

PEMANFAATAN DATA INSTAGRAM UNTUK MENGETAHUI REPUTASI TEMPAT WISATA DI LOMBOK

Muhammad Azmi¹⁾, Amiruddin Khairul Huda²⁾, dan Arief Setiyanto³⁾

¹⁾STMIK Syaikh Zainuddin NW Anjani-Lombok

^{2,3)}MTI Universitas Amikom Yogyakarta

e-mail: muhammad4zmi@gmail.com¹⁾, amirudin.2309@students.amikom.ac.id²⁾, arief_s@amikom.ac.id³⁾

ABSTRAK

Sosial media saat ini memiliki pengaruh yang sangat luar biasa dalam pengambilan keputusan, termasuk didalam industri wisata untuk mengetahui reputasi dan popularitas sebuah destinasi wisata. Saat ini penggunaan dan pemanfaatan media sosial tidak hanya sebatas media hiburan tapi juga sebagai media atau wadah untuk berbagi informasi baik informasi yang dianggap penting maupun yang tidak penting dan terabaikan. Dalam penelitian ini, kami memanfaatkan data yang tersebar melalui postingan instagram yang berkaitan dengan destinasi wisata untuk dijadikan referensi untuk mengetahui reputasi destinasi wisata yang berada di Pulau Lombok. Data yang digunakan berasal dari Instagram dengan jumlah 600 dataset, dengan menggunakan tiga kata kunci yaitu pantai, gili dan Gunung. Adapun metode yang digunakan adalah Naive Bayes Classifier yang akan digunakan untuk mengklasifikasi postingan dan komentar instagram dengan pengklasifikasian kedalam postingan positif, negatif dan netral terhadap destinasi wisata. Hasil dari penelitian ini dapat menunjukkan bahwa akurasi naive bayes sebesar 59%, sedangkan tempat wisata atau destinasi wisata yang dikategorikan bereputasi populer untuk pantai yaitu Pantai Kuta dengan presentase 85% dan untuk kategori gili yaitu gili air 47% dan kategori gunung yaitu Gunung Rinjani dengan prosentase 60%.

Kata Kunci: Python, Wisata, Naive Bayes Classifier.

ABSTRACT

Social media currently has an extraordinary influence in decision making, including in the tourism industry to find out the reputation and popularity of a tourist destination. At present the use and use of social media is not only limited to entertainment media but also as a medium or a place to share information both information that is considered important or not important and neglected. In this study, we utilize data scattered through Instagram posts related to tourist destinations to be used as references to find out the reputation of tourist destinations on the island of Lombok. The data used comes from Instagram with a total of 600 datasets, using three keywords namely beach, dyke and hill. The method used is the Naive Bayes Classifier which will be used to classify postings and Instagram comments by classifying into positive, negative and neutral posts to tourist destinations. The results of this study can show that the accuracy of naive bayes is 59%, while tourist attractions or tourist destinations which are categorized as popular for the beach are Kuta Beach with 85% percentage and for the Gili category that is 47% Gili Air namely, mountain Mount Rinjani with a percentage of 60%.

Keywords: Python, Tourism, Naive Bayes Classifier.

I. PENDAHULUAN

Sosial media saat ini telah mengubah cara orang atau masyarakat untuk berbagi dan mengungkapkan pendapat satu dengan yang lainnya[1][2]. Sosial media saat ini telah bertransformasi menjadi aspek penting sebagai media menyebarkan informasi, mencari dan memasarkan suatu produk[3]. Sosial media juga menjadi populer seiring dengan pemerataan jaringan internet dan hadirnya platform- platform sosial media. Instagram[4].

Pengguna aktif sosial media setiap tahunnya mengalami peningkatan yang progressif. Berdasarkan data dari halaman web Hootsuite (*We are social*) pengguna internet aktif pada tahun 2019 sebesar 4.388 miliar yang tersebar diseluruh dunia. Penelitian ini berfokus pada sosial media instagram[2]. Oleh sebab itu, bahkan klasifikasi otomatis sangat diperlukan untuk menangani data tersebut.

Analisis sentiment di instagram dalam prosesnya cukup rumit, disebabkan dengan adanya emotikon, kata-

kata yang tidak baku dan kesalahan pada ejaan berdampak untuk melakukan proses preprocessing sebelum melakukan ekstraksi fitur[5]. Namun dalam penelitian sebelumnya telah dilakukan penelitian dengan menggunakan naive bayes sebagai salah satu metode yang akurat[5]. Naive bayes merupakan salah satu model probabilistik yang sederhana dan bekerja dengan baik pada proses klasifikasi teks[6]. Naive bayes banyak digunakan dan diterapkan dalam proses klasifikasi sentimen karena mampu melatih dataset menghasilkan sebuah decision atau keputusan[6][7][8][9].

Pulau lombok merupakan salah satu destinasi wisata yang sangat banyak dikunjungi wisatawan, baik wisatawan lokal maupun wisatawan mancanegara. Pulau lombok selain dikenal sebagai pulau seribu masjid juga dikenal sebagai daerah yang kaya akan destinasi wisata alam yang menawan dan merupakan tujuan wisata halal populer di dunia. Berdasarkan data Badan Pusat Statistik (BPS) NTB tahun 2018, kunjungan wisatawan mancanegara 151.193, sedangkan wisatawan lokal mencapai 1.473.902 orang (BPS NTB) sedangkan pada akhir tahun 2018 hingga 2019 mengalami penurunan pasca bencana gempa bumi pada pertengahan 2018.

Berdasarkan data tersebut diketahui bahwa media sosial mampu menjadi salah satu referensi untuk mengetahui popularitas destinasi wisata berdasarkan postingan pengguna di akun sosial media mereka. Pemanfaatan sosial media bisa menjadi sarana yang cukup baik dan efektif dalam peningkatan industri pariwisata. Pada era teknologi informasi seperti saat ini sering dikaitkan dengan dengan yang lain termasuk dampak sosial dan ekonomi dalam rangka peningkatan kesejahteraan masyarakat lokal. Peran teknologi informasi saat ini dapat memberikan dampak positif terhadap kemajuan industri pariwisata di indonesia[6]. Berdasarkan data yang bersumber dari Badan Pusat Statistik (BPS) Indonesia jumlah kunjungan wisatawan mancanegara ke indonesia pada periode januari-april 2019 mengalami kenaikan 3,22% meskipun tidak mengalami kenaikan signifikan seperti tahun-tahun sebelumnya.

Cabang riset yang mengalami perkembangan sebagai akibat dari situasi ledakan informasi di internet adalah Sentiment Analysis. Sentiment Analysis sendiri merupakan salah satu studi komputasional dari tanggapan atau pendapat masyarakat, apresiasi dan emosi melalui entitas, event dan atribut yang dimiliki[10]. Dalam teknologi informasi saat ini muncullah dampak positif yang dihasilkan seperti memudahkan dalam hal melakukan komunikasi, mencari dan mengakses informasi. Fokus pembahasan pada penelitian ini membahas sentiment analysis untuk menentukan popularitas destinasi wisata di pulau lombok berdasarkan data Instagram. Metode yang digunakan dalam melakukan classifier sentiment adalah naive bayes classifier.

II. STUDI PUSTAKA

A. Penelitian Terkait

Beberapa penelitian yang membahas sentiment analysis telah dilakukan pada industri pariwisata. Pada bagian ini menjelaskan perbedaan atau perbandingan dengan penelitian sebelumnya untuk membedakan inovasi yang akan dilakukan pada penelitian ini.

Beberapa diantaranya yaitu penelitian pertama [11]. Dimana didalam penelitian ini classification teknik dimasukkan kedalam domain dari blog, dan membandingkan 3 algoritma pencarian yaitu Naive bayes, SVM dan model N-gram berbasis karakter untuk perjalanan 7 negara yang ada di AS dan Eropa. Dalam kesimpulannya pendekatan SVM dan N-gram mengungguli pendekatan Naive Bayes, dan ketika rangkaian data training memiliki banyak tinjauan, semua pendekatan yang digunakan mencapai akurasi 80%.

Penelitian kedua [12] membahas mengenai perbedaan antara tempat – tempat favorit bagi turis dan penduduk lokal menggunakan profil dari instagram.

Penelitian ketiga [2] menjelaskan studi analisis sentimen yang dilakukan lebih dari 1000 postingan dari facebook tentang siaran berita. Dan membandingkan sentimen untuk Rai – layanan penyiaran public di italia – terhadap La7 perusahaan swasta yang lebih dinamis

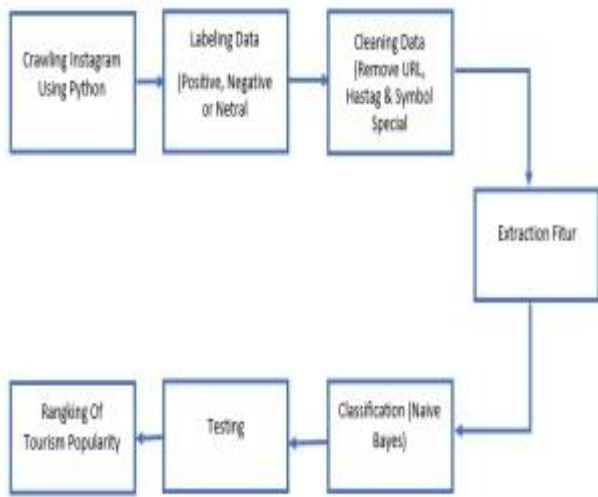
Penelitian keempat [13] membahas tentang sentimen manusia dari koleksi gambar yang ada di internet dengan skala besar berdasarkan fitur dan informasi yang ada di media sosial kontekstual (seperti : komentar teman, dan deskripsi pengguna) dengan membandingkan metode SVM dan EL dan hasilnya akurasi meningkat 6% sampai 10%.

Penelitian kelima [5] mencoba untuk menganalisis posting twitter tentang produk elektronik seperti ponsel, laptop dll, menggunakan pendekatan machine learning. Dengan melakukan analisis sentimen dalam domain tertentu, yang dimungkinkan untuk mengidentifikasi pengaruh informasi domain dalam klasifikasi sentimen.

III. METODE PENELITIAN

Situs sosial media saat ini sangat diminati dan digemari diseluruh dunia, termasuk Instagram.

Dalam penelitian ini bagian metodologi akan dibahas mengenai tahap-tahap dari penelitian yang dilakukan, seperti pada gambar 1.



Gambar 1. Alur Penelitian

Berikut penjelasan mengenai alur penelitian tersebut:

A. Mengambil Data Instagram dengan Python

Saat mengambil data di instagram menggunakan Python tanpa menggunakan API cukup memasukkan username dan password kemudian memasukkan akun instagram untuk di Crawling datanya.

B. Labeling Data

Labeling data dilakukan terhadap data yang sudah didapatkan dengan crawling dari instagram untuk dibedakan dalam kategori positif, negatif dan netral.

C. Cleaning Data

Terdapat berbagai konten non sentiment seperti URL, hastag dan lain sebagainya. Data- data tersebut merupakan keluaran dari instagram dan harus dihilangkan sehingga menghasilkan sentiment yang akurat. Cleaning seperti berikut:

a) Penghapusan URL

Data instagram merupakan sumber informasi utama, dalam postingan instagram banyak pengguna yang menyertakan tautan yang tidak bisa digunakan untuk sentimen analisis, maka URL harus dihapus dari postingan.

b) Penghapusan Simbol Khusus

Penghapusan simbol- simbol khusus yang tidak bisa digunakan untuk sentimen analisis, seperti tanda baca, angka dan lainnya.

c) Penghapusan Hastag

Penghapusan tanda hastag atau tagline, biasanya digunakan untuk memberikan nama subjek atau frasa yang populer.

D. Ekstraksi Fitur

Ekstraksi fitur dilakukan dalam menentukan bobot dari kalimat yang telah melewati *cleaning data* dengan TF-IDF.

E. Naïve Bayes Classifier

Penggunaan metode *Naïve Bayes Classifier* karena dalam penelitian ini akan melakukan proses klasifikasi, dimana *Naïve Bayes Classifier* merupakan salah satu algoritma yang mewakili metode pembelajaran serta statistik untuk klasifikasi[14]. Metode naïve bayes classifier (NBC) memiliki Kelebihan sederhana namun memiliki akurasi yang tinggi[15]. Klasifikasi naïve bayes juga mewakili metode statistik untuk klasifikasi[16]. Naive bayes juga merupakan model probabilistik dan memungkinkan untuk mendapatkan ketidakpastian tentang model algoritma probabilitas. Naïve bayes didalamnya tersedia algoritma yang dapat berguna dan data yang di amati dapat dikombinasikan. Hal ini dapat membantu dalam memahami dan mengevaluasi algoritma lainnya[17]. Teorema Naive Bayes mengacu pada konsep probabilitas bersyarat[18]. Secara umum naïve bayes dapat diformulasikan dalam persamaan berikut[19]:

$$(PAB) = \frac{PBA \times PA}{P(B)} \quad (1)$$

Keterangan :

PA = Probabilitas Kemunculan Label

PBA = Probabilitas Kemunculan Label Terhadap fitur

Metode klasifikasi yang dapat digunakan adalah NBC (Naive Bayes Classifier). NBC memiliki kelebihan sederhana namun memiliki akurasi yang cukup tinggi[15]. Dalam naïve bayes classifier setiap postingan yang berupa teks pada instagram disajikan dalam pasangan atribut $\langle a_1, a_2, a_3, \dots, a_n \rangle$ dimana a_1 adalah kata pertama, a_2 kata kedua dan seterusnya sampai a_n , sedangkan V merupakan himpunan kelas [20]. Pada saat proses klasifikasi, metode ini akan menghasilkan sebuah kategori atau kelas yang memiliki probabilitas paling tinggi (V_{MAP}) dengan memasukkan atribut $\langle a_1, a_2, a_3, \dots, a_n \rangle$. Rumus V_{MAP} dapat dilihat pada persamaan berikut.

$$V_{MAP} = \text{argmax}_{V \in V} P(V_j | a_1, a_2, a_3, \dots, a_n) \quad (2)$$

Dengan menggunakan teorema bayes dapat ditulis dengan formula berikut:

$$V_{MAP} = \text{argmax}_{V \in V} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) \times P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad (3)$$

$P(a_1, a_2, a_3, \dots, a_n)$ diberikan nilai konstan terhadap semua V_j sehingga persamaan 3 dapat diubah menjadi persamaan (4)

$$V_{MAP} = \text{argmax}_{V \in V} P(a_1, a_2, a_3, \dots, a_n | V_j) \times P(V_j) \quad (4)$$

Naive bayes classifier menyederhanakan ini dengan asumsi bahwa setiap kategori, setiap atribut bebas bersyarat dengan yang lainnya[21]:

$$VMAP = \text{argmax}_{vj} \prod_{i=1}^n P(V_j | x_i) \times P_i(a_i | V_j) \quad (5)$$

$P(V_j)$ dan probabilitas kata a_i untuk setiap kategori $P(a_i | V_j)$ dihitung pada training. Dimana

$$P(V_j) = \frac{\text{docs}_j}{\text{training}} \quad (6)$$

$$P(a_i | V_j) = \frac{n_i + 1}{n + \text{vocabulary}} \quad (7)$$

Dimana docs_j merupakan jumlah kategori dokumen j dan training adalah jumlah dokumen yang digunakan dalam proses training. Sedangkan n_i merupakan jumlah kemunculan kata pada a_i pada kategori V_j , n adalah kosakata yang muncul pada kategori V_j dan kosakata adalah jumlah kata yang unit pada semua training[20].

F. Testing

G. Perangkingan Wisata Populer

Setelah melakukan semua proses, maka akhir dari penelitian ini adalah memberikan ranking popularitas destinasi wisata di Pulau Lombok berdasarkan tanggapan pengunjung di Instagram.

IV. HASIL DAN PEMBAHASAN

A. Tipe Pengumpulan Data Instagram

Proses pengumpulan data dilakukan dengan memanfaatkan data dari instagram yang memiliki postingan terkait dengan wisata lombok. Dalam pengambilan data postingan instagram terlebih dahulu memasukkan akun berupa username dan sandi. Pengumpulan data dilakukan menggunakan Python, dengan memasukkan salah satu akun instagram yang banyak mempostingan tentang wisata lombok. Perintah code untuk collecting data seperti gambar 2.

```
F:\TEMPLATE\Big Data\instagram-crawler-master>crawler.py hashtag -t pantailombok -n 100 -o ./pantailombok1 --fetch_detail
DevTools listening on ws://127.0.0.1:49962/devtools/browser/cd354ebb-e30b-4f4d-a66d-87789243d3f7
```

Gambar 2. Collecting Data menggunakan Python

Dari proses pengambilan data tersebut, dicoba melakukan pengambilan data terhadap 100 Postingan instagram untuk setiap kata kunci yang berkaitan dengan pantailombok, gilirombok dan gunung lombok, kemudian data tersebut akan digunakan sebagai dataset.

B. Labeling Data

Proses labeling data dilakukan dengan cara manual dengan menambahkan kalimat label positive, negative dan netral. Proses labeling data dilakukan untuk membedakan postingan instagram yang termasuk dalam tiga kategori positive, negative dan netral.

```
_label_positive Kangen pantai, komen dong pantai mana favorite kalian gengs? Gili Kapal, East Lombok
_label_positive Bahagia pokoknya Lokasi: Batu Payung, Pujut - Lombok Tengah
_label_positive Sembulan Lombok
_label_positive this lombok NTB Indonesia kalau anda datang ke lombok jangan sampai ketinggalan di pantai yang satu ini dan keunika
_label_positive Pantai-pantai pesisir Utara pulau Lombok Kalau ada jalan 10 KM ditempuh dalam 5 jam, itu bukan karena macet....seti
_label_negative Pantai Teluk Lombok, Sangatta, Kutai Timur, Kalimantan Timur
_label_positive Satu, dua, tiga sayang semuanyaGili: Pantai Aan
_label_positive Harus dewasa... Lombok kuta mandalika
_label_positive Pantai Selong Belanak Bukit Lebah Lombok Nusa Tenggara Barat Indonesia
```

Gambar 3. Labeling Data

C. Cleaning Data

Setelah dilakukan proses labeling data, tahapan selanjutnya adalah melakukan proses cleaning data. pada proses cleaning data dilakukan untuk menghapus postingan ganda. Dimana pada tahap sebelumnya pada proses pengumpulan data yang berkaitan dengan pantai lombok, gili lombok dan gunung lombok rata-rata mengambil 100 data. selanjutnya semua kalimat dibersihkan dari semua tanda baca dan lainnya dengan perintah seperti pada gambar 4.

```
def form_sentence(instagram):
    instagram_blob = TextBlob(instagram)
    return ' '.join(instagram_blob.words)
print(form_sentence(train_instagrams['instagram'].iloc[10]))
print(train_instagrams['instagram'].iloc[10])
```

Gambar 4. Kode Cleaning Data

Setelah dilakukan pembersihan. Terdapat beberapa kata yang berubah maupun dihapus seperti pada tabel 1.

Tabel 1. Cleaning Data

Old Word	New Word	Status
Kau, kamu,	""	Remove
Tmpat	Tempat	Update
Dmn	Dimana	Update
Jgn	Jangan	Update
#, @, ", ' , _ , ! , ?	""	Remove

Setelah proses cleaning data selesai, selanjutnya melakukan proses fitur extraction untuk mengetahui frekuensi kemunculan kata.

D. Ekstraksi Fitur

Berdasarkan proses cleaning data, dihasilkan tabel frekuensi kemunculan terhadap label yang berkaitan dengan semua kata kunci pantai, gili dan gunung, seperti disajikan pada tabel 2.

Tabel 2. *Extraction Feature*

Label Filter	Kemunculan Positif	Kemunculan Negatif	Kemunculan Netral	Kemunculan Kata	Dataset	Positive	Negative	Netral
Lombok	72	37	60	169	600	0.426036	0.218935	0.35503
senggigi	60	10	22	92		0.652174	0.108696	0.23913
Pantai	89	29	29	147		0.605442	0.197279	0.197279
yang	48	17	32	97		0.494845	0.175258	0.329897
beach	42	7	13	62		0.677419	0.112903	0.209677
di	39	39	40	118		0.330508	0.330508	0.338983
pink	32	9	15	56		0.571429	0.160714	0.267857

Berdasarkan tabel 2, diketahui bahwa berdasarkan dataset kata lombok yang muncul 169 kali, dimana 72 berlabel positif, 60 Netral dan 37 berlabel negatif dengan probabilitas kemunculan kata sebesar 0.42 atau 42% untuk label positif, 0.21, atau 21% probabilitas pada kata negatif, dan 0.35 atau 35% untuk yang berlabel netral. Hal yang sama dapat juga dilakukan untuk kata kunci gili dan gunung.

E. Classification Naïve Bayes

Selanjutnya tahap naïve bayes dilakukan dengan beberapa tahap atau proses, yaitu:

1. Pengumpulan Data

Tahap ini dilakukan proses pengumpulan data yang akan dilakukan klasifikasi, terlihat pada gambar berikut:

```

>crawler.py hashtag -t pantaisenggigi -n 200 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t pantaikuta -n 200 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t pantaipink -n 100 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t gilitrawangan -n 200 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t gilitrasi -n 200 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t gilihonda -n 200 -o ./testing --fetch_details --fetch_comments
>crawler.py hashtag -t gunungrinjani -n 200 -o ./testing --fetch_details --fetch_comments

```

Gambar 5. Pengumpulan data testing

Gambar 5 diatas merupakan perintah yang digunakan untuk proses pengumpulan data testing, setiap kata kunci dibatasi sebanyak 200 data.

2. Klasifikasi Pantai, Gili dan Gunung

Dalam proses ini digunakan 5 kata kunci yang terkait pantai 3 kata kunci dan gili 1 kata kunci dan gunung 1 kata kunci yang akan digunakan untuk mengumpulkan data dan selanjutnya data tersebut akan melewati proses klasifikasi dengan naïve bayes, seperti yang disajikan pada tabel 3 berikut:

Tabel 3 Kata Kunci

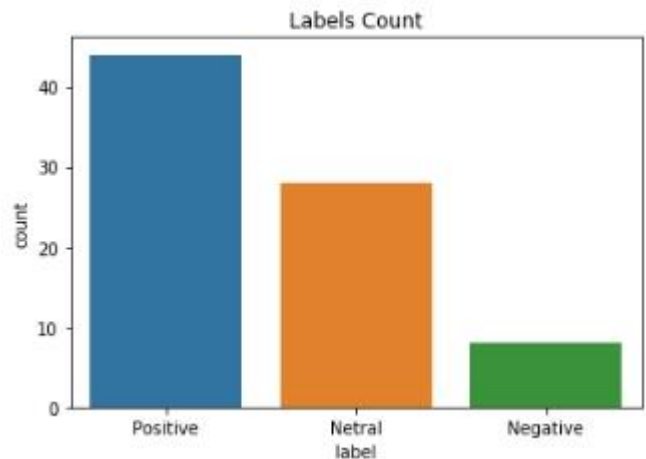
Kata Kunci	Jenis	Jumlah Data
Pantai Senggigi	Pantai	128
Pantai Kuta	Pantai	80
Pantai Pink	Pantai	48
Gili Trawangan	Gili	79
Gunung Rinjani	Gunung	83

Berdasarkan tabel 3 diatas menunjukkan bahwa jumlah postingan instagram pantai senggigi dan gili trawangan merupakan destinasi wisata terpopuler

dengan masing-masing 128 postingan untuk pantai senggigi dan 79 postingan untuk gili trawangan dan 82 postingan untuk gunung rinjani dari proses pencarian data.

3. Klasifikasi Sentiment menggunakan naïve bayes

Pada proses ini dilakukan klasifikasi sentiment dengan menggunakan naïve bayes, berikut hasil klasifikasi masing-masing kata kunci dengan naïve bayes, berikut kode program yang digunakan seperti pada gambar 6.



Gambar 6. Hasil Klasifikasi

Gambar 5 merupakan contoh hasil klasifikasi pada data set pantai senggigi, dari 128 data, 45 postingan positif, 29 netral dan 9 postingan negatif.

Sedangkan untuk hasil akurasi, presisi, recall dan F1-Score untuk percobaan klasifikasi pantai senggigi dengan Naïve Bayes seperti ditunjukkan kode program pada gambar 7 berikut.

```

#Machine Learning Pipeline
pipeline = Pipeline([
    ('bow',CountVecorizer(analyzer='text_processing')), # strings to token integer counts
    ('tfidf',TfidfTransformer()), # integer counts to weighted TF-IDF scores
    ('classifier',MultinomialNB()), # train on TF-IDF vectors w/ Naive Bayes classifier
])
pipeline.fit(msg_train,label_train)

predictions = pipeline.predict(msg_test)

print(classification_report(predictions,label_test))
print('\n')
print(confusion_matrix(predictions,label_test))
print(accuracy_score(predictions,label_test))

```

Gambar 7. Kode Program Akurasi, Presisi, Recall F1-Score

Hasil akurasi, presisi, recall dan F1-Score untuk percobaan klasifikasi Pantai Senggigi dengan Naïve Bayes seperti pada gambar 8 berikut.

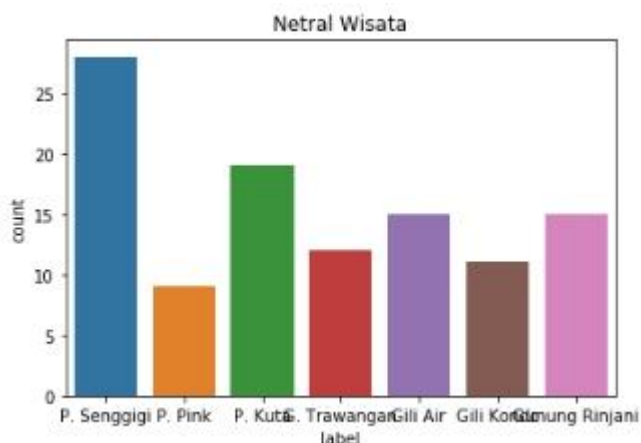
	precision	recall	f1-score
Netral	0.25	0.20	0.22
Positive	0.67	0.73	0.70
accuracy	0.56		

Gambar 8. Hasil Akurasi, Presisi, Recall F1-Score

Berdasarkan hasil diatas di tunjukkan bahwa nilai akurasinya untuk wisata pantai senggigi sebesar 56%.

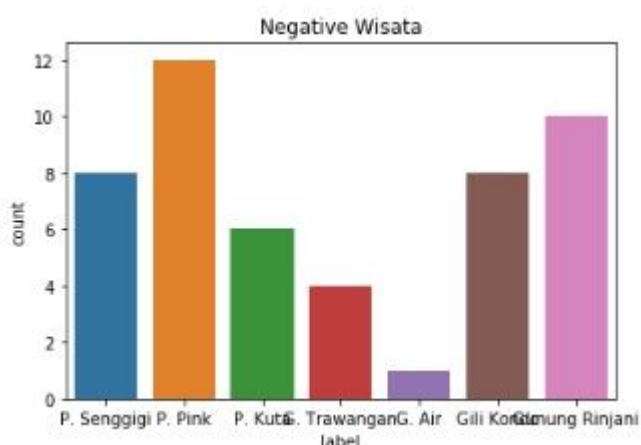
4. Testing

Tahapan testing dilakukan sehingga hasilnya dapat menunjukkan dari sentiment positive, negative dan netral dari setiap kata kunci pantai, gili dan gunung.



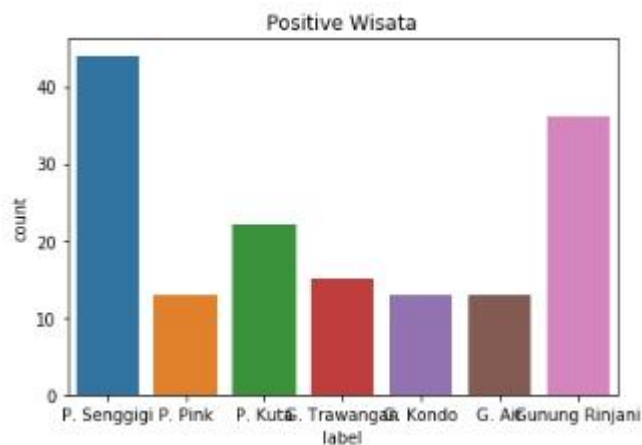
Gambar 9. Sentiment Netral

Gambar 9 diatas menunjukkan hasil sentiment netral dengan kata kunci pantai, gili dimana ditunjukkan hasil sentiment netral tertinggi yaitu pantai senggigi dan sentiment netral terendah pantai pink.



Gambar 10. Sentiment Negative

Gambar 10 merupakan hasil sentiment negative dari instagram, dimana pantai pink memiliki sentiment negative tertinggi dan Gili air paling rendah sentiment negativenya.



Gambar 11. Sentiment Positive

Gambar 11 diatas menunjukkan hasil sentiment positive dari pantai, gili dan gunung, dimana pantai senggigi menjadi destinasi yang memiliki sentiment positive tertinggi dan pantai terendah pantai pink.

Tahap selanjutnya adalah tahap pengujian, hasil tahap pengujian dapat dilihat pada table 3 di bawah ini:

Table 4. perbandingan manual vs aplikasi

Kalimat	Hasil Aplikasi	Hasil Manual	Kecocokan
Menikmati keindahan pantai dengan angin semilir membuat ku semakin bersyukur akan ciptaanNya	Positive	Positive	True
Senggigi, Guru terbaik adalah alam, setiap petualangan pasti akan mengajarkan sesuatu yang bernilai kepadamu.	Netral	Positive	False
Pantai Pink berpasir putih, rumput hijau, dan air berwarna biru. Lalu dari mana pink nya?	Netral	Netral	True
Angkernya GOA JEPANG di Pantai Pink Lombok.	Negative	Negative	True

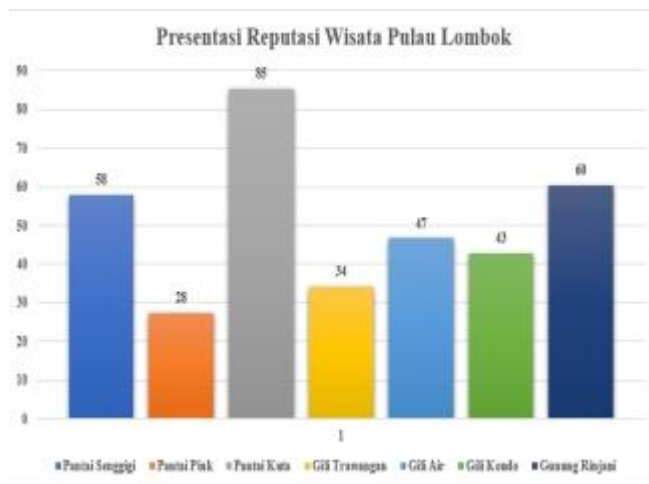
Tabel 4 diatas merupakan contoh hasil pengujian dengan membandingkan klasifikasi manual dan klasifikasi dengan aplikasi. Hasil perbandingan terhadap seluruh dataset sebagai berikut:

Tabel 5. Pengujian

Ob- jek Wisata	Jumlah Pelabelan ap- likasi dan manual			Nilai Akurasi (keco- cokan/semua data)
	Posi- tif	Negatif	Ne- tral	
Pantai Senggigi	45	8	29	/128
Pantai Kuta	22	5	19	/80
Pantai Pink	13	12	9	/48
Gili Tra- wangan	15	4	12	/79
Gili Kondo	13	8	11	60
Gili Air	13	1	15	/56
Gunung Rinjani	35	10	15	/83
Total	314/534			
Akurasi	0.588= 59%			

5. Perangkingan

Tahapan perangkingan merupakan tahap terakhir untuk mencari destinasi wisata populer di pulau Lombok seperti yang ditunjukkan gambar 12 berikut.



Gambar 12. Rangking Reputasi Wisata

Gambar 12 diatas menunjukkan hasil presentase reputasi destinasi wisata di pulau Lombok dengan kata kunci pantai, gili dan gunung, dimana untuk memperoleh nilai pada masing- masing destinasi dengan menjumlahkan label positif dijumlahkan dengan label netral dan dibagi jumlah postingan masing- masing destinasi wisata atau dengan menggunakan formula:

$$\text{Presntase} = \frac{\text{Label Positif} + \text{Label Netral}}{\text{Jumlah Postingan}} \times 100$$

Berdasarkan grafik pada gambar 12 menunjukkan bahwa destinasi wisata yang memiliki presentasi tertinggi adalah pantai kuta dengan 85%, kemudian gunung rinjani dengan 60%. Berikut peringkat reputasi wisata pantai, gili dan gunung di pulau Lombok pada table 6.

Table 6. Rangking Reputasi Destinasi Wisata

Jenis	Destinasi	Peringkat
Pantai	Kuta	1
	Senggigi	2
	Pink	3
Gili	Air	1
	Kondo	2
	Trawangan	3
Gunung	Rinjani	1

V. KESIMPULAN

Berdasarkan analisis data yang telah peneliti lakukan dengan memanfaatkan data Instagram untuk mencari destinasi wisata yang memiliki reputasi populer di pulau Lombok menggunakan kata kunci pantai, gili dan gunung dengan menggunakan Naïve Bayes Classifier. Dalam penelitian tersebut didapatkan akurasi naïve bayes sebesar 59%, dimana tempat wisata yang termasuk tempat wisata dengan reputasi populer untuk kategori pantai adalah Pantai Kuta, sedangkan untuk gili adalah Gili Air dan untuk kategori gunung adalah Gunung Rinjani.

DAFTAR PUSTAKA

- [1] Y. Mejova, "Sentiment Analysis : An Overview Comprehensive Exam Paper," *Science* (80-.), pp. 1–34, 2009.
- [2] F. Neri, C. Aliprandi, F. Capecci, M. Cuadros, and T. By, "Sentiment analysis on social media," *Proc. 2012 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2012*, pp. 919–926, 2012, doi: 10.1109/ASONAM.2012.164.
- [3] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks," *ICWSM 2010 - Proc. 4th Int. AAAI Conf. Weblogs Soc. Media*, pp. 90–97, 2010.
- [4] E. Jean and P. Ed, "Local and Global Responses to Disaster : # eqnz and the Christchurch Earthquake Associate Professor Axel Bruns ARC Centre of Excellence for Creative Industries & Innovation Queensland University of Technology , Brisbane , Australia ARC Centre of Excellenc," *Disaster Emerg. Manag. Conf. Conf. Proc.*, pp. 86–103, 2012.
- [5] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," *2013 4th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2013*, 2013, doi: 10.1109/ICCCNT.2013.6726818.
- [6] V. Narayanan, I. Arora, and A. Bhatia, "Fast and accurate sentiment classification using an enhanced Naive

- Bayes model,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8206 LNCS, pp. 194–201, 2013, doi: 10.1007/978-3-642-41278-3_24.
- [7] P. Chesley, B. Vincent, L. Xu, and R. K. Srihari, “Using verbs and adjectives to automatically classify blog sentiment,” *AAAI Spring Symp. - Tech. Rep.*, vol. SS-06-03, pp. 27–29, 2006.
- [8] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, “Pulse: Mining customer opinions from free text,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3646 LNCS, pp. 121–132, 2005, doi: 10.1007/11552253_12.
- [9] “SENTIMENT CLASSIFICATION OF MOVIE REVIEWS USING LINGUISTIC PARSING Brian Eriksson CS 838 - Natural Language Processing Final Project Report,” pp. 4–9.
- [10] Liu.B, *Opinion Mining*. Chicago, United States of America, 2012.
- [11] Q. Ye, Z. Zhang, and R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” *Expert Syst. Appl.*, vol. 36, no. 3 PART 2, pp. 6527–6535, 2009, doi: 10.1016/j.eswa.2008.07.035.
- [12] K. D. Mukhina, S. V. Rakitin, and A. A. Vishneratin, “Detection of tourists attraction points using Instagram profiles,” *Procedia Comput. Sci.*, vol. 108, no. June, pp. 2378–2382, 2017, doi: 10.1016/j.procs.2017.05.131.
- [13] Y. Wang and B. Li, “Sentiment Analysis for Social Media Images,” *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1584–1591, 2016, doi: 10.1109/ICDMW.2015.142.
- [14] A. Sinaga, “Implementasi Sentiment Analysis untuk Menentukan Tingkat Popularitas Tujuan Wisata,” *Pros. Semin. Nas. Teknol. dan Rekayasa Inf. Tahun 2017*, no. November, pp. 24–25, 2017.
- [15] I. Rish, “An empirical study of the naive Bayes classifier,” *Phys. Chem. Chem. Phys.*, vol. 3, no. 22, pp. 4863–4869, 2001, doi: 10.1039/b104835j.
- [16] M. Murnawan, “Pemanfaatan Analisis Sentimen Untuk Peningkatan Popularitas Tujuan Wisata,” *J. Penelit. Pos dan Inform.*, vol. 7, no. 2, p. 109, 2017, doi: 10.17933/jppi.2017.070203.
- [17] J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, “Naive bayes classification of uncertain data,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, no. 60703110, pp. 944–949, 2009, doi: 10.1109/ICDM.2009.90.
- [18] E. W. Sandi Fajar Rodiyansyah, “Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification,” *IJCCS*, vol. Vol. 6, No.
- [19] J. Han and M. Kambe, *Data Mining: Concepts and Techniques 2e*, 2E ed. San Fransisco: Morgan Kaufmann Publishers, 2006.
- [20] and H. S. C. D. Manning, P. Raghavan, *Introduction to Information Retrieval*. California: Cambridge University Press, 2008.
- [21] K. Tan, Steinbach, Karpatne, “Data Mining : Data Lecture Notes for Chapter 2 Introduction to Data Mining by What is Data ?,” *Ratio*, 2004.