,125	0,111	0	2	0,176
pintu	0,125	0,111	0	2	0,176
masuk	0,125	0,111	0	2	0,176
lokasi	0	0,111	0	1	0,477
hanya	0	0,111	0	1	0,477
titik	0	0,111	0	1	0,477
kalau	0	0,111	0	1	0,477
google	0	0,111	0	1	0,477
maps	0	0,111	0	1	0,477
air	0	0	0,4	2	0,176
terjun	0	0	0,2	1	0,477
ada	0	0	0,2	1	0,477

Setelah melakukan tahap *Inverse Document Frequency (IDF)*, selanjutnya menghitung nilai *TF-IDF*, yaitu perkalian antara term atau kata frekuensi yang sudah di normalisasikan dengan *Inverse Document Frequency (IDF)* pada setiap dokumen. Tabel 13 memperlihatkan hasil dari *TF-IDF* dari setiap dokumen.

Tabel 13.Hasil *TF-IDF*

Text	*TF* Normalisasi			*IDF*	*TF-IDF*		
	c1	*c2*	*c3*		*c1*	*c2*	*c3*
enggak	0,125	0	0,2	0,176	0,022	0	0,035
salah	0,125	0	0	0,477	0,06	0	0
bekas	0,125	0	0	0,477	0,06	0	0
danau	0,125	0	0	0,477	0,06	0	0
samba	0,125	0	0	0,477	0,06	0	0
beda	0,125	0,111	0	0,176	0,022	0,02	0
pintu	0,125	0,111	0	0,176	0,022	0,02	0
masuk	0,125	0,111	0	0,176	0,022	0,02	0
lokasi	0	0,111	0	0,477	0	0,053	0
hanya	0	0,111	0	0,477	0	0,053	0
titik	0	0,111	0	0,477	0	0,053	0
kalau	0	0,111	0	0,477	0	0,053	0
google	0	0,111	0	0,477	0	0,053	0
maps	0	0,111	0	0,477	0	0,053	0
air	0	0	0,4	0,176	0	0	0,07
terjun	0	0	0,2	0,477	0	0	0,095
ada	0	0	0,2	0,477	0	0	0,095

Adapun untuk data lainnya, perhitungan *TF-IDF* dilakukan dengan cara yang sama. Hasil perhitungan *TF-IDF* kemudian digunakan untuk tahap berikutnya, yakni model sentiment analisis.

3.4. Model Sentimen Analisis Dengan *Naive bayes*

Untuk membangun model klasifikasi sentimen analisis dengan *naive bayes*, pada penelitian ini digunakan 1000 data postingan atau komentar. Data dibagi menjadi dua bagian yaitu data *training* dan data *testing* dengan perbandingan 80% data *training* dan 20% data testing. Tabel 14 merupakan pembagian menjadi data training dan testing.

Tabel 14.Hasil Pembagian Data *Training* dan *Testing*

Data	**Prosentase**	**Positif**	**Negatif**	**Netral**
Training	80%	503	103	194
Testing	20%	127	28	45
Total	100%	630	131	239
		63%	13,1%	23,9%

Setelah melakukan pembagian dua data, selanjutnya masuk ke dalam tahap model klasifikasi sentimen analisis *naive bayes*. Klasifikasi *naive bayes* dilakukan untuk menentukan hasil prediksi dan aktual pada *class* positif, negatif dan netral menggunakan bahasa pemrograman R yang tersedia di dalam web *kaggle.com*. Tabel 15 merupakan hasil dari klasifikasi data *testing*.

Tabel 15.Hasil Klasifikasi Data *Testing*

Prediksi	**Aktual**			**Row Total**
	Negatif	**Netral**	**Positif**	
Negatif	23	6	10	39
Netral	1	25	13	39
Positif	4	14	104	122
Total	28	45	127	200

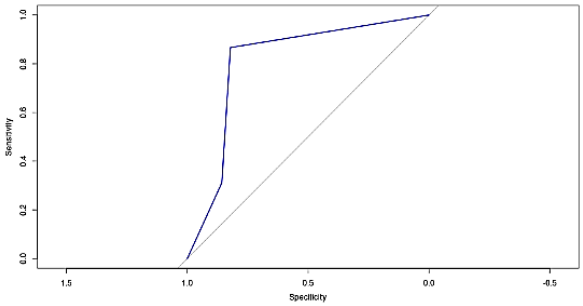
Setelah melakukan klasifikasi *naive bayes*, selanjutnya masuk ke dalam tahap uji akurasi. Tahap ini dilakukan dengan membuat *confusion matrix*, yaitu sebuah pengukuran performa untuk masalah klasifikasi *text mining* dengan menggunakan dua kelas atau lebih. *Confusion matrix* digunakan untuk mengetahui nilai akurasi, *recall*, presisi dan F1 *score* yang ada pada data *testing*. Tabel 16 merupakan hasil dari klasifikasi data *testing*. *Confusion matrix* yang sudah dibuat di dalam web *kaggle.com.*

Tabel 16.Hasil Keseluruhan *Confusion Matrix*

	Presisi	***Recall***	**F1 *Score***
Negatif	58,9%	82%	68,6%
Netral	64%	55%	59,5%
Positif	85%	81%	83,5%

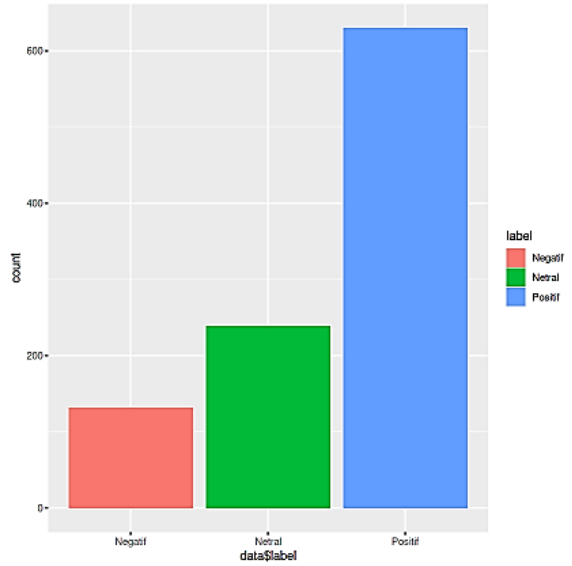
	Presisi	***Recall***	**F1 *Score***
Akurasi		76%	

Setelah melakukan klasifikasi, selanjutnya mengukur kinerja akurasi menggunakan visualisasi kurva *ROC* (*Receiver Operating Characteristic*). *ROC* digunakan untuk menghitung kinerja algoritma klasifikasi yang sudah dibuat dengan *confusion matrix*. Gambar 3 merupakan kurva *ROC* yang dibuat di *kaggle.com* dengan nilai *AUC* klasifikasi *naive bayes* sebesar 80,1% sehingga dapat dikategorikan baik.



Gambar 3. Kurva *ROC* Klasifikasi *Naive bayes*

Setelah melakukan pengukuran akurasi, baik akurasi keseluruhan maupun *ROC*, selanjutnya pada Gambar 4 menyajikan grafik visualisasi hasil proses *labelling* data. Berdasarkan Gambar 4 dapat dilihat bahwa postingan atau komentar yang disampaikan masyarakat melalui media sosial *facebook* lebih didominasi oleh opini positif dibandingkan dengan opini netral dan negatif. Dengan jumlah yang diperoleh opini positif sebesar 63%, netral sebesar 23,9% dan negatif sebesar 13,1%.

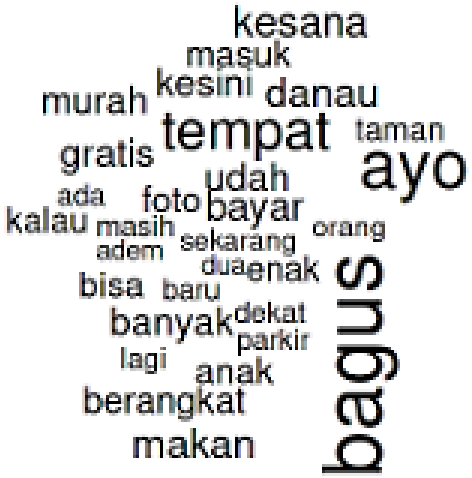


Gambar 4. Grafik Hasil *Labelling* Data

Setelah melakukan visualisasi grafik, selanjutnya mendeteksi kata yang sering muncul dari setiap kalimat menggunakan *wordcloud* yang tersedia pada Bahasa pemrograman R di web *kaggle.com*. *Wordcloud* digunakan untuk menampilkan kata secara visual [10]. Kata tersebut dapat terbentuk dengan menarik dan informatif. Semakin sering kata digunakan, maka semakin besar ukuran kata yang keluar pada *wordcloud*. Gambar 5 merupakan bentuk visual *wordcloud* dua puluh kata yang sering muncul.

Gambar 5. *Wordcloud* Dua Puluh Kata Sering Muncul

Hasil *wordcloud* seperti Gambar 5 memvisualisasikan dua puluh kata yang sering muncul memperlihatkan bahwa, sebagian besar postingan pengguna media sosial *facebook* memuat kata bagus, enggak, ayo, tempat, makan, masuk, anak, banyak, bayar dan danau. Selanjutnya memvisualisasikan kata yang sering muncul berdasarkan label positif pada web *kaggle.com*. Gambar 6 merupakan bentuk visual *wordcloud* kata yang sering muncul pada label positif.



Gambar 6. *Wordcloud* Label Positif Sering Muncul

Hasil *wordcloud* seperti Gambar 6 memvisualisasikan kata yang sering muncul pada label positif memperlihatkan bahwa sebagian besar postingan berlabel positif memuat kata bagus, ayo, tempat, kesana dan berangkat. Selanjutnya memvisualisasikan kata yang muncul berdasarkan label negatif pada web *kaggle.com*. Gambar 7 merupakan bentuk visual *wordcloud* kata yang sering muncul pada label negatif.

Gambar 7. *Wordcloud* Label Negatif Sering Muncul

Hasil *wordcloud* seperti Gambar 7 memvisualisasikan kata yang sering muncul pada label negatif memperlihatkan bahwa, sebagian besar postingan berlabel negatif memuat kata enggak, mahal, panas dan makan. Selanjutnya memvisualisasikan kata yang puncul berdasarkan label netral pada web *kaggle.com*. Gambar 8 merupakan bentuk visual *wordcloud* kata yang sering muncul pada label netral.

Gambar 8. *Wordcloud* Label Netral Sering Muncul

Hasil *wordcloud* seperti Gambar 8 memvisualisasikan kata yang sering muncul pada label netral memperlihatkan bahwa, sebagian besar postingan berlabel netral memuat kata enggak, masuk, berapa dan kalau.

3.4. Pembahasan

Hasil yang telah dilakukan pada model sentimen analisis *naive bayes* menggunakan 1000 data. Mendapatkan hasil opini positif sebesar 63%, netral sebesar 23,9% dan negatif sebesar 13,1%. Hal ini terjadi karena jumlah data positif, negatif dan netral tidak mendekati (simetris). Maka sebaiknya menggunakan F1 *score* sebagai acuan [11] seperti pada Tabel 16.

Berdasarkan Tabel 16 merupakan hasil dari perhitungan *recall* yang didapat dari pengujian data testing dari setiap masing-masing *class* yaitu positif sebesar 81%, negatif sebesar 82% dan netral sebesar 55% dengan kesuluruhan data *training* sebanyak 200 dan hasil prediksi benar sebanyak 152. Hasil perhitungan presisi pada data *testing* dari setiap *class* yaitu positif sebesar 85%, negatif sebesar 58,9% dan netral sebesar 64% dan hasil perhitungan F1 *score* didapat dari penjumlahan *recall* dan presisi dari setiap *class* yaitu F1 *score* positif sebesar 83,5%, F1 *score* negatif sebesar 68,6% dan F1 *score* netral sebesar 59,5% dengan total keseluruhan akurasi sebesar 76%..

Selanjutnya menguji model sentimen analisis *naive bayes* dengan menambahkan 100 data baru. Data tersebut sudah melalui tahap pelabelan dan *text preprocessing*. Pengujian model sentimen analisis *naive bayes* ini dilakukan dengan cara menambahkan data baru kedalam model sentimen analisis di dalam web *kaggle.com*. Gambar 9 merupakan *source code* dan hasil dari menambahkan data baru ke dalam model sentimen analisis *naive bayes.*

```
data2NB=data.frame(databaru[,2])
head(data2NB)
```

A data.frame: 6 × 1

	databaru...2.
	<fct>
1	akhir kesampaian tempat foto bagus
2	apakah ada warung makan
3	ada
4	dekat dari rumah
5	terima kasih kunjungan
6	pintu masuk kampung kita taman lansia semakin keren

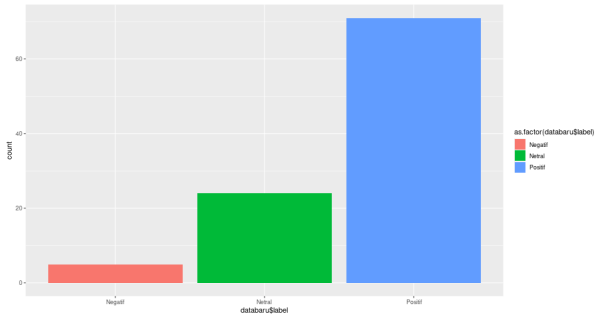
Gambar 9. *Source Code* dan Hasil Pengujian Model Sentimen Analisis

Setelah menambahkan data baru ke dalam model sentimen analisis *naive bayes*, selanjutnya menampilkan hasil dari klasifikasi sentimen analisis menggunakan data baru dan tingkat akurasi yang didapatkan. Tabel 17 merupakan hasil dari hasil klasifikasi sentimen analisis dan hasil akurasi menggunakan data baru.

Tabel 17.Hasil Klasifikasi Sentimen Analisis Menggunakan Data Baru

Prediksi	**Aktual**			***Row Total***
	Negatif	**Netral**	**Positif**	
Negatif	0	0	0	0
Netral	0	0	0	0
Positif	5	24	71	100
Akurasi	71%			

Berdasarkan Tabel 17 dapat dilihat bahwa opini positif lebih mendominasi dibandingkan opini negatif dan netral dengan hasil yang diperoleh opini positif sebesar 71%, negatif sebesar 5% dan netral sebesar 25% dengan total keseluruhan akurasi sebesar 71%. Gambar 10 merupakan visualisasi grafik hasil klasifikasi 100 data.



Gambar 10. Grafik Hasil 100 Data

Setelah melakukan tahap pengujian akurasi pada data baru, selanjutnya memvisualisisasikan *wordcloud* dua puluh kata yang sering muncul pada 100 data baru pada web *kaggle.com*. Gambar 11 merupakan visualisasi dua puluh kata sering muncul pada 100 data baru.

Gambar 11. *Wordcloud* Dua Puluh Kata Sering Muncul Pada Data Baru

Hasil *wordcloud* seperti Gambar 10 memvisualisasikan dua puluh kata yang sering muncul pada data baru memperlihatkan bahwa sebagian besar postingan di media sosial *facebook* memuat kata bagus, tempat, ada, banyak, gratis, makan, danau dan taman.

4. Kesimpulan

Berdasarkan sentimen analisis yang telah dilakukan oleh peneliti, maka dapat menarik kesimpulan sebagai berikut:

1. Sentimen analisis destinasi wisata Kabupaten Bekasi berdasarkan opini masyarakat di media sosial facebook menunjukkan bahwa opini positif mendominasi dibandingkan dengan opini negatif dan netral dengan hasil yang di dapat pada F1 *score* positif 83,5%, F1 *score* negatif 68,2% dan F1 *score* netral 59,5% dengan *recall* positif 81%, negatif 82% dan netral 55% presisi positif 85%, negatif 58% dan netral 64% dengan tingkat akurasi 76%.
2. Pengujian model sentimen analisis *naive bayes* dapat bekerja dengan baik. Hal ini dibuktikan dengan menambahkan 100 data baru pada model *naive bayes*, sehingga mendapatkan hasil yang diperoleh opini positif 71%, negatif 24% dan netral 5% dengan total keseluruhan akurasi sebesar 71%.

Daftar Rujukan

[1] BPS, “Badan Pusat Statistik Kabupaten Bekasi,” 2021. https://bekasikab.bps.go.id/statictable/2021/07/06/2066/jumlah-kunjungan-wisatawan-ke-obyek-wisata-di-jawa-barat-menurut-

kabupaten-kota-2018.html.

[2] I. Diana and Widiastuti, "Sentiment Analysis Review Novel 'Goodreads' Berbahasa Indonesia Menggunakan Naïve Bayes Classifier Sentiment Analysis Review Novel 'Goodreads' Berbahasa Indonesia Menggunakan Naive Bayes Classifier," *Semnas Ristek*, pp. 760–765, 2021.

[3] S. Sigit, U. Ema, and L. Emha Taufiq, "Analisis Sentiment Pada Twitter Dengan Menggunakan," pp. 9–15, 2018.

[4] bhatia Partek, *Data mining and data warehousing*, vol. 47. 2007.

[5] R. Matthew A and K. Mikhail, *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, and More*. 2019.

[6] M. TB Ai, "Bahan Ajar Data Mining Dengan Bahasa R Edisi Revisi 3," 2019.

[7] H. Jiawei, K. Micheline, and P. Jian, *Data mining: Data mining concepts and techniques*. 2012.

[8] G. Salvador, L. Julian, and H. Francisco, *Data Preprocessing in Data Mining*, vol. 72. 2015.

[9] A. Muhammad Iqbal, G. Dudih, and S. Falentino, "Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 807–814, 2021.

[10] RPubs, "Wordcloud,", Accessed [Online] 12 May 2022, https://rpubs.com/aswinjanuarsjaf/611448. 2020.

[11] P. V. Oddy, H. Triana, and S. Ahmad, "Outlier Detection On Graduation Data Of Darussalam Gontor University Using One-Class Support Vector Machine," *Procedia Eng. Life Sci.*, vol. 2, no. 2, pp. 89–92, 2021.