E-ISSN: 2715-2731

Sistem Rekomendasi Destinasi Wisata Menggunakan Data Demografis Berbasis Klasifikasi *Support Vector Machine*

Aldilla Qurrata A'yun[1,*], Yunifa Miftachul Arif[2], Mochamad Imamudin[3]

Program Studi Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia

[1]aldillaqurrataayun@gmail.com; [2]yunif4@ti.uin-malang.ac.id; [3]imamudin@ti.uin-malang.ac.id;

* penulis korespondensi

INFO ARTIKEL

Sejarah Artikel
Diterima: 26 November 2025
Direvisi: 30 Desember 2025
Diterbitkan: 31 Desember 2025

Kata Kunci
Data Demografis
Destinasi Wisata
Sistem Rekomendasi
Support Vector Machine

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi destinasi wisata berbasis data demografis menggunakan metode klasifikasi *Support Vector Machine* (SVM). Sistem dirancang untuk memberikan rekomendasi destinasi yang sesuai dengan karakteristik wisatawan, seperti usia, jenis kelamin, dan status sosial. Dataset yang digunakan terdiri dari sepuluh variabel demografis dan empat belas kategori destinasi wisata. Analisis awal menunjukkan bahwa dataset memiliki ketidakseimbangan kelas yang sangat tinggi, dengan kelas Jatim Park 1 mendominasi lebih dari separuh data sementara banyak kelas lain hanya memiliki 1–6 sampel. Untuk mengurangi dampak ketidakseimbangan ini, dilakukan teknik oversampling pada data training. Model SVM kemudian dilatih menggunakan beberapa kombinasi parameter dan kernel, serta diuji menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil eksperimen menunjukkan bahwa pada data training yang sudah diseimbangkan, performa model meningkat signifikan, ditunjukkan oleh nilai F1-macro pada cross-validation sebesar 0.84. Namun, ketika diuji pada data testing yang mencerminkan kondisi distribusi asli, performa model menurun, dengan akurasi sebesar 54% dan nilai F1-macro yang rendah pada sebagian besar kelas minoritas. Temuan ini menunjukkan bahwa meskipun SVM efektif pada data yang seimbang, performanya masih belum optimal pada dataset rekomendasi wisata yang sangat timpang. Penelitian ini merekomendasikan pengayaan data, terutama untuk kelas-kelas minoritas, serta eksplorasi metode penanganan ketidakseimbangan kelas lainnya pada penelitian lanjutan.

PENDAHULUAN

Sektor pariwisata merupakan salah satu penggerak ekonomi yang penting di Indonesia, dengan pertumbuhan signifikan pada jumlah wisatawan domestik maupun mancanegara dalam beberapa tahun terakhir [1]. Kota Batu, Jawa Timur, telah berkembang menjadi salah satu destinasi wisata unggulan nasional karena kekayaan alam, wahana rekreasi modern, serta daya tarik budaya yang dimilikinya. Pemerintah Kota Batu juga terus melakukan penguatan promosi pariwisata untuk meningkatkan jumlah kunjungan wisatawan dan memperkuat citranya sebagai destinasi wisata utama di Indonesia. Namun demikian, keberagaman karakteristik dan preferensi wisatawan menimbulkan tantangan dalam menyediakan pengalaman berwisata yang personal dan sesuai kebutuhan individu [2]. Banyaknya pilihan destinasi wisata di Kota Batu membuat wisatawan sering mengalami kesulitan dalam menentukan lokasi wisata yang paling relevan dengan profil mereka [3].

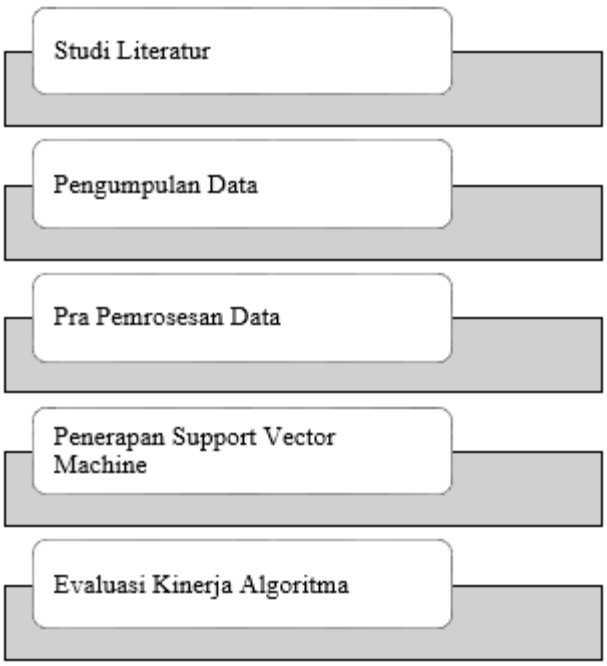
Untuk mengatasi permasalahan tersebut, sistem rekomendasi menjadi solusi penting dalam memberikan saran destinasi wisata yang sesuai dengan karakteristik pengguna. Salah

satu pendekatan yang potensial adalah memanfaatkan data demografis seperti usia, jenis kelamin, dan status sosial untuk memprediksi destinasi yang mungkin disukai wisatawan. Metode berbasis klasifikasi telah banyak digunakan dalam sistem rekomendasi, termasuk pada sektor pariwisata dan perilaku pengguna secara umum [4], [5]. Di antara berbagai algoritma klasifikasi, *Support Vector Machine* (SVM) dikenal sebagai metode yang efektif dalam menangani data berdimensi tinggi serta menghasilkan performa klasifikasi yang stabil [6].

Berbagai penelitian sebelumnya menunjukkan bahwa SVM memiliki kemampuan yang baik dalam membangun model prediksi pada konteks rekomendasi wisata. Penelitian oleh Liu et al. menunjukkan bahwa SVM dapat memberikan hasil akurasi tinggi pada rekomendasi berbasis fitur visual objek wisata [7]. Rahmawati dan Nugroho berhasil menerapkan SVM untuk memberikan rekomendasi destinasi wisata berdasarkan profil pengguna dengan hasil yang menjanjikan [8]. Hidayat dan Putri juga membuktikan bahwa kombinasi data demografis dan metode klasifikasi SVM mampu memberikan rekomendasi wisata yang lebih terarah dalam konteks provinsi Jawa Timur [9]. Selain itu, beberapa penelitian di bidang lain turut menegaskan keunggulan SVM dalam klasifikasi, seperti analisis perilaku pengguna [10], prediksi kepuasan wisatawan [11], serta klasifikasi berbasis lokasi atau perilaku check-in [12].

Melihat keberhasilan SVM pada berbagai domain rekomendasi dan klasifikasi, penerapannya pada sistem rekomendasi destinasi wisata berbasis data demografis merupakan langkah yang relevan dan potensial. Dengan memanfaatkan karakteristik pengguna, sistem rekomendasi dapat memberikan saran destinasi yang lebih terpersonalisasi, sehingga mampu meningkatkan pengalaman wisatawan serta mendukung strategi promosi pariwisata di Kota Batu. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi destinasi wisata berbasis klasifikasi SVM yang dapat memberikan rekomendasi yang lebih tepat sasaran dan mendukung pertumbuhan sektor pariwisata di Indonesia.

METODE



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan tahapan penelitian yang terdiri dari lima langkah utama, dimulai dari studi literatur untuk mengkaji teori, metode, dan hasil penelitian terdahulu yang relevan dengan sistem rekomendasi wisata dan algoritma *Support Vector Machine* (SVM). Tahap berikutnya adalah pengumpulan data, yaitu mengumpulkan data demografis pengguna beserta kategori destinasi wisata yang menjadi preferensi mereka sebagai dasar pembangunan model. Data yang diperoleh kemudian melalui proses pra-pemrosesan, meliputi pembersihan data, pengkodean data kategorikal, normalisasi, serta pembagian data menjadi data training dan testing guna memastikan kualitas dan kesiapan data untuk proses klasifikasi. Setelah itu dilakukan penerapan algoritma SVM, yaitu proses pelatihan model menggunakan data training untuk mengenali pola hubungan antara karakter demografis pengguna dan preferensi destinasi wisata, dengan pengujian beberapa konfigurasi kernel untuk mendapatkan performa terbaik. Tahap terakhir adalah evaluasi kinerja algoritma, yang dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score untuk menilai kemampuan model dalam memprediksi destinasi wisata yang sesuai dengan profil pengguna. Seluruh tahapan ini membentuk alur penelitian yang sistematis dalam mengembangkan sistem rekomendasi destinasi wisata berbasis data demografis.

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh melalui observasi langsung dan wawancara terhadap pengunjung di beberapa lokasi wisata utama Kota Batu selama periode 5 Maret hingga 6 April 2024. Proses pengumpulan data dilakukan pada titik-titik strategis yaitu area Jawa Timur Park, Alun-alun Kota Batu, dan kawasan wisata Coban Talun. Sebanyak 227 responden diwawancarai secara acak untuk memperoleh informasi mengenai karakteristik demografis dan preferensi destinasi wisata yang mereka kunjungi. Dari total data tersebut, 80% digunakan sebagai data latih dan 20% sebagai data uji dalam proses klasifikasi menggunakan algoritma SVM.

Berdasarkan hasil wawancara, diperoleh sepuluh variabel input yang merepresentasikan atribut demografis dan perilaku pengunjung sebagaimana disajikan pada Tabel 1. Variabel tersebut meliputi jenis kelamin (X1), umur (X2), pekerjaan (X3), hobi (X4), tujuan berwisata (X5), status perkawinan (X6), daerah asal (X7), teman perjalanan (X8), pendidikan terakhir (X9), serta frekuensi kunjungan atau repetition (X10). Sebagian besar atribut ini berformat kategorikal, kecuali beberapa seperti pekerjaan dan hobi yang bersifat karakter (char), namun tetap diperlakukan sebagai data kategorikal pada tahap pra-pemrosesan. Setiap atribut berperan penting dalam membentuk pola preferensi pengunjung terhadap destinasi wisata, sehingga dapat menjadi dasar bagi proses klasifikasi.

Tabel 1. Atribut Demografis

Input	Atribut	Format Data
X1	jenis_kelamin	*Kategorical*
X2	Umur	*Kategorical*
X3	pekerjaan	*Char*
X4	Hobi	*Char*
X5	tujuan_berwisata	*Kategorical*
X6	status_perkawinan	*Kategorical*
X7	daerah_asal	*Kategorical*
X8	teman_perjalanan	*Kategorical*
X9	pendidikan_terakhir	*Kategorical*
X10	repetition	*Kategorical*

Selain variabel input, penelitian ini juga memetakan empat belas jenis destinasi wisata sebagai variabel output yang menjadi target klasifikasi sebagaimana disajikan pada Tabel 2.

Item destinasi yang menjadi keluaran model terdiri dari beragam lokasi wisata populer di Kota Batu, mulai dari taman rekreasi berskala besar seperti Jawa Timur Park 1 (Y1), Jawa Timur Park 2 (Y2), dan Jawa Timur Park 3 (Y3), museum tematik seperti Museum Angkut (Y4), hingga kawasan alam seperti Pemandian Cangar (Y10), Coban Talun (Y11), dan Coban Rais (Y13). Selain itu, terdapat pula destinasi keluarga dan hiburan seperti Taman Rekreasi Selecta (Y5), Batu Night Spectacular (Y6), Eco Green Park (Y7), Alun-alun Kota Batu (Y8), Kusuma Agro (Y9), Pemandian Songgoriti (Y12), dan Predator Fun Park (Y14). Keberagaman jenis destinasi ini mencerminkan variasi preferensi wisatawan yang berkunjung ke Kota Batu.

Tabel 2. Destinasi Wisata

Output	Item Destinasi Wisata
Y1	Jawa Timur Park 1 (Jatim Park 1)
Y2	Jawa Timur Park 2 (Jatim Park 2)
Y3	Jawa Timur Park 3 (Jatim Park 3)
Y4	Museum Angkut
Y5	Taman Rekreasi Selecta
Y6	BNS (Batu Night Spectacular)
Y7	EGP (Eco Green Park)
Y8	Alun-alun Kota Batu
Y9	Kusuma Agro
Y10	Pemandian Cangar
Y11	Coban Talun
Y12	Pemandian Songgoriti
Y13	Coban Rais
Y14	PFP (Predator Fun Park)

Pra-Pemrosesan Data

Tahapan pra-pemrosesan data pada penelitian ini dilakukan untuk memastikan bahwa data yang digunakan dalam pelatihan model SVM memiliki kualitas yang baik dan siap diolah. Proses ini diawali dengan pembersihan data, yaitu mengidentifikasi dan menangani masalah seperti nilai kosong, data duplikat, serta keberadaan noise pada atribut tertentu. Baris data yang mengandung nilai kosong penting, seperti jenis kelamin atau daerah asal, diperlakukan melalui penghapusan atau imputasi menggunakan modus karena sebagian besar atribut berbentuk kategorikal. Baris data yang terdeteksi duplikasi juga dihapus untuk mencegah bias dalam proses pelatihan model. Selain itu, atribut teks seperti pekerjaan dan hobi dibersihkan dari karakter yang tidak relevan seperti simbol atau tanda baca agar konsisten dan mudah diproses.

Setelah data dibersihkan, tahap selanjutnya adalah transformasi data. Pada tahap ini, seluruh atribut kategorikal diubah menjadi nilai numerik menggunakan teknik *Encoding* seperti Label *Encoding* atau *One-Hot Encoding*, sehingga data dapat diproses oleh algoritma SVM yang hanya menerima masukan numerik. Selain proses *Encoding*, normalisasi juga dapat diterapkan pada atribut tertentu yang bersifat ordinal, seperti kategori usia atau atribut lain yang memiliki urutan nilai. Normalisasi dilakukan untuk menyamakan rentang nilai agar tidak menimbulkan bias dalam proses pelatihan model. Salah satu metode normalisasi yang umum digunakan adalah Min-Max Normalization, yang mengubah nilai data ke dalam rentang 0 hingga 1. Proses ini dapat dihitung menggunakan Persamaan (1). Meskipun sebagian besar atribut dalam penelitian ini bersifat kategorikal dan tidak memerlukan normalisasi, penerapan Min-Max tetap relevan bagi atribut yang memiliki nilai berurutan atau representasi numerik tertentu setelah proses *Encoding*.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

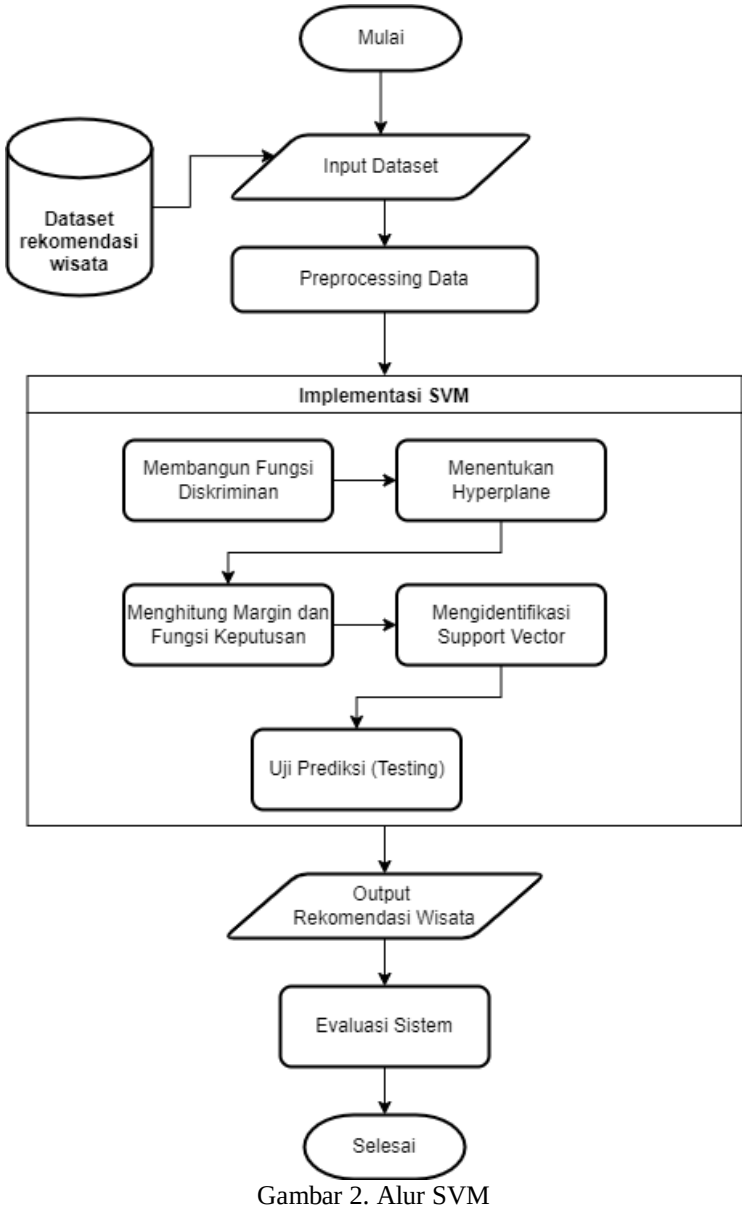
Tahapan berikutnya adalah ekstraksi dan seleksi fitur, di mana seluruh atribut demografis dievaluasi relevansinya terhadap tujuan klasifikasi. Fitur-fitur seperti tujuan berwisata, umur, hobi, status perkawinan, teman perjalanan, dan repetition dipertahankan karena dianggap memiliki pengaruh langsung terhadap preferensi destinasi wisata pengunjung. Sementara itu, fitur yang kurang relevan atau memiliki korelasi rendah terhadap output dapat dihilangkan untuk mengurangi kompleksitas model dan meningkatkan performa.

Selanjutnya dilakukan proses balancing data, terutama jika terjadi ketidakseimbangan jumlah sampel antara kategori destinasi wisata. Ketidakseimbangan data dapat menyebabkan model cenderung bias pada destinasi tertentu. Oleh karena itu, dilakukan teknik oversampling pada kelas minoritas atau undersampling pada kelas mayoritas untuk memastikan bahwa seluruh kategori memiliki representasi yang proporsional dalam proses pelatihan.

Tahap terakhir dari pra-pemrosesan adalah pembagian dataset menjadi data latih dan data uji. Dalam penelitian ini, 80% data digunakan sebagai data latih untuk membangun dan melatih model SVM, sedangkan 20% sisanya digunakan sebagai data uji untuk mengevaluasi performa model. Pembagian ini memastikan bahwa model diuji menggunakan data yang belum pernah dilihat sebelumnya sehingga akurasi yang dihasilkan benar-benar menggambarkan kemampuan model dalam melakukan generalisasi. Seluruh tahapan pra-pemrosesan ini menjadi fondasi penting dalam memastikan bahwa proses klasifikasi dapat berjalan optimal dan menghasilkan sistem rekomendasi yang akurat.

Penerapan SVM

Tahapan dalam implementasi SVM disajikan pada Gambar 2, yakni melalui beberapa tahap sistematis. Pertama, ditentukan dataset wisatawan sebanyak 227 data, dengan atribut seperti pekerjaan, hobi, tujuan berwisata, teman perjalanan, status perkawinan, hingga riwayat perjalanan. Dataset dibagi menjadi 80% sebagai data latih dan 20% sebagai data uji. Seluruh atribut kemudian diproses ke dalam bentuk numerik biner (0/1) melalui tahap preprocessing, sehingga dapat diolah oleh algoritma SVM. Selanjutnya, dibangun fungsi diskriminan awal untuk membedakan kelas destinasi wisata (Y1 hingga Y14), kemudian ditentukan *hyperplane* sebagai batas pemisah antar kelas. *Hyperplane* tersebut dirancang agar memiliki margin maksimum, sehingga memberikan pemisahan yang optimal antara profil wisatawan yang berbeda. Vektor-vektor yang berada pada batas margin kemudian diidentifikasi sebagai support vectors, karena memiliki peran penting dalam menentukan bentuk *hyperplane*. Setelah model SVM terbentuk, tahap akhir adalah uji prediksi menggunakan data uji untuk mengukur kinerja model dalam merekomendasikan destinasi wisata. Hasil pengujian kemudian dianalisis dengan melihat akurasi dan kemampuan model dalam memetakan preferensi wisatawan terhadap destinasi yang sesuai.



Gambar 2. Alur SVM

Hitung Fungsi Diskriminan Awal

Pada tahap ini, proses membangun fungsi diskriminan awal dilakukan untuk membedakan kelas atau kategori destinasi wisata berdasarkan fitur-fitur yang ada dalam dataset. Dalam konteks penelitian ini, dataset berjumlah 227 data wisatawan yang telah melalui tahap *preprocessing*, di mana setiap atribut seperti pekerjaan, hobi, tujuan berwisata, daerah asal, pendidikan terakhir, hingga teman perjalanan, telah dikodekan dalam bentuk biner. Fungsi diskriminan awal pada SVM dibentuk dari kombinasi linier antara bobot (w) dan vektor fitur (x) dengan penambahan bias (b), sehingga menghasilkan fungsi keputusan berbentuk Persamaan (2)

$$f(x) = sign(w \cdot x + b) \tag{2}$$

Fungsi ini berperan sebagai batas (*hyperplane*) yang memisahkan data wisatawan ke dalam kelas destinasi wisata tertentu, misalnya Y1 hingga Y14. Dengan adanya fungsi diskriminan awal ini, SVM dapat mengidentifikasi pola dari data latih, kemudian menentukan margin terbaik agar data dari kelas yang berbeda memiliki jarak

pemisahan maksimal. Dalam penelitian ini, fungsi diskriminan awal dibangun dari hasil perhitungan dot product antara vektor fitur biner dengan bobot awal yang diperoleh melalui optimasi, sehingga secara bertahap model dapat belajar untuk menghasilkan pemisahan yang paling optimal.

Menentukan *Hyperplane*

Pada penelitian ini, *hyperplane* berperan sebagai batas kelas yang digunakan untuk memisahkan data wisatawan berdasarkan preferensi destinasi wisata. Setelah fungsi diskriminan awal dibentuk dari kombinasi bobot (w) dan bias (b), langkah berikutnya adalah menentukan posisi *hyperplane* yang paling optimal. *Hyperplane* ini bukan sekadar garis pemisah biasa, melainkan sebuah bidang keputusan dalam ruang berdimensi tinggi yang terbentuk dari hasil *encoding* biner seluruh fitur pada dataset. Dengan jumlah 227 data yang terdiri dari berbagai atribut (pekerjaan, hobi, tujuan, pendidikan, teman perjalanan, dsb.), SVM akan mencari *hyperplane* yang mampu memisahkan kelas-kelas destinasi wisata (Y1 hingga Y14) dengan margin terbesar. Margin adalah jarak antara *hyperplane* dengan data titik terdekat dari masing-masing kelas (dikenal sebagai *support vector*). Hasil dari penentuan *hyperplane* inilah yang menjadi dasar sistem rekomendasi untuk memutuskan destinasi wisata mana yang paling sesuai dengan profil wisatawan. Dengan demikian, *hyperplane* bukan hanya sekadar pemisah matematis, melainkan inti dari mekanisme rekomendasi berbasis SVM dalam penelitian ini.

Hitung Margin dan Fungsi Keputusan

Tahap berikutnya dalam implementasi metode SVM adalah menghitung margin dan fungsi keputusan. Margin merupakan jarak antara *hyperplane* dengan titik data terdekat dari masing-masing kelas (*support vector*). Pada penelitian ini, margin menggambarkan seberapa jelas pemisahan profil wisatawan yang memiliki preferensi destinasi berbeda (misalnya Y1 hingga Y14) berdasarkan atribut biner yang sudah diproses sebelumnya. Semakin lebar margin, semakin baik model dalam memisahkan data, sehingga rekomendasi destinasi wisata akan lebih akurat dan dapat digeneralisasi dengan baik pada data baru. Secara matematis, margin dihitung dari parameter bobot (w) yang diperoleh pada proses optimasi SVM. Besarnya margin sebagaimana Persamaan (3).

$$\text{Margin} = \frac{2}{\|w\|} \tag{3}$$

Di sisi lain, fungsi keputusan dibangun untuk menentukan kelas suatu data baru. Fungsi keputusan inilah yang nantinya dipakai sistem rekomendasi untuk memetakan profil wisatawan ke dalam destinasi wisata yang paling sesuai. Fungsi ini ditulis seperti Persamaan (3.1). Di mana x adalah vektor fitur wisatawan (hasil *encoding* pekerjaan, hobi, tujuan, teman perjalanan, dsb.), w adalah bobot hasil pembelajaran, dan b adalah bias. Jika hasil f(x) bernilai positif (+1), maka data wisatawan dikategorikan ke dalam kelas tertentu (misalnya merekomendasikan destinasi Y1). Sebaliknya, jika bernilai negatif (−1), maka dikategorikan ke kelas lainnya.

Mengidentifikasi Support Vector

Tahap penting dalam metode SVM adalah mengidentifikasi *support vector*s, yaitu titik data yang berada paling dekat dengan *hyperplane* (batas pemisah kelas). *Support*

*vector*s inilah yang secara langsung menentukan posisi dan arah *hyperplane*, sehingga berperan sentral dalam membentuk model prediksi. Dalam konteks penelitian ini, *support vector*s adalah data wisatawan dengan kombinasi atribut tertentu (misalnya pekerjaan, hobi, tujuan, teman perjalanan, dsb.) yang letaknya berada di batas margin antar kelas destinasi wisata (Y1 hingga Y14).

Uji Prediksi

Setelah model dilatih menggunakan 80% data latih, maka 20% data uji digunakan untuk mengevaluasi kemampuan model dalam memberikan rekomendasi pada data yang belum pernah dilihat sebelumnya. Pada tahap ini, setiap profil wisatawan dalam data uji yang telah dikodekan ke bentuk biner (misalnya pekerjaan, hobi, tujuan, teman perjalanan, dsb.) dimasukkan ke dalam fungsi keputusan SVM pada Persamaan (3). Hasil dari fungsi ini akan menunjukkan ke kelas destinasi mana seorang wisatawan diprediksi akan cocok (misalnya Y1 hingga Y14). Jika nilai fungsi keputusan positif, wisatawan diarahkan pada kelas tertentu, sedangkan jika negatif, diarahkan pada kelas lainnya.

Evaluasi Kinerja Algoritma

Evaluasi kinerja SVM menggunakan beberapa parameter yaitu a*ccuracy, precision* dan *recall* dengan menggunakan Confusion Matrix untuk mengetahui seberapa banyak data yang berhasil di klasifikasi dan yang gagal diklasifikasi. Untuk bentuk dasar Confusion Matrix dapat dilihat pada Tabel 3.

Tabel 3. Confusion Matrix

Prediksi	Destinasi Disukai (Positive)	Destinasi Tidak Disukai (Negative)
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

Adapun metrik evaluasi yang dihitung antara lain:

1. *Accuracy*
   Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data dengan perhitungan, $accuracy = (TP + TN) / (TP+FP+FN+TN)$.
2. *Precision*
   Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Perhitungannya adalah $precision = (TP) / (TP+FP)$.
3. *Recall (Sensitivitas)*
   Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Recall menjawab pertanyaan dengan hitungan $recall = (TP) / (TP + FN)$.

HASIL DAN PEMBAHASAN

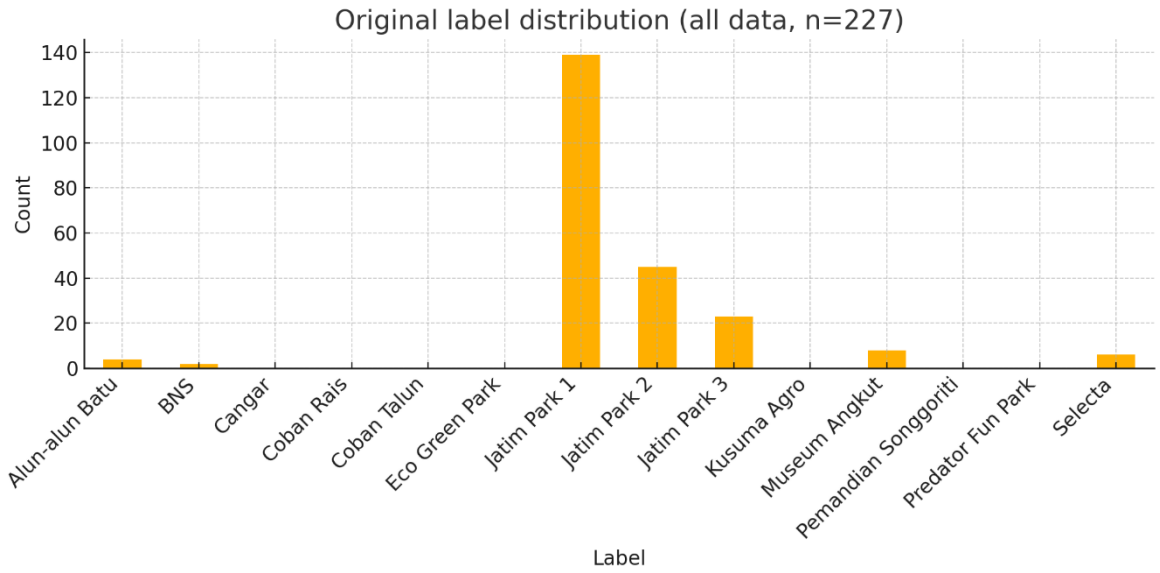
Dataset awal penelitian berjumlah 227 data responden yang diperoleh dari observasi dan wawancara pengunjung berbagai destinasi wisata di Kota Batu. Data mentah tersebut terdiri atas variabel demografis (X1–X10) dan preferensi destinasi wisata yang direpresentasikan dalam bentuk label multilabel (Y1–Y14). Setiap baris menggambarkan profil satu pengunjung beserta destinasi wisata yang paling sesuai dengan preferensinya. Sebelum digunakan untuk proses pelatihan algoritma SVM, seluruh data melewati beberapa tahapan pra-pemrosesan. Pada tahap pembersihan, nilai yang tidak valid, duplikat, serta karakter yang tidak relevan telah dihilangkan. Tahap transformasi dilakukan dengan mengonversi

seluruh variabel kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma SVM. Selain itu, seluruh variabel output Y1–Y14 dipastikan hanya berisi nilai biner (0 atau 1), sehingga bisa diolah lebih lanjut. Selanjutnya, karena SVM membutuhkan label tunggal (single-label classification), maka setiap baris data dikonversi dari format multilabel menjadi satu label kelas Y, yaitu dengan mengambil indeks kategori destinasi wisata yang bernilai 1. Jika pada satu responden terdapat lebih dari satu nilai 1, maka dipilih label pertama yang muncul (urutan Y1–Y14). Dengan cara ini, diperoleh label akhir dalam bentuk Y_label yang berisi kode kelas 1–14.

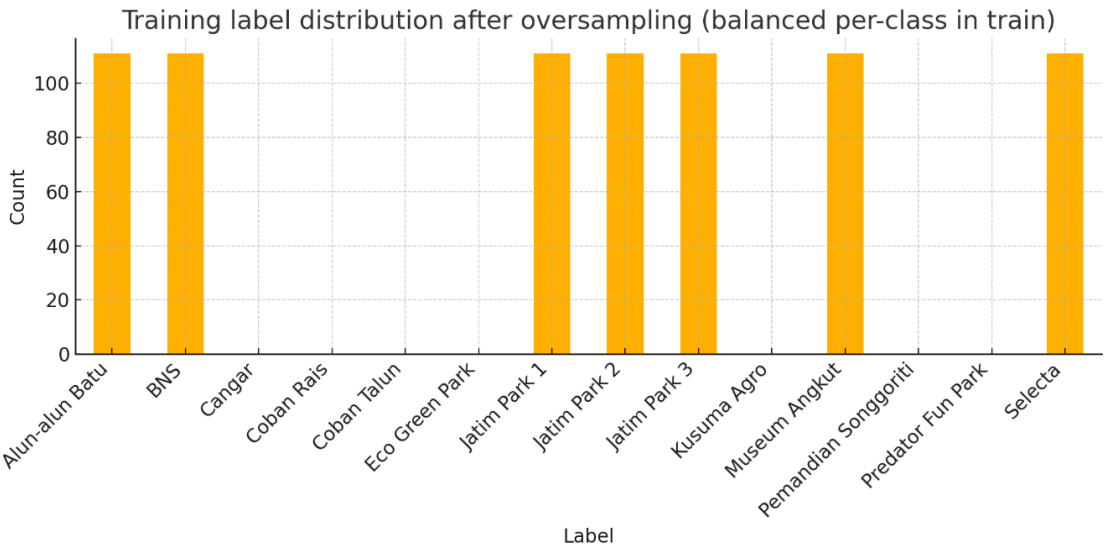
Tabel 4. Dataset Hasil Pra-Pemrosesan Data

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14
1	2	3	4	1	2	2	4	3	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	2	1	2	1	2	2	2	3	2	1	1	1	0	1	0	0	1	0	0	0	0	0	0
2	2	3	2	1	2	2	1	3	1	0	0	1	0	1	1	0	1	0	0	0	1	0	0
2	2	3	2	1	2	2	2	3	2	0	0	1	0	1	1	0	1	0	1	0	0	0	0
1	2	3	4	1	2	2	2	3	1	0	0	1	1	0	1	1	1	0	0	0	0	0	0
2	2	3	4	1	2	2	2	3	1	1	1	1	0	0	1	0	0	0	0	1	0	0	0
1	2	3	4	1	2	2	2	3	2	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	2	3	2	1	2	2	1	3	2	0	0	1	1	0	1	0	1	1	0	0	0	0	0
2	2	3	2	1	2	2	2	3	2	0	1	0	0	0	0	0	0	0	1	1	1	1	0
2	2	3	1	1	2	2	2	3	2	1	1	1	0	0	1	0	1	0	0	0	0	0	0

Jumlah total data setelah pra-pemrosesan tetap 227 baris, kecuali beberapa baris yang tidak memiliki label valid (semua Y=0). Baris seperti itu secara otomatis dihapus untuk menghindari noise dalam proses pelatihan model. Hasil akhir menunjukkan bahwa dataset siap diproses dengan struktur fitur (X1–X10) dan label tunggal Y_label. Tabel 4 hanya menampilkan 10 sampel data dari keseluruhan 227 dataset, dengan tujuan memberikan gambaran representatif mengenai bentuk data setelah melalui tahapan pra-pemrosesan. Pada tabel tersebut, terlihat bahwa seluruh fitur demografis telah terstandardisasi dalam format numerik, dan kolom output Y1–Y14 sudah berbentuk biner.



Gambar 3. Distribusi Label pada Dataset



Gambar 4. Oversampling untuk balancing data

Distribusi label sebagaimana disajikan pada Gambar 3 menunjukkan pola ketidakseimbangan yang sangat kuat. Kelas Jatim Park 1 mendominasi dataset dengan 139 sampel (61%), diikuti Jatim Park 2 sebanyak 45 sampel, dan Jatim Park 3 sebanyak 23 sampel. Sementara itu, beberapa kelas lain seperti Selecta, BNS, dan Alun-alun Batu hanya memiliki antara 1 hingga 6 sampel. Ketimpangan ini berdampak langsung pada kemampuan model karena algoritma cenderung mempelajari pola khas kelas mayoritas dan mengabaikan kelas minoritas. Untuk mengurangi dampak tersebut, dilakukan teknik oversampling pada data training sehingga jumlah sampel setiap kelas dinaikkan hingga seimbang sebelum model dilatih. Hasil oversampling disajikan pada Gambar 4. Setelah data dipisahkan menjadi training (181 sampel) dan testing (46 sampel), proses pelatihan model dilakukan menggunakan algoritma SVM dengan *pipeline* yang terdiri dari *One-Hot Encoding* untuk seluruh fitur kategorikal, StandardScaler, serta pemilihan hyperparameter melalui GridSearchCV dengan lima lipatan StratifiedKFold. Proses pencarian parameter menunjukkan bahwa kombinasi terbaik adalah kernel RBF, C = 10, dan gamma = 'scale', menghasilkan skor rata-rata f1_macro sebesar 0.8451 pada data training yang telah di-oversample. Hasil ini menunjukkan bahwa model mampu mempelajari pola dari data sintetis dengan baik, meskipun belum tentu mencerminkan performa pada data nyata.

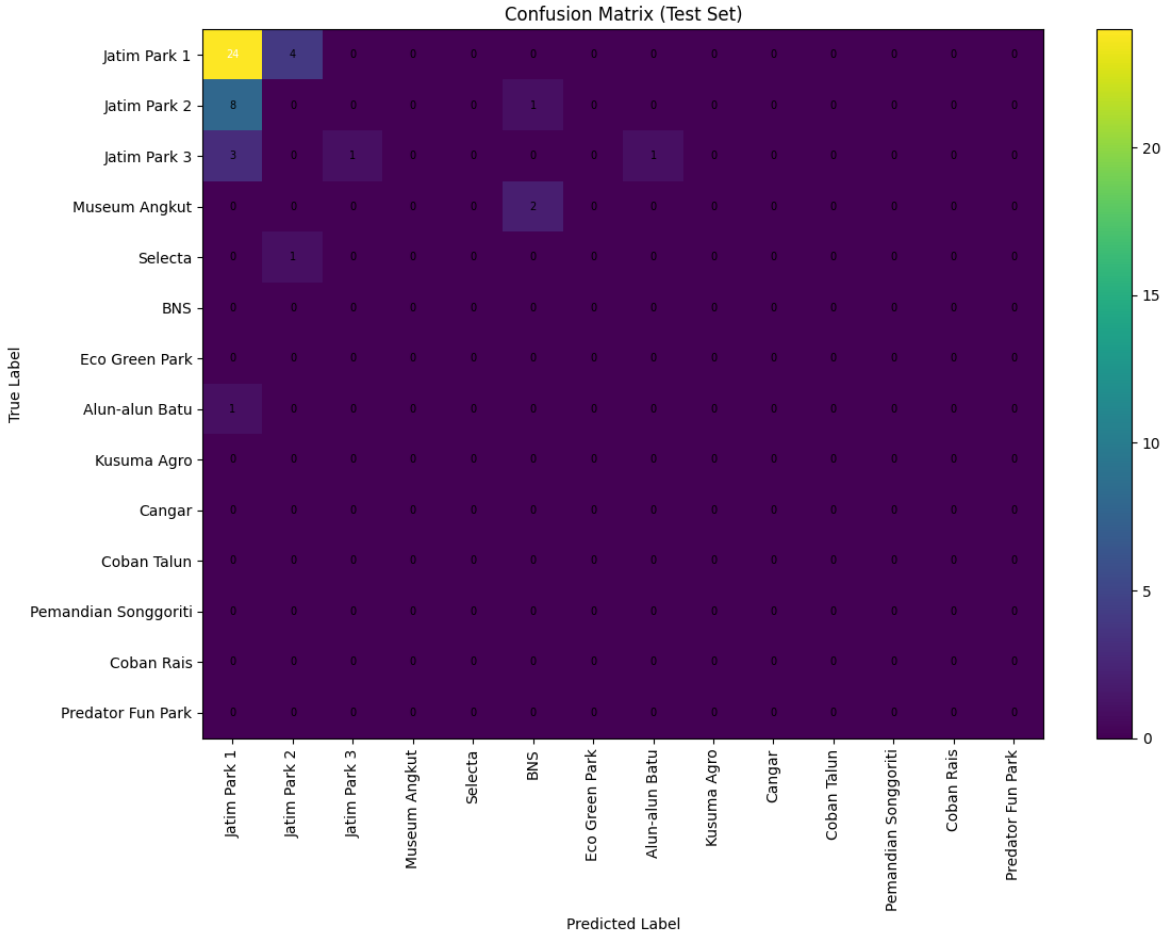
Classification report for all 14 classes:

	precision	recall	f1-score	support
Jatim Park 1	0.67	0.86	0.75	28
Jatim Park 2	0.00	0.00	0.00	9
Jatim Park 3	1.00	0.20	0.33	5
Museum Angkut	0.00	0.00	0.00	2
Selecta	0.00	0.00	0.00	1
BNS	0.00	0.00	0.00	0
Eco Green Park	0.00	0.00	0.00	0
Alun-alun Batu	0.00	0.00	0.00	1
Kusuma Agro	0.00	0.00	0.00	0
Cangar	0.00	0.00	0.00	0
Coban Talun	0.00	0.00	0.00	0
Pemandian Songgoriti	0.00	0.00	0.00	0
Coban Rais	0.00	0.00	0.00	0
Predator Fun Park	0.00	0.00	0.00	0
accuracy			0.54	46
macro avg	0.12	0.08	0.08	46
weighted avg	0.51	0.54	0.49	46

Gambar 5. Laporan Evaluasi

Berdasarkan Gambar 5, evaluasi pada data testing menunjukkan hasil yang jauh lebih moderat. Akurasi yang dicapai sebesar 54.35%, atau lebih rendah dibandingkan baseline

yang sekadar menebak kelas paling sering (≈61%). Namun demikian, akurasi bukan satu-satunya ukuran yang relevan dalam dataset yang tidak seimbang. Metrik lain yang lebih berfokus pada performa lintas kelas, yakni Precision Macro (0.2381), Recall Macro (0.1510), dan F1 Macro (0.1548), menunjukkan bahwa model hanya mampu menangani sebagian kecil kelas secara efektif. Secara keseluruhan, nilai-nilai ini mengindikasikan bahwa meskipun model mampu mengenali kelas mayoritas dengan cukup baik, ia gagal melakukan generalisasi yang memadai pada kelas minoritas.



Gambar 6. Heatmap confusion matrix

Hal tersebut terlihat lebih jelas pada laporan klasifikasi per kelas sebagaimana disajikan pada Gambar 6. Untuk Jatim Park 1, model menunjukkan performa yang sangat baik, dengan recall mencapai 0.86 dan F1-score 0.75. Sebaliknya, banyak kelas dengan jumlah sampel kecil seperti Jatim Park 2, Museum Angkut, Selecta, dan beberapa kelas wisata lain memiliki nilai *precision*, *recall*, dan *F1-score* sebesar 0.00. Ini menunjukkan bahwa tidak ada satu pun sampel dari kelas-kelas tersebut yang berhasil diprediksi dengan benar. Model cenderung mengalihkan prediksi ke kelas yang memiliki pola lebih jelas, yakni Jatim Park 1. Hasil confusion matrix memperlihatkan bahwa sebagian besar kesalahan klasifikasi berasal dari salah prediksi menjadi kelas 1, bahkan untuk kelas yang seharusnya memiliki karakteristik berbeda.

Rendahnya performa model pada kelas minoritas dapat dijelaskan oleh beberapa faktor. Selain jumlah data yang sangat tidak seimbang, fitur prediktor X1–X10 merupakan fitur kategorikal dengan rentang nilai yang relatif sempit. Kemungkinan besar fitur-fitur ini belum cukup informatif untuk membedakan secara tajam antara destinasi wisata yang secara

karakteristik atau konteks perilaku wisatawan sangat mirip. Meskipun teknik oversampling meningkatkan peluang model untuk melihat sampel dari kelas minoritas, data sintetis yang tercipta belum mampu meniru variasi nyata sehingga model berpotensi mengalami overfitting pada data hasil oversampling namun gagal mengeneralisasi pada data testing.

Perbandingan dengan baseline memperkuat temuan tersebut. Baseline most-frequent predictor masih lebih tinggi akurasinya dibandingkan model, yang berarti model tidak memperoleh keuntungan signifikan dari fitur yang tersedia. Hal ini mengindikasikan perlunya peningkatan kualitas fitur, misalnya dengan menambahkan variabel perilaku, faktor geospasial, atau preferensi wisata yang lebih rinci. Alternatif lain adalah menggabungkan beberapa kelas minoritas menjadi kelompok "wisata lain" apabila granularitas klasifikasi tidak menjadi prioritas, sehingga model dapat lebih fokus membedakan kategori besar alih-alih 14 kelas terpisah.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa SVM dengan kernel RBF dan strategi oversampling mampu memberikan pembelajaran yang kuat pada kelas mayoritas, tetapi belum efektif dalam mendeteksi kelas-kelas dengan jumlah sampel terbatas. Untuk aplikasi nyata, kinerja seperti ini belum ideal karena risiko bias prediksi yang tinggi. Upaya lanjutan yang direkomendasikan meliputi: penambahan data nyata untuk kelas minoritas, eksplorasi fitur baru yang lebih deskriptif, penerapan teknik imbalance yang lebih maju seperti SMOTE-ENN, serta penggunaan model berbasis ensemble atau gradient boosting.

KESIMPULAN

Penelitian ini telah mengembangkan sebuah sistem rekomendasi destinasi wisata berbasis data demografis menggunakan metode klasifikasi SVM. Sistem dirancang untuk memetakan karakteristik wisatawan, seperti usia, jenis kelamin, dan status sosial ke dalam rekomendasi destinasi yang paling sesuai. Hasil eksperimen menunjukkan bahwa distribusi label pada dataset sangat tidak seimbang, dengan kelas Jatim Park 1 mendominasi lebih dari separuh data, sementara banyak kelas lain memiliki jumlah sampel yang sangat sedikit. Ketidakseimbangan ini berdampak langsung pada performa model. Upaya penyeimbangan data menggunakan teknik oversampling berhasil meningkatkan performa pada tahap pelatihan (cross-validation mencapai F1-macro 0.84), namun pengujian pada data uji yang tetap mencerminkan distribusi asli menunjukkan performa yang masih rendah. Model hanya mencapai akurasi 54%, dengan nilai precision, recall, dan F1-macro yang juga rendah, terutama pada kelas-kelas minoritas yang gagal diprediksi dengan baik. Temuan ini mengindikasikan bahwa meskipun metode SVM dapat bekerja dengan baik pada data yang seimbang, sistem rekomendasi ini belum mampu memberikan prediksi optimal pada dataset yang sangat timpang seperti kasus ini. Kondisi ini menegaskan perlunya strategi peningkatan kualitas data, seperti penambahan sampel nyata untuk kelas minoritas, penggunaan metode pembobotan kelas yang lebih adaptif, atau eksplorasi algoritma alternatif yang lebih tahan terhadap ketidakseimbangan kelas. Secara keseluruhan, penelitian ini tetap menunjukkan potensi awal dalam mengembangkan sistem rekomendasi destinasi wisata berbasis demografi. Namun, akurasi dan kemampuan generalisasi model masih perlu ditingkatkan pada tahap penelitian selanjutnya agar sistem dapat memberikan rekomendasi yang lebih akurat dan bermanfaat bagi wisatawan maupun pelaku industri pariwisata.

REFERENSI

[1] Kemenparekraf, "Statistik Pariwisata Indonesia," Jakarta, 2023.
[2] N. Javadian, M. Shafiei, and R. Ali, "THOR: Hybrid recommender for personalized travel," Tourism Management Perspectives, vol. 40, p. 100902, 2021, doi: 10.1016/j.tmp.2021.100902.
[3] W. Chen and J. Gao, "A comprehensive pipeline for hotel recommendation using hybrid machine learning models," Expert Systems with Applications, vol. 149, p. 113280, 2020, doi: 10.1016/j.eswa.2020.113280.
[4] D. Rahmawati and S. Nugroho, "Penerapan SVM untuk rekomendasi tempat wisata berbasis profil pengguna," Jurnal Teknologi Informasi dan Terapan, vol. 8, no. 2, pp. 99–107, 2021, doi: 10.21009/jtit.082.05.
[5] I. M. Hidayat and W. P. Putri, "Rekomendasi wisata Jawa Timur menggunakan SVM dan PCA," Jurnal Sistem Cerdas, vol. 5, no. 1, pp. 12–20, 2023.
[6] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan penerapan algoritma *Support Vector Machine* (SVM) dalam data mining untuk menunjang strategi promosi," JUITA: Jurnal Informatika, vol. 7, no. 2, p. 71, 2019, doi: 10.30595/juita.v7i2.4378.
[7] X. Liu, H. Zhang, and L. Chen, "BoVW combined with SVM for tourism recommendation," International Journal of Computer Applications, vol. 118, no. 7, pp. 21–26, 2015, doi: 10.5120/20737-3301.
[8] D. Rahmawati and S. Nugroho, "Penerapan SVM untuk rekomendasi tempat wisata berbasis profil pengguna," Jurnal Teknologi Informasi dan Terapan, vol. 8, no. 2, pp. 99–107, 2021, doi: 10.21009/jtit.082.05.
[9] M. A. Hidayat and W. P. Putri, "Rekomendasi wisata Jawa Timur menggunakan SVM dan PCA," Jurnal Sistem Cerdas, vol. 5, no. 1, pp. 12–20, 2023.
[10] N. P. Sari, R. Hermanto, and R. Defriani, "Sentiment analysis of tourist reviews using K-NN and SVM," Procedia Computer Science, vol. 198, pp. 987–993, 2022, doi: 10.1016/j.procs.2022.01.134.
[11] A. Damayanti, A. Santoso, and F. Wibowo, "SVM model for Baturraden tourism visitor satisfaction," Tourism and AI Review, vol. 6, no. 2, pp. 55–63, 2024.
[12] Y. Yuan, Z. Zhang, and J. Lin, "SVM-based POI recommendation using user check-in behavior," Journal of Location-Based Services, vol. 16, no. 3, pp. 185–201, 2022, doi: 10.1080/17489725.2022.2101467.