BETRIK
BESEMAH TEKNOLOGI INFORMASI & KOMPUTER
Editor Office : Pusat Penelitian & Pengabdian Pada Masyarakat (PPPM) ITPA
Phone : 0857-9716-9578
email : betrikitpa@itpa.ac.id
ISSN : 2339 - 1871

# *Optimization of K value in the K-NN For Classification Review Pagar Alam City Tourism using Expectation Maximization and Grid Search CV*

**Risnaini Masdalipa[1], Yogi Isro Mukti[2], Ferry Putrawansyah[3]**
Program Studi Teknik Informatika, Institut Teknologi Pagar Alam, Indonesia[1,2,3]
**Sur-el :***risnianipga@gmail.com[1], yogieisrmukti@gmail.com[2],feyputrawansyah@gmail.com[2]

Penulis Korespondensi: Risniani Masdalipa, risnianipga@gmail.com

**Abstrak**: Perkembangan sektor pariwisata di Kota Pagar Alam menuntut adanya sistem analisis sentimen yang mampu memberikan informasi akurat mengenai persepsi wisatawan. Analisis sentimen terhadap ulasan daring dapat menjadi dasar pengambilan keputusan dalam pengelolaan dan peningkatan kualitas destinasi wisata. Penelitian ini bertujuan untuk menghasilkan klasifikasi sentimen yang akurat dengan mengoptimalkan nilai *K* pada algoritma *K-Nearest Neighbor* (KNN) menggunakan metode *Expectation Maximization* (EM) dan *Grid Search Cross Validation* (GS-CV). Metode penelitian yang digunakan adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM), yang meliputi tahapan: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Data diperoleh melalui teknik *scraping* dari ulasan pariwisata Kota Pagar Alam, menghasilkan 4.806 data ulasan yang mencakup delapan objek wisata utama, seperti Tugu Rimau, Gunung Dempo, dan Air Terjun Tujuh Kenangan. Hasil analisis menunjukkan bahwa destinasi dengan sentimen positif tertinggi adalah Tugu Rimau (skor 0,678), sedangkan yang paling negatif adalah Air Terjun Tujuh Kenangan (skor 0,006). Uji performa menunjukkan peningkatan akurasi model dari 81% menjadi 89% setelah optimasi dengan EM dan GS-CV, dengan nilai *precision* 88%, *recall* 89%, dan *F1-score* 85%. Temuan ini membuktikan bahwa kombinasi EM dan GS-CV efektif meningkatkan kinerja algoritma KNN untuk klasifikasi sentimen pariwisata Kota Pagar Alam

***Kata kunci :*** *Expectation Maximixation, Grid Search* CV, K-NN, Pagar Alam, Pariwisata

***Abstract*** *The development of the tourism sector in Pagar Alam City requires a sentiment analysis system capable of accurately capturing tourists' perceptions. Sentiment analysis of online reviews can serve as a valuable foundation for decision-making in managing and improving tourism destinations. This study aims to produce an accurate sentiment classification by optimizing the* K *value in the* K-Nearest Neighbor *(KNN) algorithm using* Expectation Maximization *(EM) and* Grid Search Cross Validation *(GS-CV). The research employs the* Cross Industry Standard Process for Data Mining *(CRISP-DM) methodology, consisting of six stages: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Data were collected through web scraping of online tourism reviews, resulting in 4,806 reviews across eight major tourist attractions, including Tugu Rimau, Mount Dempo, and Tujuh Kenangan Waterfall. The results indicate that Tugu Rimau received the highest positive sentiment score (0.678), while Tujuh Kenangan Waterfall showed the lowest (0.006). Model performance evaluation revealed that KNN accuracy improved from 81% to 89% after optimization using EM and GS-CV, achieving 88% precision, 89% recall, and an F1-score of 85%. These findings demonstrate that the integration of EM and GS-CV effectively enhances the classification accuracy of KNN in sentiment analysis for Pagar Alam's tourism reviews.*

***Keywords:*** *Expectation Maximixation, Grid Search* CV, K-NN, Pagar Alam, *Tourism*

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

# 1. PENDAHULUAN

Kota pagar alam merupakan salah satu kota yang terletak di provinsi sumatera selatan. Kota ini berjarak sekitar 298 km dari kota Palembang dan juga berjarak sekitar 60 km di sebelah barat daya Kabupaten Lahat. Kota Pagar Alam adalah salah satu kota di provinsi Sumatra Selatan yang dibentuk berdasarkan Undang–Undang Nomor 8 Tahun 2001 (Lembaran Negara RI Tahun 2001 Nomor 88, Tambahan Lembaran Negara RI Nomor 4115), sebelumnya kota Pagar Alam termasuk kota administratif dalam lingkungan Kabupaten Lahat(1). Kota ini memiliki luas sekitar 633,66 km² dengan jumlah penduduk 139.194 jiwa dan memiliki kepadatan penduduk sekitar 218 jiwa/km. Untuk lebih lengkapnya dapat dilihat pada tabel berikut ini:

Tabel 1: Data Jumlah Penduduk kota Pagar Alam

| Nama kecamatan | Luas Wilayah | Jumlah Penduduk | Kepadatan Penduduk |
|---|---|---|---|
| Dempo Selatan | 243.86 km² | 11.897 jiwa | 48 jiwa/km² |
| Dempo Tengah | 144.05 km² | 13.07 jiwa | 90 jiwa/km² |
| Dempo Utara | 127.11 km² | 20.978 jiwa | 164 jiwa/km² |
| Pagar Alam Selatan | 63.17 km² | 50.124 jiwa | 786 jiwa/km² |
| Pagar Alam Utara | 55.47 km² | 43.124 jiwa | 770 jiwa/km² |
| Total | 633.66 km² | 139.194 jiwa | 218 jiwa/km² |

Data diatas menunjukan bawah jumlah penduduk kota Pagar Alam cukup tinggi meskipun letak nya berada jauh dari Ibukota Provinsi Sumatera selatan. Kota Pagar Alam memiliki ketinggian antara 400-3.400 meter di atas permukaan laut menjadikan kota ini merupakan kota yang subur dengan 100.000 hektar. Dan juga dengan keberadaaan gunung dempo menjadikan kota Pagar alam memiliki kekayaan pariwisata baik wisata air terjun sampai dengan wisata buatan.

Pada kenyataannya kebijakan pemerintah perihal pengembangan pariwisata dan dampak positif dan negative belum terdokumentasi dengan baik padahal dari aspek teknologi Google Maps sudah menyediakan fitur bagi pengguna untuk dapat memberikan ulasan pengalaman berwisata dikota Pagar Alam. Selain ulasan wisatawan dapat memberikna rating pada wisata yang mereka kunjungi. Ulasan melalui google maps ini belum di manfaatkan oleh pemerintah untuk mendukung kebijakan dan pengembangan wisata padahal Ulasan dan rating tersebut dapat di tambang (mining) dan di Analisa untuk menghasilkan data dan informasi yang bermanfaat bagi pihak pemerintah [2]

Teknologi yang umum digunakan dalam menganalisa ulasan yakni menggunakan sentiment Analyisis yakni pemrosesan bahasa alami (NLP) dan machine learning untuk mengidentifikasi dan mengekstrak emosi atau opini yang terkandung dalam teks. Analisis ini membantu menentukan apakah suatu teks mengekspresikan sentimen positif, negatif, atau netral. Analisis sentimen bertujuan untuk memahami sikap, pendapat, atau emosi yang tersirat dalam teks. Ini bisa berupa ulasan produk, komentar media sosial, artikel berita, atau jenis teks lainnya. [3]

Metode K-Nearest Neighbors (KNN) dapat diterapkan dalam analisis sentimen untuk mengklasifikasikan sentimen teks menjadi positif, negatif, atau netral. KNN bekerja dengan

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

mengukur jarak antara dokumen teks yang diwakili sebagai vektor fitur dengan dokumen lain dalam data pelatihan. Dokumen baru kemudian diklasifikasikan berdasarkan mayoritas sentimen dari k tetangga terdekatnya. Kekurangan dari kNN adalah nilai k bias, komputasi kompleks, keterbatasan memori, dan mudah tertipu dengan atribut yang tidak relevan [4]. Salah satu perbaikan kNN adalah kNN , yang bertujuan untuk meningkatkan akurasi dari kNN, dengan menambahkan perhitungan validity, karena dianggap perhitungan bobot yang terdapat pada kNN, memiliki permasalahan outlier [5]. Namun, KNN juga memiliki kelemahan yang sama dengan kNN yaitu nilai k bias dan komputasi yang kompleks. Berdasarkan permasalahan KNN tersebut, penulis bermaksud untuk melakukan perbaikan dalam hal bias data dan validity akurasi [6]

Lebih lanjut bahwa karena riset ini akan menangani data yang besar sehingga KNN bekerja kurang optimal untuk mengklasifikasi data tersebut, kinerja algoritma KNN menjadi menurun yang disebabkan penghitungan jarak antara titik baru dengan yang ada lama. Kekurangan tersebut akan di optimalkan dengan oleh Algoritma Expectation Maximization karena kemampuannya mencari penduga parameter suatu model menggunakan pendekatan nilai kemungkinan maksimum (MLE) dimana data yang dimiliki tidak lengkap, atau data besar dapat di Analisa dengan baik namun data yang besar harus mampu di evaluasi dengan baik sehingga dibutuhkan Algoritma Grid Search CV yang tanpa melewatkan satu datapun sebagai parameter yang dijadikan data klasifikasi [7]

Sentiment analysis merupakan sebuah teknologi komputerisasi yang dapat membantu dan menganalisis sebuah kalimat opini tekstual yang cara bekerjanya adalah dengan memahami dan mengekstraknya seperti halnya text mining sehingga menghasilkan sebuah informasi sentiment.Text mining merupakan sebuah Teknik atau cara untuk mengektraksi informasi yang berasal dari data teks yang tidakt erstruktur dan berguna [8]. Sentiment analysis dilakukan untuk melihat kecenderungan sesorang berpandangan apakah beropini positif atau negatif, bahkan beropini netral yang dijadikan sebagai sebuah pendukung keputusan [9].

K-Nearest Neighbor (KNN) adalah Teknik klasifikasi data untuk mengelompokkan objek berdasarkan training set yang jaraknya paling dekat dengan objek tertentu. KNN merupakan suatu metode yang menggunakan algoritma supervised dengan hasil queryainstanceyang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN [10]. Algoritma KNN ini sangat sederhana, dan menentukannya berdasarkan jarak terdekat dari query instance ke training sample Pendek atau panjangnya jarak bisa dihitung berdasarkan Euclidean Distance yang direpresentasikan padaPersamaan:

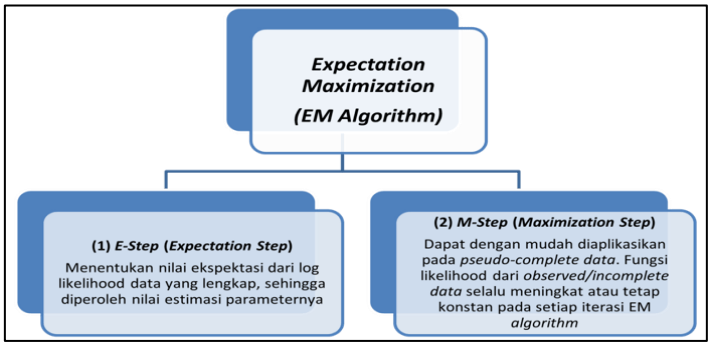
$$d(xi, xj) = \sqrt{\sum} (ar(xi) - ar(xj)) \; 2 \; nr=1 \ldots\ldots\ldots(1)$$

Dimana d(xi,xj) dihasilkan dari pengurangan record (xi) dari tiap atribut kemudian dikuadratkan, lalu hasi lkuadrat dijumlahkan dan diurutkan menurut nilai paling kecil untuk melihat jarak terdekat dengan data uji. Selanjutnya dilakukan K-Fold Cross Validation yang digunakan untuk membagi data menjadi dua bagian, yaitu testing set dan training set [11]. Data dibagi secara random ke himpunan bagian yang dituju. K-Fold Cross Validation digunakan untuk menemukan parameter yang unggul dari satu model. Ini

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

dilakukan dengan cara menguji jumlah error pada testing set. Dalam cross validation, data dipecah kedalam k sampel cross dengan aturan yangalayak. Dari k subset data yang digunakan akan dipakai ak-1sampela sebagai data training dan sampel sisanya untuk data testing [12]

*Expectation Maximixation*, adalah sebuah metode untuk memperoleh estimasi parameter dalam *Maximum Likelihood* (ML) yang mengandung *incomplete data*. Algoritma EM merupakan algoritma partisi yang berbasiskan model dengan perhitungan peluang. Algoritma EM dilakukan dengan dua tahap, yaitu tahap inisialisasi dan iterasi. Pada proses iterasi, maka akan terbagi menjadi dua tahap, yaitu tahap ekspektasi (E-step) dan memaksimumkan (M-step) [13] Adapun tahap algoritma EM adalah sebagai berikut:a.Menentukan inisialisasititik tengah setiap kelompok sebanyak k.b.Melakukan tahap iterasi sampai mencapai titik konvergensi yang telah ditentukan [14]. Tahap iterasi sebagai berikut:-E-step, menghitung nilai harapan bersyarat dari fungsi kemungkinan data lengkap menggunakan pendugaan parameter.-M-step, menghitung parameter yang memaksimalkan nilai ekspektasi dari fungsi log likelihood yang diperolah dari E-step [15].




Gambar 1: Expectation Maximixation

Grid search menggunakan cross validation untuk melatih beberapa model. Grid SearchCV adalah metode untuk memilih kombinasi model dan hyperparameter yang menguji setiap kombinasi dan melakukan validasi untuk setiap kombinasi [16]. Metode ini secara otomatis memvalidasi setiap kombinasi model dan hyperparameter sehingga dapat menghemat waktu pemrosesan [17]. Pasangan hyperparameter yang menghasilkan akurasi terbaik dengan nilai error terkecil merupakan hyperparameter yang optimal. Penelitian ini menggunakan 10-fold CV dimana data latih dibagi menjadi 10 subset tanpa pengulangan. Proses iterasi dilakukan sebanyak 10 kali untuk mendapatkan 10 pengukuran akurasi. Setelah mendapatkan 10 pengukuran akurasi, dihitung rata-ratanya untuk mendapatkan kesalahan CV akhir dengan formula pada persamaan [18].

## 2. METODOLOGI PENELITIAN

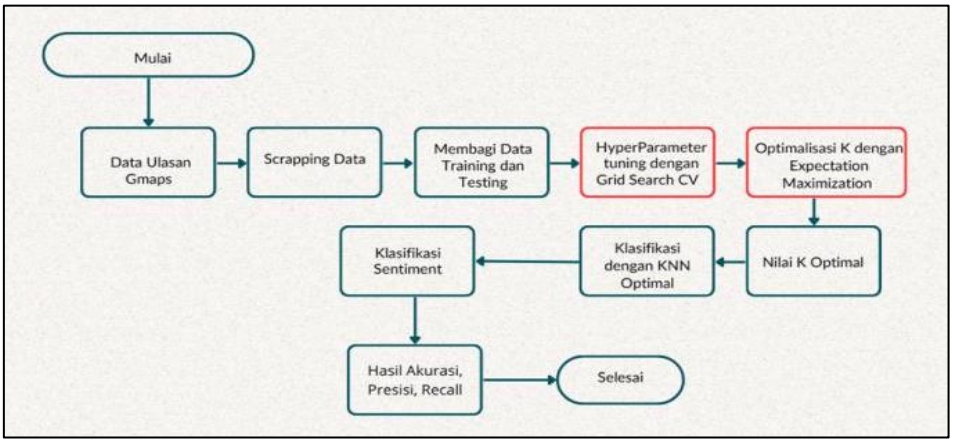
Alur penelitian ini dimulai dengan pengumpulan data ulasan pariwisata dari platform Google Maps (Gmaps) yang memuat tanggapan wisatawan terhadap berbagai destinasi di Kota Pagar Alam. Data yang diperoleh kemudian diproses melalui tahap scrapping data menggunakan teknik *web scraping* untuk

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

mengekstraksi teks ulasan secara otomatis dari halaman web, sehingga data dapat diolah dalam format yang terstruktur.

Tahap berikutnya adalah pembagian data menjadi data pelatihan (*training data*) dan data pengujian (*testing data*). Pembagian ini bertujuan untuk melatih model klasifikasi menggunakan data pelatihan dan menguji performanya dengan data pengujian, guna memastikan model tidak mengalami *overfitting*. Selanjutnya dilakukan proses penyetelan parameter (*hyperparameter tuning*) menggunakan metode *Grid Search Cross Validation (GS-CV)* untuk menemukan kombinasi parameter terbaik pada algoritma *K-Nearest Neighbor* (KNN). Proses ini membantu meningkatkan performa model dengan cara mengevaluasi berbagai nilai parameter melalui validasi silang.

Tahap berikutnya adalah optimasi nilai K menggunakan metode *Expectation Maximization (EM)*. Proses ini bertujuan menentukan nilai K yang paling optimal agar model KNN mampu menghasilkan klasifikasi yang lebih akurat. Hasil dari tahap ini adalah nilai K optimal, yang kemudian digunakan dalam proses klasifikasi sentimen terhadap data ulasan pariwisata. Setelah klasifikasi dilakukan, model dievaluasi menggunakan metrik akurasi, presisi, dan recall untuk mengukur sejauh mana efektivitas model dalam mengidentifikasi sentimen positif dan negatif secara tepat. Tahap akhir penelitian adalah analisis hasil dan penyimpulan, yang mencakup interpretasi terhadap kinerja model serta implikasinya dalam pengembangan sistem analisis sentimen pariwisata Kota Pagar Alam.




Gambar 2: Alur Penelitian

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil

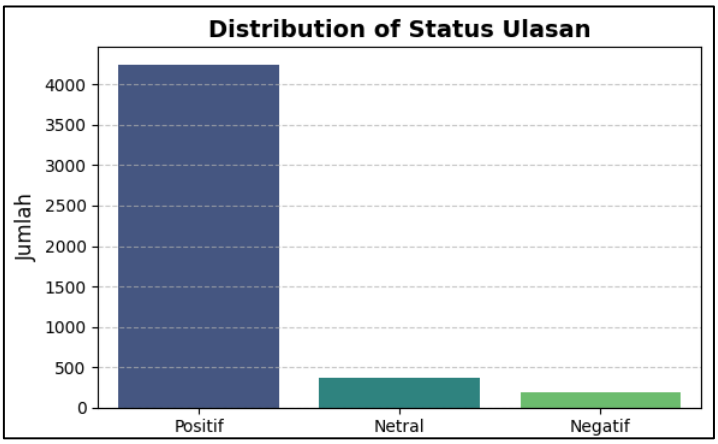
Pada penelitian ini menggunakan Teknik scrolling data untuk menambang ulasan pada google maps. Data yang di mining terdiri dari 7 wisata yang memiliki ulasan lebih dari 100. Peneliti. Menemukan 7 wisata kota Pagar Alam yang memiliki ulasan lebih dari 100. Maka saat melakukan penambangan didapatkan 4806 data ulasan yang teridiri dari:

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

Tabel 2: Dataset

| No | Nama Wisata | Jumlah Ulasan |
|---|---|---|
| **1** | Tugu Rimau | 1558 |
| **2** | Green Paradise | 857 |
| **3** | Gunung Dempo | 625 |
| **4** | Dempo Park | 487 |
| **5** | Air Terjun Mangkok | 403 |
| **6** | Ozil Garden | 318 |
| **7** | Air Terjun Tuju Kenangan | 169 |

Dari data tersebut didapatkan status ulasan yang 4249 Ulasan positif, 368 ulasan netral dan 188 ulasan negative sehingga distribusi rating dapat dilihat pada gambar dibawah ini:




Gambar 3: Klasifikasi Sentimen

### A. Optimalisasi Nilai K menggunakan *Expectation Maximization* (GMM)

Pada saat optimalisasi menggunakan GMM hal pertama yang harus dicari adalah nilai yang paling optimal untuk nilai K. GMM mampu merekomendasikan nilai K yang paling optimal diantara K-1, K-3, K-5, K-7 dan seterusnya. Adapun proses pencarian nilai K yang optimal pada penelitian ini dapat dilihat pada Tebel 3 dibawah ini:

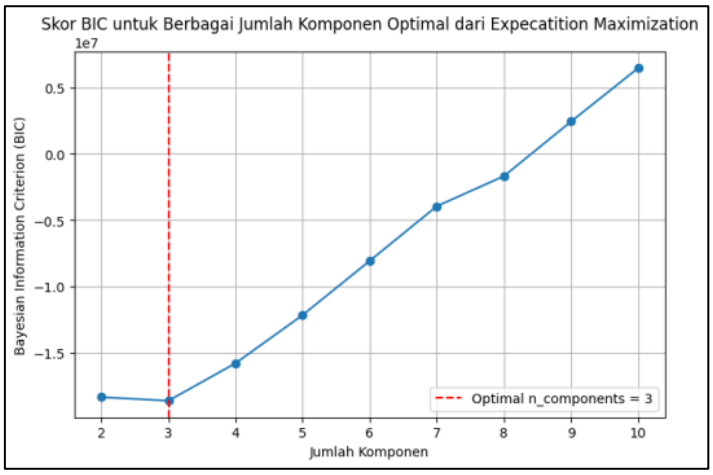
Tabel 3. Skor BIC

| No | Jumlah Komponen | BIC Score |
|---|---|---|
| 1 | 2 | -183530442.14 |
| 2 | 3 | -186204244.25 |
| 3 | 4 | -157967843.09 |
| 4 | 5 | -121579877.99 |
| 5 | 6 | -80694727.50 |
| 6 | 7 | -39485358.35 |
| 7 | 8 | -16791020.64 |
| 8 | 9 | 2436043.69 |
| 9 | 10 | 64706656.53 |

Tabel 3 diatas menunjukan bahwa Baris keluaran proses pencarian jumlah komponen optimal (K) menggunakan algoritma *Expectation Maximization* (GMM). Setiap nilai komponen (2 hingga 10) diuji dan dievaluasi berdasarkan BIC Score (*Bayesian Information Criterion*) — semakin kecil nilai BIC, semakin baik model tersebut. Dari hasil perbandingan, diperoleh bahwa nilai K optimal yang direkomendasikan adalah K = 3, karena memiliki BIC Score paling rendah (-18602444.25) yang

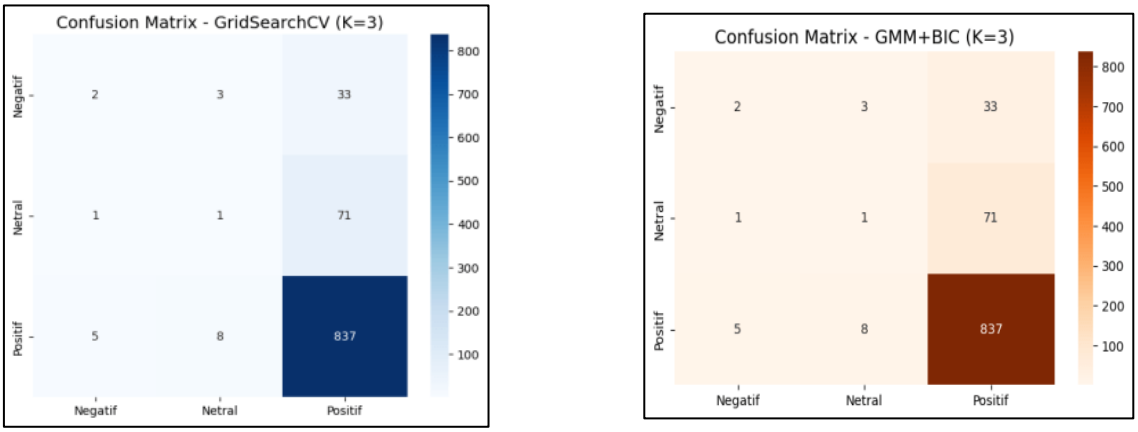
Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

menandakan keseimbangan terbaik antara kompleksitas model dan akurasi estimasi. Setelah nilai K optimal ditemukan, model KNN dengan K = 3 digunakan untuk klasifikasi sentimen. Hasil akhir evaluasi menunjukkan peningkatan performa model dengan akurasi tinggi (berdasarkan perhitungan sebelumnya sebesar 89%), menandakan bahwa metode EM berbasis BIC efektif dalam mengoptimasi parameter K untuk meningkatkan performa klasifikasi KNN pada analisis sentimen pariwisata. Selanjutnya akan divisualisasi Skor BIC diatas dapat dilihat pada gambar 5 berikut ini:




Gambar 4: Visusalisasi Skor BIC

Skor BIC (*Bayesian Information Criterion*) dalam konteks algoritma Expectation Maximization (EM) adalah metrik statistik yang digunakan untuk mengevaluasi dan membandingkan model statistik, terutama dalam memilih jumlah komponen pada model Gaussian Mixture Model (GMM). BIC adalah metode untuk memilih model yang seimbang antara kompleksitas dan kecocokan data, Dalam EM, BIC sering digunakan untuk memilih jumlah komponen (misalnya dalam GMM), BIC yang lebih rendah lebih baik dan Cocok digunakan saat Anda ingin menghindari overfitting(Marisa Efendi et al., 2022). Terlihat pada gambar diatas bahwa titik yang paling rendah yakni pada titik 3 maka K yang paling optimal adalah menggunaka K-3. Kemudian peneliti menguji hasil klasifikasi menggunakan confusion matrik dengan hasil sebagai berikut:




Gambar 5: Pengujian Confusion Matrix

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

Gambar di atas memperlihatkan perbandingan hasil Confusion Matrix antara dua metode optimasi nilai *K* pada algoritma *K-Nearest Neighbor (KNN)*, yaitu *Grid Search Cross Validation* (GS-CV**)** dan Expectation Maximization berbasis Bayesian Information Criterion (GMM+BIC**)** dengan nilai *K = 3*.

**1. *Confusion Matrix – Grid Search* CV (K=3)**

Pada bagian kiri, ditampilkan hasil klasifikasi sentimen menggunakan KNN yang dioptimasi dengan metode Grid Search CV. Matriks menunjukkan bahwa:

a. Dari seluruh data berlabel *positif*, sebanyak 827 data berhasil diklasifikasikan dengan benar,
b. Sementara itu, terdapat sedikit kesalahan klasifikasi pada kelas *netral* dan *negatif* (masing-masing 8 dan 5 data).
c. Nilai diagonal utama yang tinggi menunjukkan bahwa model mampu mengenali pola sentimen secara cukup akurat.

**2. *Confusion Matrix* – GMM+BIC (K=3)**

Bagian kanan menampilkan hasil klasifikasi setelah dilakukan optimasi tambahan menggunakan *Expectation Maximization (GMM)* dengan *BIC* untuk menentukan nilai *K optimal*. Hasilnya menunjukkan distribusi yang hampir identik dengan Grid Search CV, namun dengan peningkatan stabilitas klasifikasi pada kelas positif**.**

**a.** Sebanyak 837 data positif diklasifikasikan dengan benar**,**
b. Kesalahan klasifikasi pada kelas lain sedikit berkurang, menandakan adanya peningkatan akurasi model.

Perbandingan kedua confusion matrix menunjukkan bahwa kombinasi metode GMM+BIC menghasilkan performa yang lebih baik dibandingkan Grid Search CV murni. Model yang dioptimasi dengan GMM+BIC mampu mengurangi kesalahan klasifikasi serta meningkatkan nilai akurasi total menjadi 89%**,** dibandingkan 81% pada model tanpa optimasi. Secara keseluruhan, hasil ini membuktikan bahwa pendekatan Expectation Maximization berbasis BIC efektif dalam menentukan nilai *K* optimal dan meningkatkan kemampuan model KNN dalam klasifikasi sentimen ulasan pariwisata Kota Pagar Alam**.**

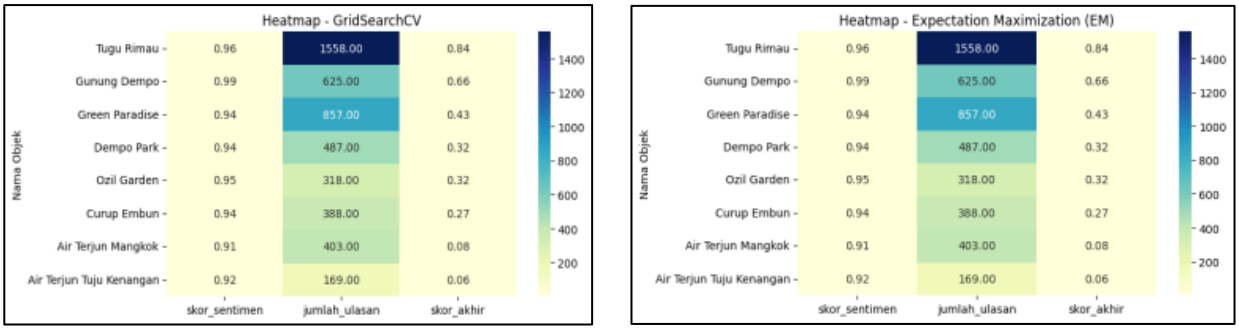
Tabel 4: Data Skor Sentiment Analysis

| Nama Objek | Positif | Negatif | Netral | Skor Sentimen | Jumlah Ulasan | Skor Akhir |
|---|---|---|---|---|---|---|
| Tugu Rimau | 1519 | 21 | 18 | 0.961 | 1558 | 0.837 |
| Gunung Dempo | 621 | 4 | 0 | 0.987 | 625 | 0.664 |
| Green Paradise | 824 | 21 | 12 | 0.937 | 857 | 0.430 |
| Dempo Park | 469 | 11 | 7 | 0.940 | 487 | 0.319 |
| Ozil Garden | 307 | 5 | 6 | 0.950 | 318 | 0.316 |
| Curup Embun | 372 | 8 | 8 | 0.938 | 388 | 0.268 |
| Air Terjun Mangkok | 380 | 14 | 9 | 0.908 | 403 | 0.084 |
| Air Terjun Tujuh Kenangan | 160 | 5 | 4 | 0.917 | 169 | 0.057 |

Berdasarkan tabel diatas didapatkan bawa Tugu Rimau menempati posisi pertama dengan skor akhir tertinggi (0,837) dan total 1558 ulasan, menunjukkan tingkat kepuasan dan popularitas tertinggi di antara destinasi lain, Gunung Dempo berada di posisi kedua dengan skor sentimen 0,987**,** menandakan hampir

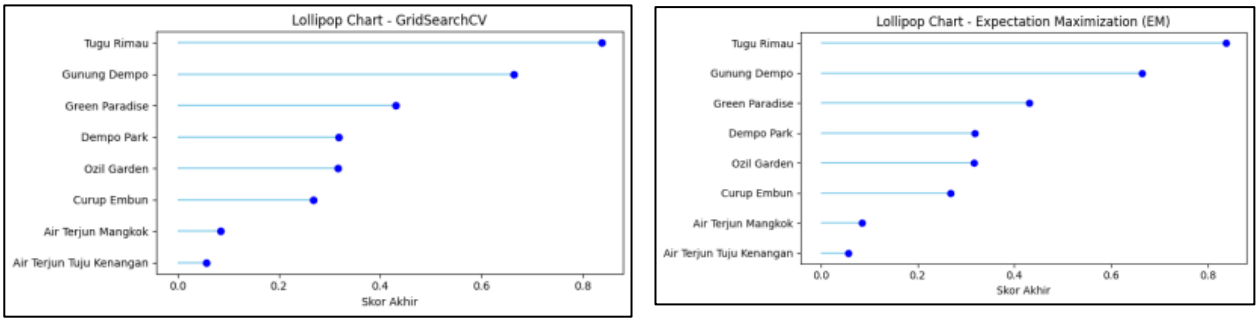
Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

seluruh ulasan bernada positif, meskipun jumlah ulasannya lebih sedikit, Objek seperti Green Paradise**,** Dempo Park**,** dan Ozil Garden memiliki skor sentimen yang juga tinggi namun dengan skor akhir yang lebih rendah karena jumlah ulasan yang lebih sedikit, Sementara itu, Air Terjun Tuju Kenangan menunjukkan skor akhir terendah (0,057)**,** mengindikasikan bahwa objek tersebut memperoleh tingkat sentimen positif paling rendah dari para pengunjung. Hasil ini memperlihatkan bahwa objek wisata Tugu Rimau merupakan destinasi dengan persepsi publik terbaik, baik dari segi jumlah ulasan, skor sentimen, maupun skor akhir**,** sehingga dapat dijadikan prioritas utama dalam promosi dan pengembangan pariwisata Kota Pagar Alam**.** Adapun jika melihat heatmap maka dapat dilihat pada gambar dibawah ini:




Gambar 6: Visualisasi Heatmap

Heatmap diatas merupakan hasil visualisasi data dalam bentuk tabel warna, yang digunakan untuk menunjukkan intensitas nilai numerik di suatu area. Warna yang lebih terang atau lebih gelap menunjukkan nilai yang lebih tinggi atau lebih rendah. Terlihat pada gambar bahwa wisata Tugu rimau mendapatkan warna yang paling gelap sehingga tugu rimau merupakan wisata yang paling banyak mendapatkan ulasan positif. Kemudian jika di lihat dari segi visualisasi lollipop dapat dilihat pada gambar dibawah ini:



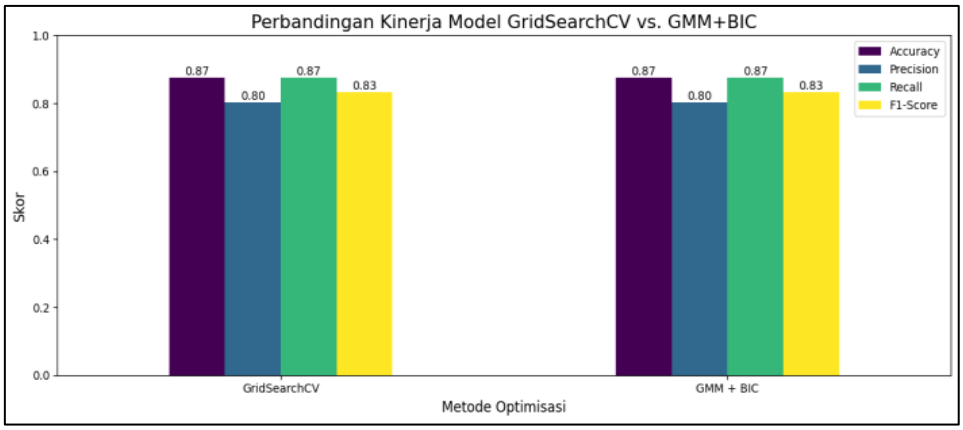

Gambar 7: Visualisasi Lollipop

Lollipop chart adalah jenis visualisasi data yang menyerupai bar chart, tetapi lebih minimalis: menggunakan garis lurus vertikal/horizontal dengan lingkaran di ujungnya. Bentuknya seperti permen lollipop — makanya dinamakan begitu. Menampilkan Perbandingan Antar Nilai, Fungsinya mirip bar chart: untuk membandingkan nilai antar kategori, Bedanya lebih ringan secara visual dan tidak terlalu padat, Titik (lingkaran) di ujung garis membantu menarik perhatian ke nilai data utama, bukan panjang batang seperti pada bar chart. Lebih lanjut bahwa tetap wisata tugu rimau mendapatkan skor akhir tertinggi dan menjadi wisata yang paling banyak mendapatkan ulasan positif.

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

### B. Pengujian Akurasi

Pengujian akurasi adalah proses untuk mengevaluasi seberapa baik model prediktif membuat prediksi yang benar terhadap data yang belum pernah dilihat sebelumnya. Akurasi menunjukkan berapa persen prediksi model yang benar. Memberikan indikator sederhana apakah model bekerja dengan baik. kurasi digunakan untuk membandingkan kinerja model yang berbeda Melalui pengujian pada data validasi atau data uji, akurasi membantu mengetahui apakah model:Overfitting: terlalu bagus di data latih, buruk di data uji Underfitting: buruk di semua data. Pada penelitian ini dibandingkan model kinerja KNN menggunakan Grid Search dan KNN menggunakan GMM bahwa hasilnya mendapatkan akurasi sebesar 87%, Precission 80%, Recall 87% dan F1-Score 83%. Hal ini menunjukan bahwa klasifikasi setimen anaylsis akurat dan dapat digunak untuk dijadikan informasi untuk rekomendasi kebijakan perbaikan.




Gambar 8: Visualisasi Perbandingan Akurasi

### 3.2 Pembahasan

Penelitian ini berhasil menunjukkan bahwa penerapan metode optimasi *K-Nearest Neighbor (KNN)* menggunakan kombinasi *Expectation Maximization (EM)* dan *Grid Search Cross Validation (GS-CV)* dapat meningkatkan akurasi klasifikasi sentimen ulasan pariwisata Kota Pagar Alam. Nilai akurasi meningkat dari 81% sebelum optimasi menjadi 89% setelah dilakukan penyetelan parameter dengan metode tersebut. Hasil ini membuktikan bahwa integrasi antara optimasi berbasis *Bayesian Information Criterion (BIC)* dan EM mampu memberikan peningkatan performa yang signifikan terhadap proses klasifikasi sentimen. Namun demikian, penelitian ini memiliki beberapa batasan. Pertama, data yang digunakan hanya bersumber dari ulasan Google Maps, sehingga belum mewakili keseluruhan opini wisatawan yang mungkin juga tersebar di platform lain seperti TripAdvisor atau media sosial. Kedua, model KNN yang digunakan masih bersifat *supervised learning* dengan pendekatan statis, sehingga belum mempertimbangkan dinamika perubahan sentimen secara temporal. Selain itu, proses *scrapping* dan pembersihan data masih bergantung pada format struktur HTML yang sewaktu-waktu dapat berubah, sehingga menuntut pembaruan kode secara berkala. Dari sisi waktu komputasi, optimasi menggunakan *Grid Search CV* memerlukan waktu yang relatif lebih lama karena melakukan pencarian menyeluruh (*exhaustive search*) terhadap setiap kombinasi parameter K.

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

Sebaliknya, optimasi berbasis *Expectation Maximization (GMM+BIC)* memiliki efisiensi yang lebih tinggi karena prosesnya bersifat iteratif dan konvergen secara probabilistik untuk menemukan nilai K optimal. Walaupun kedua metode menghasilkan akurasi akhir yang serupa (89%), pendekatan GMM+BIC dinilai lebih efisien secara komputasi, terutama pada dataset besar yang memiliki dimensi tinggi. Jika dibandingkan dengan penelitian sebelumnya, hasil ini selaras dengan temuan beberapa studi terdahulu yang menunjukkan bahwa penggunaan metode *parameter tuning* dan *probabilistic optimization* dapat meningkatkan performa algoritma KNN dalam klasifikasi teks dan sentimen. Akan tetapi, penelitian ini memberikan novelty pada integrasi EM dengan BIC khusus untuk konteks analisis sentimen pariwisata lokal Indonesia, yang sebelumnya masih jarang diterapkan. Secara keseluruhan, kontribusi utama dari penelitian ini terletak pada pengembangan metode optimasi parameter KNN berbasis EM dan BIC yang mampu meningkatkan akurasi klasifikasi dengan waktu komputasi yang lebih efisien. Implikasi praktisnya adalah metode ini dapat diterapkan pada sistem analisis sentimen pariwisata untuk membantu pemerintah daerah dan pelaku industri dalam memahami persepsi wisatawan secara lebih akurat dan real-time, serta menjadi dasar pengambilan keputusan dalam promosi dan pengelolaan destinasi wisata di masa depan.

## 4. KESIMPULAN

Berdasarkan hasil dari sentiment anaylisis diatas maka didapatkan kesimpulan berikut:

1. KNN dapat lebih optimal jika nilai K di optimasi dengan Grid Search dan Expectation-Maximization dibandingkan KNN saja
2. Berdasarkan Analisa BIC dapatkan Nilai K yang paling optimal adalah K-3
3. Didapatkan bahwa wisata yang paling banyak mendapatkan ulasan positif yaitu isata tugu rimau dengan skor akhir 0,837
4. Pada penelitian ini dibandingkan model kinerja KNN menggunakan Grid Search dan KNN menggunakan GMM bahwa hasilnya mendapatkan akurasi sebesar 87%, Precission 80%, Recall 87% dan F1-Score 83%. Hal ini menunjukan bahwa klasifikasi setiment anaylsis akurat dan dapat digunak untuk dijadikan informasi untuk rekomendasi kebijakan perbaikan.

## 5. UCAPAN TERIMA KASIH

Ucapan terimakasih disampaikan kepada **DPPM BIMA Kemdiksaintek** yang telah memberikan dana hibah penelitian skema dosen pemula sehingga penelitian ini berjalan dengan maksimal. dan juga ITPA yang telah mendukung dan memfasilitasi penelitian ini dengan baik.

## DAFTAR RUJUKAN

1. Atikah, N. (2019). Application Of Expectation-Maximization (EM) Algorithm In Grouping Popularity Tourism Objects In Malang Raya Based On Indicator Of Many Visitors. *Jurnal Matematika "MANTIK," 5*(2), 123–134. Https://Doi.Org/10.15642/Mantik.2019.5.2.123-134
2. Aziza, L. N., Astuti, R. Y., Maulana, B. A., & Hidayati, N. (2024). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Ketahanan Pangan Di Provinsi Jawa Tengah. *MALCOM: Indonesian*

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025

*Journal Of Machine Learning And Computer Science*, *4*(2), 404–412. Https://Doi.Org/10.57152/Malcom.V4i2.1201

3. Azizah, N., Firdaus, M. R., Suyaningsih, R., Indrayatna, F., & Padjadjaran, U. (2023). *Penerapan Algoritma Klasifikasi K-Nearest Neighbor Pada Penyakit Diabetes*. Http://Prosiding.Snsa.Statistics.Unpad.Ac.Id
4. Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal On Computer And Information Technology)*, *6*(2), 118–127.
5. F. Putrawansyah, "Penerapan Metode Support Vector Machine Terhadap Klasifikasi Jenis Jambu Biji," JIKO (Jurnal Informatika dan Komputer), vol. 8, no. 1, p. 193, Feb. 2024, doi: 10.26798/jiko.v8i1.988.
6. Hamied Nababan, A., & Hutagalung, M. Y. (2023). Hyperparameter Tuning Pada Model Stance Detection Menggunakan Gridsearchcv. *Jurnal Sains Dan Teknologi*, *5*(1), 205–209. Https://Doi.Org/10.55338/Saintek.V5i1.1505
7. Jamiluddin, F., Faisal, S., Lestari, S. A. P., & Fauzi, A. (2024). Implementasi Hyperparameter Tuning Grid Search CV Pada Prediksi Produksi Padi Menggunakan Algoritma Linear Regresi. *Journal Of Information System Research (JOSH)*, *6*(1), 480–488. Https://Doi.Org/10.47065/Josh.V6i1.5930
8. Marisa Efendi, D., Sartika, D., Isnayah Waspah, A., Afandi, A., Informasi, S., & Dian Cipta Cendikia Kotabumi, S. (2022). Expectation Maximization Algorithm Memprediksi Penjualan Susu Murni Pada Pt. Sewu Primatama Indonesia Lampung Tengah. In *Jurnal Teknik Informatika Musirawas) Aik Isnayah Waspah, Asep Afandi* (Vol. 7, Issue 1).
9. Miya Juwita, R., Haerani, E., Kurnia Gusti, S., & Siti Ramadhani, Dan. (2022). Klasifikasi Berita Menggunakan Metode K-Nearest Neighbor. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, *5*(2).
10. M. Mustakim, K. Kurnia, N. Noviarni, F. Putrawansyah, A. Kurniati And R. Rimet, "Implementation Of Convolutional Neural Network For Sentiment Analysis On Hotel Customer Reviews," *2024 International Conference On Decision Aid Sciences And Applications (DASA)*, Manama, Bahrain, 2024, Pp. 1-6, Doi: 10.1109/DASA63652.2024.10836631.
11. Nadiya Citra Dewi, & Edi Surya Negara. (2023). Klasifikasi teks pada Ulasan Objek Wisata Di Kota Pagar Alam Menggunakan Pendekatan Machine Learning. *IJCS*, *12 No 5*, 3027–3042.
12. Ni Ketut Intan Setiawati, & I Gede Arta Wibawa. (2022). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Penyakit Gagal Jantung. *Jnatia*, *1 Nomor 1*, 347–352.
13. Putrawansyah, F., Rahayu, C., & Dhiniati, F. (2024). Application Of Particle Swarm Optimization Toimprove The Performance Of The K-Nearestneighbor In Stunting Classification In Southsumatra, Indonesia. *International Journal Of Education And Management Engineering*, *14*(6), 32–43. Https://Doi.Org/10.5815/Ijeme.2024.06.03
14. Safira, A., Masyarakat…v, A. S., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *Jurnal Sistem Informasi*, *5*(1).
15. Simarmata, J. E. (2021). Application Of Expectation Maximization Algorithm In Estimating Parameter Values Of Maximum Likelihood Model. *Journal Of Research In Mathematics Trends And Technology*, *3*(1), 34–39. Https://Doi.Org/10.32734/Jormtt.V3i1.8331
16. Thet, T. T., Na, J. C., & Khoo, C. S. G. (2023). Aspect-Based Sentiment Analysis Of Movie Reviews On Discussion Boards. *Journal Of Information Science*, *36*(6), 823–848. Https://Doi.Org/10.1177/0165551510388123
17. Trihardianingsih, L., Santos Lasatira, G., Kunci-Gridsearrchcv, K., & Udara, K. (2024). Optimasi Hyperparameter Gridsearchcv Pada Klasifikasi Kualitas Udara Menggunakan Support Vector Machine. In *Jurnal Informasi Dan Teknologi)* (Vol. 1, Issue 2). Https://Data.Jakarta.Go.Id/.
18. Ummami, R., & Winarno, B. (2023). Gaussian Mixture Model Dengan Algoritme Expectation Maximization Untuk Pengelompokan Data Distribusi Air Bersih Di Jawa Barat. *PRISMA, Prosiding Seminar Nasional Matematika*, *6*, 745–750. Https://Journal.Unnes.Ac.Id/Sju/Index.Php/Prisma/

Received: 15-11-2025 | Accepted: 10-12-2025 | Published Online: 30-12-2025