

**PERBANDINGAN ALGORITMA *DECISION TREE*, *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* MENGGUNAKAN
DATA PARIWISATA**

SKRIPSI

Skripsi diajukan untuk memenuhi persyaratan memperoleh gelar sarjana



Disusun oleh:

Bagas Febrianto

210111471026

**PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS GLOBAL JAKARTA**

2024

PERNYATAAN ORISINALITAS SKRIPSI

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa berdasarkan hasil penelusuran berbagai karya ilmiah, gagasan dan masalah ilmiah yang diteliti dan diulas di dalam Naskah Skripsi ini adalah asli dari pemikiran saya. Tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik disuatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini dan disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah Skripsi ini dapat dibuktikan terdapat unsur-unsur jiplakan, saya bersedia Skripsi dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UUNo. 20 Tahun 2003, pasal 25 ayat 2 dan pasal 70).

Depok, 10 Juli 2024

Mahasiswa,



Bagas Febrianto

NIM. 210111471026

HALAMAN PENGESAHAN PEMBIMBING

HALAMAN PENGESAHAN PEMBIMBING

Skripsi ini diajukan oleh :
Nama : Bagas Febrianto
NIM : 210111471026
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Algoritma *Decision Tree*, *Naive Bayes* dan *Support Vector Machine* Menggunakan Data Pariwisata

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Global Jakarta.

DEWAN PEMBIMBING

Pembimbing 1 : Anindya Ananda Hapsari S.ST, M.IT ()

Pembimbing 2 : Halimatuz Zuhriyah S.KOM., M.IT ()

Ditetapkan di : Depok, Jawa Barat
Tanggal : 15 Agustus 2014

HALAMAN PENGESAHAN DEWAN PENGUJI

HALAMAN PENGESAHAN DEWAN PENGUJI

Skripsi ini diajukan oleh :
Nama : Bagas Febrianto
NIM : 210111471026
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Algoritma *Decision Tree*, *Naive Bayes* dan *Support Vector Machine* Menggunakan Data Pariwisata

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Global Jakarta.

DEWAN PENGUJI

Penguji 1 : Dian Nugraha, S.S.T., MIT

()

Penguji 2 : Ummy Gusti Salamah, S.S.T., MIT

()

Penguji 3 : Risna Oktaviati, S.S.T., MIT

()

Ditetapkan di : Depok, Jawa Barat

Tanggal : 15 Agustus 2019

KATA PENGANTAR/UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat- Nya, saya dapat menyelesaikan skripsi ini. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Teknik Jurusan Teknik Informatika pada Fakultas Teknik dan Ilmu Komputer. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

- (1) Anindya Ananda Hapsari S.ST, M.IT dan Halimatuz Zuhriyah S.KOM, M.IT, selaku dosen pembimbing yang telah menyediakan waktu, tenaga, dan pikiran untuk mengarahkan serta membimbing saya dalam penyusunan skripsi ini;
- (2) Pihak PT Inovasi Lentera Cipta Kreasi yang telah banyak membantu dalam usaha memperoleh data yang saya perlukan;
- (3) Orang tua dan keluarga saya yang telah memberikan bantuan dukungan material dan moral;
- (4) Umi Sjarifah selaku PIC Studi Independen di PT Inovasi Lentera Cipta Kreasi yang memberikan saya izin menggunakan data yang dimiliki oleh PT Inovasi Lentera Cipta Kreasi;
- (5) M. Sofyan S. Thayf, ST, M.Cs selaku mentor penulis saat melakukan studi independent dan sebagai orang yang memperkenalkan dan mengajarkan penggunaan *tools* yang penulis gunakan di skripsi ini; dan
- (6) Teman dan sahabat yang telah banyak membantu serta mendukung saya dalam menyelesaikan skripsi ini. Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu pengetahuan.

Depok, 10 Juli 2024

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI AKADEMIS

Sebagai sivitas akademika Universitas Global Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Bagas Febrianto
NPM : 210111471026
Program Studi : Teknik Informatika
Jenis Karya Ilmiah : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Global Jakarta **Hak Bebas Royalti Non-eksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

PERBANDINGAN ALGORITMA *DECISION TREE*, *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* MENGGUNAKAN DATA PARIWISATA

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Non-eksklusif ini Universitas Global Jakarta berhak menyimpan, mengalih-media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Depok, 10 Juli 2024

Yang menyatakan


Bagas Febrianto

NIM. 210111471026

ABSTRAK

PERBANDINGAN ALGORITMA *DECISION TREE*, *NAIVE BAYES* DAN *SUPPORT VECTOR MACHINE* MENGGUNAKAN DATA PARIWISATA

Oleh

Bagas Febrianto

Teknik Informatika

Pariwisata di Indonesia adalah sektor penting dengan berbagai destinasi. Pada tahun 2022, jumlah wisatawan mancanegara mencapai 5,9 juta dan meningkat menjadi 8,51 juta pada tahun 2023, dengan bandara Ngurah Rai sebagai pintu masuk utama. Untuk memaksimalkan potensi wisata, pemahaman preferensi wisatawan berdasarkan karakteristik pribadi seperti jenis kelamin dan pekerjaan sangat penting. Teknik analitik data seperti *Decision Tree*, *Naive Bayes*, dan *Support Vector Machine (SVM)* berguna untuk mengidentifikasi pola preferensi ini. Penggunaan perangkat lunak seperti *Orange Data mining* membantu peneliti mengolah data secara efisien. Penelitian menunjukkan bahwa *Orange Data mining* efektif dalam analisis data preferensi wisatawan. Setelah perbandingan 5 kali dengan data split yang berbeda, hasil menunjukkan bahwa *Decision Tree* memiliki *Accuracy*, *Precision*, dan *Recall* terbaik dengan *Confusion Matrix*, sedangkan *SVM* unggul dalam *Cross Validation*.

Kata kunci: *Decision Tree*; *Naïve Bayes*; *Support Vector Machine*.

ABSTRACT
COMPARISON OF *DECISION TREE*, *NAIVE BAYES* AND *SUPPORT*
***VECTOR MACHINE* ALGORITHMS USING TOURISM DATA**

By
Bagas Febrianto

Informatics Engineering

Tourism in Indonesia is an important sector with various destinations. In 2022, the number of foreign tourists reached 5.9 million and increased to 8.51 million in 2023, with Ngurah Rai airport as the main entry point. To maximize tourism potential, understanding tourist preferences based on personal characteristics such as Gender and occupation is essential. Data analytics techniques such as Decision Tree, Naive Bayes, and Support Vector Machine (SVM) are useful for identifying these preference patterns. The use of software such as Orange Data mining helps researchers process data efficiently. Research shows that Orange Data mining is effective in analyzing traveler preference data. After a 5-time comparison with different split data, the results show that Decision Tree has the best accuracy, precision, and Recall with Confusion Matrix, while SVM excels in Cross Validation.

Keywords: *Decision Tree; Naïve Bayes; Support Vector Machine.*

DAFTAR ISI

PERNYATAAN ORISINALITAS SKRIPSI	ii
HALAMAN PENGESAHAN PEMBIMBING	iii
HALAMAN PENGESAHAN DEWAN PENGUJI	iv
KATA PENGANTAR/UCAPAN TERIMA KASIH	v
PERNYATAAN PERSETUJUAN PUBLIKASI AKADEMIS	vi
ABSTRAK.....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II KAJIAN PUSTAKA.....	4
2.1 Landasan Teori.....	4
2.1.1 Pariwisata di Indonesia	4
2.1.2 <i>Machine Learning</i>	6
2.2 Tinjauan Penelitian Yang Berkaitan	14
BAB III METODE PENELITIAN	18
3.1 Alur penelitian	18
3.1.1 <i>Literatur review</i>	19
3.1.2 Pengumpulan Data	19

3.1.3 Pra-Pemrosesan Data	20
3.1.4 Analisis Data	20
3.1.5 Evaluasi dan Pengujian	21
3.1.6 Interpretasi Hasil	21
3.2 Lokasi & Obyek Penelitian	22
3.2.1 Lokasi Penelitian.....	22
3.2.2 Objek Penelitian.....	22
3.3 <i>Sample</i>	22
BAB IV HASIL DAN PEMBAHASAN	24
4.1 Pengumpulan Data	24
4.1.1 Responden Berdasarkan <i>Gender</i>	26
4.1.2 Responden Berdasarkan Usia	27
4.1.3 Responden Berdasarkan Cakupan Pekerjaan.....	27
4.1.4 Responden Berdasarkan Gaji	28
4.1.5 Responden Berdasarkan Kunjungan	29
4.1.6 Responden Berdasarkan Negara	29
4.1.7 Responden Berdasarkan Pilihan Tempat Wisata	30
4.3 Analisis Data	32
4.4 Evaluasi dan Validasi.....	34
4.5 Interpretasi Hasil.....	40
BAB V KESIMPULAN.....	42
5.1 Kesimpulan	42
5.2 Saran	42
DAFTAR PUSTAKA	43
LAMPIRAN.....	46

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu yang berkaitan.....	14
Tabel 3.1 Pertanyaan dan jawaban untuk responden	19
Tabel 4.1 <i>Dataset CSV Indonesian</i>	24
Tabel 4.2 Tabel perbandingan <i>Accuracy Confusion Matrix</i>	33
Tabel 4.3 Tabel perbandingan <i>Precision Confusion Matrix</i>	33
Tabel 4.4 Tabel perbandingan <i>Recall Confusion Matrix</i>	33
Tabel 4.5 Tabel Perbandingan <i>Algoritma dengan Cross Validation</i>	40
Tabel 4.6 Tabel Perbandingan <i>Algoritma dengan Confusion Matrix</i>	40

DAFTAR GAMBAR

Gambar 2.1 Jumlah Kunjungan Wisman	4
Gambar 2.2 <i>Decision Tree</i>	7
Gambar 2.3 Logo <i>Orange Data mining</i>	10
Gambar 2.4 Tabel <i>Confusion Matrix</i>	12
Gambar 2.5 Kurva <i>ROC</i>	12
Gambar 3.1 <i>Research Flowchart</i>	18
Gambar 4.2 Diagram Batang berdasarkan <i>Gender</i>	26
Gambar 4.3 Diagram Batang berdasarkan usia.....	27
Gambar 4.4 Diagram Batang berdasarkan cakupan pekerjaan	27
Gambar 4.5 Diagram Batang berdasarkan gaji	28
Gambar 4.6 Diagram Batang berdasarkan kunjungan	29
Gambar 4.7 Diagram Batang berdasarkan negara	29
Gambar 4.8 Diagram Batang berdasarkan pilihan tempat wisata	30
Gambar 4.10 Fitur File pada <i>orange</i>	31
Gambar 4.11 Fitur Preprocess pada <i>orange</i>	32
Gambar 4. 12 Kurva <i>ROC Decision Tree</i>	34
Gambar 4.13 Tabel <i>Confusion Matrix Decision Tree</i>	35
Gambar 4.14 Kurva <i>ROC Naive Bayes</i>	36
Gambar 4.15 Tabel <i>Confusion Matrix Naive Bayes</i>	37
Gambar 4.16 Kurva <i>ROC Support Vector Machine (SVM)</i>	38
Gambar 4.17 Tabel <i>Confusion Matrix Support Vector Machine (SVM)</i>	39
Gambar 4.18 Perbandingan Hasil <i>Cross Validation</i> dan <i>Confusion Matrix</i>	41

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pariwisata di Indonesia adalah sektor penting yang mencakup beragam destinasi dari pantai, gunung, hingga situs budaya. Negara ini terkenal dengan keindahan alamnya, keragaman budaya, dan warisan sejarahnya menjadikan Indonesia menjadi salah satu negara tujuan wisatawan asing untuk berlibur menurut Badan Pusat Statistik jumlah wisatawan yang mengunjungi Indonesia pada tahun 2022 sebanyak 5,9 juta dan pada tahun 2023 bertambah sebanyak 8,51 juta berdasarkan jenis pintu masuk, jumlah kunjungan wisatawan mancanegara paling banyak didominasi oleh mode angkutan udara berdasarkan data jumlah wisatawan mancanegara pada tahun 2022 hingga 2023 bandara Ngurah Rai memiliki jumlah kedatangan terbanyak dengan total 5,99 juta di susul oleh Soekarno Hatta sebanyak 1,99 juta dan Juanda sebanyak 184 ribu (BPS, 2023). Seiring dengan perkembangan teknologi dan kemajuan informasi, industri pariwisata berkembang pesat. Namun, untuk memaksimalkan potensi wisata, penting bagi pemangku kepentingan industri untuk memahami preferensi dan perilaku wisatawan.

Preferensi wisatawan sangat pribadi dan dipengaruhi oleh banyak hal, termasuk karakteristik pribadi, budaya, dan lingkungan sosial. Jenis kelamin dan pekerjaan wisatawan adalah salah satu faktor yang dapat mempengaruhi preferensi ini. Pria dan wanita dapat memiliki preferensi perjalanan yang berbeda, demikian pula pelancong dari berbagai profesi seperti pelajar, karyawan, pengusaha, atau profesional (Fadhillah, 2022).

Selama beberapa dekade terakhir, analitik data telah menjadi alat yang sangat berguna untuk memahami preferensi dan perilaku manusia. Metode pengelompokan seperti *Decision Tree* dan *Naive Bayes* adalah dua teknik analisis data yang populer untuk mengklasifikasikan dan mengidentifikasi pola berdasarkan atribut tertentu (Alim, 2021). Algoritma *Decision Tree* adalah teknik pembelajaran mesin yang memecah data menjadi cabang-cabang berdasarkan aturan Keputusan *Decision Tree* merupakan alat statistik yang ampuh untuk klasifikasi, prediksi, interpretasi, dan *clustering* (Kusanti, 2021) sedangkan algoritma *Naive Bayes*

merupakan metode klasifikasi berbasis probabilitas (Jackins, 2021) dan algoritma *Support Vector Machine* adalah algoritma yang dapat mendeteksi hyperplane dengan nilai terbaik yang membedakan antara dua kelas di ruang *input*. Pada dasarnya algoritma *Support Vector Machine* digunakan untuk proses klasifikasi linier, lalu dalam perkembangannya *SVM* juga dapat digunakan untuk menyelesaikan permasalahan non linier (Oktafiani & Rianto, 2023). Penggunaan teknik klasifikasi ini dapat membantu untuk memberikan referensi kelompok wisatawan sesuai dengan jenis kelamin, pekerjaan dan negara asal, sehingga memudahkan pengambilan Keputusan wisatawan dalam memilih tujuan destinasi wisata yang sesuai dengan karakteristik para wisatawan. Namun, menganalisis data yang kompleks seringkali menjadi tantangan bagi peneliti dan praktisi. Menangani data yang besar dan beragam, serta menggunakan teknik agregasi yang sesuai, membutuhkan pendekatan yang hati-hati dan alat yang efisien. Oleh karena itu, diperlukan penggunaan teknologi informasi yang tepat guna dan perangkat lunak yang handal untuk menganalisis data preferensi wisatawan berdasarkan jenis kelamin dan pekerjaan.

Orange Data mining adalah perangkat lunak analisis data *open source* yang memiliki banyak fitur untuk analisis data, seperti penggunaan metode *clustering* dan klasifikasi. Terdapat berbagai penelitian yang menerapkan *orange data mining* untuk mengolah data penelitiannya. Salah satu penelitian yang dilakukan oleh Rian Oktafiani dengan judul “*Perbandingan Algoritma Support Vector Machine (SVM) dan Decision Tree untuk Sistem Rekomendasi Tempat Wisata*” lalu penelitian yang dilakukan oleh Jaka Kusuma judul “*Komparasi Algoritma Support Vector Machine Dan Naive Bayes Pada Klasifikasi Ras Kucing*”.

Dengan menggunakan *Orange Data mining*, peneliti dapat mengolah data secara lebih efisien dan mendapatkan hasil analisis yang lebih akurat, membantu mengidentifikasi pola preferensi wisatawan secara lebih mendalam.

Dengan menggunakan teknologi informasi dan perangkat lunak *Orange Data mining*, penelitian ini bertujuan untuk menyajikan analisis yang lebih luas dan mendalam tentang preferensi pariwisata di Indonesia berdasarkan klasifikasi *Gender* dan pekerjaan. Penelitian ini diharapkan akan membantu industri pariwisata mengembangkan strategi pemasaran yang lebih terarah dan berhasil. Selain itu,

penelitian ini akan membantu kemajuan dalam ilmu pengetahuan dan teknologi informasi di bidang pariwisata dan analisis data.

1.2 Rumusan Masalah

Adapun rumusan masalah untuk penelitian ini adalah sebagai berikut:

1. Bagaimana hasil perbandingan *Accuracy* algoritma *Naive Bayes*, *Decision Tree* dan *Support Vector Machine* ?
2. Di antara algoritma *Naive Bayes*, *Decision Tree* dan *Support Vector Machine* algoritma apa yang memberikan hasil terbaik dalam hal *Precision* dan *Recall* ?

1.3 Tujuan Penelitian

Berikut adalah beberapa tujuan penelitian ini, antara lain:

1. Membandingkan hasil *Accuracy* algoritma *Decision Tree*, *Naïve Bayes* dan *Support Vector Machine* menggunakan data pariwisata.
2. Mengetahui apakah Algoritma *Decision Tree*, *Naïve Bayes* dan *Support Vector Machine* cocok dan efisien hasilnya terhadap penelitian yang dilaksanakan.

1.4 Manfaat Penelitian

Penelitian ini dapat Menunjukkan potensi penggunaan metode *Decision Tree*, *Naive Bayes* dan *Support Vector Machine* (SVM) dalam analisis preferensi wisata serta membantu dan mempermudah pengambilan keputusan dalam menentukan algoritma yang lebih sesuai untuk dipilih.

1.5 Batasan Masalah

Batasan masalah yang diidentifikasi dalam penelitian ini, adalah:

1. Penelitian ini akan menggunakan *Decision Tree*, *Naive Bayes* dan *Support Vector Machine* (SVM).
2. *Dataset* yang digunakan di dapat dari PT Inovasi Lentera Cipta Kreasi yang merupakan *dataset* dengan tema pariwisata.
3. *Tools* atau aplikasi yang digunakan dalam penelitian ini yaitu *Orange Data mining*.