Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 263-272
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

Perbandingan Algoritma *Naïve Bayes* Dan *K-Nearest Neighbor* pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi *Covid 19*

Defrimont Era[1], Septi Andryana[2], Albaar Rubhasy[3]
[1,2,3]Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Indonesia
e-mail: defrimontera2020@student.unas.ac.id[1], septi.andryana@civitas.unas.ac.id[2], albaar.rubhasy@civitas.unas.ac.id[3]

Abstract

Covid-19 is a trending topic of conversation on social media today. one of the trending topics on social media in 2021 in the midst of the covid 19 pandemic related to tourism destinations that will reopen amid the increase in daily cases of covid 19 in a number of areas of Indonesia in particular. The implementation of the government's policy of opening tourist destinations has drawn various arguments or opinions among the public. especially social media users by providing arguments or tweets related to opinions through social media such as Instagram, Twitter, and Facebook. In this study, an analysis of the sentiments of the Indonesian people, especially social media users, was carried out regarding government policies related to the opening of tourism in the midst of the COVID-19 pandemic using the Naive Bayes and K-Nearest Neighbor (KNN) algorithms and knowing the comparison of the performance of the two algorithms. This dataset used is taken through various social media platforms with a data crawling process. The dataset used is a sql file. the number of datasets used to process data processing in the public sentiment analysis application about the opinion of opening tourism in the midst of the covid 19 pandemic on twitter as many as 570 tweets using 12 hashtags namely #Desabisa, #desawisata, #DiIndonesiaAja, #EraNewNormal and #LabuanBajo #MulaiDariDesa #PPKMDiPerpanjang #SobatParekraf #PlacesTourism #Bali Tours #Banyuwangi Tours #Jogja Tours. In the research conducted, the implementation of the results of the application of the Naive Bayes algorithm and the K-Nearest Neighbor (KNN) algorithm in the sentiment analysis of the opening of tourism destinations during the covid-19 pandemic was successfully carried out by producing the highest Naive Bayes accuracy rate of 75.53%, precision the highest positive is 71%, the highest negative precision is 25%, the highest positive recall is 99% and the highest negative recall is 14%. As for the K-Nearest Neighbor (KNN) algorithm, the highest accuracy rate is 48.66%, the highest positive precision is 69%, the highest negative precision is 14%, the highest positive recall is 69% and the highest negative recall is 28%. The level of accuracy in the comparison of the two algorithms shows that the Naive Bayes algorithm has the best performance with an accuracy level compared to the K-Nearest Neighbor algorithm in analyzing the sentiments of social media users about the opening of tourism during the COVID-19 pandemic.

Keywords*: tourism destination, naive bayes, , K-Nearest Neigbor (KNN), sentiment analysis, social media.*

Abstrak

Covid-19 menjadi trending topik perbincangan dalam media sosial saat ini. salah satu topik trending topik di media sosial pada tahun 2021 ditengah masa pandemi covid 19 yang berkaitan dengan destinasi pariwisata yang akan dibuka kembali di tengah kenaikan kasus harian covid 19 di sejumlah daerah Indonesia khususnya. Penerapan kebijakan pemerintah pembukaan destinasi wisata tersebut menuai berbagai argumentasi atau opini dikalangan publik. khususnya pengguna media sosial dengan memberikan argumentasi atau tweet terkait opini melalui media sosial seperti instagram, twitter, dan facebook. Dalam penelitian ini, dilakukan analisis sentimen masyarakat Indonesia khususnya pengguna media sosial tersebut tentang kebijakan pemerintah terkait pembukaan pariwisata di tengah masa pandemi covid-19 dengan menggunakan algoritma Naive Bayes dan K-Nearest Neighbor (KNN) serta mengetahui perbandingan peforma kedua algoritma tersebut.Dalam penelitian ini dataset yang digunakan diambil melalui berbagai media platform media social dengan proses crawling data. Dataset yang digunakan berupa file sql. jumlah dataset yang digunakan untuk

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 263-272
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

memproses pengolahan data pada aplikasi analisis sentimen masyarakat tentang opini pembukaan pariwisata di tengah masa pandemi covid 19 pada twitter sebanyak 570 tweet menggunakan 12 hashtag yaitu #DesaBisa, #desawisata, #DiIndonesiaAja, #EraNewNormal dan #LabuanBajo #MulaiDariDesa #PPKMDiPerpanjang #SobatParekraf #TempatWisata #WisataBali #WisataBanyuwangi #WisataJogja. Dalam penelitian yang dilakukan, implementasi hasil dari penerapan algoritma Naive Bayes dan algoritma K-Nearest Neighbor (KNN) pada analisis sentimen pembukaan destinasi pariwisata di masa pandemi covid-19 berhasil dilakukan dengan menghasilkan nilai tingkat akurasi Naive Bayes tertinggi sebesar 75,53%, precision positif tertinggi yaitu 71%, precision negatif tertinggi yaitu 25%, recall positif tertinggi yaitu 99% dan recall negatif tertinggi yaitu 14%. Sedangkan untuk algoritma K-Nearest Neighbor (KNN) didapatkan nilai tingkat akurasi tertinggi sebesar 48,66%, precision positif tertinggi yaitu 69%, precision negatif tertinggi yaitu 14%, recall positif tertinggi yaitu 69% dan recall negatif tertinggi yaitu 28%. Tingkat akurasi dalam perbandingan kedua algoritma tersebut menunjukan bahwa algoritma Naive Bayes memiliki peforma terbaik dengan tingkat akurasi dibandingkan dengan algoritma K-Nearest Neighbor dalam menganalisa sentimen masyarakat pengguna media sosial tentang pembukaan pariwisata di masa pandemi covid-19.

***Kata kunci**: destinasi pariwisata, naive bayes, K-Nearest Neigbor (KNN), analisis sentimen, media sosial*

1. PENDAHULUAN

Tahun 2020 Seluruh Dunia digemparkan oleh munculnya wabah global atau penyakit baru yang berkembang begitu cepat yaitu Coronavirus Disease (Covid-19) atau virus corona. Menurut Lembaga Kesehatan International yaitu WHO menyatakan bahwa dunia mengalami situasi darurat global kesehatan dimulai semenjak Januari 2020 akibat virus corona. Banyak negara telah terinfeksi oleh penyebaran virus corona salah satunya adalah Indonesia. Dilansir pada halaman resmi data covid19.go.id orang Indonesia yang terinfeksi cirus corona sampai bulan Juli 2022 saat ini dengan jumlah yang terinfeksi virus corona sebanyak 6.103.552 orang [1]. Wabah tersebut telah menimbulkan berbagai dampak multisector sehingga menganggu nilai pertumbuhan ekonomi diberbagai belahan dunia salah satunya Indonesia. Di Indonesia Presiden Joko Widodo mengeluarkan beberapa kebijakan dalam penanganan covid-19 dalam rangka menekan laju kasus penularan covid-19. Hal ini mendapatkan perhatian masyarakat Indonesia terkait kebijakan yang dikeluarkan oleh pemerintah yaitu kebijakan karantina wilayah, hingga pembatasan sosial skala besar (PSBB) yang dimulai sejak akhir bulan februari 2020 hingga bulan Mei 2021. Sejumlah penerbangan juga dihentikan sementara kondisi ini membuat aktivitas ekonomi ikut terdampak. Covid-19 memberikan dampak yang sangat tinggi bagi perputaran perekenomian masyarakat Indonesia khususnya. Dalam hal ini Pemerintah mencoba menerapkan beberapa upaya pemulihan ekonomi nasional seperti pembukaan beberapa destinasi wisata Indonesia di tengah masa pandemi untuk menghidupkan kembali roda ekonomi masyarakat.

Penerapan kebijakan pemerintah terkait pembukaan destinasi pariwisata mendapat banyak tanggapan dari masyarakat. Tanggapan atau opini dari masyarakat sangat beragam yaitu ada yang berpendapat positif, netral, dan ada juga yang berpendapat negatif. Pendapat atau tanggapan dari masyarakat banyak dikemukakan berbagai pada media sosial seperti twitter, Instagram dan facebook. Penggunaan media sosial membuat orang bebas memberikan pendapat terkai hal apapun. Berdasarkan pembahasan tersebut, pada penelitian ini dilakukan analisa sentimen pada penggunaan media sosial tersebut terhadap kebijakan pemerintah terkait pembukaan destinasi

pariwisata di tengah masa pandemi covid-19 dengan mengklasifikasi pendapat atau tanggapan masyarakat Indonesia dengan penggunaan metode *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* pada media sosial.

Algoritma *Naïve Bayes* yaitu merupakan metode pengkategorian. Dalam metode *Naïve Bayes* melakukan pengkategorian menggunakan probalitas dan statistic. Pada masa sebelumnya (teorema bayes) melakukan prediksi peluang dengan menggunakan probabilitas dan statistic, berasumsi sangat kuat (naif) dan bergantung pada masing-masing kondisi/kejadian, itu merupakan ciri utama dari algoritma *Naive Bayes*[2]. Sedangkan Algoritma KNN metode klasifikasi contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, bergantung pada label kategori melekat pada dokumen pelatihan dengan dokumen tes [3].

Penelitian tentang analisis sentimen menggunakan algoritma *KNN* telah banyak dilakukan salah satunya penelitian sistem pakar diagnosa kerusakan VGA dengan metode *certainty factor* dan algoritma *K-Nearest Neighbor (KNN)* [4]. Pada penelitian lain menggunakan algoritma naïve bayes mengenai visualisasi dan analisa data penyebaran covid-19 dengan metode klasifikasi naïve bayes [5].

Berdasarkan permasalahan yang terjadi dan penelitian sebelumnya maka peneliti membuat penelitian yang berjudul "Analisis Perbandingan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor (KNN)* pada Analisis Sentimen Media Sosial (Studi Kasus : Kebijakan Pemerintah Pembukaan Pariwisata di Masa Pandemi Covid-19".

2. METODOLOGI PENELITIAN

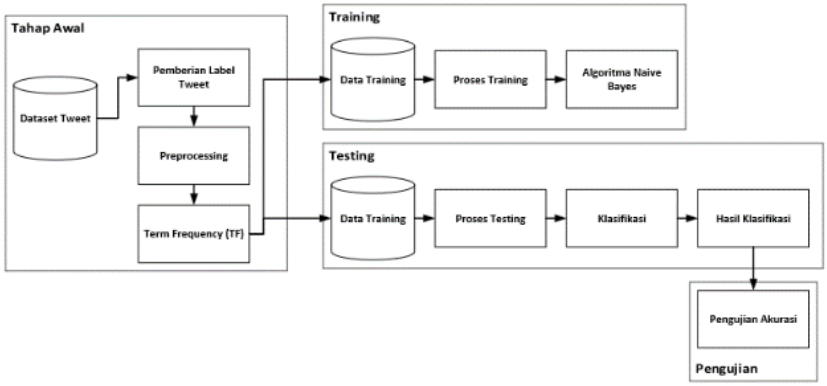
Metode penelitian yang dilakukan yaitu pengolahan data terhadap data media sosial, *twitter, facebook, dan instagram*. Sistem yang dibuat dapat mengklasifikasikan data ke dalam opini positif, netral, dan negatif. Tahapan sistem yang dilakukan sebagai berikut:

2.1. Analisis Sistem

Tahapan analisis sistem dilakukan teknik pengumpulan kebutuhan sistem *(requirement)* yang digunakan untuk mengelola data *tweet* yang akan dibuat dari hasil penelitian dan studi literature. Selain itu terdapat juga penjelasan terkait fungsi-fungsi, metode, kinerja sistem yang akan digunakan.

Pada tahap analisis sistem merupakan proses mengklasifikasikan *tweet* ke dalam label positif, netral, dan negatif. Analisis Sistem memiliki gambaran terkait proses klasisifikasi yaitu empat sub proses sehingga dapat menghasilkan satu proses klasifikasi yang mengimplementasikan algorima *Naive Bayes* dan *K-Nearest Neighbor (KNN)*. Berikut akan dijelaskan mengenai konsep alur sistem yang dibuat pada penelitian ini, hal ini terkait pada Gambar 1.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 263-272
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti



Gambar 1. Alur analisis sistem

Gambar 1 pada gambar sistem terdiri dari 4 (empat) tahap dalam analis sentimen masyarakat tentang opini pembukaan destinasi pariwisata yaitu : tahap awal yang terdiri dari pemberian label tweet, tahap *preprocessing* dan tahap proses *term frequency (TF)*. Tahap kedua yaitu training yang terdiri dari hasil data training, proses training, algoritma naïve bayes. Tahap ketiga yaitu terkait testing dimana terdiri pengujian aplikasi. Terakhir tahap ke empat yaitu tahap pengujian akurasi.

2.2. Pengumpulan Data

Pengumpulan data merupakan proses untuk pengumpulan data pendukung yang dibutuhkan dalam penelitian ini. *dataset* yang digunakan diambil melalui berbagai media *platform* media *social* dengan proses *crawling* data. *Dataset* yang digunakan berupa file sql. jumlah *dataset* yang digunakan untuk memproses pengolahan data pada aplikasi analisis sentimen masyarakat tentang opini pembukaan pariwisata di tengah masa pandemi *covid 19* pada *twitter* sebanyak 570 *tweet* menggunakan 12 hashtag yaitu #DesaBisa, #desawisata, #DiIndonesiaAja, #EraNewNormal dan #LabuanBajo #MulaiDariDesa #PPKMDiPerpanjang #SobatParekraf #TempatWisata #WisataBali #WisataBanyuwangi #WisataJogja.

2.3. Pengolahan Data

Pada tahapan pengolahan data merupakan proses *crawling twitter API* berupa data tweet langsung dari pengguna twitter. Data tersebut perlu dilakukan pengolahan terlebih dahulu agar menjadi data yang mudah digunakan dalam proses *sentiment analysis*. data *Tweet* yang didapatkan melalui proses *filter* atau penyaringan kata sehingga data *tweet* yang didapatkan menjadi lebih baku untuk proses klasifikasi kata. Pada proses ini disebut dengan tahapan *preprocessing*. Setelah melalui tahapan *preprocessing, d*ata yang berupa teks diubah menjadi bentuk angka melalui perhitungan TF (*Term Frequency*). Nilai dari TF *(term frequency)* menjadi data masukan metode algoritma *Naive Bayes.* Metode algoritma *Naive Bayes* nantinya akan menghasilkan klasifikasi untuk data yang dimasukkan pada penelitian ini.

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 263-272
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

2.3.1. Tahapan *Preprocessing*

Tahapan *preprocessing* merupakan pembersihan data yang masih terdapat karakter, *url link, hastag*, dan kalimat yang tidak diperlukan dalam pengolahan data. nantinya kata yang dihasilkan lebih menjadi kalimat baku dan *clean* sehingga dapat diklasifikasikan. Berikut tahapan *preprocessing* yaitu :

a) *Cleaning*

Pada tahap *cleaning* proses yang dilakukan adalah pembersihan kata dari simbol karakter, *url link, username* pengguna, dan kata-kata yang tidak diperlukan dalam pengolahan data sehingga memudahkan dalam proses pengolahan data nantinya.

b) *Casefolding*

Tahap *casefolding* proses perubahan huruf besar menjadi huruf kecil. Dengan proses ini akan memudahkan proses analis sentimen pada kata yang dihasilkan. Sehingga dapat memudahkan dalam pengelohan data yang dihasilkan ke dalam sistem.

c) Normalisasi Kata

Pada tahap normaslisasi kata merupakan mengubah data *tweet* yang belum sesuai dengan kaidah menurut kamus besar bahasa Indonesia (KBBI) sehingga menjadi kalimat yang baku. Tentunya data yang didapatkan hasil *crawling* masih ada kalimat penyingkatan atau tidak sesuai kaidah KBBI. Maka tahapan ini akan melakukan proses pengubahan kalimat menjadi baku untuk mempermudah sistem mengenali atau memproses kalimat tersebut.

d) *Stopword*

Pada tahap *stopword* menghilangkan kata yang tidak dibutuhkan dalam proses pengolahan data yang tidak mempunyai arti atau makna dalam kata tersebut. Proses ini disesuaikan sesuai dengan data yang tersimpan di dalam *database* yang diambil dari *stopword list* Tala [6].

e) Stemming

Pada tahap *stemming* melakukan proses menghilangkan semua kata yang memiliki imbuhan sehingga menjadi kata dasar yang sesuai dengan struktur kaidah KBBI. Penelitian ini menggunakan *referensi stemming* Sastrawi yang berbasis metode algoritma Nazief dan Andriani [7]. Dalam penelitiannya Asian, dkk [8] melakukan pengembangan terkait metode algoritma Nazief & Adriani yaitu :

1) Tersedianya data kamus kata lebih lengkap dan akurat.
2) Tersedianya penambahan aturan-aturan mengenai kata-kata yang bersifat majemuk perulangan.
3) Tersedianya penambahan aturan awalan dan akhiran, serta aturan lainnya, yaitu:

Tabel 1. Hasil Preprocessing

No	*Preprocessing*	Hasil
1	*Tweet* dasar	RESMI: PPKM akan diperpanjang, namun kabar baiknya untuk wilayah Jawa-Bali akan menerapkan PPKM level 3. Ya bisa lah turun minggu depan ke level 2, dst.. #univlox #univloxnews #PPKM #PpkmDiperpanjang
2	*Cleaning*	RESMI: PPKM akan diperpanjang, namun kabar baiknya untuk wilayah Jawa-Bali akan menerapkan PPKM level 3. Ya bisa lah turun minggu depan ke level 2, dst..

No	*Preprocessing*	Hasil
3	*Casefolding*	resmi: ppkm akan diperpanjang, namun kabar baiknya untuk wilayah jawa-bali akan menerapkan ppkm level 3. ya bisa lah turun minggu depan ke level 2, dst..
4	Normalisasi	ppkm akan diperpanjang, namun kabar baiknya untuk wilayah jawa-bali akan menerapkan ppkm level 3. bisa turun minggu depan ke level 2
5	*Stopword*	ppkm diperpanjang, kabar baiknya untuk wilayah jawa-bali menerapkan ppkm level 3. turun minggu depan ke level 2
6	*Stemming*	ppkm diperpanjang untuk wilayah jawa-bali menerapkan ppkm level 3 minggu depan level 2

2.3.2. Tahapan *Term Frequency*

Perhitungan *(TF)* melakukan proses penghitungan kemunculan pada setiap kata dalam sebuah dokumen yang melalui proses *preprocessing*. Hasil perhitungan *TF* akan disimpan pada *database* sebagai proses pengklasifikasian dengan menggunakan algoritma metode *Naive Bayes.*

2.3.3. Algoritma *Naïve Bayes*

Algoritma *Naive Bayes* dibagi menjadi dua proses, yaitu proses data *training* dan data *testing*. Pada proses training model analisis sentimen digunakan sebagai acuan dalam klasifikasi data *testing*. Proses datas *testing* membandingkan kata-kata pada dokumen uji dengan nilai probabilitas kata pada setiap kelas dokumen yang telah tersimpan dalam data *training*. contoh algoritma pengklasifikasian analis sentimen menggunakan *Naive Bayes* sebagai berikut:

a) Proses Tahapan *Training*

Pada persamaan (1) menghitung peluang kemunculan dokumen pada tiap kategori positif atau negatif $P(C_j)$.

$$P(C_j) = \frac{|category_j|}{|corpus|} \quad (1)$$

Pada persamaan (2) menghitung sebuah peluang sebuah kata yang masuk ke dalam tiap kategori data *tweet* positif atau data *tweet* negatif $P(W_i|C_j)$.

$$P(W_i|C_j) = \frac{1+ni}{n+|kosakata|} \quad (2)$$

b) Proses Tahapan *Testing*

Pada persamaan (3) menghitung untuk setiap kategori positif atau negatif lalu menentukan kategori dengan nilai tertinggi (V_{MAP}).

$$V_{MAP} = argmax\ P(C_j) \prod P(W_i|C_j) \quad (3)$$

2.3.4. Algoritma *K-Nearest Neighbors*

Algoritma K Nearest Neighbors (KNN) merupakan klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, yang bergantung pada label kategori melekat pada dokumen pelatihan mirip dengan dokumen tes [3]. Algoritma KNN merupakan metode klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat. K-Nearest Neighbors memiliki prinsip sederhana yaitu bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih[9]. MenurutSafitri[10] algoritma KNN memiliki tahapan:

Jurnal Sains Komputer & Informatika (J-SAKTI)
Volume 7 Nomor 1, Maret 2023, pp. 263-272
ISSN: 2548-9771/EISSN: 2549-7200
https://tunasbangsa.ac.id/ejurnal/index.php/jsakti

a) Nilai parameter k diambil dari jumlah tetangga paling dekat.
b) Menentukan nilai bobot pada setiap term dengan menggunakan Term Weighting TF-IDF.
c) Menghitung kemiripan antar dokumen dengan menggunakan cosine similarity (Persamaan 4).

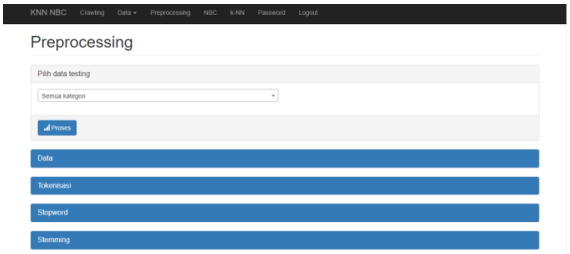
$$\cos(\theta_{ij}) = \frac{\sum_k (d_{ik} d_{jk})}{\sqrt{\sum_k d_{ik}^{\,2}} \sqrt{\sum_k d_{jk}^{\,2}}} \quad (4)$$

d) Mengurutkan hasil nilai perhitungan *cosine similarity* dari nilai besar ke kecil
e) Mengambil parameter K yang paling tinggi kemiripannya dengan dokumen yang diklasifikan

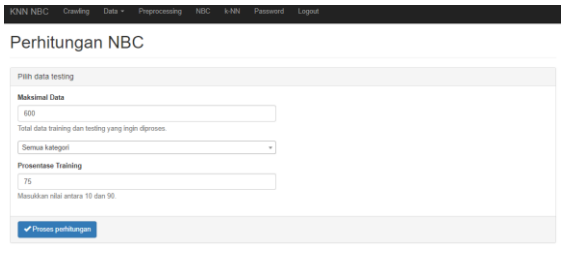
3. HASIL DAN PEMBAHASAN

Pada gambar 2. a), b); c) dan d), merupakan pengembangan aplikasi berbasis aplikasi website. Tool yang digunakan dalam pembuatan aplikasi sentimen ini menggunakan *framework PHP*, untuk desain tampilan menggunakan *boostrap dan Jquery sebagai user interface*. Dalam aplikasi ini terdapat beberapa menu meliputi *crawling data*, *Preprocessing, algoritma NBC, Algoritma K-NN*.

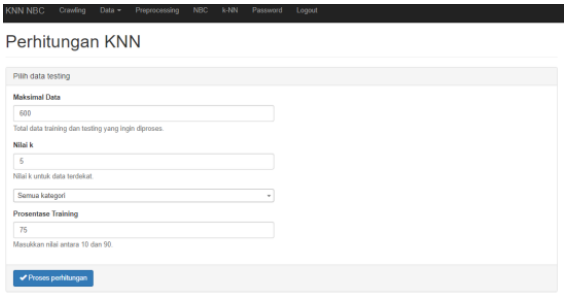
Gambar 2a



Gambar 2b



Gambar 2c



Gambar 2d

Pada pengujian akurasi sistem dilakukan pengujian terhadap hasil klasifikasi pada proses testing. Untuk mengetahui keakuratan algoritma dalam proses klasifikasi maka akan dibandingkan hasil klasifikasi menggunakan algoritma K-Nearest Neighbor. Perbandingan tersebut akan dihitung tingkat keakurasiannya

menggunakan accuracy, precision dan recall. Semakin tinggi nilai accuracy, precision dan recall pada sebuah algoritma yang digunakan maka menunjukkan bahwa algoritma tersebut berjalan dengan baik dan cocok untuk proses klasifikasi. Pengujian dilakukan dengan memilih data training dan data testing secara acak dari 570 data yang akan digunakan. Persamaan (5) merupakan rumus untuk pengujian akurasi:

$$akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (5)$$

Persamaan (6) merupakan rumus untuk pengujian *precision*:

$$precision = \frac{TP}{TP+FP} \quad (6)$$

Persamaan (7) merupakan rumus untuk pengujian *recall*:

$$recall = \frac{TP}{TP+FN} \quad (7)$$

Pada tahapan ini data akan di split menjadi 3 kali percobaan yaitu 70:30, 80:20 dan 60:40

3.1. Algoritma Naïve Bayes

Tahap pertama pada tabel 2 dilakukan percobaan 70 :30 didapatkan hasil sebagai berikut
total data testing : 141
Total prediksi yang benar : 97
Akurasi : 68,79%

Tabel 2. Perbandingan Hasil Confusion Matrix Naive Bayes 70:30

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	93	41	6	1	0,702	0,694	0,989
Negatif	1	3	115	22	0,823	0,25	0,043
Netral	3	0	117	21	0,851	1	0,125

Tahap kedua pada tabel 3 dilakukan percobaan 80:20 didapatkan hasil sebagai berikut
total data testing : 94
Total prediksi yang benar : 71
Akurasi : 75,53%

Tabel 3. Perbandingan Hasil Confusion Matrix Naive Bayes 80:20

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	93	41	6	1	0,702	0,694	0,989
Negatif	1	3	115	22	0,823	0,25	0,043
Netral	3	0	117	21	0,851	1	0,125

Tahap ketiga pada tabel 4 dilakukan percobaan 60 :40 didapatkan hasil sebagai berikut
total data testing : 188
Total prediksi yang benar : 137
Akurasi : 72,87%

Tabel 4. Perbandingan Hasil Confusion Matrix Naive Bayes 60:4 0

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	128	50	9	1	0,729	0,719	0,992
Negatif	5	0	154	29	0,846	1	0,147
Netral	4	1	162	21	0,883	0,8	0,16

3.2. Algoritma K-Nearest Neighbor

Tahap pertama pada tabel 5 dilakukan percobaan nilai K:3 dan perbandingan data training testing 70 :30 didapatkan hasil sebagai berikut
total data testing : 138
Total prediksi yang benar : 63
Akurasi : 45,65%

Tabel 5. Perbandingan Hasil Confusion Matrix KNN 70:30

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	56	27	18	37	0,536	0,675	0,602
Negatif	7	41	72	18	0,572	0,146	0,28
Netral	0	7	111	20	0,804	0	0

Tahap kedua pada tabel 6 dilakukan percobaan nilai K:5 dan perbandingan data training testing 80 :20 didapatkan hasil sebagai berikut
total data testing : 93
Total prediksi yang benar : 38
Akurasi : 40,86%

Tabel 6. Perbandingan Hasil Confusion Matrix KNN 80:20

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	34	15	10	34	0,473	0,694	0,5
Negatif	2	29	72	18	0,572	0,146	0,28
Netral	0	7	111	20	0,804	0	0

Tahap ketiga pada tabel 7 dilakukan percobaan nilai K:7 dan perbandingan data training testing 60 :40 didapatkan hasil sebagai berikut
total data testing : 187
Total prediksi yang benar : 91
Akurasi : 48,66%

Tabel 7. Perbandingan Hasil Confusion Matrix KNN 60:40

Klasifikasi	TP	FP	TN	FN	Accuracy	Precision	Recall
Positif	85	56	9	37	0,503	0,603	0,697
Negatif	1	32	129	25	0,695	0,03	0,038
Netral	5	8	140	34	0,775	0,385	0,128

4. SIMPULAN

Pada hasil penelitian yang telah dibuat terdapat beberapa kesimpulan bajwa metode Algoritma *naive bayes* dan Algoritma K-Nearest Neighbor yang dilakukan dalam

penelitian digunakan sebagai algoritma mengklasifikasikan data *tweet* positif atau data *tweet* negative terkait opini masyakat terhadap pembukaan destinasi pariwisata di tengah masa pandemi covid 19. Tahapan preprocessing sudah dilakukan seperti tahap *cleaning, casefolding,* normalisasi data, *stopword,* dan *stemming* berjalan dengan baik. Hasil dari proses tahap pengujian nanti dapat dilakukan untuk menentukan hasil klasifikasi menggunakan algoritma *NB* dan algoritma KNN. Algoritma *NB* digunakan untuk mengklasifikasikan tweet kedalam positif atau negatif terutama *tweet* mengenai pembukaan destinasi wisata di tengah pandemi *nilai tingkat akurasi Naive Bayes tertinggi sebesar* 75,53*%, precision positif tertinggi yaitu 71%, precision negatif tertinggi yaitu 25%, recall positif tertinggi yaitu 99% dan recall negatif tertinggi yaitu 14%. Sedangkan untuk algoritma KNN didapatkan nilai tingkat akurasi tertinggi sebesar 48,66%, precision positif tertinggi yaitu 69%, precision negatif tertinggi yaitu 14%, recall positif tertinggi yaitu 69% dan recall negatif tertinggi yaitu 28%.*

Sistem dapat dikembangkan dengan menggunakan metode untuk perbaikan kata tidak baku sehingga mendapatkan hasil akurasi yang lebih baik. Sistem dapat dikembangkan untuk *crawling* data menggunakan platform sosial media lain : Instagram, facebook dll

DAFTAR PUSTAKA

[1] Satuan Tugas Penanganan., 2013. "Data Covid Indonesia." https://covid19.go.id/.
[2] Mochmamad Haldi Widianto, "Algoritma Naïve Bayes" https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/.
[3] R. P. Fitrianti, A. Kurniawati, and D. Agusten, "Analisis Sentimen Terhadap Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma K-Nearest Neighbor," Semin. Nas. Apl. Teknol. Inf., pp. 27–32,2019.
[4] Efendi, Rizal Maulana Yusuf dkk. "Sistem pakar diagnosa kerusakan VGA dengan metode *certainty factor* dan algoritma *K-Nearest Neighbor (KNN)*".2021
[5] Efendi, Rizal Maulana Yusuf dkk. "Visualisasi dan analisa data penyebaran covid-19 dengan metode klasifikasi naïve bayes".2021
[6] F. Z. Tala, "A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia," *Institute for Logic, Language and Computation Universeit Van Amsterdam,* 2003
[7] B. Nazief and M. Adriani, "Stemming Indonesian: A confix-stripping approach'," *ACM Transactions on Asian Language Information Processing,* 2007
[8] J. Asian, H. E. Williams and S. M. M. Tahaghohi, "Stemming Indonesian," in *In Conferences in Research and Practice in Information Technology Series*, 2005.
[9] R.Sari, "Analisis Sentimen pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K Nearest Neighbor (K-Nn)," EVOLUSI J. Sains dan Manaj.,Vol. 8,No. 1,pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371
[10] R. N. Safitri, "Analisis Sentimen Review Pelanggan Hotel Menggunakan Metode K-Nearest Neighbor (K-NN) (Studi Kasus: Hotels.com, Booking.com, Agoda.com),"Tugas Akhir Univ. Din.,Vol. 21,No. 1, pp. 1–9, 2020, [Online]. Available: http://repositorio.unan.edu.ni/2986/1/5624.pdf.