SINTA Jurnal Sistem Informasi dan Teknologi Komputasi
Volume 1 Nomor 2 April 2024 [page: 90-97]
DOI: 10.61124/sinta.v1i2.15

CLUSTERING DAN KLASIFIKASI DATA CUACA KOTA CILACAP MENGGUNAKAN K-MEANS DAN RANDOM FOREST

Fadil Danu Rahman*[1], Mulki Indana Zulfa[2], Acep Taryana[3]
[1,2,3] Teknik Elektro, Universitas Jenderal Soedirman, Indonesia
e-mail: fadil.rahman@mhs.unsoed.ac.id

Abstrak
Pengamatan dan analisis data cuaca merupakan aspek penting dalam memahami kondisi atmosfer di suatu wilayah. Penelitian ini bertujuan untuk menganalisis data cuaca berbasis BMKG dari Stasiun Meteorologi Tunggul Wulung Cilacap menggunakan metode K-Means clustering dan algoritma Random Forest. Data cuaca dari tahun 1975 hingga 2023 diambil untuk mengidentifikasi pola dan karakteristik unik dalam kondisi atmosfer. Metode K-Means clustering digunakan untuk membentuk cluster, yang kemudian digunakan sebagai dasar untuk klasifikasi kondisi cuaca dengan algoritma Random Forest. Melalui penggunaan algoritma Random Forest, model klasifikasi berhasil memprediksi kondisi cuaca dengan tingkat akurasi yang memuaskan. Meskipun demikian, penurunan performa pada rentang tahun 2018-2023 menunjukkan adanya tantangan dalam memodelkan pola cuaca yang kompleks. Analisis menggunakan metode Elbow dan Silhouette menunjukkan jumlah cluster optimal dan evaluasi kualitas pengelompokkan. Implikasi temuan ini diharapkan dapat memberikan manfaat dalam pemahaman dan prakiraan cuaca yang lebih akurat, dengan potensi dampak positif pada berbagai sektor, seperti pertanian dan transportasi. Dengan memadukan teknik clustering dan klasifikasi, penelitian ini membuka peluang untuk pengembangan lebih lanjut dalam analisis cuaca berbasis data.

Kata kunci: Cuaca; *K-Means clustering*; *Random Forest.*

Abstract
Observing and analyzing weather data is crucial in understanding atmospheric conditions in a particular region. This study aims to analyze BMKG-based weather data from the Tunggul Wulung Cilacap Meteorological Station using the K-Means clustering method and Random Forest algorithm. Weather data from 1975 to 2023 were collected to identify unique patterns and characteristics in atmospheric conditions. The K-Means clustering method was utilized to form clusters, which were then used as a basis for weather condition classification using the Random Forest algorithm. Through the implementation of the Random Forest algorithm, the classification model successfully predicted weather conditions with satisfactory accuracy. However, a decrease in performance during the 2018-2023 period indicates challenges in modeling complex weather patterns. Analysis using the Elbow and Silhouette methods revealed the optimal number of clusters and evaluated the quality of clustering. The implications of these findings are expected to provide benefits in understanding and forecasting more accurate weather conditions, with potential positive impacts on various sectors such as agriculture and transportation. By integrating clustering and classification techniques, this research opens opportunities for further development in data-driven weather analysis.

Keywords: Weather; K-Means clustering; Random Forest.

1. PENDAHULUAN

Prakiraan cuaca memiliki peran yang sangat penting dalam kehidupan sehari-hari dan berbagai sektor seperti pertanian, transportasi, dan pengelolaan bencana alam. Stasiun Meteorologi Tunggul Wulung di Cilacap adalah salah satu lembaga yang bertanggung jawab dalam memonitor dan mengumpulkan data cuaca yang akurat. Data cuaca yang diperoleh dari lembaga ini sangat berharga dalam pemahaman dan prediksi keadaan cuaca di wilayah tersebut (1).

Demi upaya meningkatkan kualitas prakiraan cuaca, penggunaan teknik dan algoritma pemrosesan data yang canggih menjadi semakin relevan. Metode pengelompokan data *K-Means* menawarkan pendekatan yang efektif untuk mengorganisir dan memahami pola dalam data cuaca (2). Algoritma *Random Forest*, sebagai salah satu algoritma *Machine Learning*, juga menjadi alternatif menarik dalam meningkatkan akurasi prakiraan cuaca berdasarkan data BMKG dari Stasiun Meteorologi Tunggul Wulung Cilacap (3). Penelitian ini bertujuan untuk menggali potensi penggunaan *K-Means* dalam mengelompokkan data cuaca dan kemudian memanfaatkan hasil pengelompokan ini sebagai masukan dalam algoritma *Random Forest* untuk memprediksi kondisi di wilayah tersebut. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang berarti dalam akurasi prakiraan cuaca di Stasiun Meteorologi Tunggul Wulung Cilacap, serta memberikan pandangan yang lebih mendalam tentang penggunaan metode statistik dan *Machine Learning* dalam konteks ilmu meteorologi. Berdasarkan permasalahan yang ada, dibutuhkan suatu sistem yang dapat membantu proses analisis dari prakiraan cuaca. Oleh karena itu, penulis memilih bahan kajian tentang klasifikasi kondisi

DOI: 10.61124/sinta.v1i2.15

cuaca dengan judul "*Clustering* dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode *K-Means* dan *Random Forest*".

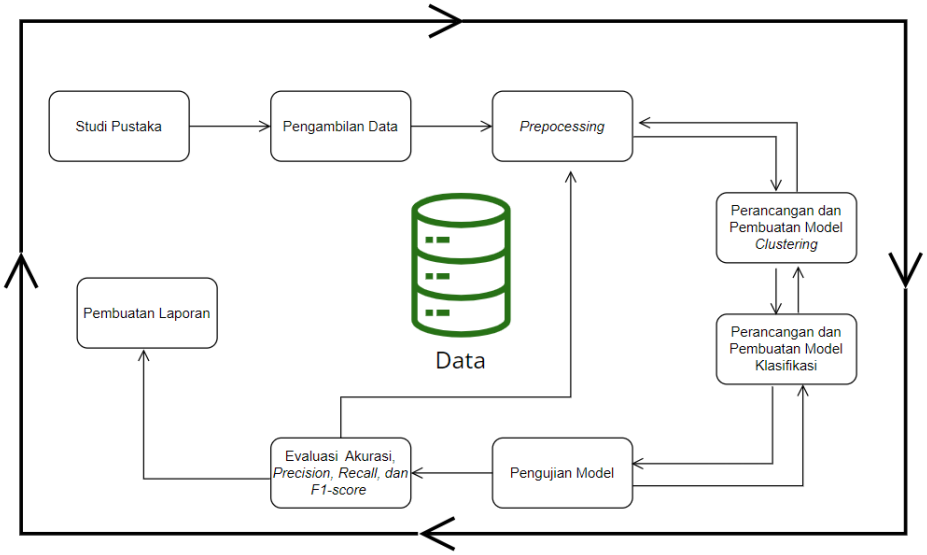
Tabel 1. Penelitian Terdahulu.

No.	Nama Peneliti	Judul Penelitian	Persamaan	Perbedaan
1.	Alvi Syahrini Utami, Dian Palupi Rini, Endang Lestari,2021 (4).	Prediksi Cuaca di Kota Palembang Berbasis *Supervised Learning* Menggunakan Algoritma *K-Nearest Neighbor*	• Menggunakan metode *Locality Sensitive Hashing* • Menggunakan algoritma *K-Nearest Neighbor (K-NN)* • Objek yang diamati adalah wilayah palembang • Menggunakan Bahasa pemrograman MATLAB	• Menggunakan metode *K-Means* untuk *clustering* • Menggunakan algoritma *Random Forest* • Objek yang diamati adalah wilayah cilacap • Menggunakan bahasa pemrograman Python
2.	Maulana Dhawangkhar, Edwin Riksakomara, 2017 (5).	Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik *Random Forest* dan CART (Studi Kasus Kota Surabaya)	• Menggunakan metode CART (*Classification and Regression Tree*) • Objek yang diamati adalah wilayah surabaya • Menggunakan Bahasa pemrograman MATLAB	• Menggunakan metode *K-Means* untuk *clustering* • Objek yang diamati adalah wilayah cilacap • Menggunakan bahasa pemrograman Python
3	Aji Primajaya, Betha Nurina Sari,2018 (6).	*Random Forest Algorithm for Prediction of Precipitation*	• Menggunakan menggunakan pengklasifikasi dari PRCP • Menggunakan Bahasa pemrograman MATLAB • Menggunakan 6 parameter	• Menggunkan *clustering K-Means* dan klasifikasi *Random Forest* • Menggunkan Bahasa pemrograman Python • Menggunakan 10 parameter
4	Nahya Nur, Farid Wajidi, Sulfayanti, Wildayani,2023 (7).	Implementasi Algoritma *Random Forest Regression* Untuk Memprediksi Hasil Panen Padi di Desa Minanga	• Objek yang diamati hasil panen • Data tidak memerlukan *clustering*	• Objek yang diamati kondisi cuaca • *Clustering K-Means* untuk label kondisi cuaca
5	JAYASRI. R, Dr. R. VIDYA,2019 (8).	*Weather Prediction Using K-Means Clustering And Naive Bayes Algorithm*	• Algoritma yang digunakan *Naive Baiyes*	• Algoritma yang digunakan *Random Forest*

DOI: 10.61124/sinta.v1i2.15

2. METODE

Penelitian ini menggunakan metodologi *Cross Industry Standard Process for Data Mining* (CRISP DM) untuk mendukung sebuah proses *clustering* dan klasifikasi dengan tahapan-tahapan seperti yang ada pada Gambar 3.1.



Gambar 1 Metodologi Penelitian

2.1 Studi Pustaka

Tahap ini akan memberikan landasan teori dan kerangka konseptual untuk mendukung penelitian. Ini mencakup pemahaman tentang algoritma *K-Means Clustering*, analisis *Elbow* dan *Silhouette* dalam pengelompokan data, penerapan Random Forest dalam machine learning.konsep data cuaca dan geofisika.

2.2 Pengambilan Data

Pengambilan data dari situs web BMKG krusial dalam penelitian ini. Data cuaca dan geofisika dari BMKG menjadi dasar utama untuk menganalisis kondisi cuaca dan faktor geofisika di wilayah penelitian. Proses ini mencakup ekstraksi informasi terkini tentang suhu, kelembaban, angin, dan parameter meteorologi lainnya. Akses langsung ke data BMKG memastikan keakuratan dan keterkinian informasi untuk analisis dan temuan penelitian. Penggunaan data dari sumber terpercaya seperti BMKG juga mendukung validitas dan reliabilitas penelitian.

2.3 *Preprocessing*

Preprocessing data dimulai dengan identifikasi dan penghapusan nilai tidak valid seperti *NaN* atau 8888, penghapusan nilai 8888 juga mencegah adanya bias atau distorsi dalam hasil analisis. Jika nilai-nilai tidak valid atau tidak tersedia tidak dihapus, mereka dapat memengaruhi proses pemodelan dan menghasilkan hasil yang tidak akurat atau bias (9). kemudian dilakukan penyesuaian format dan konversi jenis data, serta pemilihan variabel relevan. Tujuannya adalah menyusun dataset yang konsisten, akurat, dan siap digunakan dalam pemodelan menggunakan *K-Means* dengan nilai k yang ditentukan melalui *Elbow Method* atau *Silhouette Method*. Proses ini juga mencakup konversi variabel *string* menjadi *float* untuk memungkinkan penggunaan algoritma *K-Means* yang membutuhkan data numerik. Variabel yang tidak relevan dihilangkan untuk menyusun dataset yang lebih bersih dan fokus.

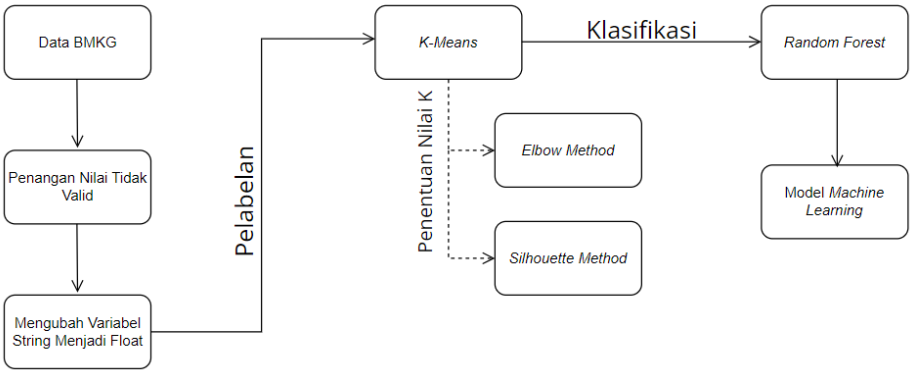
2.4 Perancangan dan Pembuatan Model *Clustering*

Dalam perancangan model menggunakan bahasa pemrograman *Python* di *Google Colab*. . *Colab* memungkinkan pengguna untuk menulis dan menjalankan kode Python secara bebas melalui *browser*, sangat sesuai untuk keperluan *machine learning* dan analisis data (10). *Python* adalah bahasa pemrograman tingkat tinggi yang sering digunakan dalam pengembangan perangkat lunak dan analisis data (11). Dengan menggunakan *Google Colab*, pengguna dapat dengan mudah mengakses sumber daya komputasi yang kuat dan menyimpan serta berbagi *notebook* secara *online*. Perancangan model *K-Means* dalam konteks ini melibatkan penggunaan *Google Colab* untuk menulis dan menjalankan kode *Python*, serta memanfaatkan *library Python* seperti *scikit-learn* dalam

implementasi algoritma *K-Means*, penentuan jumlah *cluster* optimal adalah fokus utama. *Elbow Method* dan *Silhouette Method* digunakan untuk menentukan jumlah cluster terbaik, dengan *Elbow* memperhatikan bentuk kurva inersia dan *Silhouette* mempertimbangkan kedekatan titik data dengan *cluster* lainnya (12)(13). *Elbow Method* memiliki keunggulan dalam kesederhanaannya dan memberikan plot yang mudah dipahami. Namun, terkadang kurva inersia tidak menunjukkan siku yang jelas, membuat penentuan jumlah cluster yang optimal menjadi subjektif. Selain itu, *Elbow Method* mungkin tidak efektif dalam menentukan jumlah *cluster* yang optimal dalam data yang sangat berbeda atau kompleks, dan bisa menghasilkan jumlah cluster yang tidak sesuai dengan karakteristik intrinsik dari data (14). Di sisi lain, *Silhouette Method* memberikan hasil dalam bentuk nilai *Silhouette Score* yang mudah dipahami, memungkinkan untuk mengevaluasi seberapa baik setiap titik data cocok dengan *cluster*-nya dan menentukan jumlah *cluster* yang optimal. Namun, *Silhouette Method* dapat menjadi tidak stabil terhadap *noise* dalam data dan memerlukan perhitungan yang lebih rumit. Selain itu, *Silhouette Method* tidak cukup efektif dalam menentukan jumlah *cluster* yang optimal dalam data yang sangat berbeda atau memiliki struktur yang kompleks (15). Setelah jumlah *cluster* ditetapkan, model *K-Means* diterapkan pada data latih dengan variabel cuaca relevan. Hasilnya adalah pembentukan kelompok cuaca serupa berdasarkan pola dalam data, tanpa supervisi untuk label kategori sebelumnya.

2.5 Perancangan dan Pembuatan Model Klasifiaksi

Dalam perancangan model *Random Forest*, penentuan *hyperparameter* optimal seperti jumlah pohon (*n_estimators*) dan kedalaman maksimum setiap pohon (*max_depth*) dilakukan untuk meningkatkan akurasi dan generalisasi model. Jumlah pohon mengontrol kompleksitas model, sedangkan kedalaman pohon mengatur kemampuan model untuk memahami pola kompleks dalam data. Penyesuaian *hyperparameter* ini untuk menghindari *overfitting* dan memastikan keseimbangan yang baik antara akurasi dan generalisasi model (16). Data latih yang melibatkan berbagai variabel cuaca digunakan untuk melatih model *Random Forest*, yang dapat menangani kompleksitas dan non-linearitas dalam data cuaca. Setelah melatih model, keduanya (*K-Means* dan *Random Forest*) evaluasi model dilakukan untuk memahami sejauh mana tingkat ketepatan dan kehandalan model dalam merepresentasikan pola dan perubahan cuaca. Ini memberikan wawasan tentang kemampuan model dalam melakukan klasifikasi kondisi cuaca dengan konsisten dan dapat diandalkan.



Gambar 2 Metode menentukan nilai k

2.6 Pengujian Model

Pengujian model *K-Means* dimulai dengan menentukan jumlah cluster optimal menggunakan *Elbow Method* dan *Silhouette Method*. Ini penting untuk memahami struktur internal data dan membentuk cluster yang merepresentasikan karakteristik cuaca yang berbeda. Setelah jumlah cluster ditetapkan, model *K-Means* diterapkan pada data latih untuk membentuk kelompok cuaca serupa berdasarkan variabel yang relevan.

Pengujian model Random Forest juga dilakukan dengan menentukan hyperparameter optimal seperti jumlah pohon (*n_estimators*) dan kedalaman maksimum setiap pohon (*max_depth*). Hal ini dilakukan untuk meningkatkan akurasi dan generalisasi model. Data latih digunakan untuk melatih model Random Forest dengan melibatkan berbagai variabel cuaca seperti suhu udara, kelembaban, kecepatan angin, dan lainnya.

2.7 Evaluasi Model

Model *K-Means* dievaluasi menggunakan metrik evaluasi seperti *Silhouette Score* dan *K Score* untuk *Elbow Method*. *Silhouette Score* mengukur sejauh mana pembagian data ke dalam *cluster* merepresentasikan struktur internal yang sesuai, sementara *K Score* membantu menentukan jumlah optimal *cluster*. Hasil evaluasi

DOI: 10.61124/sinta.v1i2.15

memberikan informasi tentang kemampuan model *K-Means* dalam membentuk cluster yang mencerminkan pola dan variasi cuaca yang sebenarnya.

Pada tahap evaluasi model *Random Forest*, fokusnya adalah pada akurasi klasifikasi terhadap variabel-variabel cuaca utama menggunakan metrik seperti *Confusion Matrix*, *Precision*, *Recall*, dan *F1-Score*. *Confusion Matrix* memberikan gambaran holistik tentang kinerja model dalam mengklasifikasikan kondisi cuaca, sedangkan *Precision*, *Recall*, dan *F1-Score* memberikan ukuran komprehensif tentang keseimbangan antara ketepatan dan ketelitian model. Evaluasi ini memastikan bahwa model-model tersebut dapat diandalkan untuk aplikasi di dunia nyata, seperti klasifikasi cuaca yang akurat dan informasi geofisika yang lebih baik.

2.8 Pembuatan Laporan

Proses penulisan laporan melibatkan pengolahan data hasil pengujian, analisis temuan, dan penyajian hasil penelitian melalui presentasi dan pembukuan laporan.

3. HASIL PENELITIAN

Hasil *clustering* data cuaca menggunakan *Silhouette Method* mengevaluasi seberapa baik data cuaca ditempatkan dalam *cluster* yang terbentuk. *Silhouette Method* membantu mengukur kualitas pembentukan *cluster* dengan memberikan nilai antara -1 hingga 1. Nilai positif menandakan bahwa data ditempatkan dengan baik dalam *cluster*, sedangkan nilai negatif menunjukkan bahwa data mungkin lebih cocok untuk *cluster* lain. Berdasarkan interpretasi Silhouette Score yang dijelaskan dalam tesis berjudul "An Evaluation of Clustering and Classification Algorithms in Life-Logging Devices" oleh Anton Amlinger, nilai Silhouette Score memiliki kriteria sebagai berikut: 1) 0.71 - 1.00 menunjukkan struktur yang kuat dengan pemisahan yang sangat baik antar cluster dan tingkat kesamaan yang tinggi di dalam setiap cluster, 2) 0.51 - 0.70 menandakan struktur yang cukup baik dengan hasil pengelompokan yang memadai, 3) 0.26 - 0.50 menandakan struktur yang lemah dan mungkin bersifat artifisial dengan pemisahan yang kurang jelas antar-cluster, dan 4) nilai kurang dari 0.26 menandakan bahwa tidak ditemukan struktur yang substansial dengan pemisahan yang tidak signifikan antar cluster dan tingkat kesamaan yang rendah di dalam cluster (17). Berdasarkan penelitian tersebut nilai *cluster* 2 sudah cukup baik dalam hasil pengelompokkan.

Hasil *clustering* data cuaca menggunakan *Elbow Method* membantu mengevaluasi jumlah *cluster* yang optimal untuk data tersebut. Metode ini memungkinkan penentuan jumlah *cluster* dengan mengamati kurva inersia, yang menggambarkan variabilitas dalam data sehubungan dengan jumlah *cluster* yang berbeda. Jumlah *cluster* yang optimal merupakan titik di mana penurunan inersia mulai menurun secara signifikan, dan ini ditandai oleh elbow pada grafik. Dengan demikian, Elbow Method membantu memilih jumlah cluster yang memberikan keseimbangan yang baik antara kompleksitas model dan kemampuan model untuk menjelaskan variasi dalam data.

Penelitian ini berfokus pada analisis data tingkat yang melibatkan *clustering* dan klasifikasi menggunakan algoritma *Random Forest*. Data telah dikelompokkan berdasarkan karakteristik internalnya. Setelah *clustering*, model *Random Forest* diimplementasikan untuk klasifikasi. Parameter model tidak ditentukan secara acak, melainkan melalui serangkaian eksperimen. Variasi dilakukan pada jumlah pohon dalam hutan (*n_estimators*), dengan nilai 100, 500, dan 1000, serta pada kedalaman pohon (*max_depth*) dengan nilai-nilai 0, 5, dan 10. Validasi silang (*cv*=5) digunakan untuk mengevaluasi performa model dengan baik pada data pelatihan dan data baru. Tujuan pendekatan ini adalah untuk memperoleh pemahaman yang lebih dalam tentang perilaku model *Random Forest* setelah *clustering*, serta menemukan parameter optimal untuk implementasi praktis di dunia nyata.

Tabel 2. Hasil *Clustering Silhoutte Method.*

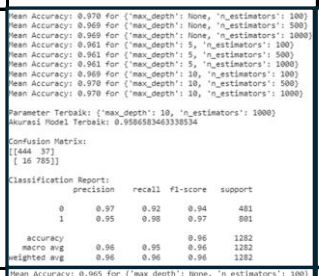
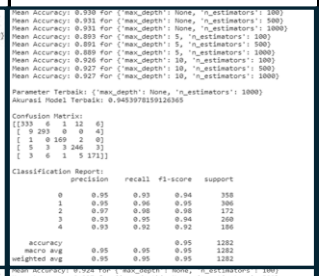
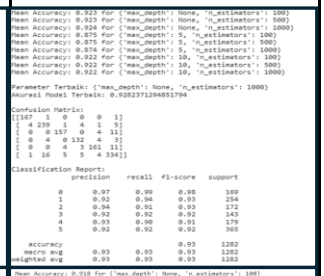
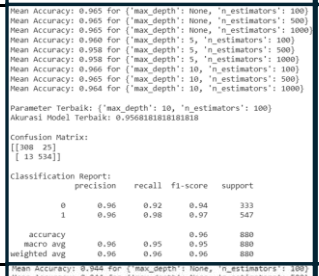
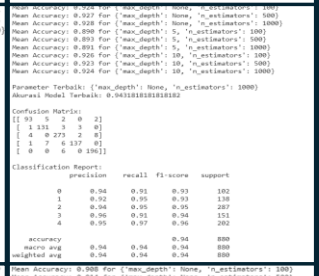
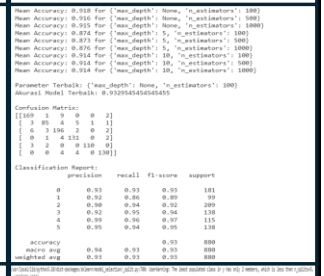
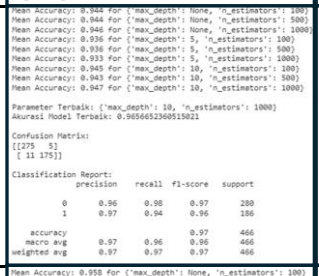
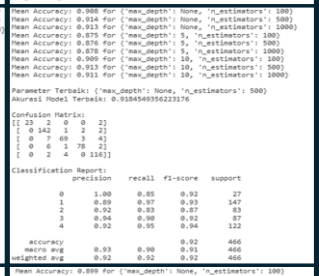
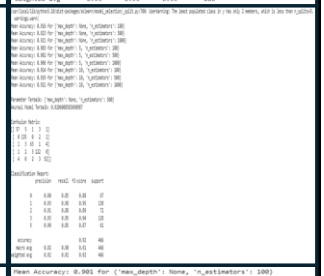
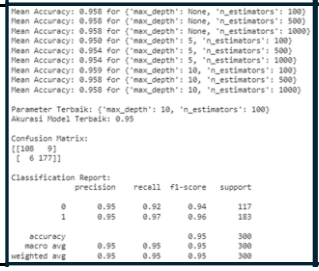
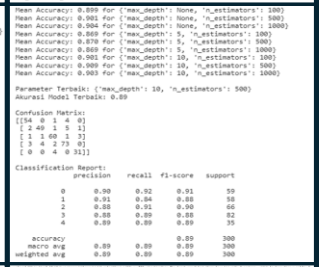
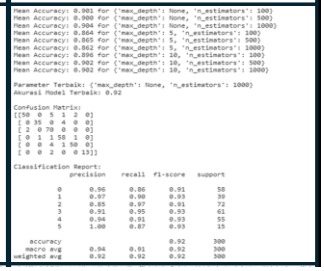
Silhouette Score					
Tahun	**2 Cluster**	**3 Cluster**	**4 Cluster**	**5 Cluster**	**6 Cluster**
2003	0,65	0,53	0,53	0,4	0,4
2008	0,65	0,55	0,51	0,39	0,4
2013	0,63	0,56	0,53	0,4	0,37
2018	0,64	0,56	0,53	0,41	0,37
2023	0,7	0,45	0,43	0,44	0,45

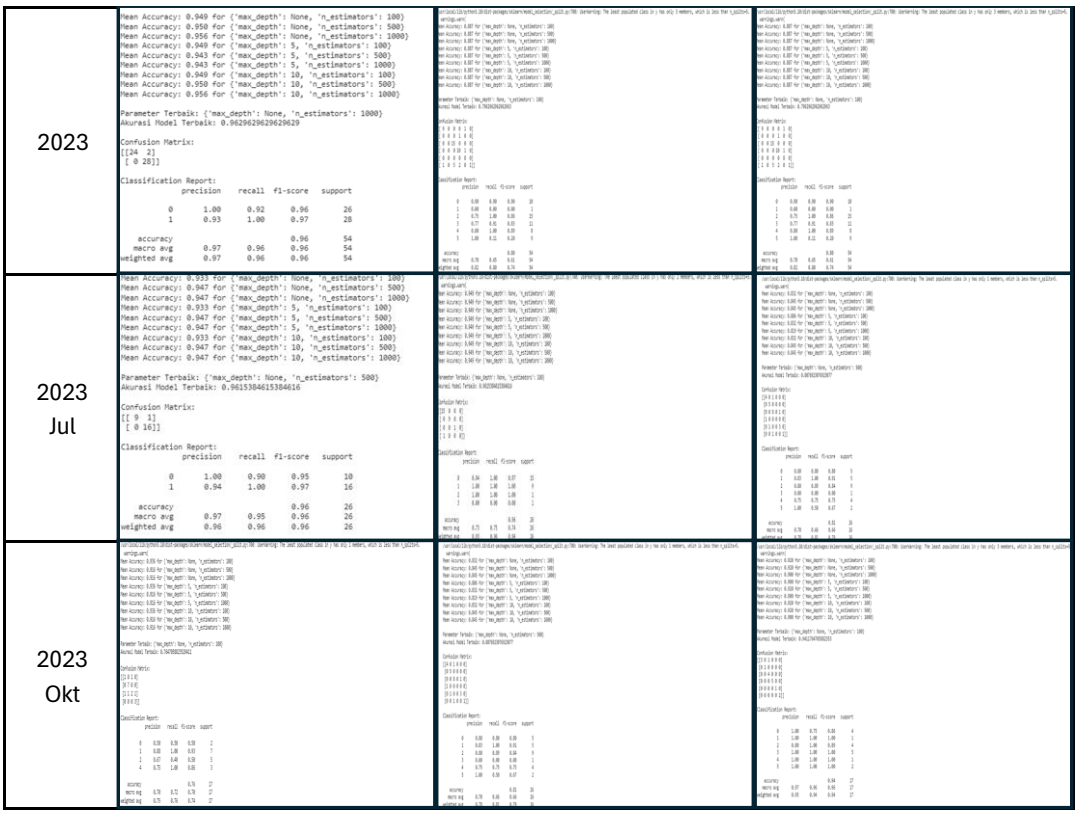
DOI: 10.61124/sinta.v1i2.15

Tabel 3. Hasil *Elbow Method*.

Hasil *Clustering Elbow Method*		
Tahun	**Jumlah *Cluster***	***K Score***
2003	5	3587083,408
2008	5	2358041,389
2013	5	1266479,463
2018	5	757279,108
2023	6	60930,768

Tabel 4. Hasil Klasifikasi *Random Forest*.

Tahun	**Hasil Klasifikasi *Clustering Silhoutte Method***	**Hasil Klasifikasi *Clustering Elbow Method***	**Hasil Klasifikasi *Clustering* BMKG**
2003			
2008			
2013			
2018			



4. DISKUSI

Salah satu masalah utama yang kami temui adalah keterbatasan dalam cara membagi cluster, yang hanya bergantung pada curah hujan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Hal ini menyebabkan hasil *clustering* kurang mewakili dan kurang informatif secara keseluruhan karena tidak mempertimbangkan parameter cuaca lainnya dengan baik. Selain itu, keberadaan data yang hilang juga menjadi kendala yang memengaruhi kualitas dari *clustering* dan klasifikasi secara keseluruhan. Namun, melalui analisis menyeluruh terhadap data yang kami peroleh dan perbandingan dengan penelitian sebelumnya, kami menemukan bahwa penggunaan *K-Means* dan *Random Forest* dapat membantu dalam menentukan jumlah *cluster* yang lebih baik dan meningkatkan akurasi. Temuan kami juga mendukung hasil penelitian dari Maulana Dhawangkhara dan Edwin Riksakomara yang berjudul "Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik *Random Forest* dan *CART* (Studi Kasus Kota Surabaya)" serta Alvi Syahrini Utami, Dian Palupi Rini, dan Endang Lestari yang berjudul "Prediksi Cuaca di Kota Palembang Berbasis *Supervised Learning* Menggunakan Algoritma *K-Nearest Neighbour*" menunjukkan bahwa kombinasi *K-Means* dan *Random Forest* memberikan hasil yang lebih baik dalam menentukan kondisi cuaca.

5. KESIMPULAN

Penelitian menunjukkan bahwa *K-Means clustering* efektif dalam menganalisis pola cuaca. Penggunaan hasil *clustering* sebagai fitur tambahan dalam integrasi dengan algoritma *Random Forest* meningkatkan kemampuan klasifikasi secara signifikan. Penentuan jumlah *cluster* optimal dengan *Elbow Method* dan *Silhouette Method* memberikan panduan yang berguna, dengan beberapa variasi dari waktu ke waktu.

Model Random Forest berhasil mengklasifikasikan kondisi cuaca berdasarkan *clustering K-Means*, meningkatkan akurasi prakiraan secara signifikan. Namun, terdapat penurunan performa pada beberapa tahun terakhir, menunjukkan variasi pola cuaca dari waktu ke waktu. Saran untuk penelitian selanjutnya mencakup mencoba metode *preprocessing* lain dan eksplorasi algoritma lain untuk prediksi cuaca yang lebih baik. Ini memberikan kontribusi penting dalam pemahaman dan prediksi kondisi cuaca yang lebih akurat dalam ilmu meteorologi.

UCAPAN TERIMA KASIH

Ucapan terima kasih kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atas izin dan akses yang diberikan untuk mengambil data cuaca dari website resmi mereka. Tanpa dukungan dan kerja sama dari pihak BMKG, penelitian ini tidak akan dapat terlaksana.

DAFTAR PUSTAKA

1. "Pengetahuan | BMKG." Accessed: Nov. 21, 2023. [Online]. Available: http://182.16.248.153/database/?p=pengetahuan.
2. "K-Means Clustering, Salah Satu Contoh Teknik Analisis Data P..." Accessed: Nov. 21, 2023. [Online]. Available: https://dqlab.id/k-means-clustering-salah-satu-contoh-teknik-analisis-data-populer
3. "What is Random Forest? | IBM." Accessed: Nov. 21, 2023. [Online]. Available: https://www.ibm.com/topics/random-forest#What+is+random+forest%3F
4. S. Utami, D. P. Rini, and E. Lestari, "Prediksi Cuaca di Kota Palembang Berbasis Supervised Learning Menggunakan Algoritma K-Nearest Neighbour," vol. 13, no. 1, 2021.
5. N. Nur and F. Wajidi, "Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," vol. 9.
6. A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *IJAIDM*, vol. 1, no. 1, p. 27, Mar. 2018, doi: 10.24014/ijaidm.v1i1.4903.
7. M. Dhawangkhara and E. Riksakomara, "Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya)," *JTITS*, vol. 6, no. 1, pp. 88–93, Feb. 2017, doi: 10.12962/j23373539.v6i1.21120.
8. R. Meenal, P. A. Michael, D. Pamela, and E. Rajasekaran, "Weather prediction using random forest machine learning model," *IJEECS*, vol. 22, no. 2, p. 1208, May 2021, doi: 10.11591/ijeecs.v22.i2.pp1208-1215.
9. V. Păpăluţă, "What's the best way to handle NaN values?," Medium. Accessed: Apr. 02, 2024. [Online]. Available: https://towardsdatascience.com/whats-the-best-way-to-handle-nan-values-62d50f738fc
10. "Google Colab." Accessed: Nov. 21, 2023. [Online]. Available: https://research.google.com/colaboratory/intl/id/faq.html
11. "What is ci? Executive Summary | Python.org." Accessed: Nov. 21, 2023. [Online]. Available: https://www.python.org/doc/essays/blurb/
12. "Elbow Method to Find the Optimal Number of Clusters in K-Means." Accessed: Nov. 21, 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning/.
13. "Selecting the number of clusters with silhouette analysis on KMeans clustering — scikit-learn 1.3.2 documentation." Accessed: Nov. 21, 2023. [Online]. Available: https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html
14. V. A. Ekasetya and A. Jananto, "KLUSTERISASI OPTIMAL DENGAN ELBOW METHOD UNTUK PENGELOMPOKAN DATA KECELAKAAN LALU LINTAS DI KOTA SEMARANG," *JDI*, vol. 12, no. 1, pp. 20–28, Aug. 2020, doi: 10.35315/informatika.v12i1.8159.
15. STMIK STIKOM Indonesia, D. A. I. C. Dewi, D. A. K. Pramita, and STMIK STIKOM Indonesia, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *MATRIX*, vol. 9, no. 3, pp. 102–109, Nov. 2019, doi: 10.31940/matrix.v9i3.1662.
16. A. Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow," O'Reilly, 2019.
17. Amlinger, Anton. *An Evaluation of Clustering and Classification Algorithms in Life-Logging Devices*. Linköpings universitet, 25 June 2015.