

Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Mancanegara Di Prov. Sulawesi Selatan Dengan K-Means Dan SVM

Nero Caesar Gosari¹, Rismayani²

¹Program Studi Teknik Informatika, Universitas Dipanegara, Makassar

²Program Studi Rekayasa Perangkat Lunak, Universitas Dipanegara, Makassar

^{1,2}Jln. Perintis Kemerdekaan KM 9, Kota Makassar, 90245, Indonesia

email: ¹nerocg03@gmail.com, ²rismayani@undipa.ac.id

Abstract – Indonesia's exchange rate can rise due to foreign tourist visits, which can also benefit the local economy. The provincial capital. South Sulawesi is Makassar which is one of the locations for tourist visits. There are 11 main tourist attractions in Prov. South Sulawesi according to sulselprov 1) Maritime Tourism, 2) Losari Beach, 3) Rotterdam Fort, 4) Somba opu Fort, 5) Takabonerate Marine Park, 6) Bantimurung National Park, 7) Malino, 8) Tanjung Bira Beach, 9) Kesu Tourism, 10) Londa Tourism, 11) Pallawa Tourism. The purpose of this study is to analyze the application of data mining in classifying the number of foreign tourists visiting the prefecture. South Sulawesi uses k-means. The data used comes from BPS Prov. South Sulawesi. The data is grouped into two clusters. That is, the most tourists as C1 with results from Malaysia, and low tourist arrivals as C0 with results from Singapore, Japan, South Korea, Taiwan, China, India, the Philippines, Hong Kong, Thailand, Australia, USA, UK, Netherlands, Germany, France, Russia, Saudi Arabia, Egypt, United Arab Emirates, Pearl of the Persian Gulf, and Switzerland then I use and process this data again with SVM to look for precision, precision and recall values and get 100.00% accuracy in the RapidMiner application.

Abstrak – Nilai tukar Indonesia bisa naik akibat kunjungan turis asing, yang juga bisa menguntungkan ekonomi lokal. Ibu kota Prov. Sulawesi Selatan adalah Makassar yang menjadi salah satu lokasi kunjungan wisatawan. Ada 11 tempat wisata utama di Prov. Sulawesi Selatan sesuai dengan sulselprov 1) Wisata Bahari, 2) Pantai Losari, 3) Benteng Rotterdam, 4) Benteng Somba opu, 5) Taman Laut Takabonerate, 6) Taman Nasional Bantimurung, 7) Malino, 8) Pantai Tanjung Bira, 9) Wisata Kese Kesu, 10) Wisata Londa, 11) Wisata Pallawa. Tujuan dari penelitian ini adalah menganalisis penerapan data mining dalam mengklasifikasikan jumlah wisatawan mancanegara yang berkunjung ke prefektur. Sulawesi Selatan menggunakan k-means. Data yang digunakan berasal dari BPS Prov. Sulawesi Selatan. Data dikelompokkan menjadi dua klaster. Artinya, wisatawan terbanyak sebagai C1 dengan hasil dari Malaysia, dan kunjungan wisatawan rendah sebagai C0 dengan hasil negara Singapura, Jepang, Korsel, Taiwan, Cina, India, Filipina, Hongkong, Thailand, Australia, USA, UK, Belanda, Jerman, Prancis, Rusia, Arab Saudi, Mesir, Persatuan Emirat Arab, Mutiara Teluk Persia, dan Swiss kemudian data tersebut saya gunakan dan diolah kembali dengan SVM untuk mencari nilai presisi, presisi dan recall serta mendapatkan akurasi 100.00% di aplikasi RapidMiner.

*) penulis korespondensi: Nero Caesar Gosari
Email: nerocg03@gmail.com

Kata Kunci – Wisata, Wisman, Sulawesi Selatan, K-means, SVM, RapidMiner.

I. PENDAHULUAN

Turis adalah orang-orang dari seluruh dunia yang datang ke suatu tempat dengan alasan tertentu. Wisatawan adalah bukan pekerja atau wisatawan yang pergi ke tempat lain atas inisiatif sendiri untuk merasakan sesuatu yang baru. Salah satu factor penentu perkembangan pariwisata adalah banyaknya jumlah wisatawan domestic dan mancanegara. [1]

Sulawesi Selatan memiliki potensi wisata yang cukup besar karena latar belakang budaya yang beragam, keragaman topografi dan keindahan alam yang beraneka ragam, serta tawaran pendukung yang cukup menjadikan wilayah Sulawesi Selatan sebagai tujuan wisata. Kegiatan pariwisata ini memberikan pemasukan bagi daerah dan masyarakat yang tentunya akan mempengaruhi pertumbuhan ekonomi Sulawesi Selatan. Pendapatan yg dihasilkan menurut pariwisata bisa diambil menurut pendapatan yg dipungut menurut para pekerja pariwisata & pelaku usaha pariwisata, pendapatan wilayah yg bersumber menurut pajak atau bea cukai barang yg pada impor sang wisatawan & pelaku usaha wisatawan yg berkunjung.

Meskipun jumlah wisman yang berkunjung ke Provinsi Sulawesi Selatan pada tahun 2018 mengalami penurunan, namun beberapa tahun terakhir mengalami peningkatan dan akan terus meningkat pada tahun 2019. Penurunan terjadi pada tahun 2020 sebagai dampak dari meningkatnya infeksi COVID-19 yang disebabkan oleh virus corona. Pandemi COVID-19 pada tahun 2021 mengakibatkan tidak ada kunjungan sama sekali. [2]

Indonesia merupakan salah satu tujuan wisata utama yang memiliki banyak perbedaan terhadap budaya dan keindahan wilayahnya masing-masing yang menjadikannya banyak wisatawan mancanegara tempat liburan, salah satunya Provinsi Sulawesi Selatan yang memiliki banyak tempat wisata. Oleh karena itu, dengan mengelompokkan jumlah wisatawan yang memasuki Prov. Sulawesi Selatan, dapat mengenali negara mana saja yang memiliki jumlah wisatawan terbanyak dan sedikit kedepannya.

Dalam pemrosesan data, ada beberapa jumlah data pada penelitian yang besar dan bervariasi, maka data tersebut diklasifikasi dengan cara melakukan proses pengelompokan. Dengan adanya proses pengelompokan, akurasi data dapat memperbaiki hasil yang diharapkan.

Dilakukan algoritma *K-Means* untuk pembuatan label baru guna memperbaiki proses klasifikasi data yang telah dilakukan sebelumnya agar tingkat akurasi model yang diusulkan menjadi lebih baik.

Setelah mendapatkan label dari proses pengelompokan *K-Means*, kemudian dilakukan proses klasifikasi, yaitu dengan menggunakan *Algoritma Support Machine* (SVM) untuk melakukan proses klasifikasi guna menentukan tingkat data jumlah wisatawan mancanegara dengan akurasi yang baik.

II. PENELITIAN YANG TERKAIT

Berikut ini daftar jurnal yg dipakai menjadi penelitian yg paling relevan:

1. Penelitian yang dilakukan oleh L. Maulida, (2018) memiliki tujuan buat melakukan pengelompokan objek wisata pada Prov. DKI Jakarta, output C1 di Taman Impian Jaya Ancol, C2 di Taman Mini Indonesia Indah dan Kebun Binatang Ragunan, dan C3 di Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Sejarah Jakarta, dan Pelabuhan Sunda Kelapa, output berdasarkan penelitian menjelaskan ada delapan catatan. [3]
2. Penelitian yang dilakukan oleh L. Surtiningsih, M. T. Furqon, and S. Adinugroho (2018), Hal ini diharapkan dapat memprediksi jumlah kunjungan wisatawan ke Bali, namun dalam penelitian menghasilkan tingkat kesalahan (MAPE) sebesar 2,513% tercapai saat mengantisipasi jumlah kunjungan wisatawan asing ke Bali. [4]
3. Penelitian yang dilakukan oleh S. Darma, S. Defit, D. Hartama, W. Robiansyah, and F. Firzada (2020), berfokus pada pencarian menurut tahun 2017-2019 bulan manakah yg mempunyai kunjungan terbanyak, output menurut penelitian menampakan bahwa wisatawan terbanyak terdapat dalam bulan Juni tahun 2017-2019. [5]
4. Penelitian yang dilakukan oleh E. Muningsih, N. Hasan, and G. B. Sulistyio (2020), melakukan penelusuran hasil analisis Davies-Bouldin Index (DBI) untuk data kedatangan wisatawan mancanegara ke Indonesia, dan ditemukan bahwa hasil dan kesimpulan pemodelan K-Means dan PCA secara signifikan lebih tinggi dibandingkan dengan menggunakan K-Means saja. DBI yang sangat baik. [6]
5. Penelitian yang dilakukan oleh I. J. Thira, N. A. Mayangky, D. N. Kholifah, I. Balla, dan W. Gata (2019), Mencari Mean Absolute Percentage Error (MAPE) dari kesalahan maksimum dan minimum untuk memprediksi kunjungan wisman ke Indonesia, menemukan bahwa dengan penggunaan fuzzy time untuk menentukan MAPE menghasilkan kesalahan terbesar sebesar 18,05% dan kesalahan minimum sebesar 0,04%. diperoleh. Mei 2017. [7]
6. Penelitian yang dilakukan oleh R. Rismayani, S. Wahyuni, M. Patasik, I. Iskandar, and S. Asnany (2021), memakai Tour Talk Share disingkat TTS, dengan menggunakan Android pada Prov. Sulsel membentuk system pencarian wisata berbasis mobile android menggunakan mengimplementasikan system wisata Tour Talk Share (TTS), khususnya menambahkan highlight Visit, Talk dan Offer. Dengan hasil jajak pendapat yang menyatakan bahwa menambahkan kerangka sangat membantu 13%, berharga 81%, tidak berharga 6%, dan sangat sia-sia 0%. [8]
7. Penelitian yang dilakukan oleh Z. A. Fatimah and A. Naryoso (2019), Cari kolaborasi antara pemanfaatan akun berita pergerakan @indtravel dan pertukaran informasi informal sehubungan dengan tingginya wisatawan asing yang berkunjung ke Indonesia. Hasil menurut penelitian memperlihatkan bahwa masih ada interaksi yg bertenaga & hubungan positif antara kekuatan akses ke data industri perjalanan pada akun @indtravel dan premi wisatawan asing menggunakan hasil pengujian rumus Kendall's Tau-b menggunakan SPSS yang membuat nilai signifikan 0,000 dan koefisien hubungan 0,497. [9]
8. Penelitian yang dilakukan oleh I. Hadiriyanto and M. Y. Darsyah (2018), ARIMA dan Musim Dingin digunakan untuk memprediksi jumlah turis asing yang berkunjung ke prefektur. Hasil penelitian ini menunjukkan bahwa model ARIMA memenuhi asumsi dan cocok untuk prediksi. Di sisi lain, model musim dingin memiliki alfa 0,2 dan tingkat kesalahan perkiraan atau MAPE 3%. [10]
9. Penelitian yang dilakukan oleh A. Indrasetianingsih and I. Damayanti (2018), Melakukan prediksi banyaknya wisatawan asing yang datang ke Indonesia menggunakan strategi ARIMA dan organisasi otak palsu. Hasil dari penelitian ini menunjukkan bahwa prediksi dengan menggunakan metode ARIMA dapat lebih baik digunakan untuk memprediksi jumlah kunjungan wisatawan asing ke Indonesia, karena memiliki nilai kesalahan MAD dan MSE yang lebih rendah daripada ANN. [11]
10. Penelitian yang dilakukan oleh D. Pramayuda, M. R. Lubis, dan I. S. Saragih (2020), melakukan pengelompokan terhadap wisatawan asing yang memasuki 34 provinsi di Indonesia dari tahun 2014 sampai 2018 menggunakan algoritma K-Means dan RapidMiner yang menghasilkan 1 provinsi mendapatkan urutan tertinggi atau C1 yaitu wisatawan

terbanyak, pada Provinsi Bali dan 33 provinsi lainnya masuk kedalam C2 atau urutan terendah. [12]

11. Penelitian yang dilakukan oleh RW Sari dan D Hartama (2018), dimana melakukan clustering terhadap jumlah tamu asing menurut 34 provinsi, dimana mendapatkan 1 provinsi pada cluster 1, yaitu Provinsi Bali, 2 provinsi pada cluster 2, yaitu Kepulauan Riau dan DKI Jakarta, dan 31 provinsi lainnya pada cluster 3. [13]
12. Penelitian yang dilakukan oleh B. S. Purnomo dan P. T. Prasetyaningrum (2021), melakukan clustering terhadap kunjungan wisatawan Kota Yogyakarta tiap bulannya pada beberapa objek wisata dengan jumlah 48 dataset, yang menghasilkan 24 output pada C1 (tinggi), 11 output pada C2, (sedang), dan 13 output pada C3 (rendah). [14]
13. Penelitian yang dilakukan oleh N. Savitri, R. Pranata, A. N. M. Clara, dan O. S. Rahajeng (2022), melakukan kluserisasi terhadap kunjungan wisata Kabupaten Kulon Progo pada tahun 2019, dan mendapatkan 20 objek wisata yang masuk kedalam kluster 1 (tinggi), 15 objek wisata yang masuk kedalam kluster 2 (sedang), dan 2 objek wisata yang masuk kedalam kluster 3 (rendah). [15]

III. METODE PENELITIAN

A. Data Mining

Arti dasar dari penambangan informasi adalah memisahkan data atau contoh yang signifikan atau menarik dari informasi dalam kumpulan data yang sangat besar. Dalam distribusi logis, penambangan data disinggung sebagai *Knowledge Discovery in a Database* (KDD).[16]

B. Pengambilan Data

Dataset yang dipakai diperoleh dari website BPS yaitu data kunjungan wisatawan mancanegara di wilayah Prov. Sulawesi Selatan dari tahun 2017-2021 berdasarkan BRS pada website <https://sulsel.bps.go.id/searchengine/result.html>. dataset yang diambil merupakan dataset bulanan dari tahun 2017-2021 dengan jumlah 22 negara yang memasuki Prov. Sulawesi Selatan sebagai wisatawan.

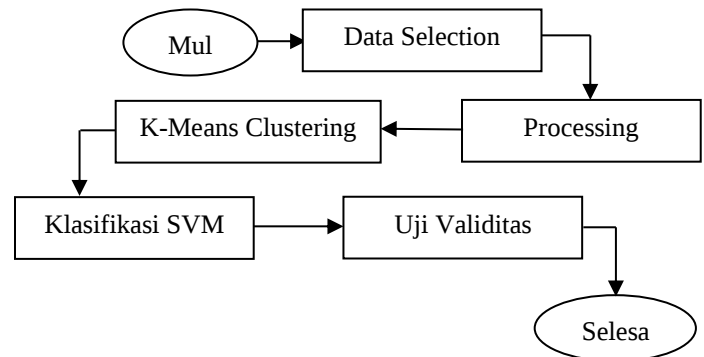
C. Preprocessing

Data Preprocessing adalah metode yang terlibat dengan mengubah informasi menjadi struktur yang lebih mudah, lebih efektif dan lebih berkualitas untuk kebutuhan klien. Penanda yang digunakan sebagai sumber perspektif menggabungkan hasil yang lebih tepat, waktu estimasi yang lebih terbatas, dan

informasi yang lebih sederhana tanpa mengubah data yang dikandungnya. [17]

D. Tahapan Penelitian

Langkah-langkah yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.



Gbr. 1 Flowchart tahapan penelitian

E. K-Means Clustering

Pengelompokkan K-Means adalah cara paling umum untuk mengumpulkan item komparatif ke dalam kumpulan diskrit, atau lebih tepatnya, memisahkan indeks informasi ke dalam himpunan bagian sehingga data di setiap himpunan bagian memiliki kepentingan yang berharga. Ketika sekelompok terdiri dari bermacam-macam item yang mirip satu sama lain dan unik dalam kaitannya dengan objek dalam kelompok yang berbeda. [18]

Proses awal perhitungan k-means dapat dilihat dibawah ini :

1. Menentukan jumlah kluster yang akan dibuat dan tentukan lagi pusat kluster k.
2. Lalu hitung semua pusat kluster dengan jarak Euclidean.

$$d(i, k) = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2}$$

3. Dan mengelompokkan data ke dalam kluster berjarak dekat menggunakan rumus.

$$\min \sum_k^i -a_{ik} = \sqrt{\sum_i^m (C_{ij} - C_{kj})^2}$$

4. Menghitung pusat kluster yang baru dengan rumus.

$$C_{kj} = \frac{\sum_k^i x_{ij}}{p}$$

Dengan : $X_{ij} \in$ Klaster ke k p = banyaknya anggota klaster ke k .

5. Ulang langkah 2-4 sehingga tidak ada data yang akan dipindahkan ke klaster yang lain. [3]

F. SVM

Model perhitungan SVM adalah suatu metode learning supervised dimana dalam suatu tahapan latihan dapat menghadirkan dua parameter antara garis dan bias b . setelah menemukan parameter yang optimal melalui optimasi programan kuadrat (quadratic programming). Jika ukuran masalahnya cukup besar, maka solver program akuadrat biasanya lama. SVM merupakan suatu algoritma strategi urutan yang bekerja untuk pembelajaran dalam suatu pengelompokan masalah yang disebut hyperplane, yang dapat memisahkan dua susunan data dari berbagai kelas. [19]

G. Uji Validitas

Uji validitas merupakan uji yang dimaksudkan untuk memutuskan apakah suatu instrumen penaksir sah atau tidak valid. Instrumen penaksir yang disinggung di sini adalah jajak pendapat. Suatu penelitian dianggap penting jika pertanyaan dalam jajak pendapat dapat mengungkap sesuatu yang diperkirakan oleh survei tersebut. [20]

H. RapidMiner

RapidMiner adalah aplikasi penambangan informasi sumber terbuka utama dan terkemuka. Ini menggabungkan aplikasi terpisah untuk pemeriksaan informasi dan motor penambangan informasi, misalnya, teknik penumpukan informasi, perubahan, tampilan dan persepsi informasi. [21]

IV. HASIL DAN PEMBAHASAN

1. Data Selection

Dalam melakukan seleksi data, data diambil dari BRS pada website BPS yang ada di <https://sulsel.bps.go.id/searchengine/result.html>. data yang diambil berupa data bulanan wisatawan mancanegara yang memasuki Prov. Sulawesi Selatan dari bulan Januari 2017 sampai Desember 2021.

2. Data View

Untuk melihat data yang telah diambil dari BPS dan di proses, maka dapat diakses melalui google spreadsheet berikut :

<https://docs.google.com/spreadsheets/d/1BLmIH6eii5lC2XGNFC2ByzDdkU3j4gh6Z3mRwSAvxc/edit?usp=sharing>.

3. Processing

Data bulanan yang telah didapatkan dari website BPS kemudian disederhanakan menjadi data tahunan dengan cara menjumlahkan data dari bulan Januari-Desember dan diberikan label tahun (2017-2021). Data ini disederhanakan menjadi data tahunan karena jika data tidak disederhanakan maka akan mengalami error pada saat ingin mencari cluster pada RapidMiner, alasan kenapa terjadi error dikarenakan data bulanan digunakan sebagai attribute dan bukan id dan menggunakan tipe data integer bukan tipe data polinomial.

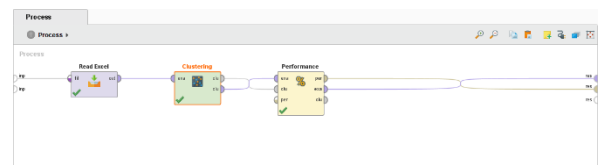
no.	negara	2017	2018	2019	2020	2021
1	singapura	850	693	600	103	0
2	malaysia	9341	7814	10056	2002	0
3	jepang	154	140	240	25	0
4	korea sela	35	1184	69	13	0
5	taiwan	68	26	37	7	0
6	tiongkok	678	239	412	44	0
7	india	331	107	185	22	0
8	philipina	237	75	82	11	0
9	hongkong	58	59	75	0	0
10	thailand	850	133	146	215	0
11	australia	222	197	315	99	0
12	amerika s	412	378	515	93	0
13	inggris	348	240	241	44	0
14	belanda	445	423	280	39	0
15	jerman	547	539	595	109	0
16	perancis	750	702	620	103	0
17	rusia	60	22	736	19	0
18	saudi arab	36	26	29	0	0
19	mesir	236	12	4	0	0
20	uni emira	0	0	2	0	0
21	barhain	0	0	0	0	0
22	switzerlar	0	0	34	38	0

Gbr. 2 Data Wisman

4. Perhitungan K-Means

Untuk mencari cluster data, maka digunakan aplikasi RapidMiner. Design K-Means pada aplikasi RapidMiner, menggunakan 3 operator. Operator pertama adalah operator 'Read Excel', operator ini berfungsi untuk membaca file pada excel. Operator kedua adalah operator 'Clustering', operator ini berfungsi untuk membagi sebuah rangkaian data menjadi beberapa kelompok data berdasarkan kesamaan-kesamaan yang sudah ditentukan sebelumnya. Operator ketiga yaitu 'Performance', Operator ini digunakan untuk menilai kinerja model yang secara otomatis memberikan daftar nilai kriteria kinerja untuk tugas tertentu. Kriteria klasifikasi tersebut misalnya presisi, akurasi, dan daya ingat.

Berikut adalah design operator K-Means pada aplikasi RapidMiner :



Gbr. 3 Design K-Means RapidMiner

Hasil clusterisasi data dari RapidMiner dapat dilihat pada gambar 3.

Row No.	negara	cluster	no.	2017.0	2018.0	2019.0	2020.0	2021.0
1	singapura	cluster_0	1	850	693	600	103	0
2	malaysia	cluster_1	2	9341	7814	10056	2002	0
3	jepang	cluster_0	3	154	140	240	25	0
4	korea selatan	cluster_0	4	35	1184	69	13	0
5	taiwan	cluster_0	5	68	26	37	7	0
6	tingkok	cluster_0	6	678	239	412	44	0
7	india	cluster_0	7	331	107	185	22	0
8	philipina	cluster_0	8	237	75	82	11	0
9	hongkong	cluster_0	9	58	59	75	0	0
10	thailand	cluster_0	10	850	133	146	215	0
11	australia	cluster_0	11	222	197	315	99	0
12	amerika serik.	cluster_0	12	412	378	515	93	0
13	inggris	cluster_0	13	348	240	241	44	0
14	belanda	cluster_0	14	445	423	280	39	0
15	jerman	cluster_0	15	547	539	595	109	0
16	perancis	cluster_0	16	750	702	620	103	0
17	rusia	cluster_0	17	60	22	736	19	0
18	saudi arabia	cluster_0	18	36	26	29	0	0
19	mesir	cluster_0	19	236	12	4	0	0
20	uni emirat arab	cluster_0	20	0	0	2	0	0
21	barhain	cluster_0	21	0	0	0	0	0
22	switzerland	cluster_0	22	0	0	34	38	0

Gbr. 4 Pemetaan K-Means pada RapidMiner

Kemudian hasil cluster yang telah didapatkan pada RapidMiner dimasukkan kedalam table data agar table cluster dapat digunakan kedalam algoritma SVM.

Dan dimasukkan kembali kedalam table menjadi :

no.	negara	2017	2018	2019	2020	2021	cluster
1	singapura	850	693	600	103	0	0
2	malaysia	9341	7814	10056	2002	0	1
3	jepang	154	140	240	25	0	0
4	korea selatan	35	1184	69	13	0	0
5	taiwan	68	26	37	7	0	0
6	tingkok	678	239	412	44	0	0
7	india	331	107	185	22	0	0
8	philipina	237	75	82	11	0	0
9	hongkong	58	59	75	0	0	0
10	thailand	850	133	146	215	0	0
11	australia	222	197	315	99	0	0
12	amerika serikat	412	378	515	93	0	0
13	inggris	348	240	241	44	0	0
14	belanda	445	423	280	39	0	0
15	jerman	547	539	595	109	0	0
16	perancis	750	702	620	103	0	0
17	rusia	60	22	736	19	0	0
18	saudi arabia	36	26	29	0	0	0
19	mesir	236	12	4	0	0	0
20	uni emirat arab	0	0	2	0	0	0
21	barhain	0	0	0	0	0	0
22	switzerland	0	0	34	38	0	0

Gbr. 5 Data Wisman Clustering

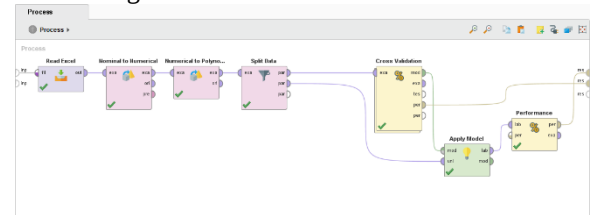
5. SVM

Data yang telah melalui proses K-Means clustering kemudian dibagi menjadi 70% untuk training dan 30% untuk testing.

Pada design SVM di aplikasi RapidMiner, digunakan 8 model operator. Operator pertama merupakan 'Read Excel', operator ini berfungsi untuk membaca file pada excel. operator kedua merupakan 'Nominal to Numerical', operator ini digunakan untuk mengkonversi jenis atribut non-numerik menjadi tipe numerik. Operator ketiga merupakan operator 'Numerical to Polynominal', operator ini digunakan untuk mengubah atribut tipe numerik menjadi tipe polinomial. Operator keempat merupakan operator 'Split Data', digunakan untuk membagi set data menjadi data training dan data test sesuai dengan partisi yang ditentukan. Operator kelima yaitu 'Cross Validation', merupakan nested operator dengan dua thread yaitu training thread yang digunakan untuk melatih model dan test thread untuk menguji dan

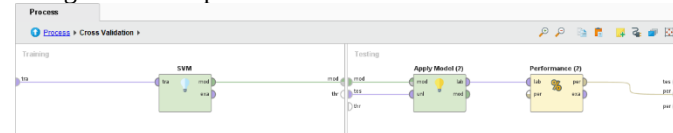
mengukur kinerja model. Operator selanjutnya merupakan operator 'SVM' didalam operator 'Cross Validation', berfungsi melakukan pembelajaran pada SVM. Kemudian digunakan 2 operator 'Apply Model', didalam operator 'Cross Validation dan diluar operator tersebut. Digunakan untuk mengeksekusi model yang sebelumnya dilatih menggunakan data pelatihan dengan data yang tidak berlabel (data testing). Dan operator terakhir adalah operator "performance", digunakan didalam dan diluar 'Cross Validation', Operator ini digunakan untuk menilai kinerja model yang secara otomatis memberikan daftar nilai kriteria kinerja untuk tugas tertentu. Kriteria klasifikasi tersebut misalnya presisi, akurasi, dan recall.

Design SVM yang digunakan pada RapidMiner adalah sebagai berikut :



Gbr. 6 Design SVM RapidMiner

Design didalam operator Cross Validation :



Gbr. 7 Design Operator Cross Validation

Parameter operator SVM yang digunakan adalah seperti table 1.

TABLE I Parameter SVM

PARAMETER	VALUE
Svm type	C-SVC
Kernel type	Rbf
Gamma	0.0
C	0.0
Cache size	80
epsilon	0.001

6. Uji Validitas

Dalam pengujian validitas, dapat diambil hasil dari proses SVM. Dari 22 data yang telah didapatkan, digunakan data sebanyak 6 data untuk mencari akurasi, presisi, dan recall.

aVector (Performance 0)			
Table View			
accuracy: 93.33% <- 14.91% (micro average: 93.75%)			
	True 0	True 1	class precision
pred. 0	15	1	93.75%
pred. 1	0	0	0.00%
class recall	100.00%	0.00%	

Gbr. 8 Hasil Data Pengujian SVM

Atau lebih jelas menggunakan table berikut :

TABLE II Hasil Data Pengujian SVM

Accuracy: 100.00 %

	True 0	True 1	Class precision
Pred. 0	6	0	100.00%

Pred. 1	0	0	0.00%
Class recall	100.00%	0.00%	

V. KESIMPULAN

Dalam melakukan penelitian terkait pengelompokan kunjungan wisman di Prov. Sulawesi Selatan, dapat diterapkan dengan menggunakan metode kluster K-Means, dan menguji apakah data tersebut valid atau tidak menggunakan SVM. Pertama-tama data diambil dari BPS dan diolah untuk memperoleh jumlah kunjungan negar pertahun yang berkunjung ke Prov. Sulawesi Selatan. Data ini diolah menggunakan aplikasi Excel dengan menggunakan rumus SUM dari data awal perbulan dari bulan Januari samapi bulan Desember. Data tersebut kemudian diolah menggunakan K-Means yang dikelompokkan menjadi dua klaster, yaitu klaster tinggi (C1) dan klaster rendah (C0).

Row No.	negara	cluster	no.	2017.0	2018.0	2019.0	2020.0	2021.0
1	singapura	cluster_0	1	850	693	600	103	0
2	malaysia	cluster_1	2	9341	7814	10056	2002	0
3	jepang	cluster_0	3	154	140	240	25	0
4	korea selatan	cluster_0	4	35	1184	69	13	0
5	taiwan	cluster_0	5	68	26	37	7	0
6	hongkong	cluster_0	6	678	239	412	44	0
7	india	cluster_0	7	331	107	185	22	0
8	philipina	cluster_0	8	237	75	82	11	0
9	hongkong	cluster_0	9	58	59	75	0	0
10	thailand	cluster_0	10	850	133	146	215	0
11	australia	cluster_0	11	222	197	315	99	0
12	amerika serik.	cluster_0	12	412	378	515	93	0
13	inggris	cluster_0	13	348	240	241	44	0
14	belanda	cluster_0	14	445	423	280	39	0
15	jerman	cluster_0	15	547	539	595	109	0
16	perancis	cluster_0	16	750	702	620	103	0
17	rusia	cluster_0	17	60	22	736	19	0
18	saudi arabia	cluster_0	18	36	26	29	0	0
19	mesir	cluster_0	19	236	12	4	0	0
20	uni emirat arab	cluster_0	20	0	0	2	0	0
21	barhain	cluster_0	21	0	0	0	0	0
22	switzerland	cluster_0	22	0	0	34	38	0

Gbr. 9 Hasil Cluster RapidMiner K-Means

Jadi hasil yang didapat dari K-Means berarti C1 terdiri dari 1 negara yaitu negara Malaysia, dan C0 terdiri dari 21 negara yaitu Singapura, Jepang, KorSel, Taiwan, China, India, Filipina, Hongkong, Thailand, Australia, USA, UK, Belanda, Jerman, Prancis, Rusia, Arab Saudi, Mesir, Uni Emirat Arab, Barhain dan Swiss menjadi catatan bagi pemerintah Prov. Sulawesi Selatan, yang kemudian dianalisis dengan data SVM dan memberikan hasil dengan akurasi 100.00%, nilai presisi sebesar 100.00%, dan nilai recall sebesar 100.00%. artinya algoritma SVM diperkirakan benar wisatawan mancanegara yang masuk ke Prov. Sulawesi Selatan sebesar 100% yang diperkirakan dengan menggunakan algoritma SVM mendapatkan perkiraan 100% benar 100% dari keseluruhan data wisatawan mancanegara yang masuk ke Prov. Sulawesi Selatan. Dengan akurasi sebesar ini, penggunaan SVM untuk memprediksi pengelompokan data bisa dilakukan dan sangat akurat. Dengan perbaikan sarana dan prasarana destinasi wisata, jumlah wisatawan mancanegara dapat meningkat, yang akan mempengaruhi penerimaan destinasi wisata dan kenaikan nilai tukar.

Saran untuk pengembangan penelitian adalah selanjutnya mengelompokkan data berdasarkan objek wisata yang ada di

Prov. Sulawesi Selatan agar tidak hanya mengetahui negara mana saja yang memiliki jumlah wisatawan tertinggi melainkan juga dapat mengetahui objek wisata mana yang memiliki kunjungan terbanyak.

UCAPAN TERIMA KASIH

Saya berterima kasih kepada pembina utama Ibu Rismayani atas informasi dan pengetahuan yang diberikan kepada saya. Saya juga mengucapkan terima kasih kepada BPS untuk menyediakan informasi yang diperlukan untuk penelitian ini. Saya juga mengucapkan terima kasih kepada orang tua saya yang selalu mendukung saya saat mengerjakan jurnal.

DAFTAR PUSTAKA

- [1] R. E. Pranata, S. P. Sinaga, and A. Wanto, "Estimasi Wisatawan Mancanegara Yang Datang ke Sumatera Utara Menggunakan Jaringan Saraf," *J. Semantik*, vol. 4, no. 1, pp. 97–102, 2018.
- [2] "Badan Pusat Statistik Provinsi Sulawesi Selatan." <https://sulsel.bps.go.id/pressrelease/2017/09/04/287/jumlah-wisatawan-mancanegara-wisman-yang-datang-melalui-pintu-masuk-makassar-pada-juli-2017-mencapai-2-049-kunjungan.html> (accessed Dec. 08, 2022).
- [3] L. Maulida, "Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means," *JISKA J. Inform. Sunan Kalijaga*, vol. 2, no. 3, pp. 167–174, 2018.
- [4] L. Surtiningsih, M. T. Furqon, and S. Adinugroho, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara Ke Bali Menggunakan Support Vector Regression dengan Algoritma Genetika," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput. E-ISSN*, vol. 2548, p. 964X, 2018.
- [5] S. Darma, S. Defit, D. Hartama, W. Robiansyah, and F. Firzada, "Penerapan Metode K-Means Dalam Pengelompokan Jumlah Wisatawan Asing Di Indonesia," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2020, vol. 2, pp. 255–261.
- [6] E. Muningsih, N. Hasan, and G. B. Sulistyoo, "Penerapan Metode Principle Component Analysis (PCA) untuk Clustering Data Kunjungan Wisatawan Mancanegara ke Indonesia," *Bianglala Inform.*, vol. 8, no. 1, pp. 58–62, 2020.
- [7] I. J. Thira, N. A. Mayangky, D. N. Kholifah, I. Balla, and W. Gata, "Peramalan Data Kunjungan Wisatawan Mancanegara ke Indonesia menggunakan Fuzzy Time Series," *J Edukasi Dan Penelit Inf.*, vol. 5, no. 1, 2019.
- [8] R. Rismayani, S. Wahyuni, M. Patasik, I. Iskandar, and S. Asnany, "Implementasi Sistem Pencarian Wisata Menggunakan Tour Talk Share (TTS) Berbasis Android Pada Provinsi Sulawesi Selatan," *Komputika J. Sist. Komput.*, vol. 10, no. 1, pp. 77–84, 2021.
- [9] Z. A. Fatimah and A. Naryoso, "Hubungan Antara Intensitas Mengakses Informasi Pariwisata Akun Instagram@ Indtravel Dan Intensitas Komunikasi Word Of Mouth Dengan Minat Wisatawan Mancanegara Berkunjung Ke Indonesia," *Interak. Online*, vol. 7, no. 1, pp. 87–98, 2018.
- [10] I. Hadiriyanto and M. Y. Darsyah, "Peramalan Jumlah Wisatawan Mancanegara di Provinsi Bali dengan Menggunakan ARIMA dan Winter," in *Prosiding Seminar Nasional Mahasiswa Unimus*, 2018, vol. 1.
- [11] A. Indrasetyaningsih and I. Damayanti, "Prediksi Jumlah Kunjungan Wisatawan Mancanegara di Indonesia dengan Menggunakan Metode ARIMA Box-Jenkins dan Jaringan Syaraf Tiruan," *J Stat. J. Ilm. Teori Dan Apl. Stat.*, vol. 10, no. 2, pp. 7–14, 2018.
- [12] D. Pramayuda, M. R. Lubis, and I. S. Saragih, "Pengelompokan Tamu Asing Ke Indonesia Berdasarkan Provinsi Dengan Algoritma K-Means," *Kesatria J. Penerapan Sist. Inf. Komput. Dan Manaj.*, vol. 1, no. 4, pp. 130–137, 2020.
- [13] R. W. Sari and D. Hartama, "Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi," in

- Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2018, vol. 1, no. 1.
- [14] B. S. Purnomo and P. T. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021.
- [15] N. Savitri, R. Pranata, A. N. M. Clara, and O. S. Rahajeng, "PENGELOMPOKAN KUNJUNGAN WISATA KABUPATEN KULON PROGO TAHUN 2019 MENGGUNAKAN K-MEANS CLUSTERING," *JUST IT J. Sist. Inf. Teknol. Inf. Dan Komput.*, vol. 12, no. 1, 2022.
- [16] A. M. Siregar, S. Kom, M. K. D. A. Puspabhuaana, S. Kom, and M. Kom, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group, 2017.
- [17] S. Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *J-SAKTI J. Sains Komput. Dan Inform.*, vol. 1, no. 2, pp. 180–185, 2017.
- [18] R. A. Indraputra and R. Fitriana, "K-Means Clustering Data COVID-19," *J. Tek. Ind.*, vol. 10, no. 3, pp. 275–282, 2020.
- [19] S. Syamsiah and A. Darmawan, "Analisa Particle Swarm Optimization Terhadap Kepuasan Taman Dadap Merah Dengan Model Svm," in *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, 2020, vol. 4, no. 1.
- [20] N. M. Janna and H. Herianto, "Konsep Uji Validitas dan Reliabilitas dengan Menggunakan SPSS," 2021.
- [21] R. Nofitri and N. Irawati, "Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang," *JURTEKSI J. Teknol. Dan Sist. Inf.*, vol. 6, no. 1, pp. 35–42, 2019.