

BAB 2

LANDASAN TEORI

2.1 Tempat Wisata

Objek wisata adalah segala sesuatu yang ada di daerah tujuan wisata yang merupakan daya tarik agar orang-orang ingin datang berkunjung ke tempat tersebut. Obyek dan Daya Tarik Wisata menurut Undang-Undang Kepariwisataaan No. 10, yaitu daya tarik wisata adalah segala sesuatu yang memiliki keunikan, keindahan dan nilai berupa keanekaragaman alam, budaya dan buatan manusia yang menjadi sasaran atau tujuan kunjungan dan wisata wisatawan dan kawasan objek wisata yang selanjutnya disebut objek wisata. [10].

2.2 Google Maps

Google Maps adalah layanan *Google* yang diberikan secara gratis dan sangat populer. *Google Maps* adalah peta dunia yang memungkinkan untuk melihat suatu wilayah. Dengan kata lain, *Google Maps* adalah peta yang terlihat di browser. Pengguna dapat menambahkan fitur *Google Maps* ke situs web yang dibuat atau bahkan blog berbayar dan gratis menggunakan *Google Maps API*. [11].

2.3 Sentimen Analisis

Analisis sentimen adalah proses pemahaman, penggalian, dan pengolahan data tekstual secara otomatis untuk mendapatkan sentimen dari informasi yang terkandung dalam sebuah kalimat opini [12]. Klasifikasi sentimen melibatkan polaritas sentimen dari kalimat yang difilter. Kalimat-kalimat ini diklasifikasikan sebagai netral, negatif dan positif sesuai dengan kasus yang diteliti[13].

2.4 Naïve Bayes Classifier

Pengklasifikasi Naive Bayes adalah algoritme yang digunakan untuk mengklasifikasikan data menggunakan metode probabilistik dan statistik. *Pengklasifikasi Naif Bayes* ditemukan oleh ilmuwan Inggris *Thomas Bayes* yang membuat prediksi tentang kemungkinan masa depan berdasarkan pengalaman masa lalu. Ciri ciri dari *Naive Bayes classifier* adalah independensi setiap kondisi, dimana

setiap kondisi atribut sistem bersifat independen dan tidak berkorelasi [14]. Berikut persamaan umum algoritma *Naive Bayes* [15] :

$$P(H_i | X) = \frac{P(H_i) \times P(X | H_i)}{P(X)} \quad (2.1)$$

Keterangan:

- $P(H_i/X)$: Sebuah peluang kata X muncul dikelas H.
- $P(H_i)$: Probabilitas kemunculan suatu kategori i.
- $P(X)$: Probabilitas kemunculan sebuah kata X.

2.5 Confusion Matrix

Confusion Matrix adalah pengukuran performa pada klasifikasi dimana output berupa dua kelas atau lebih. *Confusion Matrix* adalah tabel dengan empat kombinasi berbeda dari nilai prediksi dan nilai aktual [16]. Tujuan menggunakan *Confusion Matrix* adalah untuk melakukan pengujian terhadap sistem dan evaluasi akurasi sistem. Berikut Tabel 2.1 menampilkan Tabel *confusion matrix*:

Tabel 2.1. *Confusion Matrix*

	Nilai sebenarnya positive	Nilai sebenarnya negative
Nilai prediksi positive	True Positive (TP)	False Positive (FP)
Nilai prediksi negative	False Negative (FN)	True Negative (TN)

Komponen *confusion matrix* [17]:

- TP (*True Positive*). Jumlah data positif yang diprediksi sebagai kelas positif.
- FP (*False Positive*). Jumlah data negatif yang diprediksi sebagai kelas positif.
- FN (*False Negative*). Jumlah data negatif yang diprediksi sebagai kelas positif.
- TN (*True Negative*). Kondisi jumlah data positif dan berhasil diprediksi sebagai kelas negatif.

Berikut perhitungan dari *Confusion Matrix* untuk mendapatkan tingkat *accuracy*, *precision*, dan *recall* [18]:

Accuracy digunakan untuk pengukuran seberapa banyak prediksi yang benar (positif dan negatif) dengan total jumlah data.

$$accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + FalseNegative} \quad (2.2)$$

Precision digunakan untuk perbandingan antara *True Positive* dengan banyaknya data yang diprediksi positif (baik yang benar maupun yang salah).

$$precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (2.3)$$

Recall digunakan untuk perbandingan antara *True Positive* dengan keseluruhan sampel data yang diprediksi sebenarnya berlabel positif.

$$recall = \frac{TruePositive}{TruePositive + FalseNegative} \quad (2.4)$$

F1 Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan.

$$F1\ Score = 2 * \frac{Precision \times Recall}{Precision + Recall} \quad (2.5)$$

2.6 Term Frequency - Inverse Document Frequency (TF-IDF)

Term Frequency - Inverse Document Frequency (TF-IDF) merupakan salah satu metode yang digunakan untuk mengekstraksi ciri dari suatu text. Metode ini merupakan gabungan antara metode *term frequency (TF)* dengan metode *inverse document frequency (IDF)*. Untuk mencari ciri dari suatu text TF-IDF dapat menggunakan rumus berikut[19].

1. *Term Frequency (TF)*

Term Frequency merupakan cara yang digunakan untuk mencari bobot suatu kata dalam dokumen, dan mencari kata kunci yang hampir mirip dengan kategori yang tersedia, dengan cara seperti berikut:

$$tf(i, j) = \frac{n_{i,j}}{\sum_k n_{i,j}} \quad (2.6)$$

Keterangan:

- (a) $tf(i, j)$ = Frekuensi kata
- (b) n_{ij} = Total suatu kata yang muncul pada suatu dokumen
- (c) $\sum_k n_{i,j}$ = Total seluruh kata dalam suatu dokumen

2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) merupakan cara yang digunakan untuk menghitung jumlah dokumen yang mengandung kata yang dimaksud kemudian dibagi dengan total dokumen yang ada, dengan cara seperti berikut:

$$idf = \log \frac{N}{df_j} \quad (2.7)$$

Keterangan:

- (a) N = Total kelas
- (b) df_j = Total kelas j yang berisi kata i

3. Menghitung TF-IDF (Term Frequency Inverse Document Frequency)

Rumus 2.8 berguna untuk menjumlahkan kedua hasil TF (*term frequency*) dan IDF (*inverse document frequency*):

$$w_{ij} = tf_{ij} \times idf \quad (2.8)$$

Keterangan:

- (a) w_{ij} = Bobot kata i pada kelas j
- (b) tf_{ij} = Total kemunculan kata i pada kelas j
- (c) df_j = Total kelas j yang berisi kata i

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A