

# Analisis Sentimen Terhadap Objek Wisata Di Provinsi Jawa Timur, Jawa Tengah, Jawa Barat, Banten, Dan Dki Jakarta Pada Platform Google My Business Menggunakan Algoritma Decision Tree

1<sup>st</sup> Dwi Fickri Insan Ramdani

Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

dwifickry@student.telkomuniversity.ac.id

2<sup>nd</sup> Rachmadita Andreswari

Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

andreswari@telkomuniversity.ac.id

3<sup>rd</sup> Faqih Hamami

Fakultas Rekayasa Industri  
Universitas Telkom  
Bandung, Indonesia

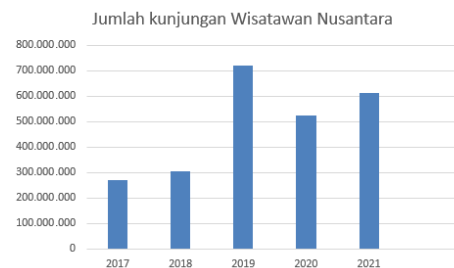
faqihhamami@telkomuniversity.ac.id

**Abstrak**—Pariwisata merupakan salah satu sektor penting dan unggulan yang memberikan kontribusi terhadap ekonomi nasional. Tingkat kepuasan wisatawan terhadap sebuah objek wisata dapat dilihat dari review yang diberikan, salah satunya platform yang dapat digunakan untuk melihat review wisatawan adalah Google My Business. Penerapan analisis sentimen menggunakan tiga sentimen, positif, negatif, dan netral. Dengan penerapan sentiment analysis dan multidimensional menggunakan metode Decision tree dapat diketahui sentimen yang diberikan wisatawan terhadap sebuah objek wisata. Seperti pada penelitian ini, review terhadap objek wisata yang ada di provinsi Jawa Timur, Jawa Tengah, Jawa Barat, Banten, dan DKI Jakarta didapatkan data sebanyak 12.680. kemudian setelah dilakukan preprocessing dan labeling menghasilkan data bersih sebanyak 8.615. pelabelan menggunakan dua library yaitu transformer dan textblob serta untuk setiap library akan dicoba dengan tiga split data yang berbeda yaitu 70:30, 80:20, dan 90:10 yang bertujuan untuk mengetahui kombinasi yang pas untuk pembuatan model machine learning ini. Dari hasil analisis sentiment menggunakan algoritma decision tree dengan pelabelan menggunakan library transformer dan split data 70:30 didapatkan nilai akurasi sebesar 78%. Hasil prediksi akan ditampilkan pada dashboard menggunakan Power BI untuk memudahkan dalam memahami data

**Kata kunci**— Sentiment Analisis, Multidimensional Analisis, Decision tree, Transformer, Textblob, Power BI

## I. PENDAHULUAN

Pariwisata merupakan salah satu sektor penting dan unggulan yang memberikan kontribusi terhadap ekonomi nasional [1]. Sebab itu, Indonesia beruntung dianugrahi keindahan alam yang sangat menarik untuk dikunjungi baik oleh wisatawan mancanegara maupun wisatawan domestik. Namun, dalam beberapa tahun terakhir terjadi pandemi virus COVID-19 yang berdampak besar pada sektor pariwisata di Indonesia. Dampak yang sangat terasa adalah penurunan jumlah pengunjung baik dari mancanegara ataupun domestik [2].



GAMBAR 1

Jumlah Kunjungan Wisatawan Nusantara dari Tahun 2017-2021

Dari gambar di atas, menunjukkan statistik perjalanan wisatawan nusantara yang mengalami penurunan yang cukup besar sekitar 200 juta pada periode COVID-19. Pada periode selanjutnya, jumlah wisatawan mulai mengalami kenaikan yang lumayan bagus sekitar 50% yang terjadi pada tahun 2022. Berdasarkan data yang dihimpun oleh BPS (Badan Pusat Statistik) tercatat ada kenaikan jumlah pengunjung atau wisatawan pada tahun 2021. Menurut [3] Selama Januari-Desember 2021, provinsi-provinsi di pulau Jawa menjadi daerah tujuan utama perjalanan wisatawan nusantara. Provinsi Jawa Timur merupakan provinsi yang paling banyak dikunjungi, tercatat sebesar 25,79% dari seluruh perjalanan di Indonesia. Provinsi paling banyak dikunjungi berikutnya adalah Provinsi Jawa Tengah dan Jawa Barat masing-masing sebesar 24,22% dan 15,79%.

Perkembangan teknologi informasi yang cepat menyebabkan industri pariwisata menjadi lebih efektif dan efisien [3]. Banyak ulasan atau opini yang bisa dijadikan penilaian terhadap objek wisata di suatu daerah. Salah satu platform yang banyak digunakan sebagai tempat beropini adalah Google My Business. Dalam Google My Business terdapat banyak sekali ulasan atau opini dari berbagai elemen masyarakat, baik itu opini yang bersifat positif, negatif, atau netral. Opini-opini tersebut akan sangat penting, selain karena dapat diolah menjadi dataset yang bisa digunakan untuk melakukan analisis sentimen terhadap sebuah objek wisata. Opini juga bisa digunakan untuk mendapatkan umpan

balik (*feedback*) dari wisatawan terhadap objek wisata di suatu daerah.

Umpan balik sangat penting bagi pengelola objek wisata untuk meningkatkan kualitas pelayanan. Banyaknya ulasan mengenai tempat wisata akan memudahkan wisatawan lain yang ingin berkunjung untuk mengetahui kondisi seperti jalan, fasilitas, harga tiket, dan informasi lain mengenai tempat tersebut. Tentunya informasi tersebut akan menjadi masukan untuk wisatawan lain untuk memutuskan akan ke tempat tersebut atau tidak [4]. Hal tersebut tentu akan berdampak pada ulasan atau opini di media sosial dan tentunya berdampak juga terhadap jumlah kunjungan wisatawan.

penelitian ini, akan melakukan analisis sentimen wisatawan terhadap objek wisata yang ada di daerah Jawa Timur, Jawa Tengah, Jawa Barat, Banten, dan DKI Jakarta dengan sumber data yaitu ulasan wisatawan yang ada pada *platform* Google My Business menggunakan algoritma Decision Tree. Pengambilan data dilakukan selama rentang waktu 3 bulan dari Januari 2023 – Maret 2023. Pada penelitian ini juga menggunakan pendekatan *multidimensional analisis* untuk mengajarkan melihat data dari berbagai sudut pandang yang berbeda. Dengan menggunakan teknik pendekatan tersebut dapat dijadikan sebagai *online analytical processing* (OLAP) untuk komposisi *query analytic cube* [5]. Adapun tujuan dari penelitian ini yaitu, untuk membandingkan sentimen pada setiap dimensi dalam *multidimensional analysis* berdasarkan metode Algoritma Decision Tree dan mengetahui hasil *performance* dari algoritma Decision Tree.

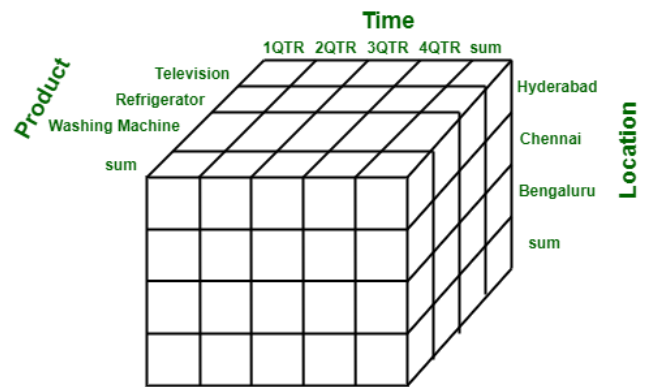
## II. KAJIAN TEORI

### A. Pariwisata

Dengan merujuk pada Undang-undang Nomor 10 Tahun 2009 tentang Kepariwisata, pengertian pariwisata adalah berbagai macam kegiatan yang terkait dengan pariwisata dan bersifat multidimensi serta multidisiplin yang muncul sebagai wujud kebutuhan setiap orang dan negara serta interaksi antara wisatawan dan masyarakat setempat, sesama wisatawan, Pemerintah, Pemerintah Daerah, dan pengusaha. Beberapa ahli juga mengemukakan pandangannya mengenai definisi pariwisata, menurut Schulalard pariwisata merupakan kegiatan yang berkaitan erat dengan ekonomi, dan juga berhubungan dengan pergerakan masuk, keluarnya orang asyik di suatu tempat tertentu[6]

### B. Multidimensional

Analisis *multidimensional* adalah sebuah teknik yang dapat digunakan untuk pengurutan data dalam *database*. Analisis *multidimensional* memungkinkan pengguna untuk melakukan analisis terkait tran pasar dan bisnis. *Online Analytical Processing* (OLAP) dan *data warehouse* menggunakan analisis ini untuk menampilkan beberapa dimensi data. Analisis *multidimensional* dapat diilustrasikan menjadi sebuah kubus data yang memungkinkan untuk memodelkan dan melihat data dari berbagai sudut pandang atau perspektif yang berbeda [7]



GAMBAR 2

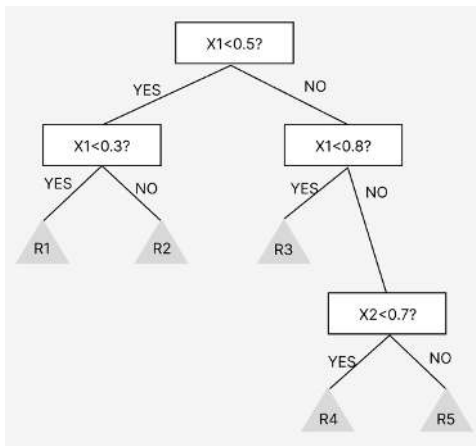
Visualisasi Cara Kerja Multidimensional Analysis

### C. Analisis Sentimen

Analisis sentimen atau yang sering disebut juga penambangan opini merupakan sebuah bidang keilmuan yang dapat digunakan untuk menganalisis opini, sentimen, evaluasi, penilaian, sikap, dan emosi seorang terhadap suatu topik [8]. Bidang keahlian dalam analisis sentimen ini dapat digunakan untuk *data mining* dan *text mining*. Analisis sentimen adalah teknik yang digunakan untuk memahami, menyelidiki, dan memproses data tekstual secara mandiri untuk mengekstraksi sentimen yang melekat pada sebuah opini. Analisis sentimen terdiri dari tiga kategori opini yaitu opini positif, opini negatif, dan opini netral. Dengan analisis sentimen, sebuah perusahaan atau lembaga dapat menentukan tanggapan suatu komunitas terhadap suatu layanan atau produk berdasarkan umpan balik komunitas atau bahkan pendapat para pakar. Sentimen mengacu pada fokus topik tertentu seperti pernyataan tentang satu topik mungkin memiliki arti yang berbeda dari pernyataan yang sama tentang topik yang berbeda. Oleh karena itu, dalam beberapa penelitian khususnya ulasan produk, karakteristik yang menentukan dari suatu produk dibahas sebelum memulai proses analisis sentimen [9]

### D. Decision Tree

Decision Tree adalah representasi yang disederhanakan dari teknik klasifikasi untuk jumlah kelas yang terbatas, di mana simpul internal dan akar diberi label dengan nama atribut, kulit diberi label dengan kemungkinan nilai atribut, dan simpul daun diberi label dengan berbagai kelas. Atribut yang diperiksa diwakili oleh setiap *node* dalam Decision Tree. Simpul daun (*leaf node*) pada setiap cabang mewakili kelas tertentu, sedangkan setiap cabang mewakili satu unit nilai ujian. Akar, yang biasanya merupakan atribut dengan pengaruh terbesar pada kelas tertentu, adalah level tertinggi dari simpul Decision Tree [10]



GAMBAR 3

Visualisasi Tree Diagram dari Decision Tree

perhitungan *gini index*. *Gini index* juga digunakan untuk menentukan perpecahan optimal meningkatkan integritas set yang dihasilkan oleh perpecahan [11].

$$Gini(L) = 1 - \sum_{i=1}^n (p_i)^2 \tag{1}$$

Keterangan:

n = jumlah term

Pi = probabilitas kemunculan term pada dokumen

$$Gini Split = \frac{n_1}{n} (L1) + \frac{n_2}{n} (L2) \tag{2}$$

Keterangan:

n<sub>1</sub> = jumlah sampel subset a

n<sub>2</sub> = jumlah sampel subset b

n = total dari jumlah sampel

L1 = hasil gini index subset a

L2 = hasil gini index subset b

E. TF-IDF

Pendekatan TF-IDF adalah metode yang paling lazim digunakan untuk menentukan bobot setiap kata dalam pencarian informasi. Metode ini juga dikenal efisien, sederhana, dan memberikan hasil yang tepat [12]. Perhitungan TF-IDF menggambarkan pentingnya suatu kata (atau frase) dalam sekumpulan dokumen. Metode ini digunakan untuk mengevaluasi bobot relevansi istilah dari setiap kata dalam satu data tekstual atau korpus. Dalam TF-IDF terdapat beberapa komponen seperti TF yang dapat digunakan untuk mengukur frekuensi kemunculan suatu istilah dalam dokumen tertentu dan dalam korpus secara keseluruhan [13]. Berikut merupakan rumus untuk TF-IDF [14]

Term frequency (TF) :

$$TF(t, d) = tf(t, d) * idf(t) \tag{3}$$

Keterangan :

TF = Frekuensi kemunculan kata pada suatu dokumen teks.

f = Jumlah kata pada suatu dokumen.

Invers Document Frequency (IDF) :

$$IDF(t) = \log \frac{1 + N}{1 + DF(t)} + 1 \tag{4}$$

Keterangan :

N = Jumlah dokumen.

DFt = jumlah dokumen yang memiliki *term* tertentu.

Vektor TF-IDF yang dihasilkan kemudian dinormalisasi dengan norma Euclidean. proses mengubah vektor dalam

ruang *Euclidean* menjadi vektor dengan panjang atau norma yang sama, yaitu 1.

Norma Euclidean :

$$v_{norm} = \frac{v}{||v||_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \tag{5}$$

Keterangan :

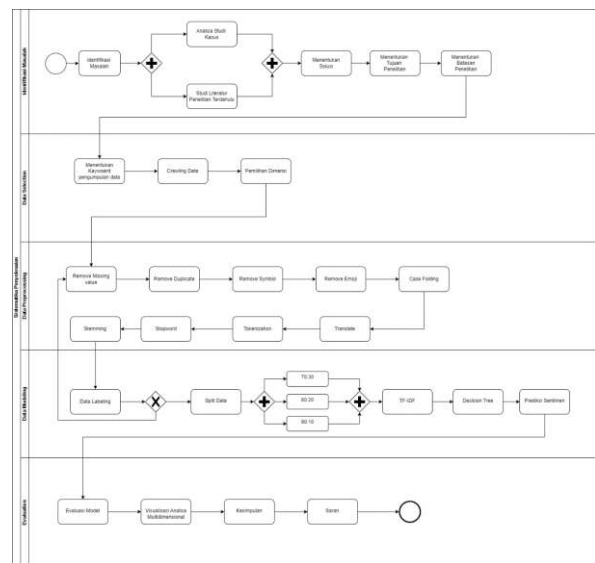
v<sub>norm</sub> = Vektor normalisasi dari v

v = Vektor yang akan dinormalisasi

||v||<sub>2</sub> = Panjang vektor v

III. METODE

Pada penelitian ini menggunakan metode algoritma Decision Tree dan analisis multidimensional untuk melihat data dari berbagai dimensi. Adapun tahapan tahapan dalam sistematika penyelesaiannya sebagai berikut:



GAMBAR 4

Sistematika Penyelesaian Masalah

Sistematika penyelesaian dalam penelitian ini melewati beberapa tahapan yaitu identifikasi masalah, *data selection*, *preprocessing*, *modeling*, dan *evaluation*. Dalam tahapan identifikasi masalah dilakukan kajian literatur dari penelitian sebelumnya yang disara relevan dengan penelitian ini. Berdasarkan hasil analisis tersebut ditentukan solusi, tujuan, dan batasan masalah pada penelitian ini.

Pada penelitian ini menggunakan data yang diperoleh dari *review* tentang objek wisata yang ada di Provinsi Jawa Timur, Jawa Tengah, Jawa Barat, Banten dan DKI Jakarta yang ada pada *platform* Google My Business dengan rentang waktu terhitung dari bulan Januari 2023 sampai Maret 2023. Data yang diperoleh akan dilakukan *preprocessing* sebelum digunakan. Ada beberapa tahapan dalam *preprocessing* seperti *Remove missing value*, *remove duplicate*, *remove symbol*, *remove emoji*, *case folding*, *translate*, *tokenization*, *stopword*, *stemming*.

Data yang telah melalui *preprocessing* menjadi data bersih dan dapat digunakan dalam pembangunan model *machine learning*. Selanjutnya data akan diberikan *label* dengan menggunakan dua skenario pelabelan yaitu dengan menggunakan *library transformer* dan *textblob*. Dan tiga

skenario *splitting* data untuk keperluan *training* dan *testing* menggunakan rasio 70:30, 80:20, dan 90:10. Data-data tersebut juga akan dilakukan pembobotan menggunakan metode TF-IDF untuk mengetahui berapa bobot sebuah *term* dalam dokumen.

Selanjutnya dari data yang sudah diberikan pembobotan menggunakan TF-IDF, akan dilakukan pembangunan model menggunakan algoritma Decision Tree dengan pemilihan parameter secara otomatis menggunakan *grid search* untuk mengetahui *hyperparameter* yang paling baik dalam pembangunan model *machine learning*.

Selanjutnya akan dilakukan *test* terhadap model *machine learning* yang telah dibuat untuk mengetahui *performa* model dalam melakukan prediksi. Evaluasi *performa* akan dilakukan menggunakan *multiclass confusion matrix* sehingga dapat dihitung *accuracy*, *precision*, *recall*, dan *f1-score* yang dihasilkan oleh model. Dari data tersebut juga akan dilakukan analisis dari berbagai sudut pandang menggunakan metode *multidimensional*.

#### IV. HASIL DAN PEMBAHASAN

Proses pengumpulan data dilakukan selama 3 bulan terhitung dari bulan Januari 2023 sampai Maret 2023 dengan metode *web scraping* menggunakan *tools extension chrome* yaitu DataMiner. Data yang diambil merupakan kumpulan *review* atau komentar dari wisatawan yang menuliskan *review* atau komentarnya terhadap sebuah objek wisata yang ada diberapa daerah di Indonesia pada *platform* Google My Business. Total ada 5 provinsi dan 26 kota yang dijadikan sampel penelitian. Data yang berhasil dikumpulkan adalah sebanyak 12.680 *row* data. Kumpulan data tersebut akan dijadikan *dataset* untuk digunakan pada penelitian ini. Data tersebut tentunya masih memiliki banyak *noise* yang harus dibersihkan terlebih dahulu pada tahapan *preprocessing* sebelum data dapat digunakan untuk kebutuhan penelitian. Ada beberapa tahapan dalam *preprocessing* seperti *Remove missing value*, *remove duplicate*, *remove symbol*, *remove emoji*, *case folding*, *translate*, *tokenization*, *stopword*, *stemming*.

##### 1. Remove missing value

Proses ini tentunya sangat penting untuk dilakukan, karena data kosong tentunya tidak memberikan kontribusi terhadap proses pembangunan model *machine learning*. Setelah proses *remove missing value* data menalami pengurangan yang awalnya 12.680 menjadi 12.441

##### 2. Remove duplicate

Proses ini menjadikan semua data bersifat unik. Karena, apabila dalam *dataset* memiliki nilai yang sama maka akan menyebabkan bias dalam analisis sentiment. Setelah melalui tahap *remove duplicate*, *dataset* berkurang sekitar 29 % yang awalnya 12.441 berkurang menjadi 8.615.

##### 3. Remove symbol

Proses *remove symbol* bertujuan menghapus seluruh objek selain huruf seperti angka dan simbol atau tanda baca lainnya. Proses ini juga akan menghapus spasi yang berlebihan menjadi hanya satu spasi.

##### 4. Remove emoji

Pada proses *remove emoji* bertujuan menghapus karakter atau simbol yang berupa *emoji*. *Emoji* ini ada banyak bentuknya seperti gambar *love*, *smile*, dan *sad*.

##### 5. Case folding

Proses *case folding* berguna untuk mengubah semua huruf kapital diganti menjadi huruf kecil. Hal ini membuat *dataset* lebih konsisten karena semua data teks yang ada diubah menjadi *lowercase*.

##### 6. Translate

*Dataset* yang didapat pada proses pengumpulan data tentunya masih sangat mentah. Masih banyak gangguan salah satunya masih banyak *review* yang ditulis dalam bahasa asing. Pada proses ini data yang mengandung bahasa asing akan diterjemahkan kedalam bahasa Indonesia.

##### 7. Tokenization

*Tokenization* pada proses *preprocessing* bertujuan untuk memecah kalimat menjadi kata-kata yang lebih kecil atau biasa disebut sebagai token.

##### 8. Stopword

Sering kali teks mengandung kata-kata yang sering muncul berulang dengan frekuensi yang cukup sering yang dimana kata tersebut tidak memiliki arti yang jelas dan tidak memberikan informasi yang penting.

##### 9. Stemming

Proses terakhir dari tahapan *preprocessing* data adalah *stemming*. Pada proses ini kata-kata yang masih memiliki imbuhan akan diubah menjadi kata dasar.

Setelah data melalui tahapan *preprocessing* dan menghasilkan data bersih yang sudah siap digunakan untuk pemodelan *machine learning*. Data mengalami penyusutan yang awalnya 12.680 menjadi 8.615 *row* data.

Data bersih yang dihasilkan setelah proses *preprocessing* akan diberikan *label*. Pelabelan dilakukan secara otomatis dengan menggunakan dua skenario pelabelan yaitu *library transformer* dan *textblob*. Dari masing-masing label akan dilakukan *splitting* data untuk *data training* dan *data testing* dengan tiga skenario *splitting* menggunakan rasio 70:30, 80:20, dan 90:10.

TABLE 1  
Hasil Dari Proses split Data Training dan Testing

Rasio	Data training	Data testing
70:30	6030	2585
80:20	6892	1723
90:10	7753	862

Tabel 1 di atas, menampilkan hasil *splitting* data dari ketiga rasio yang dipilih. *Data training* dan *data testing* selanjutnya akan dilakukan pembobotan menggunakan metode TF-IDF yang bertujuan untuk mengetahui seberapa penting sebuah kata pada sebuah dokumen.

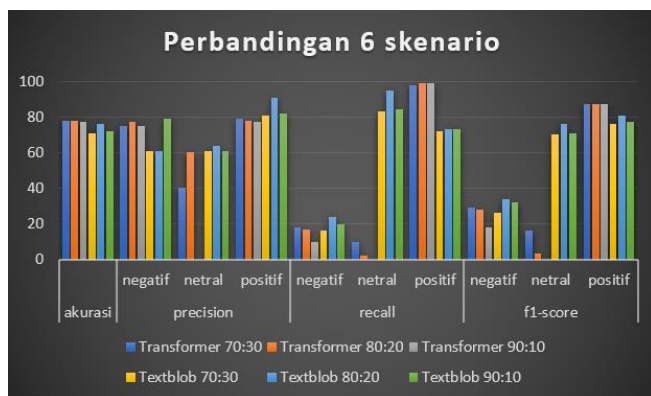
Setelah menerapkan metode *Trem Frequency-Inverse Document Frequency* (TF-IDF), data siap untuk digunakan dalam pembangunan model *machine learning*. Dalam konteks ini, algoritma Decision Tree digunakan untuk mengolah *dataset* tersebut dan melakukan pendekatan untuk membangun model klasifikasi. Pembangunan model dimulai dengan tahap pemilihan parameter yang optimal. Proses ini dilakukan secara otomatis menggunakan metode *grid search*, yang bertujuan untuk mencari *hyperparameter* terbaik yang akan digunakan dalam pembangunan model. *Hyperparameter* yang dihasilkan dari proses *grid search* tersebut nantinya akan digunakan untuk mengoptimalkan kinerja model klasifikasi. Berikut merupakan *hyperparameter* yang dihasilkan dari proses *grid search*.

TABLE 2  
Hyperparameter Yang Dihasilkan Dari Proses Grid Search

Hyperparameter	Value hyperparameter
Criterion	Gini
Max_depth	13
Min_samples_leaf	1
Min_samples_split	6

Tabel diatas menampilkan *hyperparameter* hasil dari proses *grid search*. *Hyperparameter* tersebut yang akan digunakan untuk membangun model *machine learning*. Parameter pertama yang dihasilkan dari proses *grid search* adalah *criterion* 'gini'. Parameter ini digunakan untuk menentukan metode pengukuran ketidakmurnian yang akan digunakan dalam pembangunan model *machine learning* menggunakan algoritma Decision Tree. Semakin rendah nilai gini maka semakin bersih data pada *node* tersebut. Nilai *max\_depth* yang dihasilkan dari proses *grid search* adalah 13 yang mengatur batas kedalaman maksimal dari model *machine learning* dalam melakukan prediksi untuk menghindari *overfitting*. *Min\_samples\_leaf* dengan nilai 1 menunjukkan bahwa pada setiap *leaf* setidaknya harus memiliki 1 *sample*. Dan parameter terakhir yang digunakan adalah *min\_samples\_split* dengan nilai 6 yang menunjukkan untuk setiap *node* harus memiliki jumlah minimal 6 *sample*. Apabila dalam sebuah *node* memiliki kurang dari 6 *sample* maka *node* tersebut akan dianggap sebagai *leaf*.

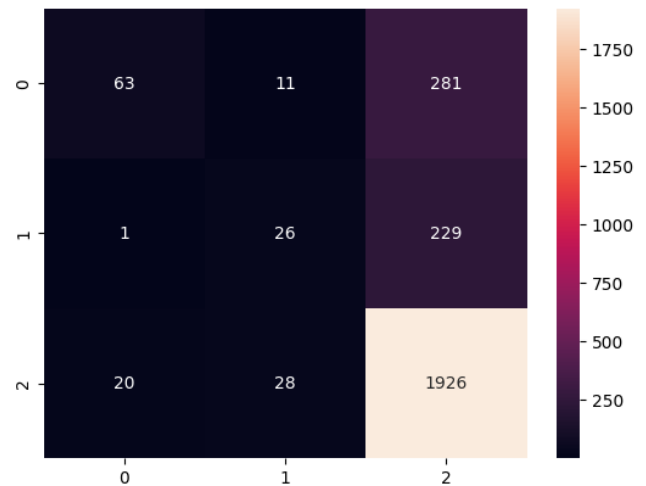
Pembangunan model *machine learning* menggunakan algoritma Decision Tree dengan melakukan 6 skenario pemodelan. Dari setiap pemodelan akan dilihat dan dipilih yang terbaik dari akurasi, *precision*, *recall*, dan *f1-score* yang nantinya akan digunakan untuk pebangunan model *machine learning*. Berikut merupakan hasil dari perbandingan 6 skenario yang telah dilakukan.



GAMBAR 5  
Hasil Perbandingan Skenario

Gambar di atas, menunjukkan hasil perbandingan dari 6 skenario yang dilakukan. Nilai akurasi menjadi penentu pertama dalam pemilihan skenario yang akan digunakan dalam pembangunan model *machine learning*. Hasil perbandingan menunjukkan terdapat dua akurasi yang memiliki nilai paling baik dan sama persis yaitu *transform 70:30* dan *transform 80:20* dengan nilai akurasi sebesar 78%. Dari dua skenario tersebut akan dilakukan analisa kembali untuk mengetahui skenario mana yang memiliki performa lebih baik. Berdasarkan hasil analisa lanjutan yang dilihat dari nilai *f1-score* menunjukkan perbedaan terutama pada

*class* netral. Nilai *f1-score* pada *class* netral untuk skenario *transformer 70:30* sebesar 16% dan untuk skenario *transformer 80:20* sebesar 3%. Dari hasil analisa tersebut menunjukkan skenario *transformer 70:30* memiliki performa yang paling baik dan akan digunakan untuk pembangunan model *machine learning*. Selanjutnya akan dilakukan evaluasi menggunakan *confusion matrix* terhadap model yang telah dipilih yaitu *transformer 70:30*. Berikut merupakan hasil *confusion matrix* dari hasil pengujian model.



GAMBAR 6  
Evaluasi Performance Model Menggunakan Confusion Matrix

merupakan tabel hasil evaluasi model *machine learning* menggunakan *confusion matrix*. Label 0 merupakan indikasi untuk label negatif, label 1 merupakan indikasi untuk label netral, dan label 2 merupakan indikasi untuk label positif. Dari *confusion matrix* tersebut, dapat dihitung *precision*, *recall*, dan *f1-score*. Berikut merupakan perhitungannya.

1. Kelas Negative

$$Precision (P) = \frac{63}{63 + 21} = 0,75$$

$$Recall (R) = \frac{63}{63 + 294} = 0,18$$

$$F1 Score = \frac{2 * 0,75 * 0,18}{0,75 + 0,18} = 0,29$$

2. Kelas Neutral

$$Precision (P) = \frac{26}{26 + 40} = 0,40$$

$$Recall (R) = \frac{26}{26 + 230} = 0,10$$

$$F1 Score = \frac{2 * 0,40 * 0,10}{0,40 + 0,10} = 0,16$$

3. Kelas Positive

$$Precision (P) = \frac{1926}{1926 + 510} = 0,79$$

$$Recall (R) = \frac{1926}{1926 + 48} = 0,98$$

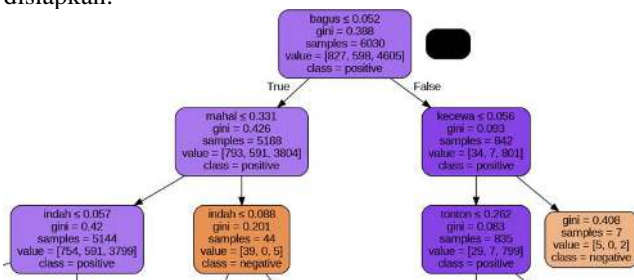
$$F1 Score = \frac{2 * 0,79 * 0,98}{0,79 + 0,98} = 0,87$$

TABLE 3

Hasil Perhitungan Precision, Recall, dan F1-score untuk setiap label

Label	Precision	Recall	F1-score
Negatif	75%	18%	29%
Netral	40%	10%	16%
Positif	79%	98%	87%

Model *machine learning* menggunakan algoritma Decision Tree akan menghasilkan *tree diagram* yang merupakan *visualisasi* dari algoritma Decision Tree itu sendiri. *Tree diagram* sangat membantu dalam melakukan analisa bagaimana proses data dipisahkan dan pengambilan keputusan pada setiap cabangnya. Pada *tree diagram* akan menampilkan struktur kirarki dari setiap node yang terhubung dengan setiap cabangnya. Berikut merupakan *tree diagram* yang dihasilkan dari pengetesan model algoritma *machine learning* dengan menggunakan data test yang telah disiapkan.



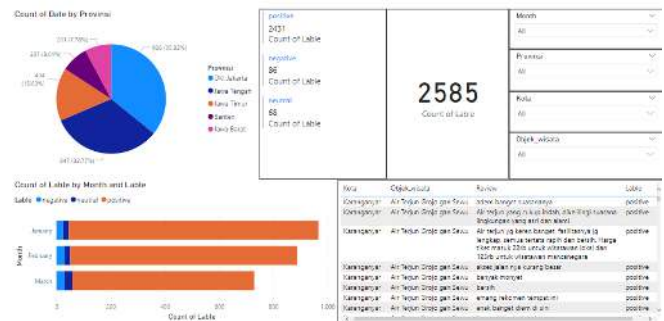
GAMBAR 7

Visualisasi Tree Diagram Yang Dihasilkan Dari Pengetesan Model Machine Learning Menggunakan Data Test

Gambar di atas, menampilkan prediksi yang dihasilkan dari penelitian ini. Data menunjukkan kata ‘bagus’ dengan nilai  $\leq 0,052$  dan nilai *gini Impurity* 0,388 merupakan variabel terbaik untuk membagi data dan menjadikan kata ‘bagus’ dengan nilai  $\leq 0,052$  sebagai *root node*. Diikuti oleh kata ‘kecewa’ sebagai cabang *true* dan kata ‘kecewa’ sebagai cabang *false*. kata ‘bagus’ menjadi *root node* berdasarkan dari perhitungan *gini splitting* yang dilakukan oleh model *machine learning*.

Selanjutnya hasil prediksi akan dilakukan visualisasi *dashboard* menggunakan Power BI. Penggunaan *dashboard* sebagai visualisasi dapat memudahkan penyampaian informasi kepada user. Pada *dashboard*, informasi lebih mudah dipahami dengan bantuan beberapa gravik yang lebih detail. Pada *analysis multidimensional*, penggunaan *dashboard* akan sangat membantu melihat data dari berbagai sudut pandang atau dimensi tertentu. Seperti pada penelitian ini, visualisasi *dashboard* mengandung beberapa *chart* atau gravik dan filter yang dapat diatur sesuai kebutuhan. Berikut merupakan model *dashboard* yang berisi data hasil prediksi model *machine learning*.

DASHBOARD SENTIMEN ANALISIS



GAMBAR 8

Visualisasi Dashboard Hasil Prediksi

Gambar di atas, merupakan visualisasi data yang dihasilkan dari proses prediksi model *machine learning* menggunakan algoritma Decision Tree. Dengan adanya *dashboard* ini diharapkan informasi dari data lebih mudah dipahami oleh pihak-pihak yang membutuhkan. Ada beberapa komponen pendukung dalam *dashboard* yang dapat memudahkan untuk informasi yang dibutuhkan. Dari hasil visualisasi *dashboard* di atas, terdapat beberapa informasi yang dapat digunakan oleh pihak-pihak yang bersangkutan seperti dinas pariwisata atau pengelola tempat wisata. Dari data *dashboard* tersebut ada beberapa sentimen negatif dari wisatawan yang dapat dijadikan bahan evaluasi untuk memperbaiki layanan waktu mendatang. Berikut beberapa informasi yang dapat dilihat dari visualisasi *dashboard*.

1. Terjadi penurunan jumlah pengunjung wisata setiap bulannya, yang diduga disebabkan oleh berakhirnya musim liburan awal tahun. Namun, menariknya, terdapat peningkatan pada sentimen negatif dan netral pada bulan Februari dan Maret. Hasil analisa menunjukkan bahwa rata-rata sentimen negatif menyoroti keluhan terkait harga tiket yang tidak memuaskan serta masalah sampah yang mengganggu.
2. Provinsi DKI Jakarta menjadi provinsi yang paling banyak dilakukan *review* oleh wisatawan sedangkan provinsi yang paling banyak dikunjungi oleh wisatawan yaitu Jawa Timur berada di urutan nomor tiga. Dari hasil analisa, yang menjadi pembeda di antara dua provinsi tersebut adalah jenis wisata yang ditawarkan oleh masing-masing provinsi. Provinsi DKI Jakarta menawarkan objek wisata buatan seperti Dufan atau wisata sejarah seperti museum yang memang telah dikelola dengan baik. Sedangkan Provinsi Jawa Timur menawarkan objek wisata alam seperti pantai klayar dan kawah ijen yang mungkin belum dikelola dengan baik

V. KESIMPULAN

Dari proses pengembangan model *machine learning* mengenai analisis sentimen terhadap objek wisata di provinsi Jawa Timur, Jawa Tengah, Jawa Barat, Banten, dan DKI Jakarta pada *platform* Google My Business menggunakan algoritma Decision Tree menghasilkan beberapa poin kesimpulan:

Pada penelitian ini melakukan pemodelan *machine learning* untuk analisis sentimen menggunakan algoritma Decision Tree dengan *sample* data yang diambil dari *review* objek wisata provinsi Jawa Tengah, Jawa Tengah, Jawa Barat, Banten, dan DKI Jakarta pada *platform* Google My Business

dengan rentang waktu mulai dari Januari 2023 sampai Maret 2023. *Sample data* yang berhasil didapatkan sebanyak 12.680 *row data* dengan kolom yang digunakan provinsi, kota, objek wisata, review, dan *date*. Proses pemodelan dilakukan dengan dua skenario *labeling* yaitu dengan menggunakan *library transformer* dan *textblob* dan pada masing-masing skenario dilakukan pengujian menggunakan *rasio split data* yang sama yaitu 70:30, 80:20, dan 90:10. Dari skenario yang telah dilakukan, didapatkan hasil akurasi paling tinggi yaitu 78% dengan menggunakan kombinasi pelabelan *transformer* dan *split data* 70:30.

Dari hasil analisis dapat dilihat perbandingan sentimen pada setiap dimensi yang dihasilkan. Pada dimensi waktu, menunjukkan 97% sentiment positif pada bulan Januari, sedangkan pada bulan Februari mengalami penurunan menjadi 95%, dan naik lagi pada bulan Maret menjadi 96%. Akan tetapi sentimen negatif juga terus naik pada setiap bulannya. Hasil analisis menunjukkan bahwa rata-rata sentimen negatif menyoroti keluhan terkait harga tiket yang tidak memuaskan serta masalah sampah yang mengganggu. Untuk dimensi lokasi menunjukkan 98,25% sentimen positif pada provinsi Jawa Timur, 96,68% sentimen positif pada provinsi Jawa Tengah, 96,63% sentimen positif pada provinsi Jawa Barat, 96,31% sentimen positif pada provinsi DKI Jakarta, dan 94,93% sentimen positif pada provinsi Banten. Provinsi DKI Jakarta dan Jawa Tengah merupakan provinsi yang paling banyak mendapatkan sentimen negatif sebesar 4%.

Evaluasi pemodelan *machine learning* menggunakan algoritma Decision Tree menghasilkan nilai akurasi yang cukup bagus. Dengan menggunakan *library transformer* dan *split data* 70:30 menghasilkan nilai akurasi sebesar 78%.

## REFERENSI

- [1] M. Sujai, "Strategi Pemerintah Indonesia Dalam Menarik Kunjungan Turis Mancanegara," *Kajian Ekonomi dan Keuangan*, vol. 20, no. 1, pp. 61–76, 2016, doi: 10.31685/kek.v20i1.181.
- [2] Egsaugm, "Pariwisata Indonesia di Tengah Pandemi," *egsaugm*, 2021. <https://egsa.geo.ugm.ac.id/2021/02/11/pariwisata-indonesia-di-tengah-pandemi/> (accessed Dec. 04, 2022).
- [3] H. Wulandari, A. Ruslani, C. Pravitasari, D. Apresiyanti, and M. Muin, "Statistik Wisatawan Nusantara 2020," *BPS*, vol. 1999, no. December, pp. 1–6, 2022, [Online]. Available: <https://www.bps.go.id/publication/download.html?nrbvfeve=ZmFiMWMxOWE3OGI1MzdkMTljZWQyNWRi&xzmn=aHR0cHM6Ly93d3cuYnBzLmdvLmlkL3B1YmxpY2F0aW9uLzlwMjEvMTIvMTQvZmFiMWMxOWE3OGI1MzdkMTljZWQyNWRiL3N0YXRpc3Rpay13aXNhdGF3YW4tbnVzYW50YXJhLTIwMjAuaHRtbA%3D%3D&twoadfn>
- [4] R. Sari, "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *EVOLUSI : Jurnal Sains dan Manajemen*, vol. 8, no. 1, pp. 10–17, 2020, doi: 10.31294/evolusi.v8i1.7371.
- [5] I. Kovacic, C. G. Schuetz, B. Neumayr, and M. Schrefl, "OLAP Patterns: A pattern-based approach to multidimensional data analysis," *Data Knowl Eng*, vol. 138, no. November 2021, p. 101948, 2022, doi: 10.1016/j.datak.2021.101948.
- [6] P. Diah and S. Pitanatri, "SEJARAH PERKEMBANGAN PARIWISATA DAN DEFINISI PARIWISATA," 2017.
- [7] Nehashrirudra, "MultiDimensional Data Model," *www.geeksforgeeks.org*, 2021. <https://www.geeksforgeeks.org/multidimensional-data-model/> (accessed Dec. 13, 2022).
- [8] B. Liu, "Sentiment Analysis and Mining of Opinions," *Studies in Big Data*, vol. 30, no. May, pp. 503–523, 2012, doi: 10.1007/978-3-319-60435-0\_20.
- [9] E. M. Sipayung, H. Maharani, and I. Zefanya, "PERANCANGAN SISTEM ANALISIS SENTIMEN KOMENTAR PELANGGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," *Jurnal Sistem Informasi (JSI)*, VOL. 8, NO. 1, April 2016, vol. 8, no. December, pp. 118–138, 2016.
- [10] A. H. Nasrullah, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PRODUK LARIS," vol. 7, no. 2, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [11] S. Tangirala, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm\*," 2020. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [12] Abdul Azis Maarif, "PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA ILMIAH," 2015.
- [13] M. N. Saadah, R. W. Atmagi, D. S. Rahayu, and A. Z. Arifin, "Sistem Temu Kembali Dokumen Teks dengan Pembobotan Tf-Idf Dan LCS," 2013.
- [14] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: <http://scikit-learn.sourceforge.net>.