

METODE NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN STUDI KASUS MEGA WISATA COASTARINA BATAM

Mayada Dwi Andini¹, Rika Harman²

¹Mahasiswa Program Studi Sistem Informasi Universitas Putera Batam

²Dosen Program Studi Sistem Informasi Universitas Putera Batam

email: pb201510027@upbatam.ac.id

ABSTRACT

Tourism is one of the destinations for tourists to visit a city. Usually, to visit a tourist spot, individuals check the ratings of the place through Google reviews. In addition to ratings, some people also read reviews to assess the quality of the tourist spot. However, if the reviews given do not match the rating, there is a need for sentiment analysis to identify positive and negative reviews. This study takes a case study from Batam city, namely Mega Wisata Coastarina Batam. The method used in this research is Naive Bayes. Naive Bayes is chosen for its popularity in classification methods with good accuracy. Based on the research results, there are more positive sentiments than negative sentiments, namely 200 positive sentiments and 141 negative sentiments. The testing results of the research show the performance of sentiment analysis using the Naive Bayes method, with an F1-score accuracy of 69.49% and a margin of error of +7.67%. The precision is 71.00% in positive predictions and 67.39% in negative predictions. The recall value is 75.53% for positive predictions and 62.09% for negative predictions. The F1-score results are 86.06% for positive predictions and 64.63% for negative predictions.

Keywords: Sentiment Analysis, Google Reviews, Naïve Bayes, RapidMiner

PENDAHULUAN

Wisata adalah kegiatan perjalanan yang dilakukan oleh seseorang atau kelompok untuk mengunjungi suatu tempat untuk tujuan rekreasi, pengembangan pribadi, atau untuk memahami keunikan daya tarik wisata yang dikunjungi dalam jangka waktu tertentu. Sebaliknya, objek wisata adalah tempat yang menjadi pusat daya tarik dan dapat memberikan kepuasan, terutama bagi para pengunjung. Tempat wisata ini memiliki banyak sumber daya alami, seperti pantai, pegunungan, flora, dan fauna, serta sumber daya buatan

manusia, seperti bangunan bersejarah, monumen, candi, tarian, atraksi, dan elemen kebudayaan unik lainnya. Mega Wisata Coastarina adalah salah satu tempat terbaik untuk berlibur di Kota Batam. Konsep Taman Impian Jaya Ancol melibatkan pengelolaan kawasan pariwisata (pusat rekreasi dan resort) untuk mewujudkan perpaduan taman bermain dan keindahan pantai Coastarina. Mega Wisata Ocarina Batam telah menerima banyak pengunjung selama akhir pekan dan libur sekolah sejak didirikan pada tahun 2009. Mega

Wisata Coastarina mudah diakses karena berada di pusat kota. Mega Wisata Coastarina Batam menawarkan harga tiket masuk murah sebesar Rp. 25.000,00.

Seiring berjalannya waktu Mega Wisata Coastarina mulai sepi pengunjung. Padahal di lihat dari rating digoogle review banyak yang memberikan nilai tinggi. Setelah dilihat lebih lanjut ternyata banyak yang memberikan komentar tidak sesuai dengan rating seperti pengunjung yang memberikan rating 1 tetapi isi komentarnya positif yaitu memberikan saran agar Mega Wisata Coastarina menjadi lebih baik. Hal tersebut membuat orang terkecoh jika hanya melihat dari ratingnya saja. Oleh karena itu perlu adanya pengelompokan komentar positif dan negatif untuk mengetahui apakah wisata mega ocarina benar benar recommended. Berdasarkan permasalahan yang telah dijabarkan maka penulis ingin mengklasifikasikan komentar positif dan komentar negatif dari goole review Mega Wisata Coastarina. Salah satu metode yang dapat digunakan untuk mengatasi hal tersebut adalah analisis sentimen menggunakan metode naive bayes.

Analisis sentimen adalah suatu metode penerapan analitika teks yang bertujuan untuk mengakses berbagai sumber data dari internet dan platform media sosial. Tujuannya adalah untuk mengumpulkan pandangan atau opini dari pengguna aktif di platform tersebut. Fungsinya sebagai alat integratif memungkinkan perusahaan mendapatkan masukan inti dari pengguna atau konsumen dengan lebih efisien. (ADMINLP2M 2022).

Penggunaan metode naive bayes juga dilakukan oleh Apriani & Gustian (2019) juga menggunakan metode naive bayes, dengan judul penelitian "Analisis Sentimen dengan Metode Naive Bayes

terhadap Komentar Aplikasi Tokopedia." Penelitian ini menunjukkan bahwa teknik Naive Bayes efektif untuk menganalisis sentimen secara otomatis. Uji coba dilakukan menggunakan data testing real-time menggunakan aplikasi yang mereka buat dan RapidMiner. Setiap kata diklasifikasikan dengan sentimen positif atau negatif. Hasil pengujian dengan RapidMiner terhadap 1.500 data testing menunjukkan nilai akurasi sebesar 97,13% dengan nilai precision 1, sementara nilai pada Class Recall menunjukkan nilai sebesar 95,49% (kelas positif: negatif), dengan nilai AUC 0,980. Sangat menarik bahwa data komentar menunjukkan sentimen negatif sebesar 63,53%, sementara sentimen positif hanya 36,47%, menunjukkan perbedaan yang signifikan antara jumlah rating yang diberikan kepada aplikasi tokopedia. Tokopedia Selain itu, peneliti menganalisis jumlah kata-kata yang sering muncul dalam dokumen data komentar kategori negatif. Kata "Tokopedia" muncul paling sering dalam 294 komentar/dokumen, sebanyak 580 kali. Kata "barang" muncul paling sering, sebanyak 408 kali, dalam 211 komentar/dokumen. Hasil menunjukkan bahwa banyak komentar negatif terkait dengan merek atau nama Tokopedia itu sendiri, disusul oleh masalah dengan barang yang tidak sesuai saat diterima, rusak saat diterima, atau tidak sesuai dengan pesanan. Dalam 153 dokumen, kata "aplikasi" muncul sebanyak 254 kali.

KAJIAN TEORI

2.1. Analisis sentimen

Analisis sentimen adalah upaya untuk mengidentifikasi sentimen dan mengkategorikan polaritas dalam teks atau kalimat sebagai sentimen positif, negatif, atau netral. Analisis sentimen

sekarang digunakan secara luas oleh para peneliti sebagai salah satu cabang ilmu komputer. Analisis sentimen untuk menilai persepsi publik biasanya dilakukan melalui jaringan sosial seperti Twitter. Karena fokusnya pada pendapat yang disampaikan sebagai positif atau negatif, analisis sentimen juga sering disebut opinion mining. Analisis sentimen melibatkan penambangan data untuk menganalisis, memproses, dan mengekstrak data teks yang berkaitan dengan entitas tertentu, seperti produk, layanan, individu, fenomena, atau topik tertentu. Teks ulasan, forum, tweet, atau blog dapat digunakan untuk menganalisis. Tokenisasi, penghapusan stopword, stemming, dan identifikasi dan klasifikasi sentimen adalah semua bagian dari proses preprocessing data. Analisis sentimen sekarang digunakan secara luas oleh para peneliti sebagai salah satu cabang ilmu komputer. Analisis sentimen untuk menilai persepsi publik biasanya dilakukan melalui jaringan sosial seperti Twitter. Karena fokusnya pada pendapat yang disampaikan sebagai positif atau negatif, analisis sentimen juga sering disebut opinion mining. Analisis sentimen melibatkan penambangan data untuk menganalisis, memproses, dan mengekstrak data teks yang berkaitan dengan entitas tertentu, seperti produk, layanan, individu, fenomena, atau topik tertentu. Teks ulasan, forum, tweet, atau blog dapat digunakan untuk menganalisis. Tokenisasi, penghapusan stopword, stemming, dan identifikasi dan klasifikasi sentimen adalah semua bagian dari proses preprocessing data. Langkah-langkah ini membantu menyederhanakan dan mengorganisasikan data teks, yang memungkinkan interpretasi yang lebih baik untuk mengetahui sentimen yang

terkandung dalam teks. (Samsir et al. 2021).

2.2 Text Mining

Istilah "text mining" atau "text analytics" dalam ilmu data mining mengacu pada teknologi yang dapat menganalisis data teks, baik semi-terstruktur maupun tidak terstruktur. Ini membedakannya dari data mining, yang lebih berfokus pada pengolahan data yang lebih terstruktur. Text mining menggunakan metode data mining, machine learning, natural language processing (NLP), information retrieval (IR), dan knowledge management untuk mengatasi masalah overload informasi. Proses text mining terdiri dari beberapa langkah sebelumnya, seperti mengkategorisasi teks, mengekstraksi informasi, dan mengekstraksi istilah. Selain itu, termasuk penyimpanan representasi menengah serta teknik untuk menganalisis representasi tersebut, seperti peraturan asosiasi, analisis distribusi, pengelompokan, dan analisis tren. Selanjutnya, hasil analisis dapat dipresentasikan. Persiapan teks sebelum proses utama sangat penting, dan proses ini disebut text preprocessing. Text preprocessing digunakan untuk mengubah data teks yang tidak terstruktur atau tidak terstruktur menjadi data yang terstruktur, yang memungkinkan proses analisis yang lebih efisien.

2.3 Klasifikasi

Klasifikasi adalah proses menilai objek data untuk menemukan kelas tertentu di antara banyak kelas yang tersedia. Klasifikasi dilakukan dalam dua tahap utama. Pertama, membuat model sebagai prototipe yang disimpan dalam memori. Kedua, menggunakan model

untuk mengidentifikasi, mengklasifikasikan, atau memprediksi objek data lain untuk menentukan kelas mana objek data tersebut termasuk dalam model yang mudah disimpan. Mengklasifikasikan jenis hewan berdasarkan karakteristik tertentu adalah contoh aplikasi klasifikasi yang umum. Dalam klasifikasi, model memiliki arti ketika mereka menerima input dan kemudian dapat memproses input tersebut dan menghasilkan keluaran berdasarkan pemikiran mereka. Setiap algoritma klasifikasi memiliki keuntungan dan kelemahan. Prinsip dasar yang digunakan oleh semua algoritma klasifikasi, bagaimanapun, tetap sama. yaitu melibatkan fase pelatihan di mana model dapat menghubungkan setiap vektor masukan ke label kelas keluaran dengan laju error yang rendah atau akurasi tinggi. Meskipun model yang dibangun biasanya dapat memprediksi dengan benar pada semua data latihnya, kinerja model klasifikasi baru diukur dengan data uji. Ini adalah saat di mana tingkat efektivitas suatu algoritma klasifikasi ditentukan. (Ariyanti & Iswardani, 2020).

2.4 Naive Bayes

Menurut Fikri et al., (2020) Salah satu algoritma yang paling banyak digunakan dalam teknik klasifikasi adalah algoritma Naive Bayes, yang didasarkan pada teori Bayes dan didasarkan pada model probabilistik dan statistik yang disederhanakan. Algoritma Naive Bayes menganggap bahwa setiap atribut bersifat independen satu sama lain dalam konteks kelas tertentu; dengan kata lain, algoritma ini menganggap bahwa keberadaan atau ketiadaan suatu fitur atau atribut dalam kelas tertentu tidak bergantung pada keberadaan atau

ketiadaan fitur atau atribut lainnya dalam kelas yang sama. Meskipun asumsi ini seringkali tidak terpenuhi sepenuhnya dalam dunia nyata, keunggulan algoritma Naive Bayes terletak pada kemampuannya untuk memberikan hasil yang baik dengan data yang cukup besar dan dalam waktu komputasi yang relatif cepat. Algoritma tersebut digunakan oleh banyak aplikasi, termasuk klasifikasi teks, pengenalan spam email, dan analisis sentimen. Dengan asumsi independensi yang sederhana, kinerjanya yang baik bahkan dengan banyak fitur adalah kelebihan yang mencolok.

Berikut persamaan dari teorema Bayes :

dimana:

X : Data dengan class yang belum diketahui

C : Hipotesis data X merupakan suatu kelas spesifik

$P(C|X)$: Probabilitas hipotesis C berdasarkan kondisi X (posterior probability)

$P(C)$: Probabilitas hipotesis C (prior probability)

$P(X|C)$: Probabilitas X berdasarkan kondisi pada hipotesis C (likelihood)

$P(X)$: Probabilitas X (predictor prior probability)

2.5 RapidMiner

RapidMiner adalah perangkat lunak atau software yang dirancang untuk pengolahan data dan digunakan dalam berbagai teknik, seperti machine learning, data mining, text mining, dan predictive analytics. Dengan menggabungkan metode statistika, kecerdasan buatan, dan basis data, RapidMiner mampu mengekstrak pola-pola dari dataset yang besar. RapidMiner menawarkan grafik untuk menunjukkan hasil ekstraksi

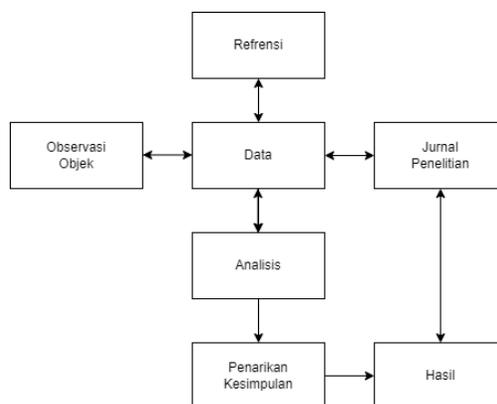
datanya, menjadikannya salah satu perangkat lunak yang disukai untuk metode ekstraksi data. (Menurut Fikri et al., 2020)

2.6 Google Review

Fitur ulasan Google Review, yang terintegrasi dengan layanan Google My Business (GMB) dan Google Maps, memungkinkan pelanggan untuk memberikan ulasan publik mengenai pengalaman mereka dengan suatu bisnis, baik terkait layanan maupun produk. Fitur ini juga berfungsi untuk menampilkan tanggapan pelanggan terhadap berbagai detail layanan yang ditawarkan oleh suatu bisnis. Dalam konteks pencarian lokal, ulasan ini dapat ditampilkan dalam hasil pencarian Google Maps, disebut sebagai "paket 3 lokal", di mana Google menampilkan tiga perusahaan teratas dalam hasil pencarian. Bisnis yang memiliki ulasan yang baik akan lebih dikenal dan diberi simbol bintang oleh Google. Sebuah bisnis harus memiliki setidaknya lima ulasan yang baik untuk mendapatkan penilaian bintang, yang memerlukan halaman web. Selain itu, informasi ini dapat ditemukan di daftar GMB, panel Google Knowledge, dan Google Maps. Ini memberikan gambaran umum tentang kualitas dan reputasi suatu perusahaan.

2.7 Kerangka Pemikiran

Adapun kerangka pemikiran penelitian ini adalah sebagai berikut:



Gambar 1. Kerangka Pemikiran

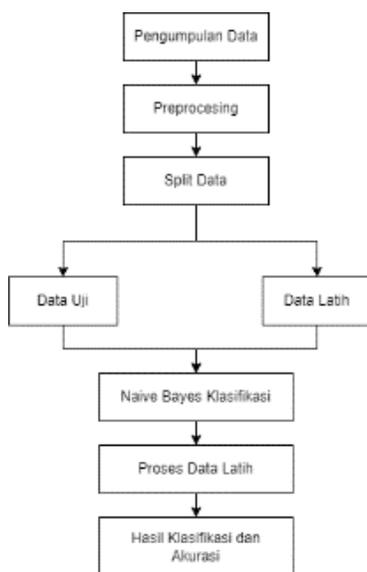
Penjelasan untuk masing-masing tahapan adalah sebagai berikut:

1. Referensi pada penelitian ini dilakukan dengan cara membaca jurnal, buku, dan dokumen terkait untuk melakukan analisis sentimen menggunakan metode naive bayes.
2. Observasi objek pada penelitian ini dilakukan dengan cara observasi di Mega Wisata Coastarina Batam untuk mengetahui analisis sentimen terhadap google review para pengunjung.
3. Data yang digunakan dalam penelitian ini berdasarkan refrensi, observasi objek, dan jurnal penelitian. Data tersebut adalah review google dari pengunjung Mega Wisata Coastarina Batam dari januari tahun 2020 hingga september 2023.
4. Jurnal penelitian nantinya akan dijadikan dasar untuk menghasilkan analisis. Jurnal yang digunakan adalah 5 tahun terakhir untuk menjaga penelitian tetap relevan.
5. Analisis yang dilakukan pada penelitian ini yaitu analisis sentimen Mega Wisata Coastarina Batam berdasarkan google review menggunakan metode Naive Bayes.

6. Penarikan kesimpulan pada penelitian ini didapatkan dari hasil analisis. Kesimpulan disini merupakan jawaban dari rumusan masalah dan menjelaskan hasil penelitian secara keseluruhan.
7. Hasil pada penelitian ini berupa klasifikasi berupa komentar positif dan negatif. Positif merupakan ulasan yang bersifat membangun atau patut dipertahankan sedangkan negatif merupakan ulasan yang bersifat kritik dan harus dibenahi.

METODE PENELITIAN

Pada penelitian ini akan diuraikan design penelitian. Pada desain penelitian, penulis memaparkan hasil dari yang telah dibuat dengan cara terstruktur agar dapat diambil sebagai acuan dalam melakukan penelitian dan dapat membantu penulis pada saat proses penelitian, desain penelitian ini juga dapat berperan sebagai pedoman bagi penulis.

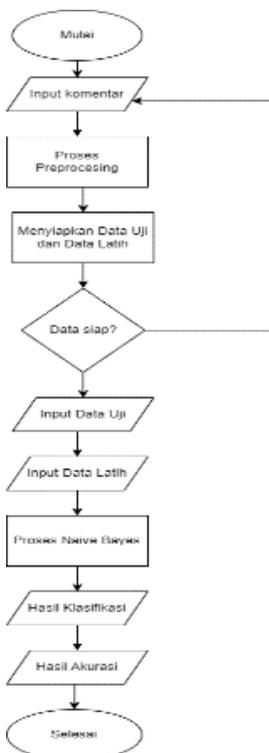


Gambar 2. Design Penelitian

Adapun design penelitian yang dapat dijabarkan pada penelitian ini dijelaskan sebagai berikut:

1. Pengumpulan Data
Data ini diambil dari Goole Review pada tempat wisata Mega Wisata Coastarina Batam.
2. Preprocessing
Pada Penelitian ini tahan preprocessing dilakukan mulai dari tokenizing, filtering, stemming, tagging, analyzing, dan cleaning.
3. Split Data
Split data terbagi menjadi dua, yaitu:
 - a. Data uji
 - b. Data Latih
4. Naïve Bayes
Proses klasifikasi data menggunakan tools RapidMainer yang diambil dari data komentar pada google review.
5. Proses Data Latih
Proses data latih yaitu proses data dengan memberikan data label sentiment berdasarkan data uji.
6. Hasil Klasifikasi dan Akurasi
Hasil penelitian yang telah diuji melalui klasifikasi teknik teks mining dengan metode Naive Bayes Serta diuji menggunakan aplikasi Rapidminer, yang akan menunjukkan hasil komentar positif dan negatif.

Adapun model penelitian, Pada penelitian ini model yang digunakan bisa dilihat dari Flowchart dibawah ini:



Gambar 2. Model Penelitian

Berdasarkan flowchart diatas dapat dijelaskan bahwa penelitian ini dimulai dari pengumpulan data, setelah data terkumpul maka data akan di proses pada tahapan preprocessing. Selanjutnya data akan dibagi menjadi data uji dan data latih untuk dilakukan proses klasifikasi menggunakan metode naive bayes. Impementasi metode tersebut digunakan untuk memberikan label pada data training. Setelah proses berhasil dilakukan maka sistem akan menghasilkan hasil klasifikasi dan hasil akurasi.

HASIL DAN PEMBAHASAN

4.1 Hasil

Hasil yang didapat pada penelitian ini adalah sebagai berikut:

1. Setelah melakukan pengujian dengan RapidMiner maka didapatkan hasil sebagai berikut:
2. Nilai akurasi yang di dapat algoritma naive bayes adalah 69.49% artinya sejumlah 69.49% model naive bayes dapat diklasifikasikan data yang benar.
3. Margin eror artinya jumlah kesalahan dalam pengambilan sampel sebesar +-7.67% terdapat 200 sentimen positif dan 141 sentimen negatif.
4. Presisi merupakan perbandingan antara hasil prediksi dengan data yang diminta yaitu dengan sejumlah 71.00% pada prediksi positif dan 67.39% pada prediksi negatif.
5. Nilai recall menggambarkan keberhasilan model dalam menemukan kembali informasi dengan perbandingan antara rasio prediksi benar positif dan keseluruhan data prediksi positif. Didapat hasil pada data prediksi positif 75.53% dan pada data negatif 62.09%.
6. F1-score merupakan perbandingan rata rata antara presisi dan recall. Didapatkan hasil dari f1-score yaitu 86.06% pada prediksi positif dan pada prediksi negatif 64.63%.

4.2 Pembahasan

Kemudian tahap cleaning. proses membersihkan. Ini menghapus fitur yang tidak penting dan tidak penting, seperti hastag, mention, retweet, whitespace, simbol, dan sebagainya. Gambar berikut menunjukkan contoh perawatan data set.

Replace adalah operator yang fungsinya untuk menghapus simbol yang tidak diperlukan. Kata kunci dari replace yang digunakan yaitu

[!~?.,,:\"#*%@%\$&()?) 🤔👍❤️👉👎🙏
👉♂️!❤️👉👉👉...?...?..?..?..?^?^🤔👉👉👉
👉👉👉👉👉👉👉👉👉👉👉👉].

Replace Missing Values adalah operator yang fungsinya mengganti nilai yang hilang dalam contoh atribut yang dipilih dengan mengganti yang ditentukan.

Remove duplicates adalah operator yang fungsinya untuk menghapus kalimat yang sama dengan membandingkan satu sama lain berdasarkan atribut yang ditentukan.

Tabel 1. Hasil Cleaning

| Sebelum Cleaning | Sesudah Cleaning |
|---|---|
| Tidak bisa berkata-kata,,, semuanya buruk. Ada biaya masuknya,,,,,, menyebalkan | tidak bisa berkata-kata semuanya buruk Ada biaya masuknya menyebalkan |
| Lokasinya Sangat Bagus...👍👍👍 ... | lokasinya sangat bagus |
| Toilet yang dekat pantai sana, tidak ada lampu, tidak ada air, toiletnya kotor banget. | toilet yang dekat pantai sana tidak ada lampu tidak ada air toiletnya kotor banget |
| situasi pasca covid: wahana dan warung makan hanya buka pada siang hari. pada malam hari, kawasan ini terasa seperti taman hiburan yang terbengkalai. waktu terbaik adalah berkunjung pada tengah hari. Ada biaya masuk sebesar Rp 25.000 per ... | situasi pasca covid wahana dan warung makan hanya buka pada siang hari pada malam hari kawasan ini terasa seperti taman hiburan yang terbengkalai waktu terbaik adalah berkunjung pada tengah hari ada biaya masuk sebesar rp per |
| Costarina akhir akhir ini sepi tapi tidak dengan keistimewaan dan keindahan yang di milikinya... Artinya wisata costarina The best👍👍 ... | costarina akhir akhir ini sepi tapi tidak dengan keistimewaan dan keindahan yang di milikinya Artinya wisata costarina the best |
| Asek bener... seharian di Costarina, Anak2 maen gembira. ... | asek bener seharian di costarina anak maen gembira |
| Sarana tempat rekreasi keluarga yg ada di daerah bengkong sadai/batam centre.. dengan aneka permainan yg tersedia dan fasilitas taman utk bermain dan santai bersama keluarga yg luas dengan view laut... Utk tiket memang mahal 25k/orang... | sarana tempat rekreasi keluarga yg ada di daerah bengkong sadaibatam centre dengan aneka permainan yg tersedia dan fasilitas taman utk bermain dan santai bersama keluarga yg luas dengan view laut Utk tiket memang mahal korang |

(Sumber : Data Penelitian, 2024)

Setelah proses cleaning, tahap berikutnya adalah proses dokumen dari data. Yaitu, proses tokenisasi, juga dikenal sebagai tokenisasi adalah memecahkan kalimat menjadi potongan kata atau token untuk mengetahui dari mana kata itu berasal.

Dalam transformasi cases, huruf diubah dari huruf kapital menjadi huruf kecil atau sebaliknya. Karena tweet yang sudah diambil sebagian besar menggunakan huruf kecil, semua data set dalam penelitian ini diubah menjadi huruf kecil.

Berdasarkan hasil crawling sebelumnya jumlah data yang akan digunakan yaitu 458 data komentar. Data latih yaitu data yang sudah melalui preprocessing kemudian melakukan pelabelan secara manual dan otomatis. Pelabelan secara manual dilakukan pada sebagian komentar. Sedangkan pelabelan otomatis mengacu pada dictionary opini positif dan negatif yang disebut data uji. Pada penelitian ini data set dibagi dua bagian yaitu data uji dan data latih dengan rasio test dan train yaitu 20% data latih dan 80% data uji. Maka Nilai akurasi yang di dapat algoritma naive bayes adalah 69.49% artinya sejumlah 69.49. Margin eror sebesar +-7.67% terdapat 200 sentimen positif dan 141 sentimen negatif. Presisi 71.00% pada prediksi positif dan 67.39% pada prediksi negatif. Nilai recall pada data prediksi positif 75.53% dan pada data negatif 62.09%. F1-score yaitu 86.06% pada prediksi positif dan pada prediksi negatif 64.63%.

SIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat diambil simpulan sebagai berikut:

1. Cara atau tindakan yang dilakukan untuk mengambil data ulasan google review Mega Wisata Coastarina Batam untuk diolah menjadi analisis sentimen yaitu dengan cara crawling data menggunakan data scraper chrome extantion instant data scraper.
2. Tahapan preprosessing data yang dilakukan sebelum data diolah menjadi analisis sentimen berdasarkan data ulasan google review Mega Wisata Coastarina Batam yaitu cleaning data untuk menghapus kata yang tidak

diperlukan, menghapus data yang sama, dan menghapus data yang kosong. Dengan menggunakan operator replace, replace missing values, remove duplicates, tokenize, dan transform cases pada aplikasi rapidminer.

3. Cara melakukan analisis sentimen berdasarkan komentar pengunjung Mega Wisata Coastarina Batam sehingga dapat di klasifikasikan menjadi positif dan negatif menggunakan aplikasi rapidminer yang mengimplementasikan metode naive bayes pada design rapidminer.
4. Cara membagi data uji dan data latih sebelum membuat ulasan positif dan negatif google review Mega Wisata Coastarina Batam sehingga hasil penelitian sesuai dengan yang diinginkan yaitu data set dibagi dua bagian yaitu data uji dan data latih dengan rasio test dan train yaitu 20% data latih dan 80% data uji.
5. Hasil performa analisis sentimen yang didapatkan mulai dari akurasi, precision, recall dan F1-score setelah diimplementasikan menggunakan metode naive bayes yaitu nilai akurasi yang di dapat algoritma naive bayes adalah 69.49% Margin eror sebesar +-7.67% terdapat 200 sentimen positif dan 141 sentimen negatif. Presisi sejumlah 71.00% pada prediksi positif dan 67.39% pada prediksi negatif. Nilai recall didapat hasil pada data prediksi positif 75.53% dan pada data negatif 62.09%. Hasil dari f1-score yaitu 86.06% pada prediksi positif dan pada prediksi negatif 64.63%.

DAFTAR PUSTAKA

- ADMINLP2M. (2022). Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe dan Cara Kerjanya. Retrieved August

12, 2023, from
<https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/>

Apriani, R., & Gustian, D. (2019). Analisis Sentimen dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, 6(1), 54–62. Retrieved from
<https://rekayasa.nusaputra.ac.id/article/view/86>

Samsir, Ambiyar, Verawardina, U., Edi, F., & Watrianthos, R. (2021). Analisis Sentiman Pembelajaran Daring pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(1), 149.
<https://doi.org/10.30865/mib.v5i1.2604>

Ariyanti, D., & Iswardani, K. (2020). Teks Mining untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma Naïve Bayes. *Jurnal IKRA-ITH Informatika*, 4(3), 125–132.

Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(2), 71–76.

Novianti, D. (2019). Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner. *Paradigma - Jurnal Komputer Dan Informatika Akademi Bina Saran Informatika*, 21(1), 49–54.
<https://doi.org/10.31294/p.v21i1.4979>

| | |
|---|--|
|  | Informasi Universitas Putera Batam. |
|  | Biodata Penulis kedua, Rika Harman, S.Kom., M.SI. merupakan Dosen Prodi Sistem Informasi Universitas Putera Batam. |

| | |
|--|---|
| | Biodata Penulis pertama, Mayada Dwi Andini merupakan Mahasiswa Prodi Sistem |
|--|---|