

TEKNIK DATA MINING UNTUK MENGLASIFIKASIKAN DATA ULASAN DESTINASI WISATA MENGGUNAKAN REDUKSI DATA *PRINCIPAL COMPONENT ANALYSIS (PCA)*

Alven Safik Ritonga¹, Isnaini Muhandhis²

¹ Universitas Wijaya Putra
Surabaya, Indonesia
alvensafik@uwp.ac.id

Abstrak

Peningkatan kunjungan wisatawan ke suatu destinasi wisata, dipengaruhi oleh kepuasan wisatawan waktu berkunjung. Untuk mengetahui suatu destinasi pariwisata sudah sesuai dengan yang diharapkan wisatawan, perlu dilakukan klasifikasi terhadap kepuasan wisatawan. Metode yang digunakan untuk pengklasifikasian ulasan kepuasan destinasi wisata adalah kombinasi antara metode *Principal Component Analysis (PCA)* sebagai metode reduksi dimensi data, dengan tiga metode data mining berikut ini; Support Vector Machine (SVM), Jaringan Saraf Tiruan (JST), dan Decision Trees. Tujuan penelitian ini adalah mendapatkan model klasifikasi yang mempunyai akurasi tinggi dalam melakukan klasifikasi ulasan kepuasan destinasi wisata. Penelitian ini menggunakan data kedua yang diambil dari *UCI Machine Learning Repository*. Hasil pengujian dengan mengkombinasikan PCA pada ketiga metode memperlihatkan bahwa akurasi klasifikasi lebih baik untuk beberapa metode. Dari ketiga metode yang dipakai, SVM-PCA mempunyai akurasi yang lebih baik dengan 91,50% disusul oleh metode ANN-PCA sebesar 89,46% dan metode Decision-PCA sebesar 88,78%.

Kata Kunci: *Data Mining, Klasifikasi, PCA, Tujuan Wisata.*

Abstract

The increase in tourist visits to a tourist destination is influenced by the satisfaction of tourists when visiting. To find out that a tourism destination is in accordance with what tourists expect, it is necessary to classify tourist satisfaction. The method used for classifying tourist destination satisfaction reviews is a combination of the Principal Component Analysis (PCA) method as a method of reducing data dimensions, with the following three data mining methods; Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Decision Trees. The purpose of this study is to obtain a classification model that has high accuracy in classifying tourist destination satisfaction reviews. This study uses second data taken from the UCI Machine Learning Repository. The test results by combining PCA in the three methods show that the classification accuracy is better for several methods. Of the three methods used, SVM-PCA has better accuracy with 91.50%, followed by ANN-PCA method at 89.46% and Decision-PCA method at 88.78%.

Keywords: *Data Mining, Classification, PCA, Tourist Destinations*

PENDAHULUAN

Sektor pariwisata nasional sangat penting karena perkembangan dan kontribusi yang diberikan sektor pariwisata melalui penerimaan devisa, pendapatan daerah, pengembangan wilayah dan penyerapan investasi. Sektor pariwisata sebagai sektor yang penting untuk dikembangkan dan dibina secara sinergi sebagai sektor unggulan. Melalui pendekatan pariwisata berkelanjutan (*sustainable tourism*) perlu sinergi antara upaya pelestarian alam dan budaya beserta warisannya untuk menunjang percepatan pertumbuhan ekonomi. Arah pembangunan pariwisata nasional dalam pasal 2 ayat 8 menjadi arah kebijakan strategi dan indikasi program pembangunan kepariwisataan nasional dalam kurun waktu tahun 2010 sampai dengan 2025 (Peraturan Pemerintah RI no. 50 thn 2015 tentang Rencana Induk Pembangunan Kepariwisata Nasional Tahun 2010 – 2025). Salah satunya adalah mendorong tingkat pertumbuhan kedatangan wisatawan mancanegara dengan cara mengembangkan destinasi pariwisata (*Analisis Kesiapan Destinasi Dalam Rangka Pencapaian Target 20 Juta Wisman Pada Tahun 2019*).

Suatu destinasi harus memiliki berbagai fasilitas kebutuhan yang diperlukan oleh wisatawan agar kunjungan seorang wisatawan dapat terpenuhi dan merasa nyaman. Berbagai kebutuhan wisatawan tersebut antara lain, fasilitas transportasi, akomodasi, biro perjalanan, atraksi (kebudayaan, rekreasi, dan hiburan), pelayanan makanan, dan barang-barang cinderamata (Pitana, 2005). Tersedianya berbagai fasilitas kebutuhan yang diperlukan akan membuat wisatawan merasa nyaman, sehingga semakin banyak

wisatawan yang berkunjung. Ulasan wisatawan dapat diketahui dengan menggunakan platform informasi wisata seperti Tripadvisor. Pada platform ini bisa dilihat ratusan juta ulasan dan opini tentang akomodasi, restoran, pengalaman, maskapai penerbangan, dan atraksi (kebudayaan, rekreasi, dan hiburan).

Untuk mengetahui suatu destinasi pariwisata sudah sesuai dengan yang diharapkan wisatawan, perlu dilakukan evaluasi terhadap kepuasan wisatawan. Karena ulasan kepuasan merupakan salah satu faktor yang dapat menentukan keputusan seorang wisatawan untuk pergi ke suatu tempat wisata, menunjukkan bahwa wisatawan dapat mengambil jumlah ulasan sebagai indikator popularitas tempat wisata atau nilai dari suatu tempat wisata yang akan mempengaruhi kemauan untuk berkunjung ke tempat tersebut.

Ada beberapa studi yang berfokus pada sistem pendukung pengambilan keputusan untuk membantu mengevaluasi kepuasan wisatawan. Ada banyak teknik data mining yang telah digunakan untuk memodelkan klasifikasi data ulasan kepuasan wisatawan. Dalam beberapa dekade terakhir, metode pembelajaran yang diawasi (*supervised learning*) menarik lebih banyak perhatian dari banyak peneliti. Di antaranya Jaringan Saraf Tiruan (JST), *Support Vector Machines* (SVM), dan *Decision Trees*. Dalam proses klasifikasi data ulasan destinasi wisata, data yang tidak penting atau data yang berlebihan harus direduksi dari data set. Jika tidak direduksi, kinerja klasifikasi menjadi bermasalah pada data dengan dimensi tinggi berupa menurunnya akurasi klasifikasi. Banyak metode yang bisa digunakan untuk reduksi dimensi. Salah satunya adalah metode

Principal Component Analysis (PCA). PCA dapat mereduksi dimensi data yang tinggi menjadi dimensi data yang lebih rendah dengan resiko kehilangan informasi yang sangat kecil. Penelitian ini akan mengembangkan teknik Data Mining untuk mengklasifikasi data ulasan wisatawan mengenai destinasi wisata dengan menggabungkan metode *Principal Component Analysis (PCA)*.

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola, atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Santosa,2007). Data mining memiliki peranan yang sangat penting dalam beberapa bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, dan teknologi. Teknik data mining yang berbeda digunakan untuk klasifikasi data ulasan destinasi wisata, seperti menerapkan *Support Vector Machine (SVM)*. Untuk pembandingan peneliti menggunakan Jaringan Saraf Tiruan (JST) dan *Decision Tree*, kedua algoritma tersebut membangun model menggunakan dataset training kemudian menguji model dengan menggunakan dataset testing (Higa,2018).

Principal Component Analysis (PCA) adalah salah satu teknik paling populer untuk pengurangan dimensi. Ini adalah teknik statistik standar yang dapat digunakan untuk mengurangi dimensi suatu set data. PCA adalah metode analisis multivariat berdasarkan transformasi linear yang sering digunakan untuk mengurangi dimensi data, untuk mengambil informasi penting dari data besar, dan untuk

menganalisis struktur variabel (Rahman dkk, 2020).

Problem dalam PCA adalah menemukan eigenvalue dan eigenvector. PCA adalah transformasi orthogonal (tegak lurus) dari sistem koordinat dimana kita mendeskripsikan data. Koordinat baru merepresentasikan data dinamakan principal component (PC). Koordinat ini dipilih dimana variansi dari data mencapai maksimum. Jika data dalam dimensi asli sulit untuk direpresentasikan melalui grafik, maka dengan dua PC atau satu PC bisa digambarkan melalui grafik. Diberikan X adalah data input berukuran $m \times n$ dimana m adalah banyaknya variabel data dan n banyaknya data set. Terlebih dahulu hitung nilai rata-rata (mean) dari X data input, dengan menggunakan rumus (1) berikut ini.

$$\bar{X} = \frac{\sum_{i=1}^m X_i}{m} \quad (1)$$

Kemudian tentukan matriks kovarians, berikut ini.

$$C_Y = \frac{1}{m} Y Y^T \quad (2)$$

$$Y = (X - \bar{X})$$

Menentukan nilai eigen dan vektor eigen dari C_Y . Nilai eigen yang diperoleh lalu diurutkan berdasarkan nilai terbesar sedemikian $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n$, nilai eigen yang dipilih adalah yang terbesar. Vektor eigen yang terkait dimasukkan ke dalam matriks transformasi W . Transformasi dari data set dilakukan dengan memproyeksikan data asli X , ke dalam sub-ruang Y sebesar n dimensi, dengan persamaan berikut,

$$Y = W^T X \quad (3)$$

Jumlah dari *principal component* ditentukan jumlah nilai eigen yang dipilih.

Misalnya jika kita hanya menginginkan satu *principal component* yang mengandung variabilitas paling banyak, kita memilih yang memiliki nilai eigen terbesar. Memilih jumlah nilai eigen yang optimal bisa dibilang merupakan suatu seni dari sains, (Irawan, 2017).

METODE PENELITIAN

Data set yang digunakan diperoleh dari website UCI (University of California, Irvine) ulasan wisatawan mengenai destinasi wisata. Jumlah data sebanyak 980 data, dibagi dua yaitu; 70% sebanyak 686 data training dan 30% sebanyak 295 data testing. Jumlah atribut data adalah 11, terdiri dari:

- a. Id Pengguna (user)
- b. Rating rata-rata tiap pengguna pada Galeri Seni. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- c. Rating rata-rata tiap pengguna pada Klub Dansa. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- d. Rating rata-rata tiap pengguna pada Juice Bars. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- e. Rating rata-rata tiap pengguna pada Restoran. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- f. Rating rata-rata tiap pengguna pada Museum. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- g. Rating rata-rata tiap pengguna pada Resorts. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).

- h. Rating rata-rata tiap pengguna pada Taman/Tempat Piknik. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- i. Rating rata-rata tiap pengguna pada Pantai. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- j. Rating rata-rata tiap pengguna pada Bioskop. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).
- k. Rating rata-rata tiap pengguna pada Tempat Ibadah. Dengan nilai, Excellent (4), Very Good (3), Average (2), Poor (1), Terrible (0).

Tahap Preprocessing Data yang dilakukan yaitu:

1. Normalisasi Data

Karena data-data pada setiap atribut mempunyai jangkauan nilai yang berbeda, maka dilakukan pra-pemrosesan yaitu normalisasi. Sebelum dilakukan proses pengujian, semua data pada setiap atribut dilakukan normalisasi sehingga menjadi dalam selang [0,1] (Nasution dkk, 2019).

Normalisasi yang digunakan adalah rumus Min-max normalisasi (Han dkk, 2012). Misalkan data input X dengan jumlah data n, x_1, x_2, \dots, x_n

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \cdot (b_{\max} - b_{\min}) + b_{\min} \quad (4)$$

Dengan;

x_i' = data input

x_{\min} = data x minimum

x_{\max} = data x maksimum

b_{\min} = batas bawah interval

b_{\max} = batas atas interval

2. Reduksi Data

Data yang dipakai pada penelitian ini dimensinya cukup besar, hal ini nantinya membuat waktu komputasi untuk pengklasifikasian makin lama, membuat analisis tidak praktis atau tidak layak. Teknik reduksi data diterapkan pada penelitian ini untuk mendapatkan dimensi data yang jauh lebih kecil, namun tetap mempertahankan integritas data asli. Pada penelitian ini untuk mereduksi data dipakai metode *Principal Component Analysis* (PCA).

Penelitian ini menggunakan tiga metode pengklasifikasi, yaitu; Jaringan Saraf Tiruan (ANN), *Support Vector Machine* (SVM), dan *Decision Tree*. Langkah-langkah pengklasifikasian, sebagai berikut:

1. Pelatihan (*training*)

Data *training* 70% dari data set digunakan untuk proses *training*, tujuannya untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi. Proses *training* menggunakan 3 metode, yaitu; Jaringan Saraf Tiruan (JST), *Support Vector Machine* (SVM), dan *Decision Tree*.

2. Pengujian (*testing*)

Data *testing* 30% dari data set digunakan untuk menguji apakah fungsi atau model sudah cukup bagus performansinya ketika diterapkan untuk mengklasifikasi. Jika performansinya pada saat *testing* belum memenuhi harapan, bisa mengatur nilai

parameter model untuk mendapatkan model dengan performansi yang lebih baik.

3. Analisis

Dalam penelitian ini menggunakan aplikasi simulasi yaitu Rapidminer versi 9.6, data set yang sudah dibagi dua bagian yaitu data training dan data testing dimasukkan kedalam basis data yang berbeda pada Rapidminer.

4. Evaluasi dan Validasi

Dilakukan untuk mengukur tingkat akurasi data ulasan destinasi wisata. Dengan menggunakan Confussion Matrix untuk menghitung tingkat akurasi, seperti ditunjukkan pada Tabel 1.

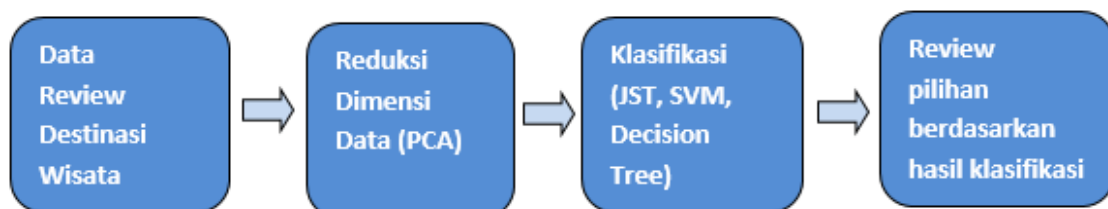
Tabel 1. Tabel confusion matrix

	Prediksi: NO	Prediksi: YES
Aktual: NO	TP	FP
Aktual: YES	FN	TN

Rumus untuk menghitung nilai akurasi berdasarkan Confussion Matrix, berikut ini;

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Alur data penelitian dapat dilihat pada Gambar 1. Dua bagian data set yang di bagi terlebih dahulu melalui proses reduksi atribut dengan menggunakan *Principal Component Analysis* (PCA) akan menghasilkan *Principal Component*. Kemudian data masuk ke dalam Metode Data Mining untuk mendapatkan hasil.



Gambar 1. Diagram blok penggunaan PCA untuk reduksi dimensi dataset dan data mining untuk klasifikasi.

HASIL PENELITIAN DAN PEMBAHASAN

Pertama yang dilakukan untuk menyelesaikan penelitian ini adalah pengolahan data. Data yang dipakai adalah

data sekunder yang diambil dari website UCI (University of California, Irvine) ulasan wisatawan mengenai destinasi wisata. Tabel 2 menunjukkan dataset ulasan wisatawan pada penelitian ini.

Tabel 2. Dataset ulasan wisatawan mengenai detinasi wisata

UserID	Galeri Seni	Klub Dansa	Juice Bar	Restoran	Museum	Resort	Taman	Pantai	Bioskop	T. Ibadah	Persepsi Label
User 1	0,93	1,8	2,29	0,62	0,8	2,42	3,19	2,79	1,82	2,42	PUAS
User 2	1,02	2,2	2,66	0,64	1,42	3,18	3,21	2,63	1,86	2,32	PUAS
User 3	1,22	0,8	0,54	0,53	0,24	1,54	3,18	2,8	1,31	2,5	TIDAK PUAS
User 4	0,45	1,8	0,29	0,57	0,46	1,52	3,18	2,96	1,57	2,86	TIDAK PUAS
User 5	0,51	1,2	1,18	0,57	1,54	2,02	3,18	2,78	1,18	2,54	TIDAK PUAS
User 6	0,99	1,28	0,72	0,27	0,74	1,26	3,17	2,89	1,66	3,66	TIDAK PUAS
User 7	0,9	1,36	0,26	0,32	0,86	1,58	3,17	2,66	1,22	3,22	TIDAK PUAS
User 8	0,74	1,4	0,22	0,41	0,82	1,5	3,17	2,81	1,54	2,88	TIDAK PUAS
User 9	1,12	1,76	1,04	0,64	0,82	2,14	3,18	2,79	1,41	2,54	TIDAK PUAS
User 10	0,7	1,36	0,22	0,26	1,5	1,54	3,17	2,82	2,24	3,12	TIDAK PUAS

Dataset di terdiri dari 11 atribut reguler yaitu tempat-tempat destinasi wisata; Galeri Seni, Klub Dansa, Juice Bar, Restoran, Museum, Resort, Taman, Pantai, Bioskop dan Tempat Ibadah. Rating rata-rata setiap destinasi wisata dinilai dengan Sangat Baik (4), Baik (3), Cukup (2), Buruk (1), Sangat Buruk (0). Terdapat 1 spesial atribut yaitu Persepsi yang terdiri dari 2 kelas yaitu Puas dan Tidak Puas. Deskripsi artibut persepsi ditunjukkan pada Tabel 3. Jumlah dataset yang dipakai adalah 980.

Tabel 3. Deskripsi atribut persepsi

Persepsi	Deskripsi
PUAS	Minimal 5 atribut reguler yang ratingnya; Sangat Baik (4), Baik (3), Cukup(2)
TIDAK PUAS	Minimal 6 atribut reguler yang ratingnya; Buruk (1), Sangat Buruk (0)

Sebelum data digunakan untuk klasifikasi, terlebih dahulu dataset di atas di normalisasi dalam selang [0,1], dataset yang sudah dinormalisasi bisa dilihat pada Tabel 4.

Tabel 4. Normalisasi dataset

User ID	Galeri seni	Klub Dansa	Juice Bar	Restoran	Museum	Resort	Taman	Pantai	Bioskop	T. Ibadah	Persepsi Label
User 1	0,205	0,495	0,619	0,143	0,228	0,630	0,984	0,577	0,444	0,361	PUAS
User 2	0,236	0,604	0,725	0,149	0,420	0,840	1,000	0,465	0,461	0,309	PUAS
User 3	0,306	0,220	0,117	0,116	0,056	0,387	0,976	0,585	0,235	0,402	TIDAK PUAS
User 4	0,038	0,495	0,046	0,128	0,123	0,381	0,976	0,697	0,342	0,588	TIDAK PUAS
User 5	0,059	0,330	0,301	0,128	0,457	0,519	0,976	0,570	0,181	0,423	TIDAK PUAS
User 6	0,226	0,352	0,169	0,036	0,210	0,309	0,968	0,648	0,379	1,000	TIDAK PUAS
User 7	0,194	0,374	0,037	0,052	0,247	0,398	0,968	0,486	0,198	0,773	TIDAK PUAS
User 8	0,139	0,385	0,026	0,079	0,235	0,376	0,968	0,592	0,329	0,598	TIDAK PUAS
User 9	0,271	0,484	0,261	0,149	0,235	0,552	0,976	0,577	0,276	0,423	TIDAK PUAS
User 10	0,125	0,374	0,026	0,033	0,444	0,387	0,968	0,599	0,617	0,722	TIDAK PUAS

Jumlah dataset yang dinormalisasi sebanyak 980. Setelah dinormalisasi dataset di atas bisa langsung dipakai menjadi data penelitian karena datanya sudah dalam bentuk bilangan real. Data penelitian dibagi 2 bagian yaitu 70% data *training* dan 30% data *testing*.

Pada tahap ini data training digunakan untuk proses pelatihan. Proses *training* dan *testing* menggunakan Jaringan Syaraf Tiruan (ANN). Pengaturan parameter untuk Jaringan Syaraf Tiruan digunakan parameter default dari rapidminer yaitu:

- Training cycles =200
- Momentum =0,9
- Decay =no
- Shuffle =yes
- Normalize =yes
- Error epsilon = 10^{-4}

Untuk PCA, *principal component* yang digunakan adalah yang memenuhi 95% dari total variance data. Hasil klasifikasi dengan menggunakan model ANN (8,6,2) bisa dilihat pada Tabel 5 confusion matriks berikut ini.

Tabel 5. Confusion matriks metode ANN-PCA (akurasi=89,46%)

	True PUAS	True TIDAK PUAS	Class precision
Predicted PUAS	24	13	64,86%
Predicted TIDAK PUAS	18	239	93,00%
Class recall	57,14%	94,84%	

Penjelasan Tabel 5 di atas adalah sebagai berikut:

- Persepsi PUAS: prediksi 24 ‘PUAS’ dan 18 ‘TIDAK PUAS’.
- Persepsi TIDAK PUAS: prediksi

239 ‘TIDAK PUAS’ dan 13 ‘PUAS’.

- Keseluruhan prediksi yang dilakukan, akurasi klasifikasi untuk Jaringan Syaraf Tiruan adalah 89,46%.

Proses training dan testing menggunakan metode *Support Vector Machine* (SVM). Pengaturan parameter yang sesuai dengan dataset dilakukan percobaan training dengan menggunakan semua parameter yang ada. Pengaturan parameter metode SVM untuk menghasilkan akurasi klasifikasi lebih bagus, adalah sebagai berikut;

- Tipe kernel : radial
- Kernel gamma : 1.0
- C : 10.0
- Convergence epsilon : 0.001
- Max iterasi : 100000
- $L_{positif}$: 1.0
- $L_{negatif}$: 1.0

Parameter PCA, *principal component* yang digunakan adalah yang memenuhi 95% dari total variance data. Hasil klasifikasi dengan menggunakan metode SVM dikombinasikan dengan PCA bisa dilihat pada Tabel 6 confusion matriks berikut ini.

Tabel 6. Confusion matriks metode SVM-PCA (akurasi=91,50%)

	True PUAS	True TIDAK PUAS	Class precision
Predicted PUAS	21	4	84,00%
Predicted TIDAK PUAS	21	248	92,19%
Class recall	50,00%	98,41%	

Penjelasan Tabel 6 di atas adalah sebagai berikut:

- Persepsi PUAS: prediksi 21 ‘PUAS’ dan 21 ‘TIDAK PUAS’.
- Persepsi TIDAK PUAS: prediksi 248 ‘TIDAK PUAS’ dan 4 ‘PUAS’.
- Keseluruhan prediksi yang dilakukan, akurasi klasifikasi metode SVM adalah 91,50%.

Sama seperti dua metode sebelumnya menggunakan dataset yang sama, untuk klasifikasi selanjutnya menggunakan metode *Decision Tree*. Pengaturan parameter metode *Decision Tree*, menggunakan kriteria *gain_ratio*. Untuk PCA, *principal component* yang digunakan adalah yang memenuhi 95% dari total variance data.

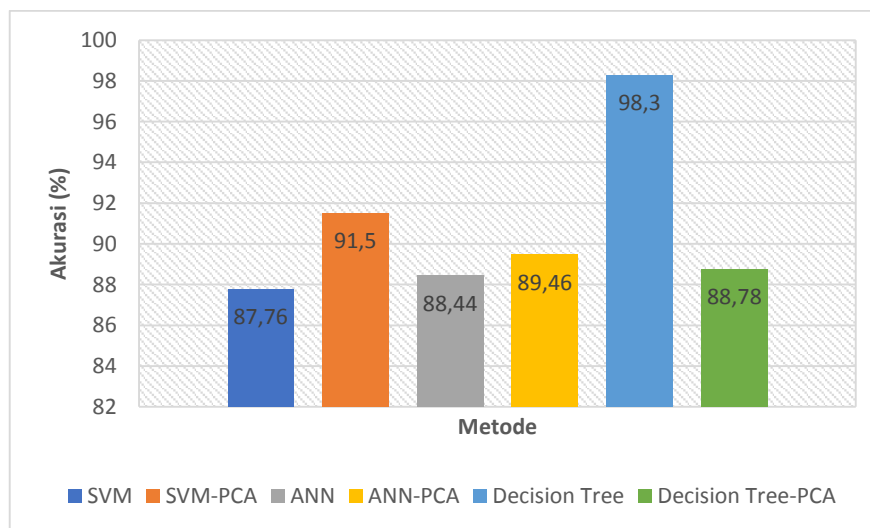
Performa hasil klasifikasi metode *Decision Tree* di kombinasikan dengan PCA, menggunakan *rapidminer*. Dapat dilihat pada Tabel 7 confusion matriks berikut ini.

Tabel 7. Confusion matriks metode *decision tree*-PCA (akurasi=88,78%)

	True PUAS	True TIDAK PUAS	Class precision
Predicted PUAS	23	14	62,16%
Predicted TIDAK PUAS	19	238	92,61%
Class recall	54,76%	94,44%	

Penjelasan Tabel 7 di atas adalah sebagai berikut:

- Persepsi PUAS: prediksi 23 ‘PUAS’ dan 19 ‘TIDAK PUAS’.
- Persepsi TIDAK PUAS: prediksi 238 ‘TIDAK PUAS’ dan 14 ‘PUAS’.
- Keseluruhan prediksi yang dilakukan, akurasi klasifikasi metode *Decision Tree*-PCA adalah 88,78%.



Gambar 2. Grafik perbandingan akurasi klasifikasi data ulasan destinasi wisata

Menggunakan reduksi dimensi dataset dengan metode PCA, memberikan akurasi klasifikasi yang berbeda setiap metode data mining yang diberikan. Tabel 8

dibawah ini memperlihatkan perbandingan performa klasifikasi metode data mining dengan menggunakan PCA dan tanpa menggunakan PCA.

Tabel 8. Metode dan akurasi untuk klasifikasi data ulasan destinasi wisata

Metode	Akurasi (%)
SVM	87,76
SVM-PCA	91,50
ANN	88,44
ANN-PCA	89,46
Decision Tree	98,30
Decision Tree-PCA	88,78

Perbandingan ini juga bisa dilihat pada Gambar 2, yaitu grafik perbandingan nilai akurasi klasifikasi. Dengan melihat dari Tabel 8 dan Gambar 2 akurasi klasifikasi dari ketiga metode, bisa dilihat bahwa dengan mereduksi dataset dengan PCA, metode SVM-PCA yang lebih baik dengan 91,50% disusul oleh metode ANN-PCA sebesar 89,46% dan metode Decision-PCA sebesar 88,78%.

KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat diperoleh suatu kesimpulan berikut ini;

1. Ulasan kepuasan wisatawan terhadap destinasi wisata dapat diklasifikasi dengan menggunakan data mining.
2. Mengurangi dataset menggunakan metode PCA, bisa meningkatkan akurasi klasifikasi untuk beberapa metode data mining.
3. Dari ketiga metode data mining dan dikombinasikan PCA, dihasilkan bahwa metode SVM-PCA mempunyai akurasi yang lebih baik metode SVM-PCA dengan 91,50% disusul oleh metode ANN-PCA sebesar 89,46% dan metode Decision-PCA sebesar 88,78%.

Diharapkan penelitian selanjutnya dibidang-bidang lain dengan menggunakan metode selain ANN, SVM, dan *Decision Tree*.

DAFTAR PUSTAKA

Analisis Kesiapan Destinasi Dalam Rangka Pencapaian Target 20 Juta Wisman Pada Tahun 2019 (n,d). Retrieved, tanggal download 22-03-2020. Kementerian Pariwisata dan Ekonomi Kreatif, <http://www.kemendparekraf.go.id/index.php/post/penelitian>.

Han, J., Kamber,M., dan Pei, J., 2012, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, Waltham.

Higa, A. (2018),” Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms”, *International Journal of Computer Applications Technology and Research*, Vol. 7, No. 1, p. 23-27.

Irawan, G. A., dan Muliantara, A.(2017), “ Prediksi Kesuburan (Fertility) dengan Menggunakan Principal Component Analysis dan Klasifikasi Naive Bayes”, *Jurnal Ilmiah Ilmu Komputer*, Vol. 10, No. 2., p. 10-14.

Nasution, A.D., Khotimah, H.H., dan Chamidah,N.(2019),“Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN”, *Journal of Computer Engineering System and Science*, Vol. 4, No.1., p.78-82.

Pitana, I G. dan Putu G. Gayatri. 2005, *Sosiologi Pariwisata*, Penerbit Andi, Denpasar.

Rahman, MA., Hossain, MF., Hossain, M., dan Ahmmed, R. (2020),”

Employing PCA and t-statistical approach for feature extraction and classification of emotion from multichannel EEG signal”, *Egyptian Informatics Journal, Vol. 21, p.23-35.*

Santosa, B., 2007, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta.