Implementasi Metode *Naïve Bayes Classifier* Terhadap Analisis Sentimen Tempat Wisata di Nusa Tenggara Barat

1st Edgarsa Bramandyo Widyarto
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
edgarsabramandyo@students.telkomuniversity.ac.id

2nd Jondri
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
jondri@telkomuniversity.ac.id

3rd Kemas Muslim Lhaksmana
Fakultas Informatika
Universitas Telkom
Bandung, Indonesia
kemasmuslim@telkomuniversity.ac.id

Abstrak*— Berwisata adalah salah satu kegiatan yang sudah menjadi sebuah kebutuhan dalam kehidupan kita. Karena dengan liburan, bisa melepas penat dari berbagai rutinitas. Sebelum menentukan tempat wisata, biasanya wisatawan mencari terlebih dahulu informasi yang dibutuhkan. Berbicara mengenai tempat wisata, di Indonesia terdapat salah satu provinsi yaitu Nusa Tenggara Barat yang terkenal akan destinasi wisatanya. Ada pantai, gunung dan juga pulau-pulau. Hadirnya media sosial, menjadikan mudah mendapatkan segala informasi dan bersifat aktual. Dengan kemudahan aksesnya, semua orang dapat berkontribusi dalam memberikan informasi, dalam hal ini adalah tempat wisata di Nusa Tenggara Barat. Wisatawan pun jika ingin mengunjungi tempat wisata, sudah memiliki gambaran mengenai tempat yang akan dikunjungi. Twitter adalah salah satu media sosial yang cukup banyak dipakai. Wisatawan yang sedang berkunjung ke tempat wisata maupun masyarakat yang berada disana dapat memberikan komentar berupa tweet. Informasi inilah yang sangat membantu untuk mengetahui kualitas tempat wisata yang akan dikunjungi. Penelitian ini dilakukan untuk mengetahui metode ekstraksi fitur TF-IDF terhadap performa algoritma Naive Bayes Classifier dalam melakukan proses klasifikasi berdasarkan tweet pengguna yang sudah pernah mengunjungi atau yang sedang berada di Nusa Tenggara Barat. Menghasilkan makro F1-score senilai 0,76 atau 76%.

***Kata Kunci*— Wisata, Sentimen, Twitter, Naïve Bayes Classifier, Bag of word, TF-IDF**

I. PENDAHULUAN

A. Latar Belakang

Berwisata adalah salah satu kegiatan yang sudah menjadi sebuah kebutuhan dalam kehidupan kita. Karena dengan liburan, bisa melepas penat dari berbagai rutinitas. Mencari suasana yang menyenangkan dan menenangkan pikiran dari keseharian yang melelahkan. Berbicara mengenai tempat wisata, di Indonesia terdapat salah satu provinsi yaitu Nusa Tenggara Barat yang terkenal akan destinasi wisatanya, baik dari wisatawan lokal hingga wisatawan mancanegara.

Sebelum menentukan tempat wisata, biasanya wisatawan mencari terlebih dahulu informasi yang dibutuhkan. Mulai dari rekomendasi tempat kemudian tempat penginapan, harga tiket wisata, wahana hiburan, harga makanan sampai kondisi terkini di tempat tersebut supaya para wisatawan bisa lebih yakin dan terbayang situasi dan kondisi yang akan dihadapi.

Dengan berkembangnya era teknologi saat ini. Dengan mudah mendapatkan berbagai informasi dari internet dan memberikan informasi. Kumpulan data-data besar tersebut tersimpan dalam basis data yang besar.Didalamnya terdapat data subjektif yang mewakili sentimen pengguna, perasaan atau penilaian yang melekat pada suatu informasi. Kumpulan data tersebut jika diterjemahkan bisa menjadi informasi yang bermakna untuk menunjang keputusan penting dalam hal pariwisata. Penambangan Data bisa menjadi sebuah alat yang sangat berguna untuk menganalisis data pariwisata yang ada[1].

Terdapat juga yang namanya analisis sentimen. Analisis sentimen juga dikenal sebagai opinion minning. Dengan salah satunya analisis perasaan pengguna seperti sikap, emosi dan pendapat dalam kata-kata menggunakan natural language processing tools[1]. Pengaruh yang dihasilkan oleh analisis sentiment serta manfaatnya membuat penelitian atau apapun yang berhubungan dengan analisis sentimen berkembang pesat, bahkan di Amerika ada sekitar 20-30 persen perusahaan yang memfokuskan pada layanan analisis sentimen[16].

Sejumlah besar informasi yang terkandung dalam situs web microblogging menjadikannya sumber data yang menarik untuk penambangan opini dan analisis sentimen [15] Data-data tersebut bisa didapatkan pada satu platfom media sosial yaitu Twitter. Twitter adalah salah satu media sosial yang cukup banyak dipakai. Pengguna biasa mengabadikan segala hal yang dialaminya. Secara tidak sengaja, tweets pengguna ini mengandung banyak informasi yang sangat berguna dan lebih bersifat subjektif. Informasi inilah yang akan sangat membantu untuk mengetahui kualitas tempat wisata yang akan dikunjungi.

Alur kerja nya akan diawali dengan pengambilan data dari twitter. Setelah data didapatkan, kemudian dilakukan preprocessing untuk mendapatkan informasi yang diperlukan saja. Kemudian dilakukan ekstraksi fitur menggunakan metode Bag Of Words (TF-IDF) untuk melakukan pembobotan dan dilanjutkan pengklasifikasian dengan metode Naïve Bayes Classifier.

Dengan adanya permasalahan dalam mencari informasi terkait wisata dan juga masih banyak tempat wisata di Nusa Tenggara Barat yang belum banyak dikenal orang. Dari twitter, sesama pengguna bisa saling berbagi terkait

informasi informasi yang sudah pasti aktual. Maka tugas akhir ini akan dibahas mengenai analisis sentimen dari data yang berasal dari twitter. Informasi yang bersifat subjektif akan diklasifikasi menjadi data positif dan negatif. Sehingga bisa diketahui kondisi terkini sebelum akhirnya memutuskan untuk berwisata ke salah satu destinasi di Nusa Tenggara Barat.

1. Topik dan Batasannya

Peneliti tugas akhir berfokus pada penerapan TF-IDF dalam klasifikasi menggunakan algoritma *Naïve Bayes Classifier* pada analisis sentiment tempat wisata di Nusa Tenggara Barat. Dengan menggunakan 5 *keyword* yang dapat mewakili tempat wisata tersebut dan menggunakan sejumlah 10.000 *tweet* dalam pemrosesannya.

2. Tujuan

Tugas akhir ini bertujuan untuk melakukan evaluasi terhadap hasil yang didapatkan dari penerapan TF-IDF dalam klasifikasi, serta melakukan analisis sentimen berdasarkan *tweet* untuk mengetahui performa algoritma *Naïve Bayes Classifier* dalam melakukan klasifikasi berdasarkan *tweet* pariwisata di Provinsi Nusa Tenggara Barat.

3. Organisasi Tulisan

Bagian selanjutnya akan menjelaskan tentang pemaparan terkait studi literatur pada Bab 2, kemudian pemaparan mengenai perancangan sistem pada Bab 3, pembahasan mengenai analisis implementasi sistem pada Bab 4 dan pada Bab 5 akan dibahas mengenai kesimpulan akhir dari penelitian tugas akhir ini.

II. KAJIAN TEORI

A. Studi Terkait

Penelitian yang sudah pernah dilakukan sebelumnya terkait sentimen analisis pada tahun 2018 oleh [10] yaitu dengan data yang berasal dari twitter. Ruang lingkup penelitian itu adalah opini publik mengenai wisata pantai. Klasifikasi ini menggunakan algoritma Support Vector Machine. Dataset yang digunakan dalam penelitian ini mengambil dari 10 pantai yang ada di Indonesia dan menggunakan 500 tweet. Kemudian data opini juga ditambahkan untuk mengelompokan wisata pantai melihat dari situasi dan kondisi terkini, ketersediaan sumber daya, fasilitas, kondisi akses perjalanannya, kondisi kesiapan masyarakat, potensi pasar dan posisi rekomendasi pariwisata yang akan dituju. Pengelompokan data ini menggunakan metode K-Means.

Selain itu pada tahun 2017 dilakukan sebuah penelitian yang dilakukan oleh [13] dengan fokus yang masih sama berkaitan dengan pariwisata namun ruang lingkup pada penelitian ini yaitu Provinsi Jawa Barat. Sumber data yang diambil menggunakan Google Maps yang berfokus pada data tempat wisata, sentimen pengunjung dan rating tempat wisata. Tahapannya ketika sudah melakukan pengambilan data dari Google Maps, dilakukan preprocessing yang bertujuan untuk membersihkan data dari hal yang tidak diperlukan. Untuk tahap klasifikasi, menggunakan metode Naïve Bayes Classifier.

Selain itu pada tahun 2018 terdapat sebuah penelitian terkait sentimen analisis untuk mengklasifikasikan tingkat kemungkinan obesitas mahasiswa oleh [8] yang juga menggunakan metode naïve bayes classifier. Dengan ruang lingkup mahasiswa dengan jurusan Sistem Informasi UIN. Mengambil sampel secara acak sebanyak 88 orang. Untuk kriteria untuk klasifikasi kemungkinan penderita obesitas antara lain lingkar perut, berat badan dan tinggi badan.

Kemudian pada tahun 2010, penelitian oleh [15] dengan metode pengumpulan korpus otomatis untuk melatih pengklasifikasian sentimen. Menggunakan Treetragger untuk POS-tagging guna mengamati perbedaan distribusi antara set positif, negatif ataupun netral. Untuk kelas netral peneliti [15] mengambil data twit sebagai data training dari akun media internasional yang berbahasa Inggris. Setelah menjalankan dan didapatkan kesimpulan menggunakan bigram adalah hasil performasi yang terbaik.

B. Sentimen Analisis

Sentimen analisis merupakan suatu bidang ilmu yang mempelajari tentang bagaimana menganalisis pendapat, sentiment dan emosi seseorang terhadap suatu produk, topik, masalah, organisasi bahkan individu. Lebih ringkasnya, Bidang ilmu ini melakukan pengelompokan dari suatu teks yang ada di dalam dokumen, kalimat atau fitur. Kemudian memberikan status terhadap apa yang sudah dikelompokan menjadi 3 kategori, diantaranya positif, netral dan negatif. Pada akhirnya bisa ditarik kesimpulan bahwa apa yang sudah dikelompokan sebelumnya bisa mewakili emosi sedih, marah atau gembira [14]

Sentimen mengacu pada fokus topik tertentu, pernyataan satu topik yang sama akan berbeda jika pada subjek yang berbeda. Sentimen analisis berguna untuk melihat kecenderungan pendapat atau opini terhadap suatu masalah, objek atau suatu keadaan yang bertujuan agar seseorang memiliki opini antara positif atau negatif. Beberapa penelitian yang terutama berfokus pada review produk, sebelum melakukan proses opinion mining dilakukan proses pengerjaan elemen dari sebuah produk yang sedang dibicarakan atau difokuskan.[2].

Tahapan yang dilakukan dalam memproses suatu dokumen diawali dengan memecah dari yang semula kumpulan karakter menjadi penggalan kata. Proses itu dinamakan tokenisasi. Data yang ditokenisasi diantaranya berupa kalimat, paragraph atau bahkan dokumen. Ketika proses tokenisasi akan ditemukan karakter delimiter. Delimiter adalah karakter tertentu yang terdiri dari spasi, tabulasi dan juga baris baru. Kemudian ada juga karakter seperti {( ) <> ! ? “} yang kadang dijadikan delimiter namun ada disuatu kondisi tidak termasuk delimiter. Karena tergantung dengan kalimat sebelumnya [3].

C. Ekstraksi Fitur

Pada tugas akhir ini menggunakan metode Term Frequency Inverse Document Frecquency (TF-IDF) yang merupakan pengembangan dari teknik Term Frequency. Cara kerja Term Frequency (TF) yaitu dengan cara melakukan pembobotan jumlah kemunculan kata t terhadap dokumen d yang dinotasikan dengan tft,d

Term Frequency Inverse Document Frecquency (TFIDF) adalah hasil perkalian dari nilai tft dengan idft yang dinotasikan dengan tf-idft,d:

$$TF - IDF(ti, dj) = tf(ti, dj) \log \frac{N}{ni} \quad (2.1)$$

Keterangan :
TF-IDF(ti,dj) didapatkan dari pembobotan kata atau *term* i pada dokumen j
tf(ti,dj) adalah banyak kata atau *term* i pada dokumen j
ni adalah total dokumen yang munculkan term i
N merupakan total dokumen pada dataset

D. Naïve Bayes Classifier

Naïve Bayes Classifier adalah sebuah metode dalam pengklasifikasian yang sederhana berdasarkan teorema Bayes dengan asumsi naif yang kuat seperti dari namanya. Naïve Bayes Classifier membuat asumsi ketika kehadiran atau ketidaan sebuah fitur tertentu dari suatu kelas tidak berhubungan dengan fitur lainnya.

Naïve Bayes Classifier termasuk dalam supervised learning. Dalam situasi dan kondisi tertentu, pemodelan ini dapat dilatih dari probabilitas yang ada. Pemodelan ini memakai likely hood maksimum sebagai metode yang dengan kata lain seseorang dapat menggunakan model naïve bayes tanpa harus melihat probabilitas Bayesian [7].

Penggunaan ini memiliki keuntungan yaitu hanya membutuhkan data latih yang tidak besar untuk menentukan parameter yang diperlukan untuk proses pengklasifikasian. Variable data bersifat independen menjadikan dalam pemrosesannya tidak membutuhkan keseluruhan data dari matriks kovarians.
Rumus Bayes [8] secara umum sebagai berikut:

$$P(H|X) = \frac{P(X|H)}{P(X)} . P(H) \quad (2.2)$$

Keterangan :
P(H|X) : Peluang hipotesis H berdasarkan kondisi x (*posteriori probabilitas*)
P(X|H) : Peluang X berdasarkan kondisi pada spekulasi tertentu
P(X) : Peluang dari X
P(H) : Peluang spekulasi H (prior prob.)
X : Data untuk class yang belum diketahui
H : Spekulasi data X merupakan suatu class spesifik

Untuk kelas klasifikasi dengan data kontinu digunakan rumus *Densitas Gauss* sebagai berikut:

$$P(X_i = x_i|Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{\frac{(x_i-\mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2.3)$$

P adalah peluang
σ adalah deviasi Standar menyatakan varian dari seluruh atribut
X_i adalah atribut dari data ke i
Y adalah kelas yang dicari
μ adalah mean, menyatakan rata-rata dari seluruh atribut
x_i adalah nilai dari atribut dari data ke i
y_i adalah bagian dari sub kelas yang akan dicari

E. Performance Evaluation Measure

Pada tahap ini memiliki fungsi untuk mengukur performa suatu sistem. PEM dalam pengimplementasiannya digunakan dalam data latih yang bertujuan untuk melihat dan mengevaluasi pemodelan yang sudah dibuat. Beberapa perhitungan dalam PEM antara lain :

1. Accuration

Accuration adalah rasio prediksi informasi yang dihasilkan oleh sistem benar dengan keseluruhan informasi yang ada.
Rumus *accuration*

$$acc = \frac{TN+TP}{FN+FP+TN+TP} \quad (2.4)$$

2. Precision.

Precision adalah rasio ukuran data yang aktualnya benar dibandingkan dengan keseluruhan data yang diprediksi oleh sistem.
Rumus *precision*

$$pre = \frac{TP}{FP+TP} \quad (2.5)$$

3. Recall.

Recall adalah rasio ukuran data yang diprediksi benar dibandingkan dengan keseluruhan data yang aktualnya nya benar.
Rumus *recall*

$$rec = \frac{TP}{FN+TP} \quad 2.6)$$

TABEL 1
Confusion Matrix

	Predicted Class	
True Class	**Negative**	**Positive**
Negative	TN	FP
Positive	FN	TP

Keterangan :
TP = *true positive* : data aktual yang bernilai positif diprediksi benar sebagai positif
FP = *false positive* : data aktual yang bernilai negatif namun diprediksi salah sebagai positif
TN = *true negative* : data aktual yang bernilai negatif diprediksi benar sebagai negatif
FN = *false negative* :data aktual yang bernilai positif namun diprediksi salah sebagai negative

F. Twitter

Twitter adalah salah satu platfom media sosial yang memberikan sebuah layanan yang menghubungkan antar pengguna untuk bertukar informasi secara cepat dan aktual. Pengguna akan menuliskan berupa tweet yang berisi foto, video atau teks. Lalu tweets yang sudah dituliskan pengguna akan di publish bisa ke profil pengguna, mengirimnya

kepada seseorang atau bahkan tweets bisa dicari dalam mesin pencarian didalam twitter [11].

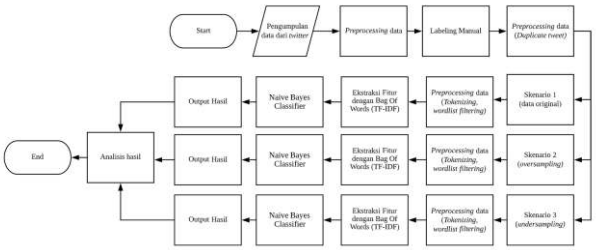
Sumber data yang diambil dari twitter. Menggunakan bahasa pemrograman python dan menggunakan library yang sudah ada.

III. METODE

A. Perancangan Sistem

1. Gambaran Umum Sistem.

Gambaran sistem secara umum dapat dilihat pada gambar 3.1.



GAMBAR 3.1
Flowchart sistem yang akan dibangun.

2. Pengumpulan Data

Pengumpulan data pada penelitian ini didapatkan dari sumber *twitter*. Data yang dikumpulkan berupa teks *tweet* menggunakan teknik *scrapping* dengan *python* dan *library* yang sudah ada. *Tweet* diambil dengan menggunakan *keyword* yang sudah bisa mewakili tempat-tempat wisata yang ada di Nusa Tenggara Barat, yaitu Pantai Senggigi, Gunung Rinjani, Pantai Kuta Lombok, Gili Trawangan dan Desa Sade. Data yang diambil dalam penelitian ini sebanyak 10.000 buah *tweet*. Untuk rentang waktu, dengan membatasi *keyword* dan 10.000 *tweet* mulai dari tanggal 2023-01-13 sampai 2020-07-08.

3. Preprocessing

Pada tahapan ini sangat dibutuhkan agar lebih optimal dalam perhitungannya sebelum memasuki tahap pengklasifikasian naïve bayes. Cara kerjanya dengan menhilangkan atau membersihkan hal yang tidak diperlukan dalam pemrosesan data.

a. Filtering Duplicate tweets

Data yang didapatkan pasti masih banyak yang duplikat. Maka dengan begitu data memasuki proses ini untuk menghilangkan data-data yang duplikat.

b. Case folding

Dengan cara kerja merubah semua huruf *uppercase* menjadi *lowecase*.

TABLE 2.
Contoh Case Folding

Input	Output
Mahal beud ! !\n\nTarif parkir di KEK Mandalika mahal, sehingga banyak dikeluhkan oleh wisatawan yg berkunjung ke pantai Kuta Lombok Tengah. parahnya, konon katanya uang hasil parkir trsb tdk jelas masuk kemana. https://t.co/janxm3fqiq	mahal beud ! !\n\ntarif parkir di kek mandalika mahal, sehingga banyak dikeluhkan oleh wisatawan yg berkunjung ke pantai kuta lombok tengah. parahnya, konon katanya uang hasil parkir trsb tdk jelas masuk kemana. https://t.co/janxm3fqiq

c. Cleaning

Tahapan ini menghilangkan karakter selain huruf dan dianggap *delimiter*

Table 3.
Contoh Cleaning

Input	Output
mahal beud ! !\n\ntarif parkir di kek mandalika ma hal, sehingga banyak dikeluhkan oleh wisatawan y g berkunjung ke pantai kuta lombok tengah. parah nya, konon katanya uang hasil parkir trsb tdk jelas masuk kemana. https://t.co/janx m3fqiq	mahal beud tarif parkir di kek mandalika mahal sehingga banyak dikeluhkan oleh wisatawan yg berkunjung ke pantai kuta lombok tengah parahnya konon katanya uang hasil parkir trsb tdk jelas masukkemana

d. Tokenizing

Memotong kalimat menjadi potongan kata kata string input.

TABLE 4
Contoh Tokenizing

Input	Output
mahal beud tarif parkir di kek mandalika mahal seh ingga banyak dikeluhkan oleh wisatawan yg berku njung ke pantai kuta lombok tengah parahnya kono n katanya uang hasil parkir trsb tdk jelas masuk ke mana	['mahal', 'beud', 'tarif', 'parkir', 'di', 'kek', 'mandalika', 'mahal', 'sehingga', 'banyak', 'dikeluhkan', 'oleh', 'wisatawan', 'yg', 'berkunjung', 'ke', 'pantai', 'kuta', 'lombok', 'tengah', 'parahnya', 'konon', 'katanya', 'uang', 'hasil', 'parkir', 'trsb', 'tdk', 'jelas', 'masuk', 'kemana']

e. Filtering

Tahapan ini menggunakan algoritma *wordlist* yaitu dengan cara kerja membandingkan dan memilih hanya kata kata yang didalam list positif atau negatif saja yang diambil hasil *tokenizing*.

TABLE 5
Contoh Filtering

Input	Output
['mahal', 'beud', 'tarif', 'parkir', 'di', 'kek', 'mandalika ', 'mahal', 'sehingga', 'banyak', 'dikeluhkan', 'oleh', ' wisatawan', 'yg', 'berkunjung', 'ke', 'pantai', 'kuta', 'l ombok', 'tengah', 'parahnya', 'konon', 'katanya', 'uang', 'hasil', 'parkir', 'trsb', 'tdk', 'jelas', 'masuk', 'kemana',]	['mahal', 'mahal', 'banyak', 'dikeluhkan', 'pantai', 'jelas', 'harusnya', 'perhatian']

4. Labeling

Tweet akan dilabelkan secara positif atau negatif tidak hanya dari sekedar kata-kata yang ada, namun juga dari makna tersembunyi didalamnya. Oleh sebab itu, dilakukan proses labeling secara manual dikarenakan mesin tidak bisa menentukan status data positif atau negatif secara tersirat. Ada disuatu kondisi ketika sebenarnya tweet bermakna positif namun dalam penyampaian atau kata-kata menggunakan konotasi negatif dan begitupula sebaliknya.

Dalam pelabelan, tweet yang bersifat positif akan ditandai dengan 1 dan untuk negatif ditandai dengan 0. Proses ini dilakukan oleh 3 orang. Hasil tersebut akan dibandingkan, jika dalam satu data terdapat 2 yang bersifat positif dan 1 negatif, maka bisa ditarik kesimpulan bahwa data tersebut positif begitu pula dengan sebaliknya.

TABEL 6
Contoh Labeling Manual

Orang 1	Orang 2	Orang 3	Hasil Perbandingan	Tweet
0	0	1	0	cuaca jelek balai taman gunung rinjani imbau pendaki hatihati
1	1	1	1	aksi clean up jalur pendakian gunung rinjani

5. Preprocessing (duplicate tweet)

Setelah dilakukan tahap labeling, ternyata masih terdapat beberapa *tweet* ganda. Karena dalam proses *preprocessing* sebelumnya beberapa tidak terdeteksi duplikat dikarenakan ada *tweet* dengan kata-kata yang sama namun dengan *attachment* yang berbeda seperti tautan *url,* gambar atau video. Dari proses ini menghasilkan 5.229 *tweet* bersih.

6. Skenario Pengujian

Dalam pengujian akan dilakukan 3 skenario. Pertama dengan data apa adanya yang masih bersifat tidak seimbang antara positif dengan negatif. Kedua, dengan metode *oversampling* dimana cara kerjanya dengan membuat replika (*resample)* data minoritas. Ketiga, menggunakan metode *undersampling*. Dengan cara kerja membuat mengurangi data mayoritas.

7. Preprocessing (tokenizing dan wordlist filtering)

Masing masing dari ketiga skenario akan menjalankan proses *tokenizing* dan *wordlist filtering*. Karena dibutuhkan untuk memasuki tahap berikutnya yaitu tahap pembobotan kata dengan TF-IDF. Proses *tokenizing* dilakukan dengan memisahkan kalimat menjadi potongan kata. Kemudian memasuki tahap *wordlist filtering* dengan cara kerja membandingkan potongan kata dengan list kata penting yang sudah disiapkan.

8. Ekstraksi Fitur

Dalam ekstraksi fitur atau pembuatan fitur, dilakukan proses pengubahan fitur teks menjadi sebuah gambaran vektor dan menggunakan TF-IDF dalam pembobotannya yang disebut *word vector.* Ketika proses pembobotan sudah dilakukan, maka sudah siap untuk lanjut ke tahap berikutnya, dataset akan ditraining menggunakan perhitungan *naïve bayes.*

9. Pengklasifikasian Naïve Bayes Classifier

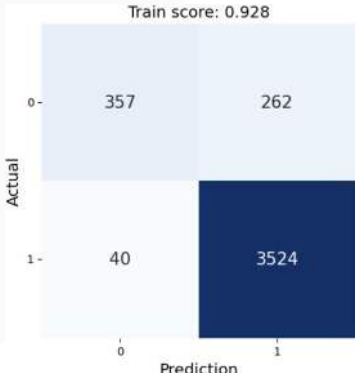
Metode *naïve bayes* adalah metode pengklasifikasikan data yang sudah ada sebelumnya. Dimana data yang sebelumya didapatkan berjumlah 10.000 setelah memasuki tahap *preprocessing* mendapatkan hasil akhir data bersih yang siap diproses berjumlah 5.229 *tweet*. Kemudian data tersebut akan terbagi menjadi dua jenis data, *data training* dan *data testing*. Dalam penelitian ini teknis pembagiannya *data training* 80% dan *data testing* 20%. Setelah data berhasil dalam proses pelatihan kemudian akan dilakukan pengujian menggunakan antara data latih dengan data tes untuk menguji hasil klasifikasi dan mengevaluasi hasil yang didapatkan.

IV. HASIL DAN PEMBAHASAN

A. Evaluasi

1. Hasil Pengujian Skenario 1

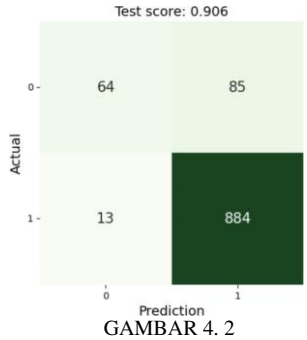
Pada scenario ini akan dilakukan pengujian dengan data apa adanya tidak ada penyesuaian apapun. Dengan data positif 85% sebanyak 4.461 data dan 15% data negatif sebanyak 768 data. Pada proses ini, 5.229 data akan terbagi menjadi *data training* 80% dari keseluruhan data yang berarti sebanyak 4.183 data dan *data testing* 20% sebanyak 1.046 data. Angka 1 mewakili *tweet* yang bersifat positif dan 0 bersifat negatif. Berdasarkan hasil TF-IDF didapatkan sebanyak 920 fitur dari 4.183 data latih. Berikut ini adalah hasil pesebaran data dari proses *training*.



GAMBAR 4. 1
Confusion matrix data training (original)

Dalam *confusion matrix* diatas terdapat 40 data dengan kondisi sebenarnya bersifat positif namun terprediksi negatif (*false negative)* dan juga ada 262 pada kondisi sebenarnya bersifat negatif namun terprediksi positif (*false positive*). Kemudian ada 3.881 *tweet* yang berhasil terprediksi sesuai dengan aslinya, dimana 3.524 *tweet* positif (*true positive*) dan 357 *tweet* negatif (*true negative*). *Train score* didapatkan dari proses yang berhasil terprediksi sesuai dengan *tweet* asli dibagi dengan total semua *data train* yang diproses. Menghasilkan 0,928 dibulatkan dan

dipersentasekan menjadi 93%. Untuk *data testing* sebagai berikut.



GAMBAR 4. 2
Confusion matrix data testing.

Pada *confusion matrix data testing* terdapat 948 *tweet* yang berhasil terprediksi sesuai dengan kondisi sebenarnya. 884 data positif (*true positive*) dan 64 data negatif (*true negative*). Sementara ada 13 *tweet* yang pada kondisi sebenarnya positif namun terprediksi negatif (*false negative)* dan 85 *tweet* terprediksi positif padahal pada kondisi sebenarnya bersifat negatif (*false positive*). *Test score* didapatkan dari proses yang berhasil terprediksi sesuai dengan *tweet* asli dibagi dengan total semua *data train* yang diproses. Menghasilkan 0,906 dibulatkan dan dipersentasekan menjadi 91% yang bisa juga dikatakan bahwa ini adalah tingkat akurasi yang kita dapatkan.

	precision	recall	f1-score	support
0	0.83	0.43	0.57	149
1	0.91	0.99	0.95	897
accuracy			0.91	1046
macro avg	0.87	0.71	0.76	1046
weighted avg	0.90	0.91	0.89	1046

GAMBAR 4. 3
Performance Evaluation Measure original

Dengan pemodelan diatas bisa dihitung tingkat accracy, precision dan recall. Accracy adalah rasio prediksi benar (positif dan negatif) dengan keseluruhan data tweet yang ada. Jadi tingkat persentase tweet yang benar diprediksi positif dan negatif sebesar 91% dari keseluruhan data.

Precision positif (1) senilai 0,91 atau 91% dan negatif (0) negatif adalah 0,83 atau 83%. Maka bisa dihitung macro avg nya senilai 0.87 atau 87%.

Recall positif (1) adalah 0,99 atau 99% dan negatif (0) 0,43 atau 43%. Kemudian bisa dihitung macro avg nya senilai 0.71 atau 71%.

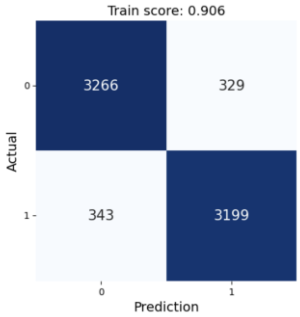
F1-score merupakan perbandingan nilai rata-rata antara precision dan recall yang dibobotkan. Indikator F1- score memperhitungkan pentingnya tweet yang terprediksi namun tidak sama dengan kondisi aktualnya. Untuk mendapatkan nilai F1-score berikut adalah rumus F1-score.

$$\text{F1score} = 2 \times \left(\frac{(\text{Precission} \times \text{Recall})}{(\text{Precission} + \text{Recall})}\right) \qquad (4.1)$$

Dengan memasukan nilai (*precision* dan *precision)* positif (1) yang sudah ada, menghasilkan *F1-score* sehesar 0,95 atau 95%. Kemudian untuk nilai *F1-score* (*precision* dan *precision)* negatif (0) adalah 0,57 atau 57%. Macro avg nya senilai 0,76 atau 76%.

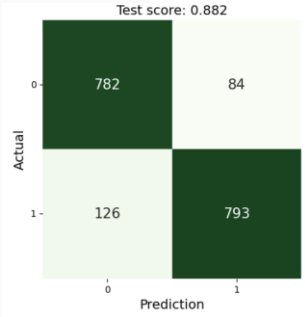
2. Hasil Pengujian Skenario 2

Setelah dilakukan *oversampling* menjadi sama antara data positif dan negatif sebanyak 4.461 data, dengan total 8.922 yang akan diproses. Kemudian akan dibagi menjadi 2 data, *data training* sebanyak 7.137 dan *data testing* sebanyak 1785. Berdasarkan hasil TF-IDF didapatkan sebanyak 920 fitur dari 7.137 data latih. Berikut ini adalah hasil pesebaran data dari proses *training.*



GAMBAR 4. 4
Confusion matrix data training (oversampling)

Dalam *confusion matrix* diatas terdapat 343 *false negative* dan juga ada 329 *false positive*. Kemudian ada 3.199 *true positive* dan 3.266 *true negative*. *Train score* 0,906 dibulatkan dan dipersentasekan menjadi 90%. Untuk *data testing* sebagai berikut.



GAMBAR 4. 5
Confusion matrix data testing

Pada *confusion matrix data testing* terdapat 793 *true positive* dan 782 *true negative*. Sementara ada 126 *false negative* dan 84 *false positive*. *Test* 0,882 dibulatkan dan dipersentasekan menjadi 88% yang bisa juga dikatakan bahwa ini adalah tingkat akurasi yang kita dapatkan.

	precision	recall	f1-score	support
0	0.86	0.90	0.88	866
1	0.90	0.86	0.88	919
accuracy			0.88	1785
macro avg	0.88	0.88	0.88	1785
weighted avg	0.88	0.88	0.88	1785

GAMBAR 4. 6
Performance Evaluation Measure oversampling

Accracy adalah rasio prediksi benar (positif dan negatif) dengan keseluruhan data *tweet* yang ada. Sebesar 88% dari keseluruhan data.

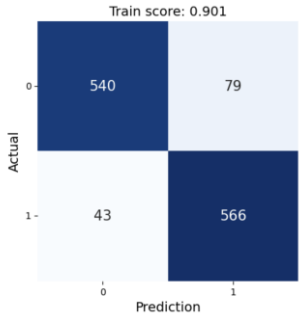
Precision positif (1) senilai 0,90 atau 90% dan negatif (0) sebesar 0,86 atau 86%. Maka bisa dihitung macro avg nya senilai 0.88 atau 88%.

Recall positif (1) senilai 0,86 atau 86% dan negatif (0) sebesar 0,90 atau 90%. Kemudian bisa dihitung macro avg nya senilai 0.88 atau 88%.

Dengan memasukan nilai (*precision* dan *precision)* positif (1) yang sudah ada, menghasilkan *F1-score* sehesar 0,88 atau 88%. Kemudian untuk nilai *F1-score* (*precision* dan *precision)* negatif (0) adalah 0,88 atau 88%. Macro avg nya senilai 0,88 atau 88%.

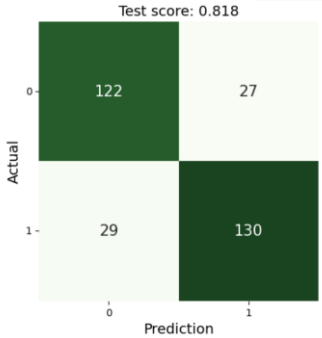
3. Hasil Pengujian Skenario 3

Setelah dilakukan *undersampling* menjadi sama antara data positif dan negatif sebanyak 768 data. Dengan total 1.536 yang akan diproses. Kemudian akan dibagi menjadi 2 data, *data training* sebanyak 1.228 dan *data testing* sebanyak 308. Berdasarkan hasil TF-IDF didapatkan sebanyak 515 fitur dari 1.228 data latih. Berikut ini adalah hasil pesebaran data dari proses *training*.



GAMBAR 4. 7
Confusion matrix data training (undersampling)

Dalam *confusion matrix* diatas terdapat 43 *false negative* dan juga ada 79 *false positive*. Kemudian ada 566 *tweet* positif (*true positive*) dan 540 *tweet* negatif (*true negative*). *Train score* seniali 0,901 dibulatkan dan dipersentasekan menjadi 90%. Untuk *data testing* sebagai berikut.



GAMBAR 4. 8
Confusion matrix data testing

Pada *confusion matrix data testing* terdapat 130 *true positive* dan 122 data *true negative*. Sementara ada 29 *false negative* dan 27 *false positive*. *Test score* senilai 82% yang bisa juga dikatakan bahwa ini adalah tingkat akurasi yang kita dapatkan.

	precision	recall	f1-score	support
0	0.81	0.82	0.81	149
1	0.83	0.82	0.82	159
accuracy			0.82	308
macro avg	0.82	0.82	0.82	308
weighted avg	0.82	0.82	0.82	308

GAMBAR 4. 9
Performance Evaluation Measure undersampling

Accracy sebesar 82% dari keseluruhan data.

Precision positif (1) senilai 0,83 atau 83% dan negatif (0) sebesar 0,82 atau 82%. Maka bisa dihitung macro avg nya senilai 0.82 atau 82%.

Recall positif (1) adalah 0,82 atau 82% dan negatif (0) senilai 0,82 atau 82%. Kemudian bisa dihitung macro avg nya senilai 0.82 atau 82%.

Dengan memasukan nilai (*precision* dan *precision)* positif (1) yang sudah ada, menghasilkan *F1-score* sehesar 0,82 atau 82%. Kemudian untuk nilai *F1-score* (*precision* dan *precision)* negatif (0) adalah 0,81 atau 81%. Macro avg nya senilai 0,82 atau 82%.

4. Analisis Hasil Pengujian

Berdasarkan pengujian yang sudah dilakukan dengan 3 skenario. Terdapat beberapa faktor yang dapat mempengaruhi hasil akhir. Dalam skenario 1, kondisi positif dan negatif yang tidak seimbang atau terpaut jauh selisihnya maka pemodelan akan cenderung bernilai terhadap data mayoritas, dimana data mayoritas ini adalah data positif. Kemudian dilakukan skenario 2 dan 3 bertujuan untuk menyeimbangkan atau mengurangi selisih data yang terpaut jauh ketimpangan sebelumnya. Didapatkan peningkatan *F1-score* lumayan signifikan dari skenario 1. Jika dilihat perbandingan antara skenario 2 dan skenario 3, bisa dilihat bahwasannya jumlah data juga akan mempengaruhi hasil akhir.

Kemudian pemilihan *keyword* dalam pengambilan data tidak tepat maka hasilnya pun tidak akan optimal. Juga tidak bisa mewakili kondisi aslinya karena ada beberapa tempat wisata di Nusa Tenggara Barat yang memiliki nama yang bermakna ganda seperti Pantai Pink dan Air Terjun Benang Kelambu. Ataupun nama tempat wisata yang tidak hanya ada di Nusa Tenggara Barat tetapi juga ada di provinsi lainnya.

V. KESIMPULAN

A. Kesimpulan

Dari hasil analisis studi terkait, dapat disimpulkan bahwa evaluasi terhadap hasil yang didapatkan dari penerapan TF-IDF dalam pembobotan *tweet* dan pengaruhnya dalam klasifikasi penggunaan metode Naïve Bayes Classifier berdasarkan *tweet* tempat wisata di Nusa Tenggara Barat dapat mengklasifikasikan kalimat *tweet* sudah cukup baik dan tepat. Dari 3 skenario yang dijalankan, masing-masing menghasilkan macro *F1-score* skenario 1 senilai 76%, skenario 2 senilai 88% dan skenario 3 senilai 82%. Terdapat beberapa faktor yang menyebabkan optimal tidaknya hasil yang didapatkan yaitu pemilihan *keyword* dan jumlah data atau *tweet.* Untuk penelitian selanjutnya, penulis menyarankan untuk menambahkan

lebih banyak data dan pemilihan *keyword* untuk mendapatkan hasil lebih optimal.

REFERENSI

[1] A. P, I. R and P. P, "Sentiment Analysis in Tourism," *IJISET - International Journal of Innovative Science, Engineering & Technology,* vol. I, no. 9, 2014.

[2] Pang, Bo and Lee, L, Vaithyanathan, S. 2002. "Sentiment Classification Using Machine Learning Techniques". Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing (EMNLP-02). USA, 2002.

[3] Triawati, C. (2009). Text Mining. Bandung, Jawa Barat, Indonesia

[4] Turban, E., & Liang, J. E. A. and T. P. (2005). Decision Support System and Intelligent Systems (7th ed). Pearson Education, Inc.

[5] Larose, D. T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.

[6] Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis.Yogyakarta: Graha Ilmu.

[7] Sumartini Saraswati, N. W. (2011). Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. Denpasar, Bali, Indonesia.

[8] Muslehatin, W., Ibnu, M., & Mustakim. (2017). Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, 250-256.

[9] Huang, Y.-A., You, Z. H., Chen, X., Chan, K., & Luo, X. (2016). Sequence-Based Prediction of Protein- Protein Interactions using Weighted Sparse Representation Model Combined with Global Encoding. BMC Bioinformatics, 17:184.

[10] Syaifudin, Yan W., and Rizki A. Irawan. "Implementasi Analisis Clustering Dan Sentimen Data Twitter Pada Opini Wisata Pantai Menggunakan Metode K-means." *Jurnal Informatika Polinema*, vol. 4, no. 3, 2018, doi:10.33795/jip.v4i3.205.

[11] Murnawan, Murnawan. (2017). PEMANFAATAN ANALISIS SENTIMEN UNTUK PEMERINGKATAN POPULARITAS TUJUAN WISATA. Jurnal Penelitian Pos dan Informatika. 7. 109. 10.17933/jppi.2017.070203.

[12] "Twitter," [Online]. Available: https://help.twitter.com/en/new-user-faq. [Accessed 19 April 2020].

[13] L. Wilianto, T. H. Pudjiantoro and F. R. Umbara, "Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naïve Bayes Classifier Studi Kasus Jawa Barat,"
Cimahi, 2017

[14] Liu, B., 2012, Sentiment Analysis and Opining Mining. Morgan & Claypool Publishers.

[15] Pak, A., dan Paurobek, P., (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining,*
Universite de Paris-Sud, Laboratoire LIMSI-CNRS.

Liu, B., 2010, Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau). Chapman and Hall/CRC, USA.