

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Artikel Penelitian

Perbandingan Algoritma Support Vector Machine (SVM) dan *Decision Tree* untuk Sistem Rekomendasi Tempat Wisata

Rian Oktafiani^{a,*}, Rianto Rianto^b^{ab}Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Jl. Siliwangi, Jombor, Sleman, D.I. Yogyakarta, 55285, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 03 Februari 2023

Revisi Akhir: 24 Agustus 2023

Diterbitkan Online: 31 Agustus 2023

KATA KUNCI

Decision Tree,
Principal Component Analysis,
Support Vector Machine,
Sistem Rekomendasi Wisata

KORESPONDENSI

E-mail: rian.oktafiani@student.uty.ac.id*

A B S T R A C T

Industri pariwisata Indonesia berkembang dari tahun ke tahun. Daerah Istimewa Yogyakarta merupakan salah satu provinsi yang memiliki banyak destinasi wisata. Pertumbuhan internet dan teknologi informasi juga menjadi faktor dalam industri pariwisata Indonesia. Dengan adanya informasi mengenai pariwisata di internet, dapat memudahkan wisatawan untuk mencari informasi. Namun, karena jumlah informasi yang sangat banyak akan membuat wisatawan kebingungan untuk menentukan tujuan wisata. Selain itu, wisata lokal memiliki potensi yang cukup tinggi untuk membantu perekonomian daerah, namun saat ini belum dieksplorasi secara maksimal. Sistem rekomendasi dan kemampuan klasifikasi tempat wisata diperlukan untuk memberikan akurasi rekomendasi yang baik. Untuk menentukan jumlah fitur yang paling menguntungkan untuk klasifikasi lokasi wisata, Teknik *Principal Component Analysis* (PCA) digunakan dalam penelitian ini untuk membandingkan metodologi *Support Vector Machine* (SVM) dan *Decision Tree* (DT). Hasilnya menunjukkan bahwa, dengan nilai akurasi 98.97% penerapan PCA dengan nilai $n=5$ dan berada pada perbandingan *Split Data* 75% : 25%, pendekatan SVM memiliki performa lebih baik daripada metode *Decision Tree*. Metode *Decision Tree* juga memiliki performa yang baik, dengan menggunakan PCA dengan nilai $n=5$, *Decision Tree* memiliki akurasi 96.55% yang berada pada perbandingan *Split Data* 85% : 15%.

1. PENDAHULUAN

Perkembangan pariwisata di Indonesia sangat berkembang dari tahun ke tahun. Menurut Badan Pusat Statistik (BPS) jumlah wisatawan mancanegara yang berkunjung ke Indonesia pada periode Januari-Oktober 2022 sebanyak 3.92 juta orang, naik 215.16% dari jumlah pengunjung pada periode yang sama tahun 2021 [1]. Salah satu dari banyak provinsi yang memiliki banyak destinasi wisata di Indonesia adalah Daerah Istimewa Yogyakarta. Yogyakarta memiliki banyak harapan dalam industri pariwisata Indonesia karena banyaknya tempat wisata. Menurut data Badan Perencanaan Pembangunan, Penelitian dan Pengembangan Daerah (Bappeda) Yogyakarta, jumlah kunjungan wisatawan yang berkunjung ke Yogyakarta pada tahun 2022 sebanyak 8.708.587 orang wisatawan [2].

Informasi tempat wisata dan pemahaman nilai budaya daerah dapat diperoleh dengan lebih mudah berkat kemajuan teknologi. Integrasi teknologi informasi ke dalam pemasaran pariwisata dilakukan untuk keuntungan penyedia layanan dan manajemen pariwisata [3]. Dengan adanya informasi mengenai pariwisata di internet, dapat memudahkan wisatawan untuk mencari informasi. Namun, karena jumlah informasi yang sangat banyak akan membuat wisatawan kebingungan untuk menentukan tujuan wisata. Selain itu, pariwisata lokal memiliki banyak potensi untuk membantu perekonomian daerah, namun belum dieksplorasi secara maksimal. Pengguna juga membutuhkan sistem yang responsif dan iteratif yang mampu membuat rekomendasi berdasarkan kebutuhan pribadi wisatawan [4]. Oleh karena itu, dibutuhkan sistem rekomendasi, untuk merekomendasikan destinasi wisata untuk wisatawan, agar wisata lokal dapat dieksplorasi dengan baik dan wisatawan tidak lagi kebingungan untuk memilih destinasi wisata.

Sistem rekomendasi adalah sistem yang dapat membuat saran mengenai item tertentu yang dapat membantu pengguna sistem membuat keputusan [5]. Informasi yang terlalu luas dapat dipersempit oleh sistem rekomendasi, yang juga dapat menyampaikan informasi berdasarkan kebutuhan pengguna. Dengan menganalisis data, informasi, dan lingkungan pengguna, sistem rekomendasi menentukan kebutuhan pengguna. Hasilnya, sistem rekomendasi akan memberikan informasi berdasarkan kebutuhan pengguna dan akan disajikan menggunakan teknik atau model sistem rekomendasi [6].

Dalam sistem pemberi rekomendasi, ada tiga pendekatan: *collaborative filtering*, *content-based filtering*, dan *hybrid*. Dalam penelitian ini, menggunakan pendekatan *content-based filtering*. Berdasarkan deskripsi item dan preferensi pengguna, *content-based filtering* merekomendasikan item yang sesuai untuk pengguna. Saat menentukan jarak antar item, *Content-Based Filtering* menggunakan data dari item yang digunakan sebagai atribut [7].

Perkembangan teknologi khususnya dalam bidang kecerdasan buatan yang berbasis *machine learning* dengan menerapkan *data mining* saat ini sangat memungkinkan untuk digunakan pada berbagai permasalahan, salah satunya untuk klasifikasi data [8]. Metode klasifikasi *data mining* untuk *machine learning* yang digunakan yaitu *Decision/Classification Tree*, *Bayesian Classifiers/Naïve Bayes Classifiers*, *K-Nearest Neighbor* dan *Support Vector Machines* [9]. Pada penelitian ini menggunakan metode klasifikasi *machine learning* menggunakan algoritma *Decision Tree* dan *Support Vector Machine*. Metode klasifikasi yang dilakukan yaitu untuk mengklasifikasikan data tempat wisata yang ada di Yogyakarta.

Metode *Decision Tree* dianggap sebagai algoritma yang banyak digunakan untuk mengkategorikan data atau objek, dengan hasil klasifikasi diperoleh dengan jelas dan mampu mengatasi *missing value*, menangani data kontinu, menangani *pruning data*, dan menggunakan nilai *gain ratio* sebagai kunci utama untuk memecahkan masalah yang ada [10]. Permasalahan yang memiliki *output* bernilai diskrit dapat dipecahkan menggunakan *Decision Tree* [11].

Support Vector Machine (SVM) adalah teknik yang dapat mengidentifikasi *hyperplane* dengan nilai terbaik yang membedakan antara dua kelas di ruang input [12]. Pada dasarnya metode *Support Vector Machine* digunakan untuk proses klasifikasi linier, kemudian dalam perkembangannya SVM juga digunakan untuk menyelesaikan permasalahan non linier [13]. Penelitian sebelumnya yang melakukan analisis sentimen terhadap ulasan *villa* di Ubud dengan membandingkan metode *Naive Bayes*, *Decision Tree* dan *k-NN*. *RapidMiner* digunakan untuk pengolahan analisis sentimen. Hasil akurasi dari penelitian ini yaitu *Naive Bayes* memiliki akurasi 86.66%, *Decision Tree* memiliki akurasi 87.78%, dan *K-NN* 91.26% [14].

Dengan menggunakan teknik *Support Vector Machine* (SVM), pada penelitian [15] membahas mengenai peramalan jumlah wisatawan yang akan berkunjung ke suatu negara. Dengan menggunakan data gabungan dari 53 negara, Selain itu, prakiraan dibuat untuk 5 tujuan wisata teratas. Hasil penelitian yaitu, SVM *Confidence* sebesar 99.13%, nilai MAPE 2.78%, dan nilai RMSE 2783.57 semuanya merupakan hasil yang sangat baik.

Penelitian selanjutnya melakukan analisis performa algoritma *Naive Bayes Classifier* (NBC), *Decision Tree* (DT), dan *Support Vector Machine* (SVM) untuk klasifikasi data ulasan pengunjung Candi Borobudur. Hasil penelitian ini yaitu, akurasi NBC sebesar 95.6%, DT 94.40% dan SVM sebesar 99.41% [16].

Kajian selanjutnya yaitu analisis sentimen terhadap *review* hotel. Hasil penelitian ini yaitu Pendekatan pembobotan TF-IDF dan metode SVM yang keduanya mencapai tingkat akurasi sebesar 95% dalam proses klasifikasi, disusul dengan metode pembobotan unigram dan metode SVM yang mencapai tingkat akurasi sebesar 94% [17].

Berdasarkan penelitian-penelitian sebelumnya, dapat dilihat metode yang memiliki hasil yang baik yaitu metode *Support Vector Machine* (SVM) dan *Decision Tree*, sehingga penelitian ini melakukan proses klasifikasi untuk rekomendasi tempat wisata, dengan menggunakan metode *Decision Tree* dan *Support Vector Machine* (SVM). Analisis perbandingan performa algoritma menggunakan aplikasi *RapidMiner*. Program perangkat lunak *open-source* yang disebut *RapidMiner* menawarkan solusi untuk penambangan data (*data mining*), penambangan teks (*text mining*), dan penelitian prediktif [18]. Hasil klasifikasi menggunakan kedua pendekatan tersebut kemudian akan dibandingkan dan diteliti untuk menentukan metode mana yang memiliki tingkat akurasi paling tinggi saat merekomendasikan tempat wisata.

2. METODE

2.1. Tahapan Penelitian

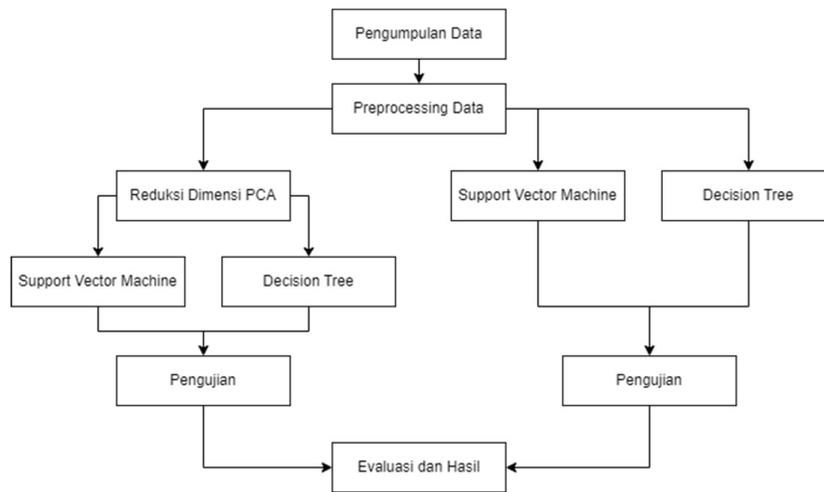
Penelitian ini dilaksanakan untuk rekomendasi tempat wisata di Yogyakarta. Rekomendasi diberikan berdasarkan hasil klasifikasi dengan menggunakan data *rating*, data tipe wisata, harga tiket masuk wisata, dan deskripsi wisata. Data yang diperoleh untuk penelitian ini didasarkan pada data sekunder dan merupakan jenis penelitian kuantitatif, serta hasil dari penelitian ini menggunakan perhitungan secara matematis menggunakan sistem. Data yang digunakan diperoleh dari Kaggle yang berjudul *Yogyakarta Tourism Place* [19]. Data yang diperoleh terlebih dahulu akan menjalani *preprocessing*. Pada tahap *preprocessing* dilakukan data *cleaning* dengan cara mengisi *missing value*. Setelah itu, dilakukan proses klasifikasi untuk rekomendasi tempat wisata menggunakan metode *Decision Tree* dan *Support Vector Machine* (SVM). Tahapan pada penelitian dapat dilihat pada Gambar 1.

2.1.1. Pengumpulan Data set

Penelitian ini memanfaatkan data sekunder yang dikumpulkan dari website Kaggle, sebuah website yang menawarkan *dataset*. (<https://www.kaggle.com/datasets/mrafif/yogyakarta-tourism-place>) tentang *dataset* tempat wisata yang berada di Yogyakarta. *Dataset* tempat wisata terdiri dari 10 kolom yaitu, no, nama, *vote average*, *vote count*, *type*, *htm weekday*, *htm weekend*, *longitude*, dan *latitude*. Total *dataset* yaitu 473 baris data.

- a. No merupakan nomor,
- b. Nama merupakan nama tempat wisata,
- c. *Vote average* merupakan data *rating*,
- d. *Vote count* merupakan jumlah orang yang melakukan *rating*,

- e. *Type* adalah jenis tempat wisata. Jenis tempat wisata pada *dataset* ini yaitu wisata alam, budaya dan sejarah, buatan, pantai, museum, wisata air, desa wisata, agrowisata, dan lainnya.
- f. *Htm weekday* merupakan harga tiket masuk pada hari kerja,
- g. *Htm weekend* merupakan harga tiket masuk diakhir pekan,
- h. *Longitude* merupakan garis bujur,
- i. *Latitude* adalah garis lintang.



Gambar 1. Tahapan Penelitian

Pada Tabel 1 merupakan *dataset* asli yang belum mengalami proses *data preprocessing*.

Tabel 1 *Dataset* Sebelum *Preprocessing*

no	vote a verage	nama	vote_count	type	htm_weekday	htm_weekend	latitude	longitude
9	1	Candi Borobudur	81922	Budaya dan Sejarah	50000	50000	-7.607.086.854	1.102.036.226
10	1	Candi Prambanan	71751	Budaya dan Sejarah	50000	50000	-7.751.834.561	1.104.915.318
24	0	Tebing Breksi	51431	Alam	10000	10000	-7.781.476.547	1.105.045.757
343	1	Gembira Loka Zoo	36337	Buatan	20000	25000	-78.062.344	1.103.967.977
346	1	The Palace of Yogyakarta (Keraton Yogyakarta)	30091	Budaya dan Sejarah	8000	8000	-78.052.845	1.103.642.031

2.1.2. Pre-Processing Data

Preprocessing adalah langkah penting dalam proses kalsifikasi untuk data. *Preprocessing* adalah proses menghilangkan kesalahan atau hal-hal lain yang dianggap tidak relevan dan dapat menurunkan nilai hasil pengolahan data. *Case folding*, *tokenization*, *filtering*, dan *stemming* adalah beberapa tahapan proses yang digunakan dalam kategorisasi dokumen yang menggunakan data bertipe teks [20]. Pemeriksaan ulang data dilakukan pada tahap *preprocessing/cleaning*, pada tahapan ini terdapat proses menghilangkan redudansi, *outlier*, dan nilai *null* (data kosong). Hal ini dilakukan untuk memastikan data input yang diproses adalah data yang “bersih”, sehingga dapat memastikan bahwa hasil perhitungan algoritma *data mining* juga akan memberikan hasil yang sesuai [21].

2.2. Reduksi Dimensi Principal Component Analysis (PCA)

Teknik untuk memecah dimensi data menjadi bentuk yang sesuai dengan beberapa bidang nilai adalah *Principal Component Analysis* (PCA). Metode PCA digunakan untuk melakukan

reduksi dimensi guna menentukan apakah fitur-fitur dalam *dataset* terhubung atau tidak [22]. PCA digunakan untuk mereduksi variabel yang dapat mempengaruhi perkiraan dan kualitas [23]. Nilai eigen dan vektor eigen matriks kovarians data digunakan untuk melakukan proses reduksi variabel [24]. Berikut persamaan untuk menghitung vektor eigen:

$$Y_i = \sum_{i=1}^p a_i x_i \tag{1}$$

Dimana:

Y_i = *eigen vector*

a_i = *eigen value*

x_i = *observation data*

p = *observation variable*

2.3. Support Vector Machine

Support Vector Machine (SVM) merupakan *supervised machine learning*, SVM dapat digunakan untuk menentukan klasifikasi dan regresi [25]. Data harus diubah dari bentuk non-linier menjadi linier melalui proses yang disebut kemelisasi. Menurut definisinya, fungsi kernel mendefinisikan fitur data mulai dari perspektif terendah dan berlanjut ke karakteristik tambahan dengan perspektif yang lebih besar [26]. Terdapat beberapa

kernel dalam *Support Vector Machine* (SVM) diantaranya kernel linear, kernel polynomial, kernel *Gaussian Radial Basis Function* (RBF), dan kernel sigmoid. Berikut persamaan untuk setiap kernel:

a. Kernel Linear

$$K(x_i, x) = x_i x \quad (2)$$

b. Kernel RBF

$$K(x_i, x) = \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right) \quad (3)$$

c. Kernel Polynomial

$$K(x_i, x) = (x_i x)^d \quad (4)$$

d. Kernel Sigmoid

$$K(x_i, x) = \tanh(\sigma(x_i x) + c) \quad (5)$$

Keterangan :

x_i dan x adalah data pada *data training*, parameter σ merupakan sigma, C adalah *complexity*, d adalah *degree*, dan $-\|x_i - x\|^2$ merupakan kuadrat jarak vektor x_i dan x .

2.4. Decision Tree

Membuat akar pohon keputusan yaitu dari atribut yang dipilih, algoritma pohon keputusan (*Decision Tree*) didasarkan pada nilai perolehan *information gain*. metode pemilihan karakteristik sesuai dengan pencapaian tujuan kelas berbasis penyelesaian objek. Atribut akar dari pohon keputusan adalah atribut dengan nilai *information gain* tertinggi [27]. Berikut persamaan untuk menghitung *entropy* dan *information gain*:

- Menyiapkan *data training*, dapat dikelaskan ke kelompok tertentu
- Menghitung *entropy* setiap atribut, menggunakan persamaan:

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (6)$$

dimana

S = Himpunan Kasus

n = Jumlah partisi S

p_i = Probabilitas ditentukan dengan membagi total kasus dengan jumlah kelas

- Menghitung *information gain* dari setiap atribut, menggunakan persamaan:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (7)$$

dimana

S = Himpunan Kasus

A = Atribut

n = Jumlah partisi S

$|S_i|$ = Jumlah partisi ke- i

$|S|$ = Jumlah kasus dalam S

- Kemudian hitunglah *gain ratio*, dengan suatu *term* yang disebut *Split Information*, rumusnya:

$$split\ information = - \sum_{t=1}^c \frac{s_1}{s} \log_2 \frac{s_1}{s} \quad (8)$$

- Kemudian hitunglah *gain ratio*

$$gainratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (9)$$

- Selanjutnya mengulangi langkah ke-2 sampai semua *record* terpartisi.

2.5. Pengujian

Model klasifikasi yang dihasilkan akan diuji. Dalam penelitian ini, *Confusion Matrix* digunakan selama pengujian. *Precision*, *recall*, dan *accuracy* membentuk *confusion matrix*. Tingkat kesamaan antara nilai aktual dan nilai yang diharapkan diukur dengan menggunakan akurasi. Sedangkan keakuratan informasi yang diminta pengguna dan *respons* yang diberikan sistem digunakan untuk mengukur nilai presisi. Tingkat keberhasilan sistem dalam memulihkan informasi ditunjukkan dengan nilai *recall*.

$$Akurasi = \frac{Tp+Tn}{Tp+Fn+Fp+Tn} \quad (10)$$

$$Presisi = \frac{Tp}{Tp+Fp} \quad (11)$$

$$Recall = \frac{Tp}{Tp+Fn} \quad (12)$$

dimana :

- TP (*True Positif*) : Prediksi kelas dan jumlah data di kelas aktual adalah positif.
- FN (*False Negative*) : Jumlah data di kelas sebenarnya positif, tetapi kelas yang diproyeksikan negatif.
- FP (*False Positive*) : Kelas yang diproyeksikan positif sedangkan kelas sebenarnya memiliki jumlah data negatif.
- TN (*True Negative*) : Kelas aktual dan prediksi memiliki jumlah data negatif.

3. HASIL

3.1. Pre-Processing Data

Dataset awalnya berisi 473 baris data dan 10 atribut berupa data sekunder, namun setelah proses *preprocessing* data selesai, hanya 388 baris data dan 10 atribut yang akan digunakan dalam penelitian. Data *preprocessing* harus diselesaikan sebelum klasifikasi. Tahap *preprocessing* dalam penyelidikan ini terdiri dari pembersihan data dan transformasi data. *Missing value* ditangani pada tahap pembersihan data (*data cleaning*), dan data yang tidak sesuai akan dihapus. Pada *dataset* tempat wisata, *data cleaning* yang dilakukan berupa pengisian *missing value*, karena pada atribut *Type* terdapat 208 baris data yang mengandung *missing value*. Pada atribut *htm_weekday* dan *htm_weekend* terdapat 206 *missing value*. Data yang tidak sesuai seperti data tempat wisata yang sudah tutup permanen, tempat wisata yang berada di luar jogja, data wisata yang memiliki *rating* tinggi namun hanya 1 orang yang melakukan *rating* akan dihapus, karena data tersebut tidak sesuai dan dapat menurunkan kualitas sistem rekomendasi.

Tahap selanjutnya adalah transformasi data yang digunakan untuk mengubah data kategorikal menjadi dalam bentuk kode angka. Pada *dataset*, dilakukan transformasi data untuk 2 atribut kategorikal berikut yaitu *type* dan label. *Type* tempat wisata diubah menjadi:

- Alam= 1,
- Budaya dan Sejarah= 2,
- Buatan = 3,
- Pantai= 4,
- Museum= 5,
- Wisata Air= 6,
- Desa Wisata= 7,
- Agrowisata= 8,
- Wisata lainnya= 9.

Sedangkan untuk atribut label, mengambil dari data *rating*. Untuk $rating < 4.5 = 0$, dan $rating \geq 4.5 = 1$. Data yang digunakan untuk

penelitian setelah dilakukan *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. *Dataset* setelah *preprocessing*

label	nama	vote average	vote count	type	htm weekday	htm weekend	latitude	longitude
1	Candi Borobudur	4.7	81922	2	50000	50000	-607.086.854	1.102.036.226
1	Candi Prambanan	4.7	71751	2	50000	50000	-751.834.561	1.104.915.318
0	Tebing Breksi	4.4	51431	1	10000	10000	-781.476.547	1.105.045.757
1	Gembira Loka Zoo	4.5	36337	3	20000	25000	-78.062.344	1.103.967.977
1	Keraton Yogyakarta	4.6	30091	2	8000	8000	-78.052.845	1.103.642.031

3.2. Pemrosesan Data

Pemrosesan data selesai setelah data telah diproses sebelumnya. Proses klasifikasi dilakukan pada tahap pengolahan data dengan menggunakan metode *Support Vector Machine* (SVM) dan *Decision Tree*. Data yang diproses sebelumnya dibagi menjadi *data training* dan *data testing* sebelum ke prosedur klasifikasi, atau proses data terpisah digunakan. Model *Support Vector Machine* (SVM) dan *Decision Tree* akan dibangun menggunakan data latih, dan keefektifannya akan dievaluasi menggunakan data pengujian. Pendekatan PCA terdiri dari *dimensionality reduction= fixed number* dan *number of component= 5*. Metode *Decision Tree* dibangun dengan *splitter='random'*, *criterion="entropy"*, dan *max_depth 300*, dan *max_features=10*. Sedangkan metode *Support Vector Machine* (SVM) dibangun dengan menggunakan *kernel type = dot* atau *linear*, *kernel cache= 200*, *convergence epsilon = 0.001*, dan *max iteration= 100000*. Hasil terbaik, diperkirakan, dapat dicapai dengan menguji poin-poin pada masing-masing pendekatan ini.

Tabel 3. Pembagian Data Latih dan Data Uji

<i>Split Data</i> Latih dan Data Uji	Data Latih	Data Uji
90% dan 10%	349	39
85% dan 15%	329	59
80% dan 20%	310	78
75% dan 25%	291	97
70% dan 30%	271	117

3.3. Pengujian

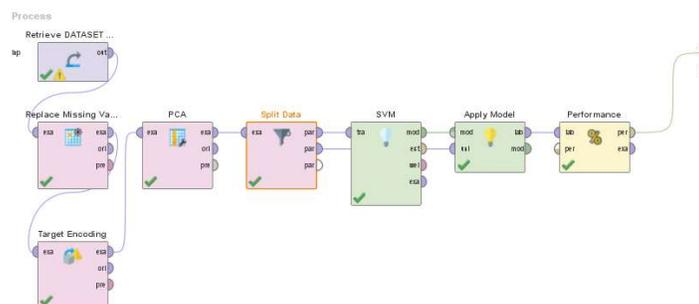
Penelitian ini menguji data latih dan data uji menggunakan lima skenario dengan menggunakan metode pengujian *percentage split*. yaitu 90%:10%, 85%:15%, 80%:20%, 75%:25% dan 70%:30%. Pembagian data latih dan data uji dengan *Percentage Split* ditunjukkan pada Tabel 3.

Pengujian pada metode SVM dan *Decision Tree* menggunakan PCA dengan $n=5$ ditunjukkan dengan hasil akurasi, presisi, dan *recall*. Hasil pengujian metode SVM ditunjukkan pada Tabel 3, sedangkan hasil pengujian metode *Decision Tree* ditunjukkan pada Tabel 4.

Tabel 4. Hasil Pengujian Metode SVM

<i>Split Data</i> Latih dan Data Uji	Akurasi	Presisi	Recall
90% dan 10%	97.44%	95.24%	94.74%
85% dan 15%	98.28%	96.77%	96.43%
80% dan 20%	98.72%	97.56%	97.37
75% dan 25%	98.97%	98.04%	97.87%
70% dan 30%	98.55%	96.15%	96.64%

Tabel 4 menunjukkan hasil dari pengujian metode SVM, dari lima skema pembagian data latih dan data uji didapatkan hasil terbaik pada skema 75%:25% dengan akurasi tertinggi 98.97%, presisi 98.04%, dan *recall* 97.87%.



Gambar 2. Rancangan Pengujian SVM dengan *RapidMiner*

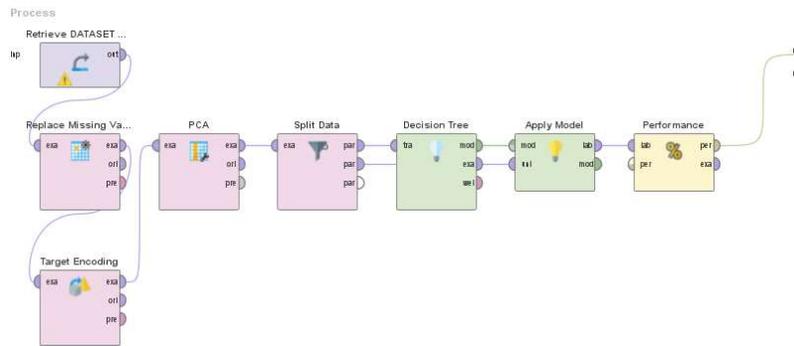
Pada Gambar 2. Merupakan rancangan pengujian algoritma *Support Vector Machine* dengan menggunakan *RapidMiner*. Rancangan pada *RapidMiner* terdiri dari inputan *dataset*, *replace missing value* untuk mengisi data yang kosong, *target encoding* berfungsi untuk menghapus atribut nominal dan mengubahnya menjadi numerik, kemudian penambahan *Principal Component Analysis*, *Split Data* untuk membagi *data training* dan *data testing*, algoritma SVM, *apply model* untuk menerapkan model SVM, dan *performance* untuk mengukur performa atau evaluasi kinerja algoritma yang diterapkan.

Pada Gambar 3 menunjukkan hasil akurasi penerapan algoritma SVM dengan penerapan PCA, dengan *Split Data* 75% dan 25%, menghasilkan akurasi 98.97%.

accuracy: 98.97%

	true 1	true 0	class precision
pred. 1	46	0	100.00%
pred. 0	1	50	98.04%
class recall	97.87%	100.00%	

Gambar 3. Hasil Akurasi SVM



Gambar 4. Rancangan Pengujian SVM dengan *RapidMiner*

Pada Gambar 4. Merupakan rancangan pengujian algoritma *Decision Tree* dengan menggunakan *RapidMiner*.

accuracy: 96.55%

	true 1	true 0	class precision
pred. 1	26	0	100.00%
pred. 0	2	30	93.75%
class recall	92.86%	100.00%	

Gambar 5. Rancangan Pengujian SVM dengan *RapidMiner*

Tabel 5. Hasil Pengujian Metode *Decision Tree*

<i>Split Data</i> Latih dan Data Uji	Akurasi	Presisi	<i>Recall</i>
90% dan 10%	94.87%	90.91%	89.47%
85% dan 15%	96.55%	93.75%	92.98%
80% dan 20%	96.16%	95.12%	94.74%
75% dan 25%	94.58%	93.75%	95.74%
70% dan 30%	96.43%	94.74%	96.43%

Tabel 5 menunjukkan hasil dari pengujian metode *Decision Tree*, dari lima skema pembagian data latih dan data uji didapatkan hasil terbaik pada skema 85%:15% dengan akurasi tertinggi 96.55%, presisi 93.75%, dan *recall* 92.98%.

Pada Gambar 5 menunjukkan hasil akurasi penerapan algoritma *Decision Tree* dengan penerapan PCA, dengan *Split Data* 85% dan 15%, menghasilkan akurasi 96.55%.

4. PEMBAHASAN

Perbandingan performa klasifikasi pada metode SVM dan *Decision Tree* (DT) menggunakan PCA memiliki perbedaan akurasi yang tipis. Berikut pada Tabel 5 merupakan tabel untuk hasil perbandingan kinerja metode SVM dan *Decision Tree*:

Tabel 6. Perbandingan Performa DT dan SVM

Data Split	DT			SVM		
	Akurasi	Presisi	<i>Recall</i>	Akurasi	Presisi	<i>Recall</i>
90% dan 10%	94.87%	90.91%	89.47%	97.44%	95.24%	94.74%
85% dan 15%	96.55%	93.75%	92.98%	98.28%	96.77%	96.43%
80% dan 20%	96.16%	95.12%	94.74%	98.72%	97.56%	97.37
75% dan 25%	94.58%	93.75%	95.74%	98.97%	98.04%	97.87%
70% dan 30%	96.43%	94.74%	96.43%	98.55%	96.15%	96.64%

Tabel 6 menunjukkan hasil perbandingan performa menggunakan *Decision Tree* dan *Support Vector Machine* (SVM) dengan menggunakan PCA. Metode SVM memiliki performa yang lebih tinggi daripada metode *Decision Tree* dengan nilai akurasi

98.97%, dengan penerapan PCA dengan nilai n=5 dan berada pada perbandingan *data split* 75%: 25%.

Metode *Decision Tree* juga memiliki performa yang baik, dengan menggunakan PCA dengan nilai $n=5$, *Decision Tree* memiliki akurasi 96.55% yang berada pada perbandingan *Split Data* 85% : 15%. Hasil dari penelitian sebelumnya menunjukkan bahwa SVM secara konsisten memberikan kinerja yang lebih baik dibandingkan *Decision Tree* [28]–[30][30]

Dalam penelitian [28], SVM dan *Decision Tree* mencapai tingkat akurasi sebesar 97% dengan diakritisasi. Namun, terjadi perubahan akurasi saat menggunakan metode normalisasi, yaitu terjadi peningkatan akurasi pada *Decision Tree* menjadi 98% dan terjadi penurunan akurasi pada SVM menjadi 84%. Temuan ini menunjukkan bahwa efek normalisasi pada akurasi bergantung pada karakteristik *dataset* dan jenis algoritma yang digunakan. Normalisasi dapat membantu meningkatkan akurasi dengan menyamakan skala fitur, mengurangi *overfitting*, atau menyeimbangkan pengaruh fitur. Namun, normalisasi juga dapat menyebabkan penurunan akurasi jika fitur-fitur sudah seimbang atau distribusi data berubah secara signifikan.

Sedangkan hasil penelitian [29] metode SVM mencapai akurasi sebesar 90.20%, sedangkan *Decision Tree* hanya mencapai akurasi sebesar 89.80%. Penelitian [30] yang menggunakan PCA menunjukkan bahwa performa akurasi SVM 98.20% dan 87.30% untuk *Decision Tree*. Perbedaan performa antara SVM dan *Decision Tree* dapat dijelaskan oleh kemampuan SVM dalam menemukan hyperplane optimal yang dapat memisahkan kelas data dengan baik di ruang fitur yang lebih tinggi akibat penerapan PCA. Sementara *Decision Tree*, sebagai metode yang berbasis pada pemisahan berulang pada fitur tertentu, mungkin mengalami kesulitan dalam menangani *dataset* dengan dimensi yang tinggi, terutama setelah dilakukan reduksi dimensi oleh PCA.

Namun, penting untuk dicatat bahwa pada penelitian sebelumnya memiliki beberapa perbedaan dalam konfigurasi eksperimen. Penelitian [28] menggunakan perbandingan *data split* 70% untuk latihan dan 30% untuk pengujian dan penelitian [30] menggunakan perbandingan *data split* 90%:10%, sementara pada penelitian ini menggunakan berbagai perbandingan data split, dengan *data splitter* baik untuk SVM 75%:25% dan *Decision Tree* 85%: 15%. Hasil temuan menyatakan perbandingan *data training* dan *testing* juga mempengaruhi hasil akurasi.

Secara keseluruhan, temuan pada penelitian ini mendukung hasil penelitian sebelumnya yang menyimpulkan bahwa SVM dengan penerapan PCA memiliki performa yang lebih baik dibandingkan *Decision Tree* dalam *dataset* ini. Namun, penting untuk selalu mempertimbangkan berbagai faktor dan memvalidasi hasil dengan cermat sebelum mengambil keputusan berdasarkan analisis klasifikasi ini.

5. KESIMPULAN

Perkembangan teknologi dapat memudahkan untuk mendapatkan informasi tentang tempat wisata. Dengan adanya informasi mengenai pariwisata di internet, dapat memudahkan wisatawan untuk mencari informasi. Namun, karena jumlah informasi yang sangat banyak akan membuat wisatawan kebingungan untuk menentukan tujuan wisata. Selain itu, wisata lokal memiliki potensi yang cukup tinggi untuk membantu perekonomian daerah, namun saat ini belum dieksplorasi secara maksimal.

<https://doi.org/10.25077/TEKNOSI.v9i2.2023.113-121>

Pengguna juga membutuhkan sistem yang responsif dan iteratif yang mampu membuat rekomendasi berdasarkan kebutuhan pribadi wisatawan. Sistem dengan menerapkan proses klasifikasi dapat menjadi salah satu alternatif untuk mempermudah dan mempercepat analisis dalam mengambil keputusan merekomendasikan tempat wisata.

Proses klasifikasi rekomendasi tempat wisata dalam penelitian ini dilakukan dengan metode *Support Vector Machine* (SVM) dan *Decision Tree* (DT). Hasil pengujian keseluruhan, didapatkan hasil perbandingan performa menggunakan *Decision Tree* dan *Support Vector Machine* (SVM) dengan menggunakan PCA. Metode SVM memiliki performa yang lebih baik daripada metode *Decision Tree* dengan nilai akurasi 98.97%, dengan penerapan PCA dengan nilai $n=5$ dan berada pada perbandingan *data split* 75%: 25%. Metode *Decision Tree* juga memiliki performa yang baik, dengan menggunakan PCA dengan nilai $n=5$, *Decision Tree* memiliki akurasi 96.55% yang berada pada perbandingan *Split Data* 85%: 15%.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik, “Jumlah Kunjungan Wisman ke Indonesia Oktober 2022,” Badan Pusat Statistik . Accessed: Jan. 30, 2023. [Online]. Available: <https://www.bps.go.id/pressrelease/2022/12/01/1879/jumlah-kunjungan-wisman-ke-indonesia-pada-oktober-2022-mencapai-678-53-ribu-kunjungan-dan-jumlah-penumpang-angkutan-udara-domestik-pada-oktober-2022-naik-10-08-persen.html#:~:text=Selama%20Januari%20Oktober%202022%2C%20jumlah,periode%20yang%20sama%20tahun%202021.>
- [2] Bappeda DI Yogyakarta, “Jumlah Wisatawan,” Dataku Bappeda DI Yogyakarta. Accessed: Jan. 30, 2023. [Online]. Available: http://bappeda.jogjaprov.go.id/dataku/data_dasar/index/214-jenis-objek-wisata?id_skpd=23
- [3] R. A. Hamid *et al.*, “How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management,” *Comput Sci Rev*, vol. 39, no. 1, p. 100337, Feb. 2021, doi: [10.1016/j.cosrev.2020.100337](https://doi.org/10.1016/j.cosrev.2020.100337).
- [4] I. P. G. H. Saputra and L. A. A. R. Putri, “Sistem Rekomendasi Perjalanan Wisata Berbasis Expert System dan Algoritma Genetika,” *Jurnal Resistor*, vol. 5, no. 1, pp. 47–54, Apr. 2022, doi: [10.31598/jurnalresistor.v5i1.761](https://doi.org/10.31598/jurnalresistor.v5i1.761).
- [5] S. Sauda and M. R. Ramadhan, “Implementasi Memory-Based dan Model-Based Collaborative Filtering pada Sistem Rekomendasi Sepeda Gunung,” *Syntax Literate: Jurnal Ilmiah Indonesia*, vol. 7, no. 6, pp. 7836–7848, Jun. 2022, doi: [10.36418/syntax-literate.v7i6.7559](https://doi.org/10.36418/syntax-literate.v7i6.7559).
- [6] Murniyati, T. Yusnitasari, E. Kurniasari, and D. Pernadi, “Clustering Relationship Berdasarkan Bobot Pembentuk Social Trust Network Untuk Sistem Rekomendasi Pada Media Sosial Instagram,” *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 3, pp. 216–228, Dec. 2022, doi: [10.35760/ik.2022.v27i3.7557](https://doi.org/10.35760/ik.2022.v27i3.7557).

- [7] D. A. Pratiwi and A. Qoiriah, "Sistem Rekomendasi Wedding Organizer Menggunakan Metode Content-Based Filtering Dengan Algoritma Random Forest Regression," *Journal of Informatics and Computer Science (JINACS)*, vol. 3, no. 03, pp. 231–239, Dec. 2021, doi: [10.26740/jinacs.v3n03.p231-239](https://doi.org/10.26740/jinacs.v3n03.p231-239).
- [8] O. Somantri, W. E. Nugroho, and A. R. Supriyono, "Penerapan Feature Selection Pada Algoritma Decision Tree Untuk Menentukan Pola Rekomendasi Dini Konseling," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, pp. 272–279, Dec. 2022, doi: [10.30865/json.v4i2.5267](https://doi.org/10.30865/json.v4i2.5267).
- [9] Erfin Nur Rohma, "Comparison of the Social Welfare Data Classification Algorithm for Bantul Regency," *Jurnal PROCESSOR*, vol. 17, no. 2, pp. 91–100, Oct. 2022, doi: [10.33998/processor.2022.17.2.1222](https://doi.org/10.33998/processor.2022.17.2.1222).
- [10] Y. Anggraiwan and B. Siregar, "Klasifikasi Harga Mobil Menggunakan Metode Decision Tree Algoritma C4.5," *Computatio: Journal of Computer Science and Information Systems*, vol. 6, no. 2, pp. 70–79, Dec. 2022, doi: [10.24912/computatio.v6i2.19994](https://doi.org/10.24912/computatio.v6i2.19994).
- [11] J. Ma'sum, A. Febriani, and D. Rachmawaty, "Penerapan Metode Klasifikasi Decision Tree Untuk Memprediksi Kelulusan Tepat Waktu," *Journal of Industrial Engineering and Technology*, vol. 2, no. 1, pp. 61–74, Jan. 2022, doi: [10.24176/jointtech.v2i1.7432](https://doi.org/10.24176/jointtech.v2i1.7432).
- [12] H. Paul, A. S. Wiguna, and H. Santoso, "Penerapan Algoritma Support Vector Machine Dan Naive Bayes Untuk Klasifikasi Jenis Mobil Terlaris Berdasarkan Produksi Di Indonesia," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 1, pp. 39–44, Jan. 2023, doi: [10.36040/jati.v7i1.5555](https://doi.org/10.36040/jati.v7i1.5555).
- [13] A. Achmad, A. Adnan, and M. Rijal, "Klasifikasi Penyakit Pernapasan berbasis Visualisasi Suara menggunakan Metode Support Vector Machine," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 78–83, Oct. 2022, doi: [10.47970/siskom-kb.v6i1.330](https://doi.org/10.47970/siskom-kb.v6i1.330).
- [14] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 205–215, Dec. 2022, doi: [10.23887/janapati.v11i3.49450](https://doi.org/10.23887/janapati.v11i3.49450).
- [15] H. Mukhtar, R. Gunawan, A. Hariyanto, Syahril, and Wide Mulyana, "Peramalan Kedatangan Wisatawan ke Suatu Negara Menggunakan Metode Support Vector Machine (SVM)," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 274–282, Dec. 2022, doi: [10.37859/coscitech.v3i3.4211](https://doi.org/10.37859/coscitech.v3i3.4211).
- [16] Y. A. Singgalen, "Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1634–1646, Dec. 2022.
- [17] A. Munir, E. P. Atika, and A. D. Indraswari, "Analisis Sentimen pada review hotel menggunakan metode pembobotan dan klasifikasi," *JNANALOKA*, vol. 3, no. 1, pp. 33–38, Mar. 2022, doi: [10.36802/jnanaloka.2022.v3-no1-33-38](https://doi.org/10.36802/jnanaloka.2022.v3-no1-33-38).
- [18] Ardi Ramdani, Christian Dwi Sofyan, Fauzi Ramdani, Muhamad Fauzi Arya Tama, and Muhammad Angga Rachmatsyah, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat Dalam Menerima Bantuan Sosial," *Jurnal Ilmiah Sistem Informasi*, vol. 1, no. 2, pp. 39–47, Jul. 2022, doi: [10.51903/juisi.v1i2.363](https://doi.org/10.51903/juisi.v1i2.363).
- [19] M. Raffif, "Yogyakarta Tourism Place," Kaggle. Accessed: Jan. 05, 2023. [Online]. Available: <https://www.kaggle.com/datasets/mraffif/yogyakarta-tourism-place>
- [20] M. Choirunnisa, N. Hidayat, and E. Susanto, "Implementasi Metode Support Vector Machine Dengan Query Expansion Pada Klasifikasi Review Di Situs Traveloka," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 5, pp. 1860–1865, May 2021, Accessed: Jan. 31, 2023. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9036>
- [21] S. Widaningsih, "Penerapan Data Mining untuk Memprediksi Siswa Berprestasi dengan Menggunakan Algoritma K Nearest Neighbor," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 3, pp. 2598–2611, Sep. 2022, doi: [10.35957/jatisi.v9i3.859](https://doi.org/10.35957/jatisi.v9i3.859).
- [22] B. N. Azmi, A. Hermawan, and D. Avianto, "Analisis Pengaruh PCA Pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression," *JUSTINDO: Jurnal Sistem dan Teknologi Informasi*, vol. 7, no. 2, pp. 94–103, Aug. 2022, doi: [10.32528/justindo.v7i2.8190](https://doi.org/10.32528/justindo.v7i2.8190).
- [23] A. Dinanti and J. Purwadi, "Analisis Performa Algoritma K-Nearest Neighbor dan Reduksi Dimensi Menggunakan Principal Component Analysis," *Jambura Journal of Mathematics*, vol. 5, no. 1, pp. 155–165, Feb. 2023, doi: [10.34312/jjom.v5i1.17098](https://doi.org/10.34312/jjom.v5i1.17098).
- [24] M. F. Azhari and F. A. Fitriani, "Coronary Heart Disease Risk Prediction Using Binary Logistic Regression Based on Principal Component Analysis," *Enthusiastic: International Journal of Applied Statistics and Data Science*, vol. 2, no. 1, pp. 47–55, May 2022, doi: [10.20885/enthusiastic.vol2.iss1.art6](https://doi.org/10.20885/enthusiastic.vol2.iss1.art6).
- [25] J. Kusuma, A. Jinan, M. Z. Lubis, R. Rubianto, and R. Rosnelly, "Komparasi Algoritma Support Vector Machine Dan Naive Bayes Pada Klasifikasi Ras Kucing," *Generic: Jurnal Ilmu Komputer dan Teknologi Informasi*, vol. 14, no. 1, pp. 8–12, Jan. 2022, Accessed: Jan. 31, 2023. [Online]. Available: <http://generic.ilkom.unsri.ac.id/index.php/generic/article/view/122>
- [26] D. E. Safitri and A. S. Fitriani, "Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Kernel Gaussian Rbf Untuk Prediksi Partisipasi Pemilu Terhadap Demografi Kota Surabaya," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 1, p. 36, Jun. 2022, doi: [10.21927/ijubi.v5i1.2259](https://doi.org/10.21927/ijubi.v5i1.2259).
- [27] I. Arfyanti, M. Fahmi, and P. Adytia, "Penerapan Algoritma Decision Tree Untuk Penentuan Pola Penerima Beasiswa KIP Kuliah," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1196–1201, Dec. 2022, doi: [10.47065/bits.v4i3.2275](https://doi.org/10.47065/bits.v4i3.2275).

- [28] M. Wibowo and R. Ramadhani, "Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan," *Jurnal Media Informatika Budidarma*, vol. 5, no. 3, pp. 913–921, Jul. 2021, doi: [10.30865/mib.v5i3.3086](https://doi.org/10.30865/mib.v5i3.3086).
- [29] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *Journal of Information System Management (JOISM)*, vol. 3, no. 1, pp. 1–7, Jan. 2021, doi: [10.24076/JOISM.2021v3i1.341](https://doi.org/10.24076/JOISM.2021v3i1.341).
- [30] M. Mayasari, D. I. Mulyana, and M. B. Yel, "Komparasi Klasifikasi Jenis Tanaman Rimpang Menggunakan Principal Component Analisis, Support Vector Machine, K-Nearest Neighbor Dan Decision Tree," *Jurnal Teknik Informatika Kaputama(JTIK)*, vol. 6, no. 2, pp. 644–655, Jul. 2022, Accessed: Jul. 24, 2023. [Online]. Available: <https://jurnal-backup.kaputama.ac.id/index.php/JTIK/article/view/878>



Dr. Ir. Rianto, M. Eng

Dilahirkan di Imogiri, Daerah Istimewa Yogyakarta 50 tahun yang lalu. Dosen di Program Studi Sains Data Universitas Teknologi Yogyakarta yang menyelesaikan program Doktor bidang Teknologi Informasi di Universitas Gunadarma Jakarta ini memiliki minat riset pada bidang *Artificial Intelligence* khususnya *Natural Language Processing*. Beberapa artikelnya sudah diterbitkan di jurnal internasional maupun nasional bereputasi. Selain mengajar pada Program Studi Sains Data, Magister Teknologi Informasi Universitas Teknologi Yogyakarta juga menggunakan jasanya untuk mengampu mata kuliah *Information System Security*, *Recommender System*, dan *Decision Support System*.

NOMENKLATUR

Y_i	<i>Eigen vector</i>
a_i	<i>Eigen value</i>
x_i	<i>Observation data</i>
p	<i>Observation variable</i>
$x_i \ x$	Data pada <i>data training</i>
σ	Sigma
C	<i>Complexity</i> ,
d	<i>Degree</i>
$-\ x_i - x\ ^2$	Kuadrat jarak vektor x_i dan x .
S	Himpunan Kasus
n	Jumlah partisi S
p_i	Probabilitas
A	Atribut
n	Jumlah partisi S
$ S_i $	Jumlah partisi ke- i
$ S $	Jumlah kasus dalam S
TP	<i>True Positif</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
TN	<i>True Negative</i>

BIODATA PENULIS



Rian Oktafiani, S.Kom
Mahasiswa program studi Magister Teknologi Informasi Universitas Teknologi Yogyakarta, yang menyelesaikan program Sarjana bidang Sistem Informasi di Universitas Teknologi Yogyakarta ini memiliki minat riset pada bidang *data mining*, *recommender system*, dan *information system*.