Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037

SENTIMENT ANALYSIS OF COMMENTS ON TOURIST ATTRACTIONS IN LAMPUNG PROVINCE USING THE NAIVE BAYES METHOD

Damayanti[1], Muhammad Arif Wirahudha[*2], Dyah Ayu Megawaty[3]

[1,2,3] Information System, Faculty of Engineering and Computer Science, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: muhammad_arif_wirayudha@teknokrat.ac.id
Received : Oct 18, 2024; Revised : Feb 25, 2025; Accepted : Mei 07, 2025; Published : May 17, 2025

Abstract

Lampung Province is a province that has so much natural beauty, this also makes Lampung Province one of the tourist destinations that are visited by many domestic and foreign tourists so that there is a problem, namely the many negative comments that are not in accordance with reality affect the number of tourist visits to Lampung Province because they are not in accordance with reality so that they affect public opinion about tourism in Lampung Province which results in tourist attractions being deserted. The method used to analyze sentiment analysis is the naive bayes algorithm by crawling data using python. The stages of the naive bayes algorithm in the study using preprocessing consist of five processes, namely cleansing, tokenization, case folding, stopword removal, and stemming. Lampung Province tourist attractions are Pahawang, Way Kambas, Krui Beach / West Coast, Mutun Beach and Kiluan Bay. The results of a fairly high level of accuracy in positive comments on Pahawang Beach. In this study, it was concluded that the impact of comments can affect the number of visitors coming to tourist attractions.

Keywords : *Naive Bayes, Opinion, Sentiment Analysis, Tourism, Travel, Tweet.*

1. PENDAHULUAN

Saat ini berwisata atau travelling bagi sebagian orang sudah termasuk dalam kebutuhan hidup yang harus dipenuhi, tujuan dari wisata itu sendiri adalah untuk mencari hiburan dan keluar dari kepenatan aktivitas. Berwisata merupakan kegiatan yang tidak bisa dipisahkan dari kehidupan masyarakat. Tujuan berwisata yaitu untuk melepaskan diri dari kejenuhan, dan mencari situasi baru sehingga dapat menenagkan pikiran dari aktivitas sehari-hari.

Provinsi Lampung merupakan sebuah Provinsi yang memiliki banyak keindahan alam, sehingga Provinsi Lampung menjadi salah satu tujuan wisata yang banyak dikunjungi wisatawan Nusantara maupun Mancanegara. Dilihat dari perkembangan wisatawan dari tahun ke tahun, Kunjungan wisatawan ke Provinsi Lampung selalu mengalami peningkatan kepopuleran dalam daya Tarik alamnya. Pada tahun 2017 kunjungan wisatawan mencapai 11,39 juta wisatawan nusantara dan 245 ribu wisatawan mancanegara, ditahun 2018 kembali mengalami peningkatan yaitu 13,93 juta wisatawan nusantara dan 274 ribu mancanegara [1].

Objek wisata Provinsi Lampung merupakan tempat yang menyenangkan sehingga selalu di *review* ataupun opini oleh para pengguna Media Sosial yang dibagikan melalui akun jejaring *online* seperti *Twitter, Instagram,* maupun *Facebook.* Opini merupakan penilaian ataupun pendapat pribadi

seseorang untuk menjelaskan tentang sesuatu hal yang mereka alami atau ketahui. Opini sendiri seringkali dijadikan sebuah acuan wisatawan lain untuk berkunjung ke suatu tempat wisata yang dapat di lakukan analisis sentimen. Berdasarkan hasil observasi yang dilakuka terdapat masalah yaitu banyaknya komentar negatif yang tidak sesuai dengan kenyataan mempengaruhi jumlah kunjungan wisatawan ke Provinsi Lampung dikarenakan tidak sesuai dengan kenyataan sehingga berpengaruh terhadap opini msyarakat mengenai wisata yang ada di Provisi lampung yang mengakibatkan tempat wisata sepi pengunjung.

Sentiment analysis dikenal sebagai opinion mining, merupakan sebuah area penelitian yang menganalisis opini publik, emosi, penilaian, sikap dan sentimen tentang suatu, objek seperti, produk, layanan, individu, peristiwa, masalah dan topik. Untuk melakukan *sentiment analysis* diperlukan sebuah metode, salah satuya *Naive Bayes*. Dalam metode ini klasifikasi dilakukan dengan menghitung probabilitas. *Naive Bayes* memiliki keunggulan dalam efisiensi dan kesederhanaan pada pengklasifikasian teks terutama dalam penerapan pada aplikasi praktis secara langsung seperti membagi kategori berita atau menyaring spam. Metode klasifikasi yang digunakan dalam penelitian ini adalah *Naive Bayes Classifier*.

Algoritma *Naive Bayes* adalah sebuah pengklasifikasian probabilitas sederhana yang dapat menghitung sekumpulan peluang atau kemungkinan kejadian dengan menjumlahkan frekuensi atau kombinasi nilai data yang ada [2]-[3]-[4]. Keuntungan menggunakan algoritma *naive bayes* karena algotirma sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

Beberapa penelitian terkait yang telah dilakukan diantaranya penelitian Sihaloho (2020) dengan judul Analisis Sentimen Objek Wisata Danau Toba berdasarkan Ulasan Pengunjung menggunakan Algoritma Support Vector Machine. Hasil penelitian [5]. Sebelumnya analisis sentimen pada objek wisata menggunakan Naive Bayes *Classifier* telah diteliti oleh Subarkah et al (2022) dengan judul Analisis Sentimen Review Tempat Wisata Pada Data *Online* Travel Agency Di Yogyakarta. Hasil yang diperoleh dari penelitian ini menghasilkan nilai analisis sentimen dari klasifikasi ini menghasilkan nilai accuracy sebesar 92,84%, dengan weighted average recall 93%, precision 92%, dan F1-Score 93%. Penelitian selanjutnya penelitian Ipmawati et al (2024) meneliti tentang Analisis Sentimen Tempat Wisata Berdasarkan ulasan pada *Google Maps*. Hasil penelitian ini adalah tingkat akurasi rata-rata sebesar 83,8% berdasarkan ulasan pengunjung di situs *Google Maps* [8]. Kemudian pada penelitian Syahlan et al (2023) dengan judul Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode *Support Vector Machine* (SVM) Studi Kasus: Taman Air Mancur Sri Baduga Purwakarta. Hasil penelitian tersebtu menunjukkan hasil akurasi sebesar 81%, nilai presisi 94%, dan recall 99%. Nilai ini menunjukan bahwa klasifikasi algoritma *Support Vector machine* dinilai cukup baik [9]. Kemudian pada penelitian Pati et al (2020) dengan judul Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K- Nearest Neighbo. Hasil penelitian menunjukkan bahwa penggunaan metode K-Nearest Neighbor diperoleh tingkat akuras sebesar 76.53% sedangkan metode Naive Bayes Classifie rsebesar 73.47% [10].

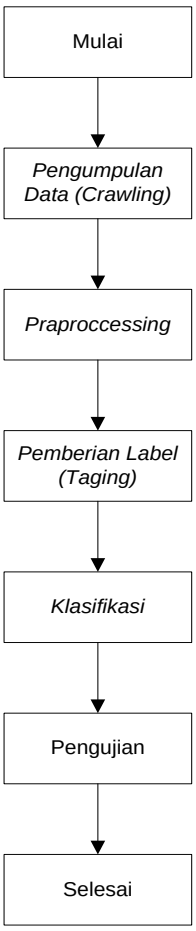
Berdasarkan penjelasan yang telah diberikan akan dilakukan penelitian tentang *media social sentiment analysis* untuk mengklasifikasikan opini netizen terhadap objek wisata di Provinsi Lampung. Penelitian ini masuk kedalam *fined grained sentiment analysis* yaitu analisis pada kalimat. Penelitian ini bertujuan untuk menganalisis sentimen komentar terhadap objek wisata di Provinsi Lampung menggunakan algoritma *Naive Bayes*. Hasil klasifikasi analisis sentimen yaitu sebuah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan pengguna untuk menentukan apakah sepotong tulisan itu bermakna positif, negatif atau netral.

2. METODE PENELITIAN

2.1. Tahapan Penelitian

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037

Tahapan penelitian merupakan alur dalam penelitian yang dilakukan secara berurutan, berikut ini gambaran tahapan penelitian dapat dilihat pada Gambar 1:



Gambar 1. Tahapan Penelitian

2.2. *Naive Bayes Classifier*

Salah satu tugas Data Mining adalah klasifikasi data, yaitu memgklasifikasikan (memetakan) data ke dalam satu atau beberapa kelas yaang sudah didefinisikan sebelumnya. *Naive Bayes Classifier (NBC)* adalah salah satu metode dalam klasifikasi data. *Naive Bayes Classifier* merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [11]-[7]-[12]. Dasar dari Naive Bayes yang dipakai dalam pemrograman dengan rumus:

$$P(A|B) = (P(B|A) * P(A))/P(B) \quad (1)$$

Keterangan :
Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi:

$$P(Ci|D) = (P(D|Ci)*P(Ci)) / P(D) \quad (2)$$

Naive Bayes Classifier atau bisa disebut sebagai *Multinomial Naive Bayes* merupakan model penyederhanaan dari Metode klasifikasi yang berakar pada Teorema bayes. Teorema Bayes sendiri dikenal dengan memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya.

Penyederhanaan Teorema bayes ke *naive bayes* didasarkan pada asumsi naïve yaitu asumsi yang sangat kuat akan independensi dari masing-masing kondisi/kejadian.
Dibawah ini adalah persamaan model penyederhanaan dari Metode Bayes adalah:

VMAP = arg max P(Vj | a1 , a2 ,.....an) (3)

2.3. Cara Kerja Analisis *Naive Bayes Classifier*

Metode yang digunakan dalam penelitian ini adalah metode *Naive Bayes Classifier* dimana data yang sudah didapatkan akan diklasifikasikan menjadi tiga kategori kelas yaitu positif, negatif dan netral. Tahapan awal untuk menganalisis data adalah sebagai berikut:

2.3.1. *Crawling* Data

Data yang digunakan dalam penelitian ini yaitu data sekunder. Dataset berisi teks berbahasa Indonesia yang diperoleh dari sosial media *twitter*, data yang didapat adalah seluruh opini masyarakat tentang komentar objek wisata. Pengumpulan data dilakukan pada rentang periode 2020-2021 dengan jumlah 4160 komentar. Dengan menggunakan bahasa pemograman *python* dan aplikasi *jupiter note book* dengan kata kunci "pahawang" dengan jenis sentiment positif, negatif, dan netral. Untuk proses penyaringan data komentar masyarakat akan disimpan menggunakan format file CSV.

2.3.2. Data *Training* dan Data *Testing*

Data yang digunakan pada penelitian ini terdiri dari dua jenis data yaitu data latih (*training*) dan data uji (*testing*). Data latih (*taining*) diambil dari kumpulan data *twitter* yang telah dilabeli sesuai sentimennya secara manual. Pelabelan tersebut dibagi kedalam sentimen positif, negatif dan netral. Sedangkan data uji (*testing*) yang digunakan yaitu kumpulan data *twitter* yang belum memiliki label.

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Metode *Naive Bayes*

Pada tahap klasifikasi ini menggunakan metode *Naive Bayes* yaitu proses *training*. Pada tahap ini dilakukan proses *training* terlebih dahulu untuk pelatiha. Tabel 1 menampilkan contoh perhitungan manual *Naive Bayes Classification* dengan sampel 6 *tweet* data training.

Tabel 1. Kasus Data Training

Tweet	Fitur
Tweet1	@ roseanne2lc pahawang kayaknya tetkenal banget gue sering dengar
Tweet2	@ keyoËŠ pulau pahawang cantikkk bgt
Tweet3	@ Jihyo krakatau tapi gak bagus.
Tweet4	@ amrin Pernah ke pahawang kecil kayanya sih pas dulu
Tweet5	@anyas Wah keren pahawang tapi jauh ya
Tweet6	@rika Pahawang island keren banget

Pada *pre-processing* ada beberapa tahapan yang harus dilakukan, tahapan tersebut sebagai berikut:

A. Tahapan *Cleaning*

Tahap *cleaning* dimana karakter selain huruf dihilangkan dan dianggap delimiter dan menghapus juga URL, mention dan hastag [13]. Tabel 2 menampilkan hasil proses *cleaning*.

Tabel 2. Hasil Proses Cleaning

Tweet	Hasil Cleaning
@ roseanne2lc pahawang kayaknya tetkenal banget gue sering denger	pahawang kayaknya tetkenal banget gue sering dengar
@ keyoËŠ pulau pahawang cantikkk bgt	pulau pahawang cantikkk bgt
@ Jihyo krakatau tapi gak bagus.	krakatau tapi gak bagus.

Tweet	Hasil Cleaning
@ amrin Pernah ke pahawang kecil kayanya sih pas dulu	Pernah ke pahawang kecil kayanya sih pas dulu
@anyas Wah keren pahawang tapi jauh ya	Wah keren pahawang tapi jauh ya
@rika Pahawang island keren banget	Pahawang island keren banget

B. *Case Folding*

Melakukan perubahan semua huruf dalam dokumen menjadi huruf kecil [13]. Tabel 2 menampilkan contoh *case folding* dari data *tweet*.

Tabel 3. Hasil Case Folding

Hasil Cleaning	Hasil Case Folding
pahawang kayaknya tetkenal banget gue sering denger	pahawang kayaknya tetkenal banget gue sering dengar
pulau pahawang cantikkk bgt	pulau pahawang cantikkk bgt
krakatau tapi gak bagus.	krakatau tapi gak bagus.
Pernah ke pahawang kecil kayanya sih pas dulu	pernah ke pahawang kecil kayanya sih pas dulu
Wah keren pahawang tapi jauh ya	wah keren pahawang tapi jauh ya
Pahawang island keren banget	pahawang island keren banget

C. Tahapan Transformasi

Transformasi merupakan tahapan untuk mengubah kata yang kurang baku atau kata gaul menjadi kata baku yang sesuai dengan KBBI. Penentuan kata baku dilakukan berdasarkan kamus bahasan gaul. http://adhitezt12.blogspot.com/2012/12/kamus-bahasa-alay-lengkap.html dan https://indowonders.com/kamus-bahasa-gaul/. Tabel 4 menampilkan hasil transformasi data.

Tabel 4. Hasil Transformasi Data

Hasil Case Folding	Hasil Transformasi
pahawang kayaknya tetkenal banget gue sering denger	pahawang sepertinya terkenal banget gue sering dengar
pulau pahawang cantikkk bgt	pulau pahawang cantik banget
krakatau tapi gak bagus.	krakatau tapi tidak bagus.
pernah ke pahawang kecil kayanya sih pas dulu	pernah ke pahawang kecil sepertinya sih pas dulu
wah keren pahawang tapi jauh ya	wah keren pahawang tapi jauh ya
pahawang island keren banget	pulau pahawang keren banget

D. Tahap *Tokenizing*

Tahapan tokenizing merupakan tahapan pemotongan string input berdasarkan tiap kata yang menyusunnya. Tabel 5 menampilkan hasil dari proses tokenizing.

Tabel 5. Hasil Proses Tokenizing

Hasil Transformasi	Hasil Tokenizing
pahawang sepertinya terkenal banget gue sering dengar	pahawang | sepertinya | terkenal banget | gue | sering | dengar |
pulau pahawang cantik banget	pulau | pahawang | cantik | banget |
krakatau tapi tidak bagus.	krakatau | tapi | tidak | bagus |

pernah ke pahawang kecil sepertinya sih pas dulu	pernah | ke | pahawang | kecil | seperti | sih | pas | dulu |
wah keren pahawang tapi jauh ya	wah | keren | pahawang | tapi | jauh | ya|
pulau pahawang keren banget	pulau | pahawang | keren | banget |

E. Tahapan *Stemming*

Tahapan menghapuskan atau menghilangkan kata imbuhan baik awalan maupun akhiran [14] bertujuan agar kalimat menjadi baku dan menjadi satu kata sama didalam data set. Tabel 6 menampilkan hasil dari proses stemming.

Tabel 6. Hasil Proses Stemming

Hasil Tokenizing	Hasil Stemming	Komentar
pahawang sepertinya terkenal banget gue sering dengar	pahawang | seperti | kenal | banget | gue | sering | dengar|	Positif
pulau pahawang cantik banget	pulau | pahawang | cantik | banget	Positif
krakatau tapi tidak bagus.	krakatau | tapi | tidak | bagus |	Negatif
pernah ke pahawang kecil sepertinya sih pas dulu	pernah | pahawang | kecil | seperti | sih | pas |dulu|	Netral
wah keren pahawang tapi jauh ya	wah | keren | pahawang | tapi jauh | ya |	Netral
Pulau pahawang keren banget	pulau | pahawang | keren | banget |	Positif

3.1.1. Perhitungan Metode *Naive Bayes*

Dari data table 6 dibuatlah sebuah model *prior probability* yang mengacu pada persamaan, jumlah nilai positif, netral dan negatif sisapat dari Tabel 1 berikut ini adalah hasil yang didapat yaitu:

$$\boldsymbol{Pior\ Probability} = \frac{\boldsymbol{P(A)}}{\boldsymbol{P(B)}} \quad (4)$$

Keterangan :
P (A) : Jumlah Komentar (negative, positif dan netral)
P (B) : Total Komentar
Contoh perhitungan :
Positif $= \frac{3}{6}$
Tabel 7 menampilkan hasil yang didapat dalam perhitungan *pior probability*.

Tabel 7. Hasil Perhitungan *Prior Probability*

Atribut Kelas	P (Class)
Positif	3/6
Negatif	1/6
Netral	2/6 atau 1/3

Kemudian dilakukan perhitungan *conditional probability* masing-masing kelas dengan menggunakan rumus berikut:

$$\boldsymbol{conditional\ probability} = \frac{\boldsymbol{P(A)}}{\boldsymbol{P(B)}} \quad (5)$$

Keterangan :
P (A) : Jumlah Kata dalam Komentar (negative, positif dan netral)

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037

P (B) : Total Kata dalam Komentar
Contoh perhitungan :

$$\text{Pahawang (positif)} = \frac{3}{3}$$

$$\text{Pahawang (negatif)} = \frac{0}{1}$$

$$\text{Pahawang (netral)} = \frac{2}{2}$$

Tabel 8 menampilkan hasil yang didapat dalam perhitungan *conditional probability*.

Tabel 8. Perhitungan Conditional Probability

Trem	**Kategori**		
	Positif	**Negatif**	**Netral**
Pahawang	$\frac{3}{3}$	$\frac{0}{1}$	$\frac{2}{2}$
Seperti	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Kenal	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Banget	$\frac{3}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Gue	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Sering	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Dengar	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Pulau	$\frac{2}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Cantik	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{0}{2}$
Krakatau	$\frac{0}{3}$	$\frac{1}{1}$	$\frac{0}{2}$
Tidak	$\frac{0}{3}$	$\frac{1}{1}$	$\frac{0}{2}$
Bagus	$\frac{0}{3}$	$\frac{1}{1}$	$\frac{0}{2}$
Pernah	$\frac{0}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Kecil	$\frac{0}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Dulu	$\frac{0}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Keren	$\frac{1}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Jauh	$\frac{0}{3}$	$\frac{0}{1}$	$\frac{1}{2}$
Pulau	$\frac{2}{3}$	$\frac{0}{1}$	$\frac{0}{2}$

Dalam proses perhitungan pada Tabel 7 kemungkinan dokumen terdapat sebuah dokumen yang tidak muncul dalam suatu kelas yang sangat besar maka akan bernilai 0. Untuk itu dibutuhakan sebuah teknik *laplace smoothing*, teknik ini yang menambahkan angka 1 untuk setiap tweet yang ditemukan. Dengan penambahan angka 1 tersebut angka 0 akan dihilangkan dan hasil perhitungan tidak rusak dan sesuai. Perhitungan teknik *laplace smooting* dapat dihitung dengan rumus dibawah ini :

$$\boldsymbol{laplace\ smooting} = \frac{\boldsymbol{P(A)}}{\boldsymbol{P(B)}} \quad (6)$$

Keterangan :
P (A) : Jumlah Kata dalam Komentar (negative, positif dan netral)
P (B) : Total Kata dalam Komentar

Tabel 9 menampilkan hasil dari perhitungan menggunakan Teknik laplace smooting

Tabel 9 Hasil Perhitungan Teknik Laplace Smoothing

Trem	**Kategori**		
	Positif	**Negatif**	**Netral**
Pahawang	$\frac{4}{9}$	$\frac{1}{7}$	$\frac{3}{8}$
Seperti	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{2}{8}$
Kenal	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Banget	$\frac{4}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Gue	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Sering	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Dengar	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Pulau	$\frac{3}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Cantik	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{1}{8}$
Krakatau	$\frac{1}{9}$	$\frac{2}{7}$	$\frac{1}{8}$
Tidak	$\frac{1}{9}$	$\frac{2}{7}$	$\frac{1}{8}$
Bagus	$\frac{1}{9}$	$\frac{2}{7}$	$\frac{1}{8}$
Pernah	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{2}{8}$

Trem	Kategori		
	Positif	**Negatif**	**Netral**
Kecil	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{2}{8}$
Dulu	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{2}{8}$
Keren	$\frac{2}{9}$	$\frac{1}{7}$	$\frac{2}{8}$
Jauh	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{2}{8}$
Pulau	$\frac{3}{9}$	$\frac{1}{7}$	$\frac{1}{8}$

Hasil perhitungan probabilitas tersebut digunakan sebagai model probabilitistik yang selanjutnya digunakan sebagai data acuan untuk menentukan data *testing*.

3.1.1. Proses *Testing*

Pada *pre-processing* ada beberapa tahapan yang harus dilakukan, tahapan testing tersebut sebagai berikut:

A. Tahapan *Cleaning*

Tahap dimana karakter selain huruf dihilangkan dan dianggap delimiter dan menghapus URL, mention dan hastag. Tabel 10 menampilkan hasil cleaning data testing.

Tabel 10. *Cleaning* Data Testing

Tweet	**Hasil *Cleaning***
@ yulian zara Pahawang terkenal cantik banget sering pernah dengar	pahawang terkenal cantik banget sering pernah dengar
@triptwigs liburan Krakatau pas dulu keren tapi jauh	liburan krakatau pas dulu keren tapi jauh

B. *Case Folding*

Melakukan perubahan semua huruf dalam dokumen menjadi huruf kecil[15]. Tabel 11 menampilkan contoh hasil *case folding* data testing.

Tabel 11. Case Folding Data Testing

Hasil *Cleaning*	**Hasil *Case Folding***
Pahawang terkenal cantik banget sering pernah dengar	pahawang terkenal cantik banget sering pernah dengar
liburan Krakatau pas dulu keren tapi jauh	liburan krakatau pas dulu keren tapi jauh

C. Tahapan Transformasi

Transformasi merupakan tahapan untuk mengubah kata yang kurang baku atau kata gaul menjadi kata baku yang sesuai dengan KBBI. Penentuan kata baku dilakukan berdasarkan kamus bahasan gaul.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037

D. Tahap *Tokenizing*

Tahap tokenizing merupakan tahapan pemotongan string input berdasarkan tiap kata yang disususun [16]. Tabel 12 menampilkan hasil dari proses tokenizing data testing.

Tabel 12. Tokenizing Data Testing

Hasil Transformasi	Hasil *Tokenizing*
pahawang terkenal cantik banget sering pernah dengar	pahawang | terkenal | cantik | banget | sering | pernah | dengar
liburan krakatau saat dulu keren tapi jauh	liburan | krakatau | saat | dulu keren | tapi | jauh

E. Tahapan *Stemming*

Tahapan menghapuskan atau menghilangkan kata imbuhan baik awalan maupun akhiran bertujuan agar kalimat menjadi baku dan menjadi satu kata sama didalam data set [16]. Tabel 13 menampilkan hasil proses stemming.

Tabel 13. Stemming

Hasil *Tokenizing*	Hasil *Stemming*
pahawang terkenal cantik banget sering pernah dengar	pahawang | kenal | cantik | banget | sering | pernah | dengar
liburan krakatau saat dulu keren tapi jauh	liburan | krakatau | saat | dulu | keren | tapi | jauh

Proses *testing* dihitung probabilitasnya dan dicari probabilitas tertinggi menggunakan persamaan sebagai berikut:

P(*Tweet*7|positif) $= P(t_{pahawang}|positif) * P(t_{kenal}|positif) * P(t_{cantik}|positif) * P(t_{banget}|positif) * P(t_{pernah}|positif) * P(t_{dengar}|positif) * P(positif)$ (7)

$= \frac{4}{9} * \frac{2}{9} * \frac{2}{9} * \frac{4}{9} * \frac{1}{9} * \frac{2}{9} * \frac{3}{6}$

$= 0.0001133799$

P(*Tweet*7|negatif) $= P(t_{pahawang}|negatif) * P(t_{kenal}|negatif) * P(t_{cantik}|negatif) * P(t_{banget}|negatif) * P(t_{pernah}|negatif) * P(t_{dengar}|negatif) * P(negatif)$ (8)

$= \frac{1}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{6}$

$= 0.0000012047$

P(*Tweet*7|netral) $= P(t_{pahawang}|netral) * P(t_{kenal}|netral) * P(t_{cantik}|netral) * P(t_{banget}|netral) * P(t_{pernah}|netral) * P(t_{dengar}|netral) * P(netral)$ (9)

$= \frac{3}{8} * \frac{1}{8} * \frac{1}{8} * \frac{1}{8} * \frac{2}{8} * \frac{1}{8} * \frac{2}{6}$

$= 0.0000063297$

P(*Tweet*8|positif) $= P(t_{krakatau}|positif) * P(t_{dulu}|positif) * P(t_{keren}|positif) * P(t_{jauh}|positif) * P(positif)$

$= \frac{1}{9} * \frac{1}{9} * \frac{2}{9} * \frac{1}{9} * \frac{3}{6}$

$= 0.00014641$

P(*Tweet*8|negatif) $= P(t_{krakatau}|negatif * P(t_{dulu}|negatif) * P(t_{keren}|negatif) * P(t_{jauh}|negatif) * P(negatif)$ (10)

$= \frac{2}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{7} * \frac{1}{6}$

$= 0.0001229312$

P(*Tweet*8|netral) $= P(t_{krakatau}|netral) * P(t_{dulu}|netral) * P(t_{keren}|netral) * P(t_{jauh}|netral) * P(netral)$ (11)

$= \frac{1}{8} * \frac{2}{8} * \frac{2}{8} * \frac{2}{8} * \frac{2}{6}$

$= 0.00061875$

Tabel 14 menampilkan hasil perhitungan probabilitas dari setiap data uji, diperoleh hasil.

Tabel 14. Nilai Probabilitas Data *Testing*

Tweet	**Probabilitas**		
	Positif	**Negatif**	**Netral**
*Tweet*7	**0.0001133799**	0.0000012047	0.0000063297
*Tweet*8	0.00014641	**0.0001229312**	0.00061875

Dari hasil analisa data *testing* diatas yaitu *Tweet*7 termasuk kategori sentimen Positif. Sedangkan *Tweet 8* termasuk kategori sentimen Negatif. Tabel 15 menampilkan hasil analisa data *testing* yang dilakukan

Tabel 15. Hasil Analisa

Tweet	**Fitur**	**Kategori**
Tweet 7	pahawang kenal cantik banget pernah dengar	Positif
Tweet 8	krakatau dulu keren jauh	Negatif

Berdasarkan Tabel 15 makan dapat dihitung nilai perangkingan berdasarkan nilai ketentuan yang diberikan oleh pakar Bahasa Indonesia yang menyatakan bahwa kata positif memiliki bobot 8, negatif memiliki bobot 2, dan netral memiliki bobot 0 sehingga pada hasil analisa yang dilakukan perangkingan objek wisata. Table 16 menampilkan hasil pembobotan untuk perankingan objek wisata.

Tabel 16. Hasil Pembobotan

Objek Wisata	**Kategori**	**Bobot**	**Rangking**
Pahawang	Positif	8	1
Krakatau	Negatif	2	2

Berdasarkan Tabel 16 maka rekomendasi tempat objek wisata berdasarkan nilai bobot yang diberikan yang direkomendasikan adalah Pahawang.

3.2. Hasil Pembahasan

3.2.1. *Splitting* Data

Spliting data adalah proses pembagian data set menjadi 2 file data yaitu data training dan data testing [17]. Data training adalah bagian dataset yang kita latih untuk membuat prediksi atau menjalankan fungsi dari sebuah algoritma. sesuai tujuannya masing-masing. Kita memberikan petunjuk melalui algoritma agar mesin yang kita latih bisa mencari korelasinya sendiri. Sedangkan data testing adalah adalah bagian dataset yang kita tes untuk melihat keakuratannya, atau dengan kata lain melihat performanya, dapat dilihat pada Gambar 2:

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037

```
# Splitting data training and data testing dan disimpan ke csv
train, test = train_test_split(df, test_size=0.2, random_state=42)
train.to_csv('Data Training/train.csv', index=False)
test.to_csv('Data Testing/test.csv', index=False)
# splitting data yang sudah ditransform dan divektorisasi
persentase_data_test = 0.8
X_train, X_test, Y_train, Y_test = train_test_split(x, y, test_size=persentase_data_test, random_state=42)
```

Gambar 2. *Source Code Python Splitting* Data

Dari source code diatas dapat dilihat bahwa adanya proses pembagian data dari data "data_uji1.csv" menjadi data train.csv dan data test.csv

3.2.2. *K-fold Cross Validation*

Cross validation adalah suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal [18]. Metode ini sering juga disebut dengan *k-fold cross validation* dimana percobaan sebanyak *k* kali untuk satu model dengan parameter yang sama, dapat dilihat pada Gambar 3:

```
# Menghitung Akurasi Model pendekatan Naive Bayes
# akurasi_model_naive_bayes = roc_auc_score(Y_test, clf.predict_proba(X_test)[:, 1])
# print("Akurasi Model pendekatan Naive Bayes : " + str(akurasi_model_naive_bayes*100) + "%")

# Penghitungan K Fold CV
nilai_k = 10
scores = cross_val_score(clf, x, y, cv=nilai_k)
print("Hasil Peformansi Menggunakan K-FOLD dengan Pendekatan Klasifikasi Naive Bayes")
print("Nilai K : " + str(nilai_k))
print(scores)
print("Sehingga, akurasi rata-rata : " + str((scores.mean())*100) + "%")
```

Gambar 3. *Source Code Python* K-Fold *Cross Validation*

Dalam perhitungan *K-fold* nilai K=10 adalah *10-fold cross validation* yang artinya adalah melakukan percobaan sebanyak 10 kali tahapan. dari 10 kali percobaan akan didapatkan nilai rata-rata dari pengujian tersebut yang akan diimplementasikan dan mendaptkan akurasi rata rata yaitu 97.45088 % untuk 10 kali pengujian nilai *fold*.

3.2.3. Pemodelan *Naive Bayes*

Metode *Naive Bayes Classifiers* yaitu salah satu metode klasifikasi teks berdasarkan probabilitas kata kunci dalam membandingkan dokumen latih dan dokumen uji. Keduanya dibandingkan melalui beberapa tahap persamaan, yang akhirnya diperoleh hasil probabilitas tertinggi yang ditetapkan sebagai kategori dokumen baru [19]-[20]. Dalam pemodelan *Naive Bayes* ini di lakukan dengan beberapa sekenario untuk mendapatkan hasil *performance* terbaik. Dari proses scenario 1 saat proses *preprocessing* dengan *steaming* dan dengan nilai *k-fold* yang di masukan direntang 1-10 dapat dilihat pada Gambar 4:

```
import numpy as np
import pandas as pd
import pandas as pd
import nltk as df
from nltk.tokenize import RegexpTokenizer
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import
TfidfVectorizer
from
Sastrawi.StopWordRemover.StopWordRemoverFactor
y import StopWordRemoverFactory
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import train_test_split
from sklearn import naive_bayes
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import roc_auc_score
df = pd.read_csv('data_uji1.csv', sep=',')
# print(df)

# menghapus value NaN karena operasi predict tidak
menerima value NaN
df = df.dropna()

# Splitting data training and data testing dan
disimpan ke csv
train, test = train_test_split(df, test_size=0.2,
random_state=42)
train.to_csv('Data Training/train.csv', index=False)
test.to_csv('Data Testing/test.csv', index=False)
```

```
# TFIDF Vectorizer, digunakan untuk melakukan
vektorisasi. Digunakan StopWords dari Sastrawi
# Agar bisa melakukan vektorisasi menggunakan
bahasa Indonesia
token = RegexpTokenizer(r'[a-zA-Z0-9]+')
factory = StopWordRemoverFactory()
stop_words = factory.get_stop_words()
cv = TfidfVectorizer(use_idf=True, lowercase=True,
strip_accents='ascii', stop_words=stop_words)

#pembagian untuk kolom data twit sebagai variabel
x, dan dilakukan transform agar bisa digunakan untuk
# klasifikasi naive bayes
y = df['nilai']
x = cv.fit_transform(df['data tele'])

# splitting data yang sudah ditransform dan
divektorisasi
persentase_data_test = 0.8
X_train, X_test, Y_train, Y_test = train_test_split(x, y,
test_size=persentase_data_test, random_state=42)

# Menjalankan proses naive bayes menggunakan
MultinomialNB
clf = naive_bayes.MultinomialNB()
clf.fit(X_train, Y_train)

# Menghitung Akurasi Model pendekatan Naive
Bayes
# akurasi_model_naive_bayes =
roc_auc_score(Y_test, clf.predict_proba(X_test)[:, 1])
# print("Akurasi Model pendekatan Naive Bayes : " +
str(akurasi_model_naive_bayes*100) + "%")
```

```
# Penghitungan K Fold CV
nilai_k = 10
scores = cross_val_score(clf, x, y, cv=nilai_k)
print("Hasil Peformansi Menggunakan K-FOLD
dengan Pendekatan Klasifikasi Naive Bayes")
print("Nilai K : " + str(nilai_k))
print(scores)
print("Sehingga, akurasi rata-rata : " +
str((scores.mean())*100) + "%")

y_pred = clf.predict(X_test)
print("Confusion matrix")
cm = confusion_matrix(Y_test, y_pred)
print(cm)
print("")
print(classification_report(Y_test, y_pred))
```

```
Hasil performa menggunakan k-fold dengan pendekatan klasifikasi
Naive Bayes
Nilai K : 10
[0.48414851 0.5841582 0.55445545 0.55445545 0.58415842
0.4950495 0.51485149 0.56435644 0.47524752 0.57  ]
Sehingga, akurasi rata-rata : 53.81881188118813%
  [[ 539   26    0]
   [   0 2715    0]
   [   0   47    0]]

              precision    recall  f1-score   support

         neg       1.00      0.95      0.98       565
     neutral       0.97      1.00      0.99      2715
         pos       0.00      0.00      0.00        47

    accuracy                           0.98      3327
   macro avg       0.66      0.65      0.65      3327
weighted avg       0.96      0.98      0.97      3327
```

Gambar 4. *Source Code Python Pemodelan Naive Bayes*

4. DISKUSI

Hasil dikusi yang didapat bahwa klasifikasi secara otomatis komentar positif, negatif, dan dan netral pada opini media sosial tentang objek wisata menggunakan Algoritma *naive* bayes dengan akurasi rata rata yaitu 97.45088 % untuk 10 kali pengujian nilaai fold. Berdasarkan hasil penelitian yang telah dilakukan terlihat bahwa algoritma *naive bayes classifier* dapat mengklasifikasikan suatu opini berupa komentar ke dalam dua kelas yaitu positif dan negatif dengan akurat. Tingkat keakurasian dari pengkasifikasian tersebut sangat dipengaruhi oleh proses training. Sehingga dapat disimpulkan dari pengklasifikasian yang dihasilkan dapat terlihat dengan jelas informasi sentimen public terhadap tempat wisata. Hasil penelitian ini menghasilkan komentar positif yang sinigfikan sehingga dapat merekomendasikan tempat wisata terbaik sesuai pengunjung.

Penelitian ini didukung oleh penelitian Subarkah *el al* (2022) yang menyatakan bahwa analisis menggunakan algoritma *naive bayes* dapat menghasilkan tingkat akurasi yang baik. Dampak dari penelitian ini terhadap pengembangan ilmu pengetahuan dapat merekomendasikan tempat wisata terbaik khususnya di bidang pariwisata dan analisis *sentiment*.

5. ACKNOWLEDGEMENT

Penulis mengucapkan terima kasih kepada Universitas Teknokrat Indonesia atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian ini. Selain itu, kami juga

mengucapkan apresiasi yang mendalam kepada para reviewer atas masukan, kritik, dan saran yang sangat membantu dalam meningkatkan kualitas tulisan ini.

6. KESIMPULAN

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa dalam mengekstraksi data dari media sosial tentang objek wisata di Provinsi Lampung menggunakan *python* dengan melakukan *crawling* data twiter dengan tahapan *cleaning, tokenizing, filtering, case folding,* dan *steaming.*

Hasil klasifikasi secara otomatis komentar positif, negatif, dan dan netral pada opini media sosial tentang objek wisata di provinsi Lampung menggunakan Algoritma *naive* bayes dengan akurasi rata rata yaitu 97.45088 % untuk 10 kali pengujian nilai *fold.*

Hasil perhitungan klasifikasi dapat digunakan untuk menentukan rekomendasi terbaik pada tempat wisata di Provinsi Lampung dengan persentase semosi true positif 539 data, false netral 26, sedangkaan false negatif 0 data, sedangkan false Positif mendapat 0 data dan True netral mendapatkan 2715 data, serta false negatif 0. false positif 0, false netral 47 dan true negative 0.

Rekomendasi untuk penelitian selanjutnya dapat menggunakan algoritma lain seperti Random Forest atau XGboost untuk membandingkan hasil nilai akurasi yang lebih baik.

DAFTAR PUSTAKA

[1] BPS, "Badan Pusat Statistik," 2022. .

[2] D. Alita, I. Sari, A. R. Isnain, and S. Styawati, "Penerapan Naïve Bayes Classifier Untuk Pendukung Keputusan Penerima Beasiswa," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 1, p. 17, 2021, doi: 10.33365/jdmsi.v2i1.1028.

[3] R. Situmorang, U. M. Husni Tamyis, and L. S. Andar Muni, "Analisis Sentimen Destinasi Wisata Di Jawabarat Pada Twitter Menggunakan Algoritma Naive Bayes Classifier," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 339–342, 2023, doi: 10.51876/simtek.v8i2.287.

[4] W. Khofifah, D. N. Rahayu, and A. M. Yusuf, "Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Review Masyarakat Terhadap Tempat Wisata Pantai Di Kabupaten Karawang Pada Ulasan Google Maps," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 16, no. 4, pp. 28–38, 2022, doi: 10.35969/interkom.v16i4.192.

[5] I. Tri, P. Sihaloho, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Objek Wisata Danau Toba berdasarkan Ulasan Pengunjung menggunakan Algoritma Support Vector Machine," vol. 6, no. 9, pp. 4204–4209, 2022, [Online]. Available: http://j-ptiik.ub.ac.id.

[6] H. Sulistiani, I. A. Rahman, B. M. Hurohman, A. Nurkholis, and Styawati, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 8, no. 2, pp. 299–303, 2022, doi: 10.37859/jf.v13i3.6303.

[7] A. R. Isnain *et al.*, "Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis," *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.

[8] M. Z. Subarkah, M. Hilda, and E. Zukhronah, "Analisis Sentimen Review Tempat Wisata Pada Data Online Travel Agency Di Yogyakarta Menggunakan Model Neural Network IndoBERTweet Fine Tuning," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 543–552, 2022, doi: 10.34123/semnasoffstat.v2022i1.1246.

[9] M. S. Syahlan, D. Irmayanti, and S. Alam, "Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Support Vector Machine (Svm)," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 8, no. 2, pp. 315–319, 2023, doi: 10.51876/simtek.v8i2.281.

[10] G. K. Pati and E. Umar, "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2309, 2022, doi: 10.30865/mib.v6i4.4635.

[11] A. Amolik, "Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques," *Int. J. Eng. Technol.*, vol. 7, no. 6, pp. 2319 – 8613, 2016.
[12] A. R. Isnain, H. Sulistiani, B. M. Hurohman, A. Nurkholis, and S. Styawati, "Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen," *J. Edukasi dan Penelit. Inform.*, vol. 8, no. 2, p. 299, 2022, doi: 10.26418/jp.v8i2.54704.
[13] J. Mantik, Y. R. Saputri, and H. Februariyanti, "Sentiment Analysis on Shopee E-Commerce Using the Naïve Bayes Classifier Algorithm," *J. Mantik*, vol. 6, no. 2, pp. 1349–1357, 2022, doi: https://doi.org/10.35335/mantik.v7i3.4020.
[14] R. Ulgasesa, A. B. P. Negara, and T. Tursina, "Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal," *J. Sist. dan Teknol. Inf.*, vol. 10, no. 3, p. 286, 2022, doi: 10.26418/justin.v10i3.53880.
[15] R. Julianto, E. D. Bintari, and I. Indrianti, "Analisis Sentimen Layanan Provider Telepon Seluler pada Twitter Menggunakan Metode Naïve Bayesian Classification," *J. Big Data Anal. Artif. Intell.*, vol. 3, no. 1, pp. 23–30, 2017, doi: https://doi.org/10.52436/1.jutif.2024.5.5.2004.
[16] N. Nofiyani and W. Wulandari, "Implementasi Electronic Data Processing Untuk meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1621, 2022, doi: 10.30865/mib.v6i3.4332.
[17] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *J. Sains dan Inform.*, no. August, pp. 19–28, 2023, doi: 10.34128/jsi.v9i1.622.
[18] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
[19] B. Saputra, S. Anwar, E. Tohidi, H. Susana, and D. Pratama, "Penerapan Algortima Naïve Bayes Dalam Klasifikasi Harga Ponsel," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3587–3594, 2024, doi: 10.36040/jati.v7i6.8281.
[20] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 935-950
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4037