



## Computer Based Information System Journal

ISSN (Print): 2337-8794 | E- ISSN : 2621-5292  
web jurnal : <http://ejournal.upbatam.ac.id/index.php/cbis>



# ANALISIS SENTIMEN OBJEK WISATA DI *GOOGLE MAPS* MENGUNAKAN METODE *DECISION TREE*

**Tukino<sup>1</sup>, Arif Rahman Hakim<sup>2</sup>**

<sup>1</sup>Universitas Putera Batam, Indonesia

<sup>2</sup>Universitas Nagoya Indonesia, Indonesia

### INFORMASI ARTIKEL

Diterima Redaksi: Tanggal  
Diterbitkan Online: Tanggal

### KATA KUNCI

Analisis Sentimen, *Google Review*,  
*Decision tree*, *RapidMiner*

### KORESPONDENSI

E-mail:  
[tukino@puterabatam.ac.id](mailto:tukino@puterabatam.ac.id)  
[Arif.ibn06@gmail.com](mailto:Arif.ibn06@gmail.com)

### A B S T R A C T

*Tourism is a popular choice for travelers exploring a city. Typically, people rely on Google Reviews to assess the quality of tourist spots by checking ratings and reading reviews. However, discrepancies between reviews and ratings can occur. In this case study, we focus on Taman Wisata Taman Laut Batam in Batam city. Using the Decision tree method, we conducted sentiment analysis. The results revealed more positive sentiments (200) than negative sentiments (141). The F1-score accuracy was 79.59%, with a margin of error of  $\pm 8.78\%$ . Precision for positive predictions was 71.00%, while negative predictions had 77.49% precision. Recall values were 75.53% for positive predictions and 62.09% for negative predictions. Overall, positive sentiment predictions achieved an F1-score of 87.07%, while negative predictions scored 74.73%.*

## I. Latar Belakang

Wisata merupakan kegiatan perjalanan yang dilakukan oleh seseorang atau kelompok untuk mengunjungi suatu lokasi dengan 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 berbagai sumber daya, baik yang bersifat alami seperti pegunungan, pantai, flora, dan fauna, maupun yang dibuat oleh manusia seperti bangunan bersejarah, monumen, candi,

tarian, atraksi, dan unsur kebudayaan khas lainnya. Salah satu destinasi wisata yang dapat dijadikan tujuan berlibur di Kota Batam adalah Taman Wisata Taman Laut Batam. Perpaduan taman bermain dan keindahan Taman Wisata Taman Laut Batam direalisasikan serta diwujudkan melalui konsep Taman Impian Jaya Ancol dengan pengelolaan kawasan pariwisata (pusat rekreasi dan resort). Sejak resmi berdiri tahun 2009 silam, Taman Wisata Taman Laut Batam senantiasa ramai kunjungan saat akhir pekan dan libur sekolah. Taman Wisata Taman Laut Batam berlokasi di Jembatan 2 Bareleng, menjadikannya mudah ditemukan dan dapat

dilalui dengan kendaraan beroda dua maupun roda empat. Harga tiket masuk yang ditawarkan oleh Taman Wisata Taman Laut Batam tergolong terjangkau, yaitu senilai Rp. 35.000.00,-.

Seiring berjalannya waktu Taman Wisata Taman Laut Batam mulai sepi pengunjung. Padahal di lihat dari rating di *Google 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 Taman Wisata Taman Laut Batam 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 Taman Wisata Taman Laut Batam benar-bener recommended. Berdasarkan permasalahan yang telah dijabarkan maka penulis ingin mengklasifikasikan komentar positif dan komentar negatif dari *google review* Taman Wisata Taman Laut Batam. Salah satu metode yang dapat digunakan untuk mengatasi hal tersebut adalah analisis sentimen menggunakan metode *decision tree*.

Analisis sentimen merupakan suatu proses penerapan analitika teks untuk mengakses berbagai sumber data dari internet dan beragam platform media sosial. Tujuan utamanya adalah untuk menghimpun pandangan atau opini dari pengguna yang aktif di platform tersebut. Analisis sentimen berfungsi sebagai alat yang mengintegrasikan seluruh data, memungkinkan perusahaan untuk mendapatkan masukan inti dari pengguna atau konsumen dengan efisiensi.(ADMINLP2M 2022).

Penggunaan *decision tree* juga dilakukan oleh Apriani & Gustian (2019) dengan judul penelitian Analisis Sentimen dengan Metode *Decision tree* terhadap komentar Aplikasi Tokopedia. Dalam penelitian ini, Metode *Decision tree* terbukti efektif untuk menganalisis sentimen secara otomatis. Uji coba dilakukan dengan menggunakan data testing secara real-

time menggunakan aplikasi yang mereka bangun dan *RapidMiner*, di mana setiap kata diklasifikasikan dengan sentimen positif atau negatif. Hasil pengujian menggunakan *RapidMiner* terhadap 1.500 data testing menunjukkan nilai akurasi sebesar 97,13%, dengan nilai precision 1, sementara pada *Class Recall* dihasilkan nilai sebesar 95,49% (*positive class: negative*). Nilai AUC yang dihasilkan mencapai 0,980. Menariknya, data komentar yang mendapat sentimen negatif mencapai 63,53%, sementara sentimen positif hanya mencapai 36,47%. Hal ini menunjukkan perbandingan yang berbeda dengan jumlah rating yang diberikan kepada Aplikasi Tokopedia. Peneliti juga melakukan analisis terhadap jumlah kata-kata yang sering muncul dalam dokumen data komentar kategori negatif. Ditemukan bahwa kata 'Tokopedia' muncul paling banyak, yaitu sebanyak 580 kali dalam 294 komentar/dokumen, diikuti oleh kata 'barang' yang muncul sebanyak 408 kali dalam 211 komentar/dokumen. Temuan ini mengindikasikan bahwa banyak komentar negatif terkait dengan brand atau nama Tokopedia itu sendiri, disusul oleh masalah kondisi barang yang tidak sesuai saat diterima, barang rusak saat diterima, atau barang yang dikirim tidak sesuai dengan pesanan. Terkait dengan komentar negatif terhadap aplikasi, kata 'aplikasi' muncul sebanyak 254 kali dalam 153 dokumen.

## II. Kajian Literatur

Analisis sentimen merupakan suatu proses yang bertujuan untuk menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat, sehingga dapat dikategorikan sebagai sentimen positif, negatif, atau netral. Saat ini, para peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti Twitter, sering digunakan dalam analisis sentimen untuk menilai persepsi publik. Analisis sentimen juga sering disebut sebagai opinion mining, karena fokusnya pada pendapat yang

dinyatakan sebagai positif atau negatif. Dalam analisis sentimen, penambangan data dilakukan untuk menganalisis, memproses, dan mengekstrak data tekstual dalam suatu entitas, seperti layanan, produk, individu, fenomena, atau topik tertentu. Proses analisis dapat mencakup teks ulasan, forum, tweet, atau blog. Tahap preprocessing data melibatkan serangkaian langkah, termasuk tokenization, penghapusan stopword, stemming, dan identifikasi serta klasifikasi sentimen. Langkah-langkah ini membantu menyederhanakan dan mengorganisir data teks sehingga dapat diinterpretasikan dengan lebih baik untuk mendapatkan wawasan mengenai sentimen yang terkandung dalam teks tersebut. (Samsir et al. 2021).

Dalam ilmu data mining, istilah text mining atau text analytics merujuk pada teknologi yang memiliki kemampuan untuk menganalisis data teks, baik yang bersifat semi-terstruktur maupun tidak terstruktur. Hal ini membedakannya dari data mining yang lebih berfokus pada pengolahan data yang lebih terstruktur. Text mining bertujuan untuk mengatasi masalah overload informasi dengan memanfaatkan metode data mining, *Machine Learning*, *Natural Language Processing (NLP)*, *Information Retrieval (IR)*, dan *Knowledge Management*. Proses *Text Mining* mencakup beberapa tahap, termasuk *preprocessing* koleksi dokumen, seperti kategorisasi teks, ekstraksi informasi, dan ekstraksi istilah. Selain itu, melibatkan penyimpanan representasi menengah dan metode untuk menganalisis representasi tersebut, seperti analisis distribusi, pengelompokan, analisis tren, dan peraturan asosiasi. Hasil dari analisis tersebut kemudian dapat divisualisasikan. Pentingnya persiapan teks sebelum proses utama ditekankan, dan proses ini disebut sebagai *text preprocessing*. Tujuan dari *text preprocessing* adalah mengubah data teks yang awalnya tidak terstruktur atau sembarang menjadi data yang terstruktur, memfasilitasi proses analisis yang lebih efektif.

Klasifikasi adalah suatu tugas yang melibatkan penilaian objek data untuk

menentukan kelas tertentu di antara sejumlah kelas yang tersedia. Dalam konteks klasifikasi, terdapat dua tahap utama yang dilakukan. Pertama, adalah pembangunan model sebagai prototipe yang disimpan sebagai memori. Kedua, adalah penggunaan model tersebut untuk melakukan pengenalan, klasifikasi, atau prediksi pada objek data lain, sehingga dapat diketahui kelas mana objek data tersebut termasuk dalam model yang mudah disimpan. Contoh aplikasi klasifikasi yang umum adalah dalam mengklasifikasikan jenis hewan berdasarkan sejumlah atribut tertentu. Model dalam klasifikasi memiliki makna di mana suatu model menerima masukan, kemudian mampu memproses masukan tersebut dan memberikan keluaran sebagai hasil dari pemikirannya. Setiap algoritma klasifikasi memiliki kelebihan dan kekurangan masing-masing. Meskipun demikian, prinsip dasar yang digunakan oleh semua algoritma klasifikasi tetap sama, yaitu melibatkan tahap pelatihan di mana model dapat memetakan setiap vektor masukan ke label kelas keluaran dengan akurasi tinggi atau laju error yang rendah. Meskipun model yang dibangun umumnya dapat memprediksi dengan benar pada semua data latihnya, kinerja sebenarnya dari suatu model klasifikasi baru diukur ketika dihadapkan dengan data uji. Itulah momen di mana efektivitas suatu algoritma klasifikasi ditentukan. (Ariyanti & Iswardani, 2020).

Menurut Fikri et al., (2020) Algoritma *Decision tree* adalah salah satu algoritma yang umum digunakan dalam teknik klasifikasi. Algoritma ini mendasarkan pendekatannya pada model probabilistik dan statistik yang disederhanakan, dengan mengacu pada Teorema Bayes. Keunikan *Decision tree* terletak pada asumsi yang disebut sebagai sederhana, yaitu asumsi bahwa setiap atribut bersifat independen satu sama lain dalam konteks kelas tertentu. Dengan kata lain, algoritma *Decision tree* mengasumsikan bahwa keberadaan atau ketiadaan suatu fitur atau atribut dari suatu kelas tidak memiliki ketergantungan terhadap keberadaan atau ketiadaan fitur atau atribut lainnya dari kelas yang sama. Meskipun asumsi ini sering kali tidak sepenuhnya terpenuhi dalam

dunia nyata, keunggulan algoritma *Decision tree* terletak pada kemampuannya untuk memberikan hasil yang baik dengan data yang cukup besar dan dalam waktu komputasi yang relatif cepat. Algoritma ini sering digunakan dalam aplikasi klasifikasi teks, pengenalan spam email, analisis sentimen, dan banyak lagi. Kelebihannya yang mencolok terletak pada kinerjanya yang baik bahkan dengan jumlah fitur yang besar, meskipun asumsi independensi yang sederhana.

Berikut persamaan dari teorema Bayes :

$$P(C|X) = \frac{P(C|X).P(C)}{P(X)}$$

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)

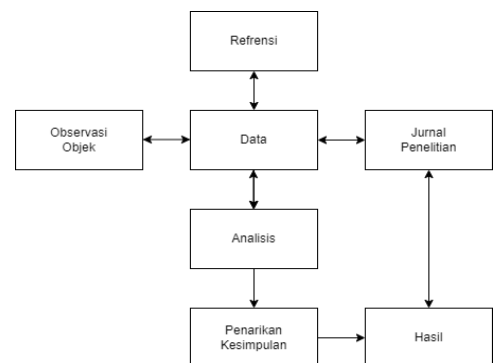
*P(X)* : Probabilitas *X* (predictor prior probability)

Menurut Novianti (2019) *RapidMiner* adalah sebuah alat atau *tools* yang digunakan dalam berbagai teknik di lingkungan machine learning, data mining, text mining, dan *predictive analytics*. *RapidMiner* merupakan perangkat lunak atau *software* yang dirancang untuk pengolahan data. Dengan memanfaatkan prinsip dan algoritma data mining, *RapidMiner* mampu mengekstrak pola-pola dari dataset yang besar dengan menggabungkan metode statistika, kecerdasan buatan, dan basis data. *RapidMiner* memberikan kemudahan bagi penggunaanya dalam melakukan perhitungan data yang sangat besar dengan menggunakan operator-operator. Operator tersebut berfungsi untuk memodifikasi data. Proses ini melibatkan penghubungan data dengan node-node pada operator, dan pengguna hanya perlu menghubungkannya ke node hasil

untuk melihat outputnya. Hasil yang dihasilkan oleh *RapidMiner* dapat ditampilkan secara visual melalui grafik, menjadikannya salah satu pilihan perangkat lunak yang populer untuk ekstraksi data dengan metode-metode data mining. -  
Brilian Rahmat C.T.I.

*Google Review* merupakan fitur ulasan yang terintegrasi pada layanan *Google My Business* (GMB) dan *Google Maps*. Fitur ini memungkinkan pelanggan untuk memberikan ulasan publik mengenai pengalaman mereka pada suatu bisnis, baik terkait layanan maupun produk. *Google Review* juga berperan dalam menampilkan tanggapan pelanggan terhadap berbagai detail layanan yang ditawarkan oleh suatu bisnis. Dalam konteks pencarian lokal, ulasan-ulasan ini dapat tampil dalam hasil pencarian di *Google Maps*, seringkali disebut sebagai "local 3 pack," di mana *Google* menampilkan tiga bisnis teratas dalam hasil pencarian. Bisnis dengan ulasan yang baik akan lebih terlihat dan diberi simbol bintang oleh *Google*. Pentingnya ulasan ini juga terlihat dalam hal penilaian, di mana biasanya sebuah bisnis perlu memiliki setidaknya lima ulasan baik agar mendapatkan penilaian bintang yang akan ditampilkan di halaman ulasan mereka. Informasi ini juga dapat muncul di daftar GMB, panel *Google Knowledge*, dan *Google Maps*, memberikan gambaran keseluruhan mengenai reputasi dan kualitas suatu bisnis.

Adapun kerangka pemikiran penelitian ini adalah sebagai berikut:



Gambar 1. Kerangka Pemikiran

Penjelasan untuk masing-masing tahapan adalah sebagai berikut:

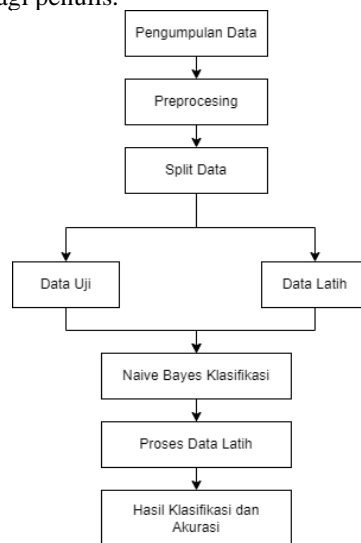
1. Referensi pada penelitian ini dilakukan dengan cara membaca jurnal, buku, dan dokumen terkait untuk melakukan analisis sentimen menggunakan metode *decision tree*.
2. Observasi objek pada penelitian ini dilakukan dengan cara observasi di Taman Wisata Taman Laut Batam untuk mengetahui analisis sentimen terhadap *Google Review* para pengunjung.
3. Data yang digunakan dalam penelitian ini berdasarkan referensi, observasi objek, dan jurnal penelitian. Data tersebut adalah review google dari pengunjung Taman Wisata Taman Laut 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 sentiment Taman Wisata Taman Laut Batam berdasarkan *Google Review* menggunakan metode *Decision tree*.
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.

### III. Metodologi

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

<http://ejournal.upbatam.ac.id/index.php/cbis>

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 Taman Wisata Taman Laut 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. *Decision tree*

Proses klasifikasi data menggunakan tools RapidMainer yang diambil dari data komentar pada *Google Review*.

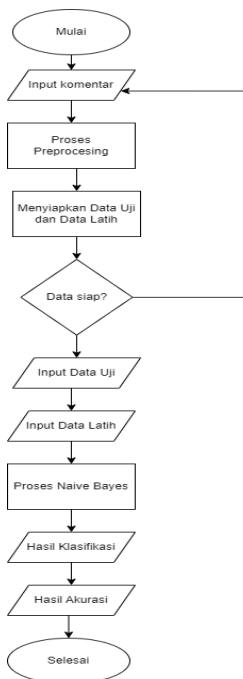
#### 5. Proses Data Latih

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 *Decision tree* 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 *decision tree*. Implementasi metode tersebut digunakan untuk memberikan label pada data training. Setelah proses berhasil dilakukan maka sistem akan menghasilkan hasil klasifikasi dan hasil akurasi.

<http://ejournal.upbatam.ac.id/index.php/cbis>

IV. Pembahasan

Data diambil dari *Google Review Taman Wisata Taman Laut Batam* yaitu data komentar pengunjung dengan menggunakan crawling data. Crawling data menggunakan data scraper chrome extension instant data scraper. Hasil dari data scraper disebut dengan data set.

Data set adalah kumpulan data yang diperoleh melalui proses crawling data. Data set kemudian akan mengalami modifikasi melalui tahap *text preprocessing*. Tujuan dari tahap ini adalah untuk meningkatkan kemampuan sistem komputer dalam mengenali struktur data set, sekaligus merapikan data yang semula tidak terstruktur menjadi data yang terstruktur dengan baik. *Text preprocessing* melibatkan beberapa proses, di antaranya adalah cleaning dan proses dokumen dari data. Dalam proses dokumen dari data, terdapat tiga tahapan utama, yaitu:

1. Tokenization: Proses pembagian teks atau kalimat menjadi unit-unit yang lebih kecil atau kata perkata, disebut dengan token. Token bisa berupa kata, frasa, atau karakter tergantung pada kebutuhan analisis.
2. Transformasi Kasus (Transform Cases): Proses ini mengubah huruf dalam teks menjadi format tertentu, seperti mengubah huruf menjadi huruf kecil atau huruf besar agar konsisten dalam analisis.

Melalui serangkaian tahapan *text preprocessing* ini, data set dapat diubah menjadi format yang lebih mudah diinterpretasikan oleh sistem komputer, memungkinkan analisis data yang lebih akurat dan efisien.

Kemudian tahap cleaning, pada proses cleaning menghapus atribut yang tidak dibutuhkan dan tidak memiliki makna, seperti hastag, mention, retweet, whitespace, simbol dll. Berikut adalah gambar contoh cleaning pada 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.

Sebelum Cleaning	Sesudah Cleaning
Tidak bisa berkata-kata,, semuanya buruk. Ada biaya masuknya,,,,, menyebalkan	Tidak bisa berkatakata 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
Taman Wisata Taman Laut Batam akhir akhir ini sepi tapi tidak dengan keistimewaan dan keindahan yang di milikinya... Artinya wisata Taman Wisata Taman Laut Batam The best🤔👍 ...	Taman Wisata Taman Laut Batam akhir akhir ini sepi tapi tidak dengan keistimewaan dan keindahan yang di milikinya Artinya wisata Taman Wisata Taman Laut Batam The best
Asek bener... seharian di Taman Wisata Taman Laut Batam,	Asek bener seharian di Taman Wisata Taman Laut Batam

Sebelum Cleaning	Sesudah Cleaning
Anak2 maen gembira. ...	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

Table 1. Hasil Cleaning

Setelah proses cleaning, tahap selanjutnya yaitu proses documents from data. Selanjutnya proses tokenize atau tokenisasi adalah proses memecahkan kalimat menjadi potongan kata atau token untuk mengetahui asal munculnya kata. Transform case merupakan proses pemerataan huruf dari huruf kapital menjadi huruf kecil atau sebaliknya. Pada penelitian ini data set diubah menjadi huruf kecil semua karena tweet yang sudah diambil sebagian besar menggunakan huruf kecil.

Berdasarkan hasil crowling 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 *decision tree* adalah 79.59% artinya sejumlah 79.59% model *decision tree* dapat diklasifikasikan data yang benar. Margin eror artinya jumlah kesalahan dalam pengambilan sampel sebesar +-8.78% terdapat 200 sentimen positif dan 141 sentimen negatif. Presisi merupakan perbandingan

antara hasil prediksi dengan data yang diminta yaitu dengan sejumlah 71.00% pada prediksi positif dan 77.49% pada prediksi negatif. 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%. F1-score merupakan perbandingan rata rata antara presisi dan *Recall*. Didapatkan hasil dari f1-score yaitu 87.07% pada prediksi positif dan pada prediksi negatif 74.73%.

## V. Kesimpulan

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

1. Cara atau tindakan yang dilakukan untuk mengambil data ulasan *Google Review* Taman Wisata Taman Laut Batam untuk diolah menjadi analisis sentimen yaitu dengan cara crawling data menggunakan data scraper chrome extantion instant data scraper.
2. Tahapan preprocessing data yang dilakukan sebelum data diolah menjadi analisis sentimen berdasarkan data ulasan *Google Review* Taman Wisata Taman Laut 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 Taman Wisata Taman Laut Batam sehingga dapat di klasifikasikan menjadi positif dan negatif menggunakan aplikasi *RapidMiner* yang mengimplementasikan metode *decision tree* pada design *RapidMiner*.
4. Cara membagi data uji dan data latih sebelum membuat ulasan positif dan negatif *Google Review* Taman Wisata Taman Laut 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 *decision tree* yaitu nilai akurasi yang di dapat algoritma *decision tree* adalah 79.59% Margin eror sebesar +8.78% terdapat 200 sentimen positif dan 141 sentimen negatif. Presisi sejumlah 71.00% pada prediksi positif dan 77.49% pada prediksi negatif. Nilai *Recall* didapat hasil pada data prediksi positif 75.53% dan pada data negatif 62.09%. Hasil dari f1-score yaitu 87.07% pada prediksi positif dan pada prediksi negatif 74.73%.

## Daftar Pustaka

- [1] ADMINLP2M. 2022. "Analisis Sentimen (Sentiment Analysis) : Definisi, Tipe Dan Cara Kerjanya." <https://lp2m.uma.ac.id/2022/02/21/analisis-sentimen-sentiment-analysis-definisi-tipe-dan-cara-kerjanya/> (August 12, 2023).
- [2] Apriani, Rita, and Dudih Gustian. 2019. "Analisis Sentimen Dengan *Decision tree* Terhadap Komentar Aplikasi Tokopedia." *Jurnal Rekayasa Teknologi Nusa Putra* 6(1): 54–62. <https://rekayasa.nusaputra.ac.id/article/view/86>.
- [3] Samsir et al. 2021. "Analisis Sentiman Pembelajaran Daring Pada Twitter Di Masa Pandemi COVID-19 Menggunakan Metode *Decision tree*." *Jurnal Media Informatika Budidarma* 5(1): 149.
- [4] Ariyanti, Dyah, and Kurnia Iswardani. 2020. "Teks Mining Untuk Klasifikasi Keluhan Masyarakat Pada Pemkot Probolinggo Menggunakan Algoritma *Decision tree*." *Jurnal IKRA-ITH Informatika* 4(3): 125–32.
- [5] Novianti, Deny. 2019. "Implementasi Algoritma *Decision tree* Pada Data Set Hepatitis Menggunakan Rapid Miner." *Paradigma - Jurnal Komputer dan Informatika Akademi Bina Saran Informatika* 21(1): 49–54.
- [6] S. A. Arnomo, "Analisa Decision Tree untuk Kepuasan Penggunaan Sinyal



- dari Base Transceiver Station (BTS),” *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 199, 2021, doi: 10.26418/justin.v9i2.43425
- [7] Baiq Nurul Azmi, Arief Hermawan, and Donny Avianto, “Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 4, no. 4, pp. 281–290, 2023, doi: 10.35746/jtim.v4i4.298.
- [7] Q. R. Cahyani et al., “Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm Article Info ABSTRAK,” *JOMLAI J. Mach. Learn. Artif. Intell.*, vol. 1, no. 2, pp. 2828–9099, 2022, doi: 10.55123/jomlai.v1i2.598.
- [8] R. Diana, H. Warni, and T. Sutabri, “Penggunaan Teknoogi Machine Laerning untuk Pelayanan Monitoring Kegiatan Belajar Mengajar pada SMK Bina Sriwijaya Palembang,” *J. Tek. Inform.*, vol. 5, no. 1, pp. 41–50, 2017, [Online]. Available: <https://jurnal.stmikdci.ac.id/index.php/jutekin/article/view/709/630>
- [9] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [10] M. A. Abubakar, M. Muliadi, A. Farmadi, R. Herteno, and R. Ramadhani, “Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung,” *J. Inform.*, vol. 10, no. 1, pp. 13–18, 2023, doi: 10.31294/inf.v10i1.14531.