

ANALISIS SENTIMEN ULASAN PENGUNJUNG WISATA LAGUNA WAIKURI PADA GOOGLE MAPS MENGGUNAKAN MULTI KELAS SUPPORT VECTOR MACHINE

Jefreydo Exelino Bata, Istiadi, Syahroni Wahyu Iriananda

Teknik Informatika, Universitas Widyagama Malang

Jl. Borobudur No.35, Kota Malang, Indonesia

exelinobata@gmail.com

ABSTRAK

Laguna Waikuri merupakan objek wisata Laguna berbentuk danau dimana air bersumber dari air laut yang masuk melalui selah-selah batu karang, Laguna yang berada di Nusa Tenggara Timur ini memiliki pemandangan yang indah. Maka dari itu perlu dilakukan analisis terhadap ulasan terhadap Laguna Waikuri untuk mengetahui Opini pengunjung wisata mengenai tempat tersebut. Data diperoleh teknik Scrapping. setelah itu data diberi label Negatif, Netral dan Positif yang disesuaikan dengan rating, selanjutnya pada tahap Praprosesing, data dibersihkan untuk mengurangi atribut kata. Selanjutnya tahap N-GRAM dan TF-IDF. Untuk proses Klasifikasi dibagi menjadi 9 kali pengujian dengan Confusion Matrix dan Algoritma Multi Kelas Support Vector Machine dengan kernel Linear, Polinomial, RBF dan Sigmoid, hasilnya diturunkan ke Akurasi, Presisi, Recall dan F1-Measure. Akurasi tertinggi pada pengujian 90:10 mendapatkan akurasi sebesar 90,67% untuk kernel Polinomial dan RBF. Presisi pada pengujian 10:90 untuk kernel linear, 20:80 kernel linear dan sigmoid, 30:70 dan 40:60 data uji dan data uji dan data latih untuk kernel sigmoid mendapat presisi sebesar 93,85%. recall tertinggi pengujian 50:50 untuk kernel polinomial dan 90:10 data uji dan data latih untuk kernel Polinomial dan RBF mendapatkan Recall sebesar 98,67%. f1-measure tertinggi pada pengujian 90:10 data uji dan data latih untuk kernel Linear, Polinomial dan RBF Mendapatkan F1-Score sebesar 95,09%.

Kata kunci : Analisis Sentimen, Laguna Waikuri, Multi Kelas Support Vector Machine, Pariwisata

1. PENDAHULUAN

Pariwisata pada saat ini sudah menjadi salah satu sektor utama dalam perekonomian global dan juga mendukung pertumbuhan ekonomi jangka panjang dengan pengaruh yang sangat besar dalam meningkatkan pertumbuhan ekonomi pada suatu daerah dan juga sebagai pendapatan negara [1].

Pariwisata didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat serta pemerintah daerah. Pariwisata yang baik harus memiliki daya tarik, kemudahan perjalanan, sarana serta promosi, Pengembangan pariwisata perlu didukung dengan perencanaan yang matang dengan sasaran untuk meningkatkan kualitas hidup [2].

Seiring berkembangnya penggunaan teknologi dan media sosial saat ini telah memberikan dampak yang signifikan terhadap bagaimana seseorang akan mengunjungi suatu destinasi wisata dan juga berbagi pengalaman yang telah dituju [3].

Dengan teknologi yang sudah maju ini tidak menjadi hambatan untuk mencari informasi atau suatu topik yang dibicarakan atau yang sedang viral di media sosial maka dengan demikian pendapat seseorang melalui sikap emosi yang diketikan dalam bahasa yang tertulis adalah bagian dari definisi sentimen [4].

Pariwisata pada saat ini menjadi bahan yang banyak diperbincangkan banyak orang mulai dari anak-anak sampai orang dewasa yang berhubungan dengan berbagai macam hal yang mendukung seperti berbagai sarana, fasilitas layanan/jasa yang disediakan pemerintah, masyarakat, pengusaha untuk

meningkatkan potensi wisata yang menarik dan dapat dikembangkan menjadi destinasi Pariwisata yang baik [5].

Dalam konteks ini analisis sentimen terhadap suatu tempat wisata sudah menjadi sangat penting karena dapat memberikan wawasan dan pengetahuan akan suatu tempat wisata bagi yang belum pernah mengunjungi daerah tersebut. Penelitian mengenai analisis sentimen telah menjadi topik yang menarik dalam memahami fakta, opini, sikap dan penilaian orang terhadap suatu tempat, Ulasan dari internet merupakan hal yang penting karena bisa dijadikan sebagai informasi mengenai suatu tempat [6].

Laguna Waikuri merupakan objek wisata dengan danau air asin yang air dari danau ini bersumber dari air laut yang masuk melalui selah - selah bebatuan karang, danau yang berada di Desa Kalena Rongo, Kecamatan Kodi, kabupaten sumba barat daya dengan jarak sekitar 60 km dari kota Tambolaka ini memiliki pemandangan yang indah dan juga masih sangat alami, Saat ini, keberadaan Laguna Waikuri sudah mulai banyak di ekspos ke kalangan wisatawan lainnya melalui wisatawan yang sudah pernah berkunjung di Laguna Waikuri entah itu dari dalam negeri maupun luar negeri. Namun bagi sebagian orang yang belum pernah berkunjung atau baru pertama kali datang ke Laguna waikuri ini akan akan mengalami kesulitan saat akan mengunjungi danau tersebut, Di karenakan mereka belum mengetahui cukup informasi menuju tempat yang dituju [7].

Berdasarkan pada latar belakang diatas, perlu dilakukan proses analisis sentimen untuk teks tersebut. Analisis sentimen digunakan untuk memahami opini publik terhadap Laguna Waikuri. Pada penelitian ini, penulis menggunakan Multi Kelas Support Vector Machine untuk memprediksi teks dengan 3 kelas, yaitu kelas Negatif, Netral dan Positif yang disesuaikan dengan rating pemberian bintang pada ulasan komentar [8].

Penelitian serupa yang juga dilakukan oleh Syahlan mengenai analisis sentimen menggunakan metode Support Vector Machine, pada penelitian analisis sentimen taman air mancur sri baduga purwakarta proses pengambilan data dilakukan dengan metode scrapping data setelah itu di preprocessing, penelitian dilakukan dengan menggunakan bahasa pemrograman python pada Google Colab. dari hasil eksekusi mengenai analisis sentimen maka didapatkan hasil komentar positif sebanyak 872 kata, negatif 156, netral sebanyak 127 dari 1155 kata yang didapatkan, dan juga persentase yang didapatkan dari hasil eksekusi adalah 81% dan untuk Akurasinya sebesar 94% [9].

Penelitian ini akan menjelaskan bagaimana analisis sentimen menggunakan Multi Kelas Support Vector Machine untuk mengklasifikasikan opini pengunjung wisata terhadap Laguna Waikuri. Adapun tujuan dari penelitian ini adalah menerapkan algoritma Multi Kelas Support Vector Machine untuk melakukan analisis sentimen komentar pengunjung untuk mengetahui seberapa banyak komentar bernilai positif, negatif dan netral yang disesuaikan dengan jumlah bintang atau pemberian rating yang diberikan pengunjung wisata di Laguna Waikuri.

Pada penelitian ini menggunakan Multi Kelas Support Vector Machine untuk analisis sentimen dikarenakan kehandalannya dalam memisahkan kelas, margin yang besar, menangani data yang berdimensi tinggi, fleksibilitas dalam menganalisis data linier maupun yang nonlinier dan juga untuk data yang digunakan untuk penelitian ini terdiri atas 3 buah kelas, jadi dengan menggunakan multi kelas support vector machine adalah pilihan yang tepat.

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Pada penelitian analisis sentimen dengan menggunakan algoritma support vector machine, pada objek taman air mancur sri baduga purwakarta di aplikasi Google Maps ini menggunakan pemrograman Python dengan tools Google Colaboratory, pada saat proses pengambilan data dilakukan dengan metode scrapping data setelah itu di preprocessing menjadi bagian-bagian kecil, hasil yang didapatkan dari hasil eksekusi mengenai analisis sentimen maka didapatkan hasil komentar positif sebanyak 872 kata, negatif 156, netral sebanyak 127 dari 1155 kata yang didapatkan, dan juga persentase yang didapatkan dari hasil eksekusi adalah 81% dan untuk Akurasinya sebesar 94% [10].

Hasil penelitian dari dengan menggunakan Support Vector Machine untuk analisis sentimen objek wisata Danau Toba berdasarkan ulasan pengunjung, dimana hasil pengujian yang dilakukan maka di dapatkan pada hasil klasifikasi sentimen dengan dua kelas sentimen dan tiga kelas sentimen pada masing-masing skenario terkait parameter dari model support vector machine dengan menggunakan kernel linear dan RBF, dimana setelah di eksekusi maka didapatkan hasil akurasi sebesar 83%, precision sebesar 84%, recall sebesar 83% dan f-measure sebesar 83% [11].

Pengujian pada objek wisata dieng menggunakan Algoritma K-Nearest Neighbor (K-NN) dimana proses dilakukan dengan menggunakan dataset untuk data training yang terbagi menjadi klasifikasi 74 positif dan 73 data negatif juga dataset yang digunakan untuk testing sebanyak 37 ulasan. Dan setelah di uji akurasi menggunakan pemrograman python dan didapatkan hasil yang sangat baik, hasil pengujian didapatkan adalah sebuah sistem analisis sentimen yang dapat mengidentifikasi dan mengklasifikasi ulasan pengunjung wisata Dieng dengan menggunakan Metode K-NN dan memperoleh akurasi sebesar 86% [12].

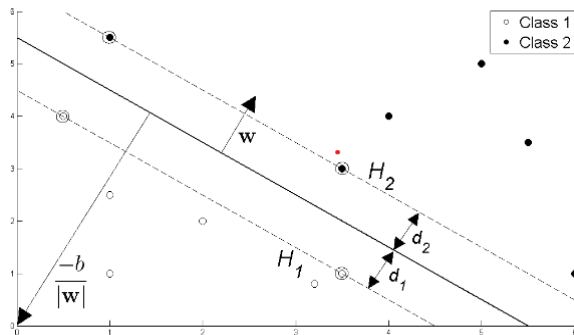
Penelitian yang dilakukan dengan menggunakan algoritma atau metode Naive Bayes Classifier dan K-Nearest Neighbor untuk analisis sentimen komentar pengunjung terhadap danau Weekuri, dengan menggabungkan kedua algoritma untuk perbandingan akurasi yang diuji menggunakan Rapidminer, metode Naive Bayes Classifier memperoleh akurasi sebesar 73.47% sedangkan metode K-Nearest Neighbor mendapat akurasi sebesar 76,53% [13].

2.2. Laguna Waikuri

Laguna Waikuri merupakan sebuah danau air asin yang berada di Desa Kalena Rongo, Kecamatan Kodi Utara, Kabupaten Sumba Barat Daya, Provinsi Nusa Tenggara Timur. Tebing karang tinggi yang memisahkan laut dengan Laguna (Lagoon) ini, dimana air laut masuk melalui celah-celah batu karang, kedalam air dari Laguna (Lagoon) tersebut tergantung pada pasang surut air laut yang ada dibalik tebing karang tersebut. Laguna waikuri memiliki keunikan dalam pembentukannya, dimana air laut dan air sungai mengalir dan terperangkap diantara dinding batu karang, Air ini sangatlah jernih sehingga dapat dengan mudah melihat terumbu karang pada dasar Laguna, selain keindahan laguna tersebut banyak pengunjung yang merasa tidak cukup dengan keindahan laguna tersebut cukup hanya mengagumi pesonanya tapi tertarik dengan merasakan kesejukan dengan mandi dan berenang. Laguna Waikuri banyak dikunjungi saat menjelang siang hingga sore hari, karena Laguna waikuri tidak dibuka selama 24 jam karena jika ingin berkunjung ke Laguna waikuri pengunjung harus berada disana pada jam operasional wisata pukul 08:00 – 17:00, wisata Laguna waikuri ini dibuka setiap hari dengan jam operasional yang sama.

2.3. Multi Kelas Support Vector Machine

Multi Kelas Support Vector Machine (SVM) merupakan algoritma yang digunakan untuk klasifikasi dan regresi yang dirancang untuk mengolah data menjadi Hyperplane yang mengklasifikasi ruang input menjadi dua buah kelas dengan pengelompokan kasus linear yang akan dipisahkan oleh Hyperplane menurut kelasnya, Hyperplane ini tepat berada di tengah-tengah kedua class dengan jarak d ke titik data terdekat untuk masing-masing class. d disebut margin, dan titik-titik data yang berada tepat pada jarak d dari Hyperplane disebut Support Vector [14].



Gambar 1. Hyperplane membagi dua buah class

Hyperplane memaksimalkan margin antara dua kelas data, sehingga SVM adalah model yang baik untuk klasifikasi ketika data dapat dipisahkan secara linear. Titik-titik di luar Support Vectors tidak mempengaruhi Hyperplane, yang membuat SVM tahan terhadap outliers hingga batas tertentu [15].

2.4. Confusion Matrix

Confusion matrix adalah alat evaluasi yang digunakan dalam machine learning untuk mengukur kinerja model klasifikasi. Matriks ini berisi informasi tentang hasil prediksi yang dibuat oleh model dibandingkan dengan nilai sebenarnya dalam dataset uji. Tujuan utama dari confusion matrix adalah untuk memberikan gambaran yang lebih lengkap tentang kinerja model dibandingkan dengan hanya menggunakan metrik akurasi. Dalam mengevaluasi performance algoritma dari machine learning (ML) khususnya supervised learning. Penggunaan Confusion Matriks dapat digunakan sebagai acuan untuk mempresentasikan prediksi dan kondisi sebenarnya dari data yang di hasilkan oleh algoritma Machine Learning (ML). Confusion Matriks bisa menentukan Akurasi, Presisi dan recall dan juga f1-measure [16].

2.4.1. Akurasi

Akurasi merupakan rasio prediksi benar positif dan negatif dan juga netral dan juga merupakan proses yang dapat melihat tingkat keakuratan dalam klasifikasi. Dengan keseluruhan data, Dengan rumus sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

Dimana TP adalah True Positif, TN adalah true negatif, FP adalah false positif dan FN adalah false negatif.

2.4.2. Presisi

Presisi (Precision) merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif untuk menghasilkan nilai yang maksimal. Berikut ini merupakan rumus presisi.

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

2.4.3. Recall

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif, recall juga diperlukan untuk mengukur tingkat keberhasilan proses dari klasifikasi. Berikut ini merupakan rumus Recall.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

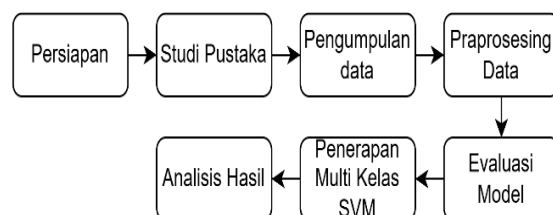
2.4.4. F1-Measure

F1-Measure merupakan perbandingan rata-rata presisi dan recall yang dibobotkan. Dimana F1 Score digunakan ketika ingin mencari keseimbangan antara presisi dan recall atau akan mengambil hasil dari rata-rata yang dihasilkan oleh presisi dan recall. Berikut ini adalah rumus dari F1-Measure.

$$\text{F-measure} = \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4)$$

3. METODE PENELITIAN

Tahapan penelitian ini, untuk data yang digunakan diambil pada ulasan komentar Laguna Waikuri pada Google Maps menggunakan teknik scrapping menggunakan Selenium dan BeautifulSoup pada Library Python dan data yang digunakan adalah data Primer.



Gambar 2. Flowchart Penelitian

Berikut merupakan tahapan yang dilakukan untuk penelitian ini yang terdapat pada gambar flowchart diatas:

3.1. Persiapan

Pada tahapan persiapan penelitian ini menggunakan Multi Kelas Support Vector Machine sebagai metode yang akan digunakan. Dan juga pada tahap persiapan ini juga akan mempersiapkan perangkat – perangkat apa saja yang dibutuhkan untuk kelancaran penelitian ini.

3.2. Studi pustaka

Pada tahap ini penulis akan berfokus untuk mencari referensi yang cocok untuk penelitian ini.

3.3. Pengumpulan data

Pada tahap pengumpulan data, dimana data yang digunakan pada penelitian ini diambil dari kolom komentar ulasan pengunjung Laguna Waikuri pada Google Maps menggunakan teknik scrapping.

3.4. Praprosesing data

Pada tahap ini akan membersihkan data dari kalimat yang kurang bermakna, emoticon dan angka, data yang diambil tidak dapat langsung untuk digunakan dan harus melewati tahap praprosesing.

3.5. Evaluasi model

Setelah melewati tahap praprosesing, selanjutnya akan membuat model klasifikasi menggunakan confusion matrix dengan 4 kernel Linear, Polinomial, RBF dan Sigmoid.

3.6. Penerapan multi kelas SVM

Pada tahap ini model yang sudah dilatih sebelumnya akan di evaluasi dan mengukur kinerja klasifikasi untuk di turunkan ke akurasi, presisi, recall dan f1-measure.

3.7. Analisis Hasil

Pada tahap ini akan menganalisis hasil evaluasi model dengan 4 kernel linear, Polinomial, RBF dan Sigmoid dan juga hasil akurasi, presisi, recal dan f1-measure.

4. HASIL DAN PEMBAHASAN

4.1. Proses Pengumpulan Data

Pengumpulan data dilakukan melalui teknik web scraping yang berisi pendapat, pengalaman dan juga tanggapan pengunjung mengenai Laguna Waikuri, sementara untuk rating pemberian bintang melambangkan tingkat kepuasan pengunjung terhadap tempat wisata tersebut. Pengumpulan data ulasan yang diambil tidak terbatas waktu sedangkan untuk jumlah ulasan dari seluruh ulasan yang diambil sebanyak 1406 ulasan. Pada tabel berikut ini merupakan sampel dari dataset.

Tabel 1. Sampel dataset

Review	Rating
Datang kesini langsung disambut oleh anak ² yang menawarkan jasa foto.....	5
Spot kedua setelah pantai mandorak. Danau cantik yg bersebalahan langsung dgn laut yg dipisahkan oleh tebing. Wajib nyebur klo suka berenang. Ada beberapa lapak yg menjual jajanan, kelapa muda serta souvenir khas sumba.	5
Saya sangat menyukai pemandangan dari tempat ini. Sungguh menakjubkan memiliki danau dan laut di sisinya. Cantik sekali 😊 ...	4

Review	Rating
5 bintang untuk danau dan pemandangan yang indah. Tapi untuk fasilitas dan jalan menuju danau mungkin bintang 3. Tapi sangat layak untuk pergi ke sana ❤️❤️	5

Dataset yang digunakan atau yang diambil hanya terdiri dari ulasan dan juga rating atau pemberian bintang yang diberikan oleh pengunjung.

4.2. Praprosesing Data

Data yang didapatkan dari proses pengumpulan data belum sepenuhnya siap untuk digunakan karena data tersebut masih terdapat emoticon, angka dan juga kata – kata lain yang kurang bermakna maka perlu dilakukan praprosesing data untuk menghilangkan karakter – karakter selain huruf dan juga mengurangi kosakata sehingga dokumen teks lebih efektif untuk dianalisis.

4.2.1. Cleaning

Pada tabel 2 ini merupakan proses cleaning sebelum dan sesudah proses cleaning.

Tabel 2. Cleaning

Sebelum	Sesudah
Danau ini adalah sebuah laguna yang dibatasi dengan laut oleh karang yang tinggi, sehingga air laut hanya bisa masuk melalui pori2 karang. Bisa direnangi. Airnya sangat jernih. Kalau belum berenang di Danau Weekuri, jangan bilang pernah pergi ke Sumba!	Danau Ini Adalah Sebuah Laguna Yang Dibatasi Dengan Laut Oleh Karang Yang Tinggi Sehingga Air Laut Hanya Bisa Masuk Melalui Pori Karang Bisa Direnangi Airnya Sangat Jernih Kalau Belum Berenang Di Danau Weekuri Jangan Bilang Pernah Pergi Ke Sumba

Ditahap ini akan menghapus karakter atau kata yang tidak relevan atau mengganggu seperti tanda baca, angka, emoticon, dan juga simbol – simbol.

4.2.2. Case folding

Pada tabel 3 ini merupakan proses case folding sebelum dan sesudah case folding.

Tabel 3. Case folding

Sebelum	Sesudah
Danau Ini Adalah Sebuah Laguna Yang Dibatasi Dengan Laut Oleh Karang Yang Tinggi Sehingga Air Laut Hanya Bisa Masuk Melalui Pori Karang Bisa Direnangi Airnya Sangat Jernih Kalau Belum Berenang Di Danau Weekuri Jangan Bilang Pernah Pergi Ke Sumba	danau ini adalah sebuah laguna yang dibatasi dengan laut oleh karang yang tinggi sehingga air laut hanya bisa masuk melalui pori karang bisa direnangi airnya sangat jernih kalau belum berenang di danau weekuri jangan bilang pernah pergi ke sumba

Ditahap ini merubah huruf yang semula kapital menjadi huruf kecil, ini dilakukan untuk

menghilangkan perbedaan antara huruf besar dan huruf kecil pada saat analisis data.

4.2.3. Tokenizing

Pada tahap ini merupakan proses tokenizing sebelum dan sesudah tokenizing.

Tabel 4. Tokenizing

Sebelum	Sesudah
danau ini adalah sebuah laguna yang dibatasi dengan laut oleh karang yang tinggi sehingga air laut hanya bisa masuk melalui pori karang bisa direnangi airnya sangat jernih kalau belum berenang di danau weekuri jangan bilang pernah pergi ke sumba	['danau', 'ini', 'adalah', 'sebuah', 'laguna', 'yang', 'dibatasi', 'dengan', 'laut', 'oleh', 'karang', 'yang', 'tinggi', 'sehingga', 'air', 'laut', 'hanya', 'bisa', 'masuk', 'melalui', 'pori', 'karang', 'bisa', 'direnangi', 'airnya', 'sangat', 'jernih', 'kalau', 'belum', 'berenang', 'di', 'danau', 'weekuri', 'jangan', 'bilang', 'pernah', 'pergi', 'ke', 'sumba']

Ditahap ini bertujuan untuk memisahkan teks atau kalimat menjadi bagian – bagian yang kecil agar setiap kata dapat lebih mudah diproses.

4.2.4. Stopword removal

Pada tahap stopwords removal ini bertujuan untuk menghapus kata – kata umum yang tidak memberikan banyak informasi penting untuk dianalisis, seperti kata “dan”, “di”, “ke”, “yang” dan lain sebagainya.

4.2.5. Filtering

Pada tahap ini merupakan proses filtering sebelum dan sesudah filtering.

Tabel 5. Filtering

Sebelum	Sesudah
['danau', 'ini', 'adalah', 'sebuah', 'laguna', 'yang', 'dibatasi', 'dengan', 'laut', 'oleh', 'karang', 'yang', 'tinggi', 'sehingga', 'air', 'laut', 'hanya', 'bisa', 'masuk', 'melalui', 'pori', 'karang', 'bisa', 'direnangi', 'airnya', 'sangat', 'jernih', 'kalau', 'belum', 'berenang', 'di', 'danau', 'weekuri', 'jangan', 'bilang', 'pernah', 'pergi', 'ke', 'sumba']	['danau', 'laguna', 'dibatasi', 'laut', 'karang', 'air', 'laut', 'masuk', 'pori', 'karang', 'direnangi', 'airnya', 'jernih', 'berenang', 'danau', 'weekuri', 'bilang', 'pergi', 'sumba']

Ditahap filtering ini bertujuan untuk menyaring dan membuang kata – kata yang tidak mempunyai makna dan juga mengurangi kata yang tidak memberikan informasi penting.

4.2.6. Normalisasi bahasa

Pada tahap ini merupakan proses normalisasi bahasa sebelum dan sesudah normalisasi bahasa.

Tabel 6. Normalisasi Bahasa

Sebelum	Sesudah
['danau', 'laguna', 'dibatasi', 'laut', 'karang', 'air', 'laut', 'masuk', 'pori', 'karang', 'direnangi', 'airnya', 'jernih', 'berenang', 'danau', 'weekuri', 'bilang', 'pergi', 'sumba']	['danau', 'laguna', 'dibatasi', 'laut', 'karang', 'air', 'laut', 'masuk', 'pori', 'karang', 'direnangi', 'airnya', 'jernih', 'berenang', 'danau', 'weekuri', 'bilang', 'pergi']

Normalisasi bahasa bertujuan menyelaraskan variasi penulisan yang berbeda namun memiliki makna yang sama.

4.2.7. Stemming

Pada tahap ini merupakan proses stemming sebelum dan sesudah stemming.

Tabel 7. Stemming

Sebelum	Sesudah
['danau', 'laguna', 'batas', 'laut', 'karang', 'air', 'laut', 'masuk', 'pori', 'karang', 'renang', 'air', 'jernih', 'renang', 'danau', 'weekuri', 'bilang', 'pergi', 'sumba']	danau laguna batas laut karang air laut masuk pori karang renang air jernih renang danau weekuri bilang pergi sumba

Pada tahap stemming ini mengubah kata – kata menjadi bentuk dasar dan mengelompokkan variasi kata dan dapat meningkatkan efisiensi analisis data teks.

4.2.8. Proses N-GRAM

Pada tahap ini merupakan proses N-GRAM untuk membuat rangkaian dari kata yang berurutan dalam teks. Pada gambar berikut ini adalah sampel proses N-GRAM.

	about	about	about	above	above	absolutely	absolutely	absolutely	absolutely	...	you
	great	to	above	and	the	absolutely	one	place	so		to
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0

Gambar 3. Proses N-GRAM

Proses N-GRAM membantu dalam menangkap konteks dari urutan kata – kata yang mungkin memberikan informasi lebih dibandingkan dengan melihat kata – kata secara individual.

4.2.9. TF-IDF

Pada tahap ini akan melakukan transformasi ke TF-IDF, pada gambar berikut ini sampel dari hasil TF-IDF.

	about	above	absolutely	adat	adik	after	air	aja	ajaib	ajak	...	wooden
0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.0
1	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.0
2	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.0
3	0.0	0.0	0.0	0.0	0.320772	0.0	0.117666	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	0.0	...	0.0

Gambar 4. Proses Transformasi ke TF-IDF

Pada tahap ini mengubah teks menjadi representasi numerik berdasarkan seberapa sering kata yang dalam dokumen (Term Frequency) dan seberapa jarang kata tersebut muncul dalam semua dokumen (Inverse Dokument Frequency).

4.3. Model Confusion Matrix

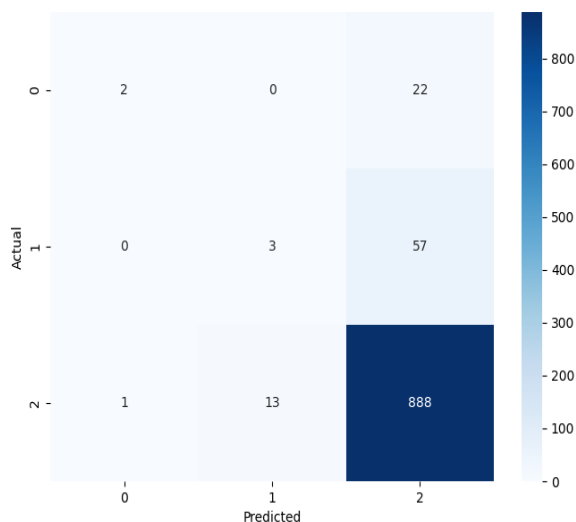
Pada tahap ini akan memaparkan hasil prediksi negatif, netral dan juga positif yang dilakukan oleh model klasifikasi dengan 4 buah kernel Linear, Polinomial, RBF, dan juga Sigmoid. Pengujian ini dilakukan sebanyak 9 kali pengujian, pada tabel ini merupakan proses pembagian data latih dan data uji.

Tabel 8. Pembagian data latih dan data uji

Data Uji	Data Latih
10% (110)	90% (985)
20% (219)	80% (876)
30% (329)	70% (766)
40% (438)	60% (657)
50% (584)	50% (547)
60% (657)	40% (438)
70% (767)	30% (328)
80% (876)	20% (219)
90% (986)	10% (109)

Proses pengujian ini lakukan sebanyak 9 kali pengujian 10% data uji dan 90% data latih sampai 90% data uji dan 10% data latih, hal ini dimaksudkan untuk mendapatkan nilai atau performa yang baik saat pengujian. Dari hasil pengujian tersebut hasil tertinggi dan juga terbaik didapatkan pada pengujian 90% data uji dan 10% data latih dengan menggunakan 4 buah kernel Linear, Polinomial, RBF dan Sigmoid.

Pada hasil pengujian 90% data uji dan 10% data latih untuk kernel Linear, seperti pada gambar berikut ini:



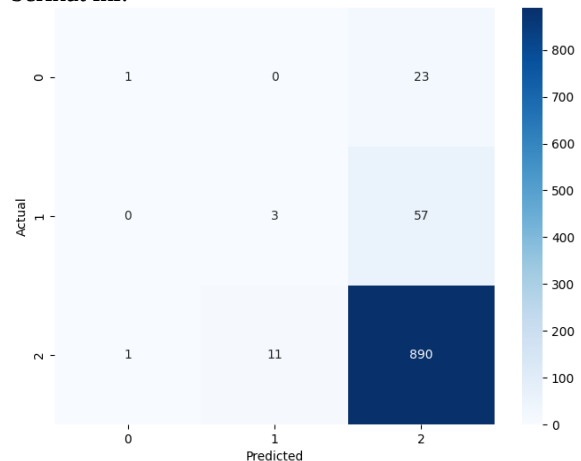
Gambar 5. kernel Linear pengujian 90% data uji dan 10% data latih

Dimana hasil dari matrix seperti dibawah ini.

- Label aktual negatif (0) yang diprediksi negatif terdiri dari 2 data

- Label aktual netral (1) yang diprediksi netral terdiri dari 0 data
- Label aktual positif (2) yang diprediksi positif terdiri dari 1 data
- Label aktual negatif (0) yang diprediksi negatif terdiri dari 0 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 3 data
- Label aktual positif (2) yang diprediksi positif terdiri atas 13 data
- Label aktual negatif (0) yang prediksi negatif terdiri dari 22 data
- Label aktual netral (1) yang prediksi netral terdiri dari 57 data
- Label aktual positif (2) yang prediksi positif terdiri dari 888 data

Pada hasil pengujian 90% data uji dan 10% data latih untuk kernel Polinomial, seperti pada gambar berikut ini:

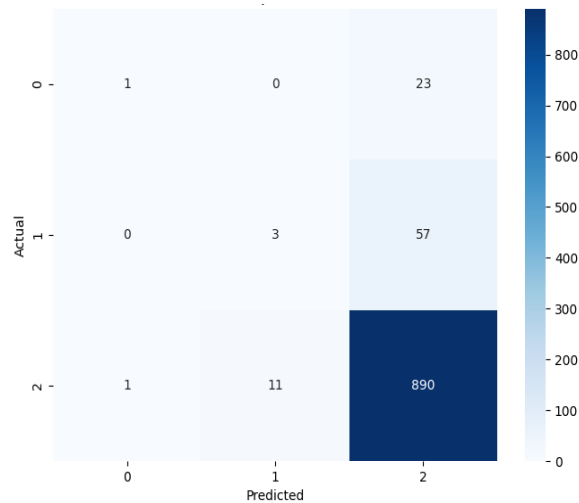


Gambar 6. Kernel Polinomial pengujian 90% data uji dan 10% latih

Dimana hasil dari matrix seperti dibawah ini.

- Label aktual negatif (0) yang diprediksi negatif terdiri dari 1 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 0 data
- Label aktual positif (2) yang diprediksi positif terdiri dari 1 data
- Label aktual negatif (0) yang diprediksi negatif terdiri dari 0 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 3 data
- Label aktual positif (2) yang diprediksi positif terdiri atas 11 data
- Label aktual negatif (0) yang prediksi negatif terdiri dari 23 data
- Label aktual netral (1) yang prediksi netral terdiri dari 57 data
- Label aktual positif (2) yang prediksi positif terdiri dari 890 data

Pada hasil pengujian 90% data uji dan 10% data latih untuk kernel RBF, seperti pada gambar berikut ini:

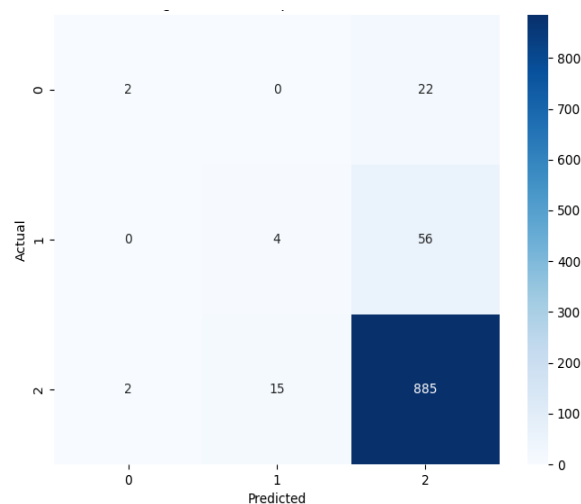


Gambar 7. kernel RBF pengujian 90% data uji dan 10% data latih

Dimana hasil dari matrix seperti dibawah ini.

- Label aktual negatif (0) yang diprediksi negatif terdiri dari 1 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 0 data
- Label aktual positif (2) yang diprediksi positif terdiri dari 1 data
- Label aktual negatif (0) yang diprediksi negatif terdiri dari 0 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 3 data
- Label aktual positif (2) yang diprediksi positif terdiri atas 11 data
- Label aktual negatif (0) yang prediksi negatif terdiri dari 23 data
- Label aktual netral (1) yang prediksi netral terdiri dari 57 data
- Label aktual positif (2) yang prediksi positif terdiri dari 890 data

Pada hasil pengujian 90% data uji dan 10% data latih untuk kernel Linear, seperti pada gambar berikut ini:



Gambar 8. kernel Sigmoid pengujian 90% data uji dan 10% data latih

Dimana hasil dari matrix seperti dibawah ini.

- Label aktual negatif (0) yang diprediksi negatif terdiri dari 2 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 0 data
- Label aktual positif (2) yang diprediksi positif terdiri dari 2 data
- Label aktual negatif (0) yang diprediksi negatif terdiri dari 0 data
- Label aktual netral (1) yang diprediksi netral terdiri dari 4 data
- Label aktual positif (2) yang diprediksi positif terdiri atas 15 data
- Label aktual negatif (0) yang prediksi negatif terdiri dari 22 data
- Label aktual netral (1) yang prediksi netral terdiri dari 56 data
- Label aktual positif (2) yang prediksi positif terdiri dari 885 data

4.4. Evaluasi

Setelah mencari model dengan confusion matrix selanjutnya pada tahap penerapan multi kelas support vector machine ini digunakan untuk mencari Akurasi, Presisi, Recall dan F1-Measure.

Dimana untuk Akurasi terbaik dan juga tertinggi terdapat pada 90:10 data uji dan data latih untuk kernel linear, polinomial, RBF dan sigmoid sebesar 90,67%.

Dimana untuk presisi tertinggi terdapat pada 10:90 data uji dan data latih untuk kernel linear kelas positif, 20:80 data uji dan data latih untuk kernel linear dan kernel sigmoid, 30:70 data uji dan data latih untuk kernel sigmoid kelas positif dan 40:60 data uji dan data latih untuk kernel sigmoid kelas positif mendapat presisi sebesar 93,85%.

Recall tertinggi terdapat pada 50:50 data uji dan data latih untuk kernel polinomial kelas positif, 90% data uji dan 10% data latih untuk kernel linear, polinomial, rbf dan sigmoid kelas positif mendapatkan Recall sebesar 98,67%.

Dan untuk F1-Measure tertinggi terdapat pada 90:10 data uji dan data latih untuk kernel linear, polinomial dan rbf untuk kelas positif mendapatkan F1-Measure sebesar 95,09%.

4.5. Analisis Hasil

Berdasarkan hasil evaluasi dengan ke 4 kernel Linear, Polinomial, RBF dan Sigmoid menunjukan bahwa performa yang baik pada ratio pembagian data, dengan performa tertinggi pada ratio 90:10 data uji dan data latih. Precision menunjukkan bahwa semua kernel sangat efektif dalam memprediksi kelas Positif, namun mengalami kesulitan dalam memprediksi kelas Negatif, kecuali untuk kernel Sigmoid yang lebih baik dalam hal ini. Recall juga tinggi untuk kelas Positif, tetapi rendah untuk kelas Netral, yang mengindikasikan tantangan model dalam mendeteksi kelas tersebut. F1-Score mengungkapkan bahwa kernel Linear dan Polynomial secara konsisten memberikan nilai tinggi untuk kelas Positif, sementara

kernel Sigmoid menunjukkan variasi yang lebih besar, dengan nilai tertinggi pada rasio 20:80 antara data uji dan data latih.

Secara keseluruhan, semua kernel menunjukkan performa yang baik dalam mendeteksi kelas Positif, tetapi mengalami kesulitan dengan kelas Negatif dan menunjukkan variasi performa dalam memprediksi kelas Netral, dengan kernel Linear dan Polynomial cenderung lebih stabil dan konsisten dibandingkan dengan kernel RBF dan Sigmoid. Akurasi tertinggi pada rasio 90:10 mencapai 90,67% untuk kernel Polynomial dan RBF. Precision tertinggi dicapai pada rasio 10:90 untuk kernel Linear, 20:80 untuk kernel Linear dan Sigmoid, serta 30:70 dan 40:60 untuk kernel Sigmoid dengan presisi 93,85%. Recall tertinggi sebesar 98,67% dicapai pada rasio 50:50 untuk kernel Polynomial dan rasio 90:10 untuk kernel Polynomial dan RBF. F1-Score tertinggi sebesar 95,09% dicapai pada rasio 90:10 untuk kernel Linear, Polynomial, dan RBF.

5. KESIMPULAN DAN SARAN

Dari hasil pengujian yang dilakukan dapat disimpulkan bahwa algoritma Multi Kelas Support Vector Machine dengan kernel Polynomial dan Linear menunjukkan performa yang konsisten lebih baik dibandingkan kernel lainnya RBF dan Sigmoid. Pengujian dilakukan dengan rasio pembagian data latih dan data uji dimulai dari 10:90 – 90:10, Akurasi Tertinggi pada kernel Polynomial dan Linear mencapai akurasi tertinggi Pada rasio 90:10 data latih, kedua kernel ini mendapatkan akurasi sebesar 90,67%, yang merupakan hasil tertinggi dibandingkan kernel Linear dan Sigmoid. Presisi tertinggi pada pengujian 10:90 untuk kernel linear, 20:80 kernel Linear dan Sigmoid, 30:70 dan 40:60 data uji dan data uji dan data latih untuk kernel sigmoid mendapat presisi sebesar 93,85% adalah hasil tertinggi dibandingkan kernel Polinomial dan RBF. recall tertinggi pengujian 50:50 untuk kernel polinomial dan 90:10 data uji dan data latih untuk kernel Polinomial dan RBF mendapatkan Recall sebesar 98,67% hasil tertinggi dibandingkan kernel Linear dan Sigmoid. f1-measure tertinggi pada pengujian 90:10 data uji dan data latih untuk kernel Linear, Polinomial dan RBF Mendapatkan F1-Score sebesar 95,09% merupakan hasil terbaiknya.

Dari hasil penelitian ini disarankan untuk menggunakan kernel Polinomial dan Linear sebagai pilihan yang tepat untuk data ulasan ini ketika mengklasifikasi data yang tidak seimbang dan juga untuk ratio pembagian data dimana ketika data uji dan data latih semakin besar maka akan semakin baik juga model dalam meningkatkan akurasi, presisi, recall dan f1-measure.

DAFTAR PUSTAKA

- [1] Afrilian, P. (2021). Pusaka is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License Analisis Peran Dinas Pariwisata Kota Bukittinggi Dalam Meningkatkan Fasilitas Pada Objek Wisata Taman Margasatwa Kinantan. In *Journal of Tourism, Hospitality, Travel and Busines Event* (Vol. 3, Issue 1). Online.
- [2] Aponno, J. C. (2022). *Penerapan Algoritma Sentiment Analysis Dan Naïve Bayes Terhadap Opini Pengunjung Di Tempat Wisata Pantai Pintu Kota, Kota Ambon*. 9(4), 3180–3188. <http://jurnal.mdp.ac.id>
- [3] Hary Jocom, Daniel Daud Kameo, Intiyas Utami, & Viktor Bungtilu Laiskodat. (2021). Rantai Nilai Pariwisata Sumba Timur dan Sumba Barat Daya. *Jurnal Kajian Dan Terapan Pariwisata*, 1(2), 1–21. <https://doi.org/10.53356/diparojs.v1i2.23>
- [4] Hendrayana, I. G., Divayana, D. G. H., & Kesiman, M. W. A. (2023). Komparasi Metode SVM, K-NN DAN NBC Pada Analisis Sentimen. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, 4(1), 191–198. <https://doi.org/10.35870/jimik.v4i1.157>
- [5] Ipmawati, J., Saifulloh, S., & Kusnawi, K. (2024). Analisis Sentimen Tempat Wisata Berdasarkan Ulasan pada Google Maps Menggunakan Algoritma Support Vector Machine. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(1), 247–256. <https://doi.org/10.57152/malcom.v4i1.1066>
- [6] Irwan, I., Khadijah, U. L., & Tahir, R. (2020). Memperkenalkan Pariwisata Budaya dan Heritage Kepada Generasi Muda Melalui Virtual Tour ke Pulau Penyengat. *Sosial Budaya*, 17(2), 133. <https://doi.org/10.24014/sb.v17i2.11010>
- [7] Kristianingrum, E., & Yuniar Rahman, A. (n.d.). *Klasifikasi presiden republik indonesia menggunakan SVM Kernel Polynomial dengan Fitur Ekstraksi Gabor..* <https://doi.org/10.25126/jtiik.202296748>
- [8] Maya Sari, A., Kunci, K., & Daerah, P. (n.d.). *Pengaruh Pariwisata Terhadap Perekonomian Daerah Kasus Bali*.
- [9] Pati, G. K., & Umar, E. (2022). Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-Nearest Neighbor. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2309. <https://doi.org/10.30865/mib.v6i4.4635>
- [10] Saniati, S., Assuja, M. A., Neneng, N., Puspaningrum, A. S., & Sari, D. R. (2022). Implementasi E-Tourism sebagai Upaya Peningkatan Kegiatan Promosi Pariwisata. *International Journal of Community Service Learning*, 6(2), 203–212. <https://doi.org/10.23887/ijcsl.v6i2.45559>
- [11] Sibyan, H., & Hasanah, N. (n.d.). Analisis sentimen pada wisata dieng dengan algoritma K-Nearest Neighbor (K-NN). *Jurnal Penelitian*

- Dan Pengabdian Kepada Masyarakat UNSIQ*, 9(1), 38–47.
- [12] Syahlan, M. S., Irmayanti, D., Alam, S., Studi, P., Informatika, T., Tinggi, S., Wastukencana, T., Cikopak, J., 53, N., Cikao, B., & Barat - Indonesia, P.-J. (2023). *Analisis sentimen terhadap tempat wisata dari komentar pengunjung dengan menggunakan metode Support Vector Machine (SVM) (Studi Kasus: Taman Air Mancur Sri Baduga Purwakarta)*. 8(2).
- [13] Tri, I., Sihalohe, P., Ratnawati, D. E., & Rahayudi, B. (2022). *Analisis Sentimen Objek Wisata Danau Toba berdasarkan Ulasan Pengunjung menggunakan Algoritma Support Vector Machine* (Vol. 6, Issue 9). <http://j-ptiik.ub.ac.id>
- [14] Tuhuteru, H., Refialy, L. P., Laturake, M., & Gildion Pattirane, S. (2024). Tourist Perceptions Through Sentiment Analysis to Support Tourism Development in Maluku Province. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 8, Issue 1). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [15] Wibowo, S. H., & Toyib, R. (n.d.). Support Vector Machine Method for Recognizing Patterns in Signatures. *Jurnal Media Infotama*, 18(2), 323