



## **SENTIMEN ANALISIS KOTA TEGAL BERBASIS ASPEK MENGUNAKAN ALGORITMA NAÏVE BAYES**

**Novita Fitria Putri, Muhammad Fikri Hidayattullah, Dwi Intan Af'idah\***

Teknik Informatika, Politeknik Harapan Bersama, Indonesia

---

**Abstrak:** Kurangnya ketentuan yang jelas mengenai penggunaan e-mail institusi oleh masyarakat Kota Tegal dalam menyampaikan informasi, saran, dan pengaduan kepada pemerintah. Sebagai gantinya, masyarakat cenderung menggunakan media sosial seperti Twitter untuk menyampaikan pendapat dan keluhan mereka, namun pemerintah mengalami kesulitan dalam menanggapi secara efektif. Penelitian ini bertujuan untuk mengembangkan sebuah aplikasi yang dapat mengklasifikasikan ulasan masyarakat Kota Tegal berdasarkan beberapa aspek seperti wisata/hiburan, pendidikan, fasilitas/layanan publik, dan kuliner. Melalui analisis sentimen berbasis aspek, aplikasi ini dapat memberikan informasi kepada pemerintah mengenai pendapat dan keluhan masyarakat secara spesifik. Penelitian ini menggunakan metode analisis sentimen dengan menggunakan algoritma Naïve Bayes untuk mengklasifikasikan ulasan masyarakat sebagai sentimen positif atau negatif. Pengujian dilakukan dengan menggunakan data ulasan masyarakat Kota Tegal yang dikumpulkan dari Twitter. Hasil penelitian menunjukkan bahwa aplikasi yang dikembangkan dapat mengklasifikasikan ulasan masyarakat Kota Tegal ke dalam beberapa aspek dengan tingkat akurasi pada data training diatas 90% dan tingkat akurasi pada data testing diatas 75% pada setiap kategori. Diharapkan aplikasi ini dapat membantu pemerintah dalam meningkatkan kualitas Kota Tegal dengan menanggapi pendapat dan keluhan masyarakat secara lebih efektif dan spesifik.

**Kata kunci:** Analisis Sentimen, Klasifikasi Aspek, Kota Tegal, Naïve Bayes

---

### **I. PENDAHULUAN**

Kota Tegal di Jawa Tengah memiliki potensi yang menarik untuk dikembangkan. Walikota Tegal telah mengambil langkah dengan menyediakan alamat email resmi pemerintah untuk memperkuat hubungan dengan masyarakat. Ini bertujuan untuk memberikan pelayanan efektif melalui teknologi informasi.

Namun, belum ada ketentuan yang jelas tentang penggunaan email oleh masyarakat

untuk mendapatkan informasi, memberikan saran, atau menyampaikan pengaduan (Khaerani & Sarai, 2020).

Dari hal tersebut banyak pendapat dan keluhan masyarakat di Kota Tegal disampaikan melalui media sosial, terutama Twitter. Pemerintah kesulitan memberikan tanggapan karena banyaknya opini yang ada. Untuk pengembangan kualitas kota, terutama dalam ekonomi, fasilitas, dan pelayanan publik, penting untuk mengklasifikasikan ulasan masyarakat ke berbagai aspek. Ini membantu pemerintah untuk lebih mendalam memahami keluhan yang ada dan

---

\*) [dwiintanafidah@poltektegal.ac.id](mailto:dwiintanafidah@poltektegal.ac.id)

memberikan tanggapan yang lebih spesifik (Graham et al., 2015).

Analisis sentimen berbasis aspek adalah mengidentifikasi sentimen terhadap berbagai aspek yang telah ditentukan dalam teks bahasa alami (Zhao & Yu, 2021). Metode umum yang digunakan untuk ini adalah *Naïve Bayes*, yang terbukti memiliki hasil akurasi dan metrik evaluasi (presisi, *recall*, F1-Score) yang lebih baik daripada SVM dalam beberapa penelitian (Fikri et al., 2020). *Naïve bayes* merupakan model klasifikasi yang paling sederhana dan paling sering digunakan, *Naïve bayes* merupakan *machine learning* yang memanfaatkan perhitungan probabilitas suatu kelas berdasarkan persebaran kata di dalam dokumen (Yulita et al., 2021). Tantangan dalam analisis sentimen sering kali muncul ketika terdapat ketidakseimbangan data antara kelas-kelas yang ada. Penggunaan model dengan data yang tidak seimbang cenderung menghasilkan akurasi prediksi yang rendah (Heranova, 2019). Tujuan dari pengujian algoritma ini adalah agar dapat otomatis mengklasifikasikan ulasan sebagai sentiment yang bernilai positif atau bernilai negatif. Sentimen positif mengindikasikan adanya pandangan yang positif terhadap suatu konteks, sedangkan sentimen negatif mencerminkan pandangan negatif terhadap suatu konteks.

Penelitian sebelumnya melakukan analisis sentimen aspek kuliner di Kota Bandung melalui ulasan TripAdvisor. Mereka menggunakan teknik klasifikasi dengan *Naïve Bayes* dan pelabelan multi aspek. Bobot ulasan dianalisis berdasarkan kontribusi pemberi ulasan di TripAdvisor. Hasil pengujian menunjukkan kombinasi metode ini menghasilkan akurasi 98,67%, jauh lebih tinggi dari penelitian sebelumnya yang hanya mencapai 88,78% (Fitri et al., 2019).

Penelitian lainnya mengevaluasi respon pelanggan terhadap layanan internet dan komunikasi untuk perusahaan. Dengan nilai f1-score mencapai 91,62%, penelitian ini menyediakan patokan yang kuat bagi perusahaan untuk menilai dan meningkatkan kualitas layanan mereka (Rahman et al., 2019). Penelitian sebelumnya menganalisis teks untuk mengidentifikasi aspek dan menilai sentimen (positif, negatif, netral) dengan TF-IDF dan klasifikasi *Naïve Bayes*. Hasil uji coba menunjukkan F1-Score sebesar 62,81% untuk data yang diterjemahkan dari Bahasa Inggris ke Bahasa Indonesia tanpa penghapusan *stopword* (Yutika et al., 2021). Penelitian sebelumnya membantu perusahaan memahami pandangan masyarakat terhadap merek melalui review menggunakan metode *Naïve Bayes*. Prediksi sentimen produk online dilakukan sesuai sistem yang disiapkan. Akurasi variasi: terendah 52,66% (pengujian 5 kelas, 80% latih, 20% uji), tertinggi 77,8% (pengujian 3 kelas, 90% latih, 10% uji) (Sulistio & Handojo, 2022). Penelitian sebelumnya mengembangkan pendukung keputusan menggunakan text mining dan sentiment analysis untuk menilai opini pengunjung wisata di Tegal. Metode *Naïve Bayes* memiliki akurasi 73,33%, lebih baik daripada Decision Tree. Studi ini menyarankan optimasi lebih lanjut untuk meningkatkan akurasi sentiment analysis dalam pengelolaan tempat wisata (Somantri & Dairoh, 2019).

## II. METODOLOGI

Naive Bayes adalah sebuah pendekatan yang berdasarkan pada teorema Bayes, yang menggabungkan pengetahuan yang didapat sebelumnya dengan informasi baru. Dengan demikian, algoritma ini merupakan salah satu metode klasifikasi yang simpel namun memiliki tingkat akurasi yang tinggi. Prinsip dasar di

balik Naive Bayes adalah mengklasifikasikan suatu data ke dalam kelas tertentu berdasarkan probabilitas yang terkait dengan kelas tersebut. Untuk melakukan ini, algoritma Naive Bayes menghitung probabilitas kelas berdasarkan data pelatihan yang diberikan. Secara khusus, ada dua tahap utama dalam penggunaan Naive Bayes untuk klasifikasi:

#### A. Tahap Pelatihan:

##### - Menghitung Probabilitas Kelas

Pada tahap ini, algoritma menghitung probabilitas setiap kelas berdasarkan jumlah data pelatihan yang termasuk dalam kelas tersebut. Misalnya, jika ada dua kelas A dan B, maka algoritma akan menghitung  $P(A)$  dan  $P(B)$  berdasarkan jumlah data yang berada di kelas A dan B.

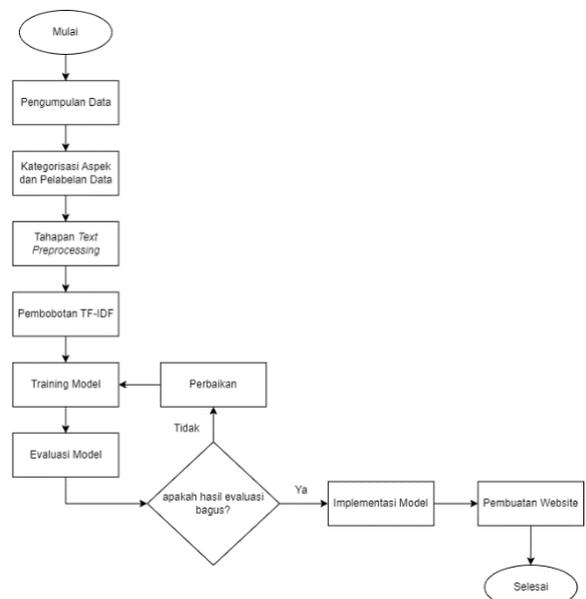
##### - Menghitung Probabilitas Fitur

Algoritma juga menghitung probabilitas munculnya setiap fitur dalam kelas tertentu. Ini melibatkan menghitung berapa kali suatu fitur muncul dalam data pelatihan dari kelas tertentu dibagi dengan total data dalam kelas tersebut. Misalnya, untuk fitur X dalam kelas A, algoritma menghitung  $P(X|A)$ .

#### B. Tahap Klasifikasi:

Setelah tahap pelatihan selesai, model Naive Bayes dapat digunakan untuk mengklasifikasikan data baru ke dalam salah satu kelas berdasarkan probabilitas yang telah dihitung sebelumnya. Ini melibatkan menghitung probabilitas kelas untuk data baru berdasarkan fitur-fiturnya dan memilih kelas dengan probabilitas tertinggi sebagai kelas prediksi untuk data tersebut.

Gambar 1 memperlihatkan alur penelitian yang dilakukan, meliputi beberapa tahapan.



**Gambar 1.** Alur Penelitian

#### A. Pengumpulan Data

Penelitian ini mengumpulkan data ulasan dari Twitter dengan metode *Crawling*. Data tersebut membahas tentang Kota Tegal dalam bahasa Indonesia dan dikategorikan ke dalam aspek Wisata/Hiburan, Fasilitas/Layanan Publik, Pendidikan, dan Kuliner. Pengumpulan data dilakukan dari Januari hingga Desember 2022, menghasilkan total 3154 data. Data kemudian dilabeli manual berdasarkan aspek yang relevan dengan kata-kata kunci tertentu.

#### B. Text Preprocessing

*Preprocessing teks* dalam penelitian ini memiliki tujuan untuk meningkatkan kualitas data dan efektivitas algoritma dengan menghilangkan data yang tidak relevan. Langkah-langkah *preprocessing* meliputi:

1. *Casefolding*: Mengubah semua huruf dalam teks menjadi huruf kecil untuk menghindari kesalahan pengolahan akibat perbedaan besar-kecil huruf.
2. *Filtering*: Menyaring data tidak relevan dan menghapus duplikasi agar data lebih terstruktur dan relevan.

3. *Tokenisasi*: Memecah teks menjadi kata-kata atau unit yang lebih kecil, sambil menghilangkan karakter tidak diperlukan seperti tanda baca, simbol, angka, dan lainnya.
4. *Stemming*: Mengganti kata tidak baku dengan bentuk yang lebih baku untuk memudahkan analisis dan mengurangi ambiguitas.
5. *Stemming*: Mengganti kata tidak baku dengan bentuk yang lebih baku untuk memudahkan analisis dan mengurangi ambiguitas.

Tujuan akhir dari preprocessing adalah mendapatkan data yang lebih terstruktur dan siap untuk dianalisis dengan algoritma yang lebih efektif.

#### C. Pembobotan TF-IDF

TF-IDF adalah teknik yang memberikan bobot pada kata dalam dokumen. Tujuannya adalah mengidentifikasi kata-kata penting atau yang sering muncul dalam dokumen (Kadhim, 2018). Dalam tahap pretraining, metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk menghasilkan representasi vektor kata-kata. Proses ini menerapkan pembobotan kata untuk mengubah teks menjadi data numerik. Ini diperlukan karena algoritma machine learning bekerja optimal dengan data numerik. Oleh karena itu, dalam penelitian ini, data teks diubah menjadi vektor numerik menggunakan pembobotan kata TF-IDF, yang merupakan salah satu metode dalam pembobotan seperti Bag of Words, N-gram, Word2Vec, dan lainnya (Wati et al., 2023). Semakin jarang istilah muncul dalam dokumen, semakin tinggi nilai IDF yang diberikan (Yulietha et al., 2017). Rumus TF:

$$TF (ta, db) = f (ta, db) \quad (1)$$

IDF dapat dihitung dengan rumus sebagai berikut:

$$IDF = \log n/(df(t)) \quad (2)$$

TF-IDF dapat dirumuskan sebagai berikut:

$$TF-IDF(ta, db) = TF(ta, db) * IDF(ta, db) \quad (3)$$

#### D. *Training* Model

Pada training model terdapat beberapa tahap, diantaranya:

##### 1. Pembagian Data

Tahap ini membagi data menjadi 80% data training dan 20% data *testing*. Data *training* digunakan untuk mengembangkan model klasifikasi Naïve Bayes. Hasil klasifikasi dari data *training* diterapkan pada data *testing* untuk evaluasi performa model. Pembagian 80:20 menghindari overfitting dan memastikan evaluasi yang akurat pada data belum terlihat.

Pada model, probabilitas kata dalam kelas dihitung tanpa mempertimbangkan fitur-fitur khusus dari contoh. Estimasi probabilitas priori dan probabilitas fitur dilakukan dengan menghitung jumlah kemunculan kelas dan fitur dalam data latih. Model ini menghasilkan 8 model untuk klasifikasi aspek dan sentimen pada aspek.

##### 2. Model Aspek

Pada tahap ini, dataset dibagi menjadi data training dan data *testing* menggunakan `train_test_split` dari library sklearn. Kemudian, data training disesuaikan dengan Random Oversampler untuk menyeimbangkan kelas minoritas dan mayoritas. Tahap *oversampling* hanya berlaku pada data *training* dan tidak digunakan pada data *testing* agar evaluasi model tetap realistis. *Oversampling* pada data *training* meningkatkan kemampuan model dalam mengenali kelas minoritas. Model klasifikasi

Multinomial NB dilatih pada data *training* yang telah mengalami *oversampling*, dan evaluasi akurasi serta presisi dilakukan pada data yang sama. Evaluasi akhir tetap dilakukan pada data testing yang tidak mengalami perubahan untuk hasil yang objektif dan andal.

### 3. Model Sentimen Aspek

Pada tahap ini, data sentimen aspek diambil bersama dengan vektor TF-IDF dari dataset dan dimasukkan ke objek `READ_DATA`. Data sentimen dengan nilai -1 dihapus, dan data training serta data testing dibentuk melalui fungsi `train_test_split` dari library `sklearn`. Model *Multinomial Naive Bayes* dilatih dengan data *training* dan diuji menggunakan data testing. Dalam klasifikasi sentimen aspek ini, tidak digunakan *oversampling*. Namun, *oversampling* bisa diterapkan untuk mengatasi ketidakseimbangan kelas dalam dataset. Teknik *oversampling* seperti *Random Oversampling* atau *Synthetic Minority Oversampling Technique (SMOTE)* dapat digunakan untuk mengatasi ketidakseimbangan kelas tersebut.

### 4. Pelatihan

Pada tahap pelatihan model, dataset dilabeli dan dijalani proses preprocessing untuk digunakan dalam metode *Naïve Bayes*. Model data ini disimpan dalam format *pickle*. Saat pelatihan, sistem membaca tweet dan mengklasifikasikannya ke dalam aspek yang sesuai. Jika kata tidak cocok dengan kriteria suatu aspek, sistem akan mencari kategori lain. Kata-kata yang tidak cocok akan dihapus otomatis. Fitur diekstraksi dari kata dan model data disimpan untuk penggunaan selanjutnya.

### E. Evaluasi Model

Pada tahap evaluasi akhir penelitian, dilakukan analisis untuk menentukan apakah algoritma yang digunakan merupakan model klasifikasi yang baik dengan akurasi tinggi dalam menganalisis sentimen ulasan terhadap

Kota Tegal. Evaluasi ini menggunakan confusion matrix, yang merupakan alat untuk mengukur kualitas dari proses klasifikasi terhadap satu atau lebih kelas.

		Actual Class	
		Yes	No
Preditd Class	Yes	TP	FN
	No	FP	TN

**Gambar 2.** Actual Class dan Predict Class

Dalam confusion matrix, terdapat 4 nilai klasifikasi yang digunakan untuk menghitung akurasi, presisi, recall, dan f1-score:

1. True Positive (TP): Hasil prediksi positif yang benar.
2. True Negative (TN): Hasil prediksi negatif yang benar.
3. False Positive (FP): Hasil prediksi positif yang seharusnya negatif.
4. False Negative (FN): Hasil prediksi negatif yang seharusnya positif.

Dengan menggunakan nilai-nilai ini, dapat dihitung akurasi, presisi, recall, dan f1-score menggunakan persamaan yang sesuai.

$$akurasi = \frac{TN+TP}{(TN+FN+TP+FP)} \quad (4)$$

$$precision = \frac{true\ positive}{true\ positive+false\ positive} \quad (5)$$

$$recall = \frac{true\ positive}{true\ positive+false\ negative} \quad (6)$$

$$f1 - score = 2 \times \frac{(precision \times recall)}{(precision + recall)} \quad (7)$$

Jumlah data yang salah positif diperoleh dari total nilai positif yang tidak benar positif dalam setiap kelas yang sesuai.

**Tabel 1.** Daftar jurnal-jurnal bereputasi nasional di Fakultas Sains dan Teknologi Unida Gontor (Dihin, 2016).

No	Nama Jurnal	Penerbit
1	Agrotech	Prodi Agroteknologi
2	FIJ	Prodi Teknik Informatika

### III. HASIL DAN PEMBAHASAN

#### 3.1 Hasil

##### 1. Hasil Pengumpulan Data

Sampel dari pengumpulan data dapat dilihat pada Gambar 3 data sampel hasil crawling berupa file csv dan selanjutnya dilabeli berdasarkan aspek yang relevan dengan kata-kata kunci tertentu.

A	B	C	D	E
1 tweet	wisata/hil pendidikan/fasilitas/lc kuliner			
2 Wisuda Universitas BSI Kampus Kabupaten Banyumas ke-16 & Kampus Kota Tegal ke-4 (06/12/2021) @ Java Heritage Hotel Purwokerto https://t.co/MSS0lvsQgW	-1	1	-1	-1
3 (03/12/2021) Mengikuti kegiatan bersih pantai yang diselenggarakan Lanal Tegal di Pantai Alam Indah Kota Tegal. Kegiatan ini diikuti oleh berbagai instansi terkait di kc	-1	-1	1	-1
4 @infopIntegal Selamat malam, mohon di tindak lanjuti di daerah saya dari jam 17.30 WIB mati listrik sampai sekarang belum menyalakan Alamat di Jl. Dcripto mangunkusu	-1	-1	0	-1
5 [TF] info banjir daerah tegal kota dong lurr, kayong pgn seblak kie, makasih	-1	-1	-1	1
6 Pemasangan portal di jalan masuk Alun-alun Kota Tegal diprotes pedagang. Wali Kota Tegal Deddy Yon Supriyono membeberkan alasan dari penutupan jalan tersebut. h	-1	-1	0	-1
7 Dalam rangka memperingati Hari Juang Kartika (HJK) TNI-AD tahun 2021 Kodim 0734/Kota Yogyakarta gelar kegiatan bhakti sosial donor darah bertempat di Jl. Tegal Gen	-1	-1	1	-1
8 Perusahaan di Kota Tegal wajib membayar upah sesuai Upah Minimum Rp 2.005.930,52 utk pekerja dengan masa kerja kurang dari 12 bulan; dan wajib menyusun Struktu	-1	-1	-1	-1
9 PPKM Level 3 Batal, Akhirnya Tegal Nyalakan Lagi Lampu-lampu Jalan di Kota #JawaTengah #KotaTegal https://t.co/kU3VzqvaTP	-1	-1	1	-1
10 Askrindo memberikan pembinaan dan bantuan Tanggung Jawab Sosial dan Lingkungan berupa sarana produksi kepada beberapa Kelompok seperti UMKM Seafood Olah	-1	-1	1	-1
11 Pembukaan Pelatihan Jabatan Fungsional Perawat Angkatan II (9 - 16 Desember 2021) dan Pelatihan Jabatan Fungsional Bidan (9 - 15 Desember 2021) BKPPD Kota Tegal	-1	1	-1	-1
12 Live Report Kasatalantas Polres Tegal Kota melaporkan pantauan Kamseltincarlantas dari Simpang Pramesthi Kota Tegal. Situasi Arus Lalu Lintas dari arah Timur maupun	-1	-1	1	-1
13 Aplikasi itu bertujuan memudahkan dan mendekatkan pelayanan PN Tegal dengan masyarakat, khususnya warga Kota Tegal. https://t.co/aTyuprYMCV	-1	-1	1	-1
14 Waktunya kulineran sate kambing muda di Sate Tegal Marem lagi kuyy! Satenya terbuat dari daging kambing muda pilihan yang didatangkan langsung dari Kota Tegal pa	-1	-1	-1	1
15 Ditentang Warga, Wali Kota Tegal Ngotot Tutup Jalan #protes #masyarakat #penutupan #jalan #tegal #alun #alun #protes https://t.co/UrOrYeVtPr	-1	-1	0	-1
16 Warga Kota Tegal Sekarang Bisa Mencetak Adminduk Mandiri https://t.co/2aQFB6XJST	-1	-1	1	-1
17 Wali Kota Tegal Dedy Yon Supriyono mengungkapkan alasan kebijakan menutup kawasan Alun-alun dan Jalan Pancasila dengan portal. https://t.co/8tmiML3Z3a	-1	-1	1	-1
18 Sejak dipadamkan mulai 1 Desember 2021, lampu penerangan jalan umum (PJU) di Kota Tegal akhirnya kembali dinyalakan bertahap mulai Rabu (8/12/2021) malam. #pi	-1	-1	1	-1
19 Meski diprotes warga dan pedagang, pemortalan akses jalan menuju kawasan Alun-alun dan Jalan Pancasila Kota Tegal tetap diberlakukan dan tetap melarang kendaraa	-1	-1	0	-1
20 Jalan Dr. Soetomo Kota Tegal. Pemadaman Lampu Jalan oleh @PemkotTegal Mohon Pak @ganjarpranowo kasih solusi https://t.co/qSp6oRQpIM	-1	-1	0	-1

**Gambar 3.** Dataset

##### 2. Hasil Preprocessing

Hasil dari tahapan preprocessing masing-masing dijelaskan sebagai berikut: Tabel 2 menunjukkan hasil proses casefolding, Tabel 3 menunjukkan hasil proses filtering, Tabel 4

menunjukkan hasil proses tokenisasi, Tabel 5 menunjukkan hasil dari proses konversi slangword, Tabel 6 menunjukkan hasil dari proses stopword.

**Tabel 2.** Sampel data sebelum dan sesudah Casefolding.

No	Data Sebelum Case Folding	Data Sesudah Case Folding
1	Menikmati pemandangan tegal dari atas, bisa jadi pengalaman dan sensasi tersendiri saat naik bianglala di rita park #tegal. Kalau pas cuaca bagus bisa melihat gunung slamet di sebelah selatan atau kalau malam, bisa juga lihat lampu2 kota	menikmati pemandangan tegal dari atas, bisa jadi pengalaman dan sensasi tersendiri saat naik bianglala di rita park #tegal. kalau pas cuaca bagus bisa melihat gunung slamet di sebelah selatan atau kalau malam, bisa juga lihat lampu2 kota
2	Sudah lama rasanya, tidak duduk2 di depan masjid agung kota tegal ini sambil nikmatin mi ayam, apalagi pas sore2 sambil menunggu maghrib tiba, rasanya jadi semakin enak. Punya kuliner favorit yang biasa mangkal di depan masjid agung juga?	sudah lama rasanya, tidak duduk2 di depan masjid agung kota tegal ini sambil nikmatin mi ayam, apalagi pas sore2 sambil menunggu maghrib tiba, rasanya jadi semakin enak. punya kuliner favorit yang biasa mangkal di depan masjid agung juga?

**Tabel 3.** Sampel data sebelum dan sesudah Filtering

No	Data Sebelum <i>Filtering</i>	Data Sesudah <i>Filtering</i>
1	Menikmati pemandangan tegal dari atas, bisa jadi pengalaman dan sensasi tersendiri saat naik bianglala di rita park #tegal. Kalau pas cuaca bagus bisa melihat gunung slamet di sebelah selatan atau kalau malam, bisa juga lihat lampu2 kota	menikmati pemandangan tegal dari atas, bisa jadi pengalaman dan sensasi tersendiri saat naik bianglala di rita park tegal. kalau pas cuaca bagus bisa melihat gunung slamet di sebelah selatan atau kalau malam, bisa juga lihat lampu kota
2	Sudah lama rasanya, tidak duduk2 di depan masjid agung kota tegal ini sambil nikmatin mi ayam, apalagi pas sore2 sambil menunggu maghrib tiba, rasanya jadi semakin enak. Punya kuliner favorit yang biasa mangkal di depan masjid agung juga?	sudah lama rasanya, tidak duduk di depan masjid agung kota tegal ini sambil nikmatin mi ayam, apalagi pas sore sambil menunggu maghrib tiba, rasanya jadi semakin enak. punya kuliner favorit yang biasa mangkal di depan masjid agung juga

**Tabel 4.** Sampel data sebelum dan sesudah Tokenisasi

No	Data Sebelum Tokenisasi	Data Sesudah Tokenisasi
1	menikmati pemandangan tegal dari atas, bisa jadi pengalaman dan sensasi tersendiri saat naik bianglala di rita park tegal. kalau pas cuaca ...	[menikmati, pemandangan, tegal, dari, atas, bisa, jadi, pengalaman, dan, sensasi, tersendiri, saat, naik, bianglala, di, rita, park, tegal, kalau, pas, cuaca]
2	sudah lama rasanya, tidak duduk di depan masjid agung kota tegal ini sambil nikmatin mi ayam, apalagi pas sore sambil ...	[sudah, lama, rasanya, tidak, duduk, di, depan, masjid, agung, kota, tegal, ini, sambil, nikmatin, mi, ayam, apalagi, pas, sore]

**Tabel 5.** Sampel data sebelum dan sesudah Stemming

No	Data Sebelum <i>Konversi Slangword</i>	Data Sesudah <i>Konversi Slangword</i>
1	[menikmati, pemandangan, tegal, dari, atas, bisa, jadi, pengalaman, dan, sensasi, tersendiri, saat, naik, bianglala, di, rita, park, tegal, kalau, pas, cuaca]	[menikmati, pemandangan, tegal, dari, atas, bisa, jadi, pengalaman, dan, sensasi, sendiri, saat, naik, bianglala, di, rita, park, tegal, kalau, pas, cuaca]
2	[sudah, lama, rasanya, tidak, duduk, di, depan, masjid, agung, kota, tegal, ini, sambil, nikmatin, mi, ayam, apalagi, pas, sore]	[sudah, lama, rasanya, tidak, duduk, di, depan, masjid, agung, kota, tegal, ini, sambil, nikmatin, mi, ayam, apalagi, pas, sore]

**Tabel 6.** Sampel data sebelum dan sesudah Stopword

No	Data Sebelum <i>Stopword</i>	Data Sesudah <i>Stopword</i>
1	[menikmati, pemandangan, tegal, dari, atas, bisa, jadi, pengalaman, dan, sensasi, sendiri, saat, naik, bianglala, di, rita, park, tegal, kalau, pas, cuaca]	[menikmati, pemandangan, tegal, dari, atas, bisa, jadi, pengalaman, sensasi, sendiri, naik, bianglala, rita, park, tegal, kalau, cuaca]
2	[sudah, lama, rasanya, tidak, duduk, di, depan, masjid, agung, kota, tegal, ini, sambil, nikmatin, mi, ayam, apalagi, pas, sore]	[sudah, lama, rasanya, tidak, duduk, depan, masjid, agung, kota, tegal, sambil, nikmatin, mi, ayam, apalagi, sore]

3. Pembobotan TF-IDF  
*Pretraining* menggunakan metode TF-IDF menghasilkan representasi vektor untuk setiap

kata. Gambar 4 memperlihatkan hasil dari vektor TF-IDF yang diperoleh.

```
vektor_tfidf
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

Gambar 4. TF-IDF

4. Hasil Implementasi Model

Proses dimulai dengan melakukan pengambilan data real-time melalui Twitter. Data tersebut kemudian mengalami tahap *preprocessing* sebelum diuji menggunakan model yang telah dibangun sebelumnya. Setelah pengujian, hasilnya akan disimpan dalam *database*. Gambar 5 menampilkan tampilan data hasil setelah melalui tahap implementasi model dan dimasukkan ke dalam database.

tanggal	tweet	wisata_hiburan	pendidikan	fasilitas_layanan_publik	kuliner
2020-01-01	Ratusan hektar bawang merah di Kota Tegal terendam...	-1	-1	-1	-1
2020-01-01	Malam tahun baru menjadi kelabu, semalam di Kota T...	-1	-1	-1	-1
2020-01-01	(Tegal, 01/01/2020) bin melaporkan dari Terminal T...	-1	-1	-1	-1
2020-01-02	Ini alun alun kota Tegal! Jadi bukan salah pak Ani...	1	-1	-1	1
2020-01-02	@NyRusmini2 @Faujijn_ @yusuf_dumdum Lampung merah ...	-1	-1	1	-1
2020-01-02	@Faujijn_ @NyRusmini2 @yusuf_dumdum RSUD Kardina...	-1	-1	1	-1
2020-01-02	@ganjarpranowo @pemkabKEBUMEN @pemkab_klaten @Disk...	-1	-1	-1	-1
2020-01-02	Kantor Dinas Sosial (Dinsos) kota Tegal kebanjiran...	-1	-1	-1	-1
2020-01-02	@lchi_fm Siang kak lchi, disistem memang ditulis...	-1	-1	-1	-1
2020-01-02	Ratusan ekor ikan tambak yang berada di Jalan Mata...	-1	-1	0	-1
2020-01-03	@infotegal Kayang enak es sagwan e mas Edi min. Jl...	-1	-1	-1	1
2020-01-03	#Pemberdayaan TEGAL, Relawan inspirasi Rumah Zaka...	-1	-1	-1	-1
2020-01-03	@CNNIndonesia "Pak bantu warga sendiri aja dulu no...	-1	-1	-1	-1
2020-01-03	kurjungi markas PMI Kota Tegal yg terkena dampak p...	-1	-1	-1	-1

Gambar 5. Implementasi Model

5. Hasil Evaluasi Model

Data training digunakan untuk melatih model, sementara data testing digunakan untuk mengevaluasi model pada data yang belum pernah dilihat sebelumnya. Pengujian melibatkan perbandingan hasil prediksi model dengan nilai sebenarnya pada data *testing* menggunakan metrik evaluasi seperti Akurasi, Presisi, *Recall*, atau *F1-Score*. Tabel berikut menunjukkan hasil pengujian model

menggunakan data training untuk setiap aspek.

Tabel 7. Pengujian Model Aspek Training

Aspek/Kategori	Accuracy
Wisata/Hiburan	96%
Pendidikan	99%
Fasilitas/Layanan Publik	94%
Kuliner	95%

Tabel 8. Pengujian Model Aspek Testing

Aspek/Kategori	Accuracy
Wisata/Hiburan	86%
Pendidikan	95%
Fasilitas/Layanan Publik	74%
Kuliner	79%

Pengujian selanjutnya yaitu pengujian pada model Sentimen. Dengan hasil akurasi dari pengujian data training pada Tabel 9, sedangkan hasil pengujian pada data testing pada Tabel 10.

Tabel 9. Hasil pengujian model sentimen aspek training

Aspek/Kategori	Accuracy
Wisata/Hiburan	99%
Pendidikan	96%
Fasilitas/Layanan Publik	100%
Kuliner	99%

Tabel 10. Hasil pengujian model sentimen aspek testing

Aspek/Kategori	Accuracy
Wisata/Hiburan	91%
Pendidikan	86%
Fasilitas/Layanan Publik	82%
Kuliner	98%

Tabel 11. Hasil Pengujian Klasifikasi Aspek

Aspek/Kategori	Precision	Recall	F1-Score
Wisata/Hiburan	94%	86%	89%
Pendidikan	85%	74%	77%
Fasilitas/Layanan Publik	97%	95%	96%
Kuliner	93%	79%	83%

**Tabel 12.** Hasil Pengujian Sentimen

Aspek/Kategori	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Wisata/Hiburan	98%	98%	98%
Pendidikan	87%	86%	86%
Fasilitas/Layanan Publik	89%	82%	83%
Kuliner	91%	91%	91%

### 3.2 Pembahasan

Akurasi pada pengujian data *testing* mengalami penurunan karena faktor seperti *overfitting* dan kurangnya variasi dalam data *training*. Meskipun demikian, model ini masih menunjukkan nilai akurasi yang baik, dengan rata-rata di atas 70% untuk pengujian data *testing* dan rata-rata di atas 90% untuk pengujian data *training*.

Terdapat perbedaan yang signifikan antara nilai akurasi, *precision*, *recall*, dan *f1-score* pada setiap aspek. Perbedaan tersebut muncul karena akurasi hanya mempertimbangkan hasil prediksi yang benar (*True Positive* dan *True Negative*). Selain itu, nilai *recall* lebih tinggi dari pada *precision*, yang menunjukkan bahwa model lebih baik dalam mengidentifikasi data positif daripada keseluruhan data yang sebenarnya positif, serta lebih berhati-hati dalam menghindari kesalahan *False Positive*. Karena ketidakseimbangan data, nilai *precision* rendah menunjukkan bahwa model sulit dalam memprediksi data yang sebenarnya positif. Sebagai akibatnya, model cenderung memprediksi kelas yang memiliki jumlah data yang lebih banyak. Karena nilai *precision* dan *recall* yang rendah, hal ini mengakibatkan nilai *f1-score* yang rendah.

### IV. KESIMPULAN

Hasil penelitian menunjukkan bahwa aplikasi yang dikembangkan dapat mengklasifikasikan ulasan masyarakat Kota Tegal ke dalam beberapa aspek dengan tingkat akurasi pada

data *training* diatas 90% dan tingkat akurasi pada data *testing* diatas 75% pada setiap kategori. Untuk hasil implementasi model berupa data yang disimpan kedalam *database* terlihat peningkatan yang signifikan pada data pada tahun 2023 dalam aktivitas terkait sentimen, terutama pada bulan Juni dan Juli.

### DAFTAR PUSTAKA

- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode *Naïve Bayes* dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(2), 71–76.
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment Analysis of Social Media Twitter with Case of Anti-Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using *Naïve Bayes*, Decision Tree, LGBT Campaign in Indonesia using *Naïve Bayes*, Decision Tree, and R. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Graham, M. W., Avery, E. J., & Park, S. (2015). The role of social media in local government crisis communications. *Public Relations Review*, 1–9. <https://doi.org/10.1016/j.pubrev.2015.02.001>
- Heranova, O. (2019). Synthetic Minority Oversampling Technique pada Averaged One Dependence Estimators untuk Klasifikasi Credit Scoring. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1, 443–450.
- Kadhim, A. I. (2018). An Evaluation of Preprocessing Techniques for Text Classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 16(December), 22–32.
- Khaerani, T. R., & Sarai, M. V. (2020). Analisis Pengembangan *Electronic Government di Kota Tegal*. 8, 79–86.

- Rahman, M. A., Budiarto, H., & Setiawan, E. I. (2019). Aspect Based Sentimen Analysis Opini Publik Pada Instagram dengan Convolutional Neural Network. *Journal of Intelligent System and Computation*, 1(2), 50–57. <https://doi.org/10.52985/insyst.v1i2.83>
- Somantri, O., & Dairoh. (2019). Analisis Sentimen Penilaian Tempat Tujuan Wisata Kota Tegal Berbasis Text Mining. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(2), 191–196. <https://doi.org/10.26418/jp.v5i2.32661>
- Sulistio, H. G., & Handojo, A. (2022). Aspect-Based Sentiment Analysis pada Ulasan E- Commerce dengan Metode Support Vector Machine untuk Mendapatkan Informasi Sentimen dari Beberapa Aspek. *Jurnal Infra*.
- Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), 84–93. <https://doi.org/10.34010/jamika.v13i1.9424>
- Yulietha, I. M., Faraby, S. Al, & Adiwijaya. (2017). Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine Sentiment Classification Of Movie Reviews. *E-Proceeding of Engineering'*, 4(3), 4740–4750.
- Yulita, W., Nugroho, E. D., & Algifari, M. H. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 1. <https://doi.org/10.33365/jdmsi.v2i2.1344>
- Yutika, C. H., Adiwijaya, & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 422–430. <https://doi.org/10.30865/mib.v5i2.2845>
- Zhao, A., & Yu, Y. (2021). Knowledge-enabled BERT for aspect-based sentiment analysis. *Knowledge-Based Systems Journal*, 227. <https://doi.org/10.1016/j.knosys.2021.107220>