

IMPLEMENTASI ALGORITMA *NAÏVE BAYES* TERHADAP ANALISIS SENTIMEN KEPUASAN PENGUNJUNG WISATA DI KABUPATEN SAMPANG

M. HASYIM ASARI

(180411100144)

ABSTRAK

Naïve Bayes merupakan algoritma yang umum dipakai untuk klasifikasi teks, dan juga merupakan metode machine learning yang menggunakan perhitungan probabilitas. *Naïve Bayes* diusulkan *Thomas Bayes*, yang merupakan seorang ilmuwan inggris, yaitu melakukan prediksi peluang masa depan berdasarkan pengalaman masa lalu dengan menggunakan statistik untuk melakukan perhitungan probabilitas hipotesis. Data yang dipakai dalam penelitian ini di dapatkan dengan melakukan *crawling ulasan google maps* pengunjung objek wisata Kabupaten Sampang. Data ulasan pengunjung yang dilakukan pengambilan merupakan ulasan yang menggunakan bahasa indonesia. Pengambilan data ulasan tersebut dilakukan dengan menggunakan *Tools Instant Data Scraping*. Jumlah data awal yang dihasilkan dari *crawling* sebanyak 1.000 yang tercatat mulai tahun 2018 hingga tahun 2024 (6 tahun terakhir). Tujuan dari penggunaan metode *Naïve Bayes Classifier* dengan jumlah data yang digunakan 1.440 ulasan setelah melakukan SMOTE. Selanjutnya dilakukan skenario ujicoba menggunakan *k-fold cross validation* dengan k-5 menghasilkan nilai akurasi 69%. Ketika menggunakan *split validation* pembagian data latih dan data uji (80%:20%), (70%:30%), (60%:40%), dan (50%:50%). Akurasi mencapai 65%, 63%, 63% dan 63%. Sehingga pada penelitian ini diketahui bahwa penggunaan teknik *k-fold cross validation* dengan k-5 menghasilkan akurasi yang lebih tinggi ataupun lebih baik yaitu dengan akurasi 69%.

Kata kunci : Objek Wisata Kabupaten Sampang, Analisis Sentimen, *Google maps*, *Naïve bayes*.

Abstract

Naïve Bayes is an algorithm commonly used for text classification and is a machine learning method based on probability calculations. Naïve Bayes was proposed by Thomas Bayes, an English scientist, to predict future probabilities based on past experiences using statistics for hypothesis probability calculations. The data used in this research were obtained by crawling Google Maps reviews from visitors to tourist attractions in Sampang Regency. The visitor reviews collected were in the Indonesian language. The data crawling was conducted using the Instant Data Scraping tool. The initial dataset comprised 1,000 reviews recorded from 2018 to 2024 (the last six years). The purpose of employing the Naïve Bayes Classifier method with the dataset expanded to 1,440 reviews after applying SMOTE was to enhance the balance of the dataset. Subsequently, experimental scenarios were conducted using k-fold cross-validation with k=5, which resulted in an accuracy of 69%. When using split validation with training and testing data splits of (80%:20%), (70%:30%), (60%:40%), and (50%:50%), the accuracy achieved was 65%, 63%, 63%, and 63%, respectively. Thus, this research concludes that using the k-fold cross-validation technique with k=5 yields higher and better accuracy, specifically 69%.

Keywords: *Tourist Attractions in Sampang Regency, Sentiment Analysis, Google Maps, Naïve Bayes*

MOTTO

**“GAGAL URUSAN NANTI YANG TERPENTING KITA BERANI UNTUK
MENCoba DAN TERUS MENCoba”**

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
KATA PENGANTAR	iv
ABSTRAK.....	v
MOTTO.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
KODE PROGRAM	xiii
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	5
1.2.1 Permasalahan	5
1.2.2 Metode Usulan	5
1.2.3 Pertanyaan Penelitian	5
1.3 Tujuan dan Manfaat	5
1.3.2 Tujuan	5
1.3.3 Manfaat	5
1.4 Batasan Masalah.....	6
1.5 Sistematika Penulisan.....	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Analisis Sentimen	8
2.2 Wisata Sampang	8
2.3 Ulasan Google Maps.....	9
2.4 Web Crawling	9
2.5 Text Preprocessing	9
2.6 Smote	10

2.7 Pembobotan TF-IDF	11
2.8 Naïve Bayes Classifier	12
2.9 Confusion Matrix	15
2.10 Bahasa Pemrograman Python.....	16
2.11 K-fold Cross Validation	17
2.12 Split validation	17
2.13 Penelitian terkait	18
BAB III METODE PENELITIAN	23
3.1 Arsitektur Sistem	23
3.2 Data Penelitian	24
3.3 Pelabelan.....	26
3.4 Preprocessing	26
3.4.1 Cleaning.....	26
3.4.2 Tokenizing.....	27
3.4.3 Normalisasi	29
3.4.4 Stopword Removal	29
3.4.5 Stemming.....	30
3.5 Ekstraksi Fitur TF-IDF	32
3.6 Klasifikasi Naïve Bayes	34
3.7 Evaluasi	38
3.8 Skenario Ujicoba	38
BAB IV HASIL DAN PEMBAHASAN.....	39
4.1 Tahap Pengumpulan Data	39
4.2 Pelabelan Data	39
4.3 Tahap Preprocessing	40
4.3.1 Cleaning.....	40
4.3.2 Tokenizing.....	41
4.3.3 Konversi Slangword.....	42

4.3.4 Stopword	43
4.3.5 Stemming.....	45
4.4 Tahap SMOTE	45
4.5 Tahap Pembobotan TF-IDF	47
4.6 Tahap Pembagian Data.....	48
4.7 Tahap Klasifikasi Dengan <i>Naïve Bayes Classifier</i>	49
4.8 Tahap Evaluasi Model	49
4.9 Hasil Uji Skenario	50
4.9.1 Split Validation.....	50
4.9.2 K-fold 5.....	52
Bab V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran.....	57
DAFTAR PUSTAKA	58
LAMPIRAN.....	60
BIODATA PENULIS	63

DAFTAR GAMBAR

Gambar 3. 1 Arsitektur Sistem.....	23
Gambar 3.2 Tahapan Preprocessing.....	26
Gambar 3.3 Tahapan Cleaning	27
Gambar 3.4 Tahapan Tokenizing	28
Gambar 3.5 Tahapan Normalisasi.....	29
Gambar 3.6 Tahapan Stopword	30
Gambar 3.7 Tahapan Stemming.....	31
Gambar 3.8 Alur proses TF-IDF	32
Gambar 3.9 Alur proses naïve bayes classifier	35
Gambar 4. 1 Pelabelan Data	40
Gambar 4. 2 Hasil proses cleaning.....	41
Gambar 4. 3 Hasil proses tokenizing	42
Gambar 4. 4 Hasil proses slangword	43
Gambar 4. 5 Hasil proses stopword.....	44
Gambar 4. 6 Hasil proses stemming.....	45
Gambar 4. 7 hasil proses smote	47
Gambar 4. 8 hasil proses tf-idf	48
Gambar 4. 9 hasil perbandingan split dan k=5	56

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	15
Tabel 3. 1 Dataset.....	24
Tabel 3. 2 Contoh Cle aning.....	27
Tabel 3. 3 Contoh Tokenizing.....	28
Tabel 3. 4 Contoh Normalisasi	29
Tabel 3. 5 Contoh Stopword.....	30
Tabel 3. 6 Contoh Steaming.....	31
Tabel 3. 7 Contoh data sesudah di pre processing	32
Tabel 3. 8 Perhitungan TF-IDF	33
Tabel 3. 9 Hasil Pembobotan TF-IDF	33
Tabel 3. 10 data uji hasil matriks transpose dari hasil pembobotan tf-idf.....	36
Tabel 3. 11 Data uji hasil maktrik transpose dari hasil bobot tf-idf.....	36
Tabel 3. 12 Hasil perhitungan mean	36
Tabel 3.13 Hasil perhitungan standar deviasi	37
Tabel 3. 14 Hasil perhitungan like lihood.....	37
Tabel 3. 15 Tabel hasil contoh penerapan <i>naïve bayes classifier</i>	37
Tabel 3. 16 Ske nario Ujicoba.....	38
Tabel 4. 1 hasil split validation.....	52
Tabel 4. 2 hasil K-fold Cross Validation	54
Tabel 4. 3 hasil perbandingan split validation dan K-fold cross validation.....	55

KODE PROGRAM

Kode Program 4. 1 Proses Cleaning	41
Kode Program 4. 2 Proses Tokenizing	42
Kode Program 4. 3 Proses Konversi Slangword.....	43
Kode Program 4. 4 Proses Stopword	44
Kode Program 4. 5 Proses Stemming	45
Kode Program 4. 6 Proses Smote	47
Kode Program 4. 7 Proses pembobotan TF-IDF	48
Kode Program 4. 8 Proses Pembagian Data	49
Kode Program 4. 9 Proses Klasifikasi Naive Bayes	49
Kode Program 4. 10 Tahap Evaluasi Model.....	50
Kode Program 4. 11 Uji Skenario Split Validation.....	52
Kode Program 4. 12 Skenario ujicoba K=5	54