Implementasi K-Means Clustering untuk Pengelompokan Pola Cuaca di Kabupaten Sambas

Khalidya[1], Muhammad Usman[2], Andri Hidayat[3], Erifa Syahnaz[4]

1,2Program Studi Manajemen Informatika, Jurusan Manajemen Informatika, Politeknik Negeri Sambas, Indonesia

e-mail: [dyalilid647@gmail.com]@poltesa.ac.id

Abstrak

Perubahan kondisi cuaca memberikan dampak yang signifikan terhadap berbagai sektor, khususnya pertanian, sehingga diperlukan pengolahan data untuk memperoleh informasi mengenai pola cuaca. Data cuaca harian yang dimiliki Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Meteorologi Paloh masih tersaji dalam bentuk data mentah sehingga karakteristik pola cuaca belum dapat dipahami secara langsung. Penelitian ini bertujuan mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan pola cuaca di Kabupaten Sambas menggunakan kerangka kerja Knowledge Discovery in Databases (KDD). Data yang digunakan merupakan data cuaca harian periode 2024–2025 yang meliputi suhu minimum (TN), suhu maksimum (TX), suhu rata-rata (TAVG), kelembapan udara rata-rata (RHAVG), curah hujan (RR), dan lama penyinaran matahari (SS). Jumlah Cluster optimal ditentukan menggunakan Elbow Method, sedangkan kualitas hasil Clustering dievaluasi menggunakan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa data cuaca dapat dikelompokkan menjadi tiga Cluster yang merepresentasikan kondisi cuaca normal, kering, dan basah, dengan pola musiman yang konsisten sepanjang tahun. Hasil pengelompokan selanjutnya diimplementasikan ke dalam sistem berbasis Web menggunakan Flask dan MySQL untuk menyajikan informasi dan visualisasi pola cuaca sehingga lebih mudah dipahami oleh pengguna, khususnya masyarakat dan petani di Kabupaten Sambas.

Kata kunci: *BMKG; Data Mining; K-Means Clustering; KDD; Pola Cuaca.*

Abstract

Changes in weather conditions have a significant impact on various sectors, particularly agriculture, so that data processing is needed to obtain information about weather patterns. Daily weather data from the Paloh Meteorological Station of the Meteorology, Climatology, and Geophysics Agency (BMKG) is still presented in raw form, so weather pattern characteristics cannot be directly understood. This study aims to implement the K-Means Clustering algorithm to classify weather patterns in Sambas Regency using the Knowledge Discovery in Databases (KDD) framework. The dataset consists of daily weather data for the 2024-2025 period, including minimum temperature (TN), maximum temperature (TX), average temperature (TAVG), average relative humidity (RHAVG), rainfall (RR), and sunshine duration (SS). The optimal number of clusters was determined using the Elbow Method, while clustering performance was evaluated using the Davies-Bouldin Index (DBI). The results show that the weather data can be grouped into three clusters representing normal, dry, and wet weather conditions, with a consistent seasonal pattern throughout the year. The clustering results were then implemented in a Web-based system built with Flask and MySQL to present weather pattern information and visualizations, making it easier for users, particularly the public and farmers in Sambas Regency, to understand the results.

Keywords: *BMKG; Data Mining; K-Means Clustering; KDD; Weather Patterns.*

1. PENDAHULUAN

Perubahan kondisi cuaca dalam beberapa tahun terakhir memberikan dampak yang signifikan terhadap berbagai aspek kehidupan, seperti sektor pertanian, perikanan, transportasi, serta aktivitas masyarakat sehari-hari. Informasi mengenai kondisi cuaca menjadi sangat penting karena berperan sebagai dasar pengambilan keputusan pada berbagai sektor, khususnya sektor pertanian yang sangat dipengaruhi oleh variasi suhu, curah hujan, kelembapan udara, dan lama penyinaran matahari [1]. Cuaca merupakan kondisi atmosfer pada waktu dan tempat tertentu yang bersifat dinamis sehingga dapat berubah dalam waktu yang relatif singkat, sehingga diperlukan informasi mengenai karakteristik pola cuaca yang mudah dipahami agar dapat mendukung proses pengambilan keputusan [2].

Kabupaten Sambas merupakan salah satu kabupaten di Provinsi Kalimantan Barat dengan karakteristik cuaca tropis dan tingkat variasi yang cukup tinggi akibat pengaruh letak geografis wilayah pesisir, kondisi atmosfer, serta perubahan musim. Data cuaca harian di Kabupaten Sambas dikelola oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melalui Stasiun Meteorologi Paloh, meliputi suhu minimum (TN), suhu maksimum (TX), suhu rata-rata (TAVG), kelembapan udara rata-rata (RHAVG), curah hujan (RR), dan lama penyinaran matahari (SS). Data tersebut masih disajikan dalam bentuk data mentah berformat Microsoft Excel maupun CSV sehingga karakteristik pola cuaca belum dapat dipahami secara langsung oleh pengguna dan proses identifikasinya masih dilakukan secara manual [3].

Salah satu pendekatan yang dapat digunakan untuk mengolah data cuaca adalah Data Mining, yaitu proses menggali pola atau informasi tersembunyi dari sekumpulan data melalui tahapan Knowledge Discovery in Databases (KDD) yang terdiri atas Selection, Preprocessing, Transformation, Data Mining, dan Evaluation [3]. Pada tahap Data Mining digunakan algoritma K-Means Clustering karena memiliki proses komputasi yang relatif sederhana, efisien dalam mengolah data numerik, serta mampu mengelompokkan data berdasarkan tingkat kemiripan karakteristik antarobjek [1]. Beberapa penelitian terdahulu, seperti klasterisasi data meteorologi wilayah Indonesia Timur menggunakan K-Means dan Fuzzy C-Means [2], identifikasi pola iklim stasiun cuaca menggunakan data NOAA GSOD [4], analisis pola cuaca di Provinsi Sumatera Utara [3], serta perbandingan K-Means dan DBSCAN pada data ulasan wisata dengan evaluasi Silhouette Index dan Davies-Bouldin Index [8], menunjukkan bahwa algoritma K-Means mampu menghasilkan pengelompokan data berdasarkan tingkat kemiripan karakteristik setiap objek. Namun, sebagian besar penelitian tersebut masih berfokus pada proses Clustering tanpa menerapkan tahapan KDD secara lengkap atau evaluasi kualitas Cluster menggunakan DBI [4], dan penelitian yang menerapkan K-Means untuk mengelompokkan pola cuaca menggunakan data harian BMKG Kabupaten Sambas masih sangat terbatas.

Berdasarkan kesenjangan tersebut, penelitian ini menerapkan tahapan KDD secara lengkap dengan menggunakan normalisasi Min-Max Scaling [7], algoritma K-Means Clustering, Elbow Method [9], dan Davies-Bouldin Index (DBI) [7] untuk menghasilkan pengelompokan pola cuaca yang optimal berdasarkan data cuaca harian BMKG Kabupaten Sambas. Penelitian ini bertujuan mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan pola cuaca di Kabupaten Sambas berdasarkan data cuaca harian BMKG periode 2024–2025, menentukan jumlah Cluster optimal serta mengevaluasi kualitas hasil pengelompokan, dan menyajikan hasil pengelompokan tersebut melalui sistem berbasis Web sebagai media informasi dan visualisasi. Hasil penelitian diharapkan dapat memberikan informasi mengenai karakteristik pola cuaca yang dapat dimanfaatkan sebagai pendukung analisis pola cuaca, khususnya pada sektor pertanian di Kabupaten Sambas.

2. METODE

Metode penyelesaian masalah yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD), yaitu serangkaian proses untuk menggali pengetahuan atau informasi yang tersembunyi dari sekumpulan data melalui tahapan Selection, Preprocessing, Transformation, Data Mining, dan Evaluation [5]. Data yang digunakan merupakan data cuaca harian Kabupaten Sambas periode Januari 2024 hingga Desember 2025 yang diperoleh dari BMKG Stasiun Meteorologi Paloh [14], [15], dengan enam parameter, yaitu TN, TX, TAVG, RHAVG, RR, dan SS. Seluruh proses pengolahan data diimplementasikan menggunakan bahasa pemrograman Python pada lingkungan Google Colaboratory, dengan bantuan Library Pandas, NumPy, Scikit-learn, dan Matplotlib.

Selection

Tahap Selection dilakukan untuk mengambil data yang relevan dari keseluruhan data operasional, yaitu atribut Tanggal, TN, TX, TAVG, RHAVG, RR, dan SS, sehingga hanya data yang berkaitan dengan tujuan penelitian yang digunakan pada proses Data Mining.

Preprocessing / Cleaning

Data BMKG menggunakan kode 8888 dan 9999 sebagai penanda data yang tidak tersedia. Nilai tersebut dikonversi menjadi NaN, kemudian missing value ditangani menggunakan Mean Imputation, yaitu mengganti nilai yang hilang dengan nilai rata-rata masing-masing parameter cuaca. Metode ini dipilih agar jumlah data tetap dipertahankan sehingga seluruh data dapat digunakan pada proses Clustering.

Transformation

Data hasil Preprocessing dinormalisasi menggunakan Min-Max Normalization melalui MinMaxScaler dari Library Scikit-learn agar seluruh atribut memiliki rentang nilai 0 sampai 1. Normalisasi diperlukan agar atribut dengan nilai lebih besar tidak mendominasi proses perhitungan jarak pada algoritma K-Means [7].

Data Mining

Proses Data Mining dilakukan menggunakan algoritma K-Means Clustering dengan parameter random_state = 42 dan n_init = 10 pada Library Scikit-learn. Algoritma bekerja dengan menentukan Centroid awal, menghitung jarak setiap data terhadap Centroid menggunakan Euclidean Distance [10], menempatkan

data pada Cluster dengan jarak terdekat, kemudian memperbarui Centroid secara iteratif hingga tidak terjadi perubahan anggota Cluster (konvergen) [11]. Jumlah Cluster optimal ditentukan menggunakan Elbow Method dengan menghitung nilai Within Cluster Sum of Squares (WCSS) pada K = 2 hingga K = 6 [9].

Evaluation

Kualitas hasil Clustering dievaluasi menggunakan Davies-Bouldin Index (DBI) melalui fungsi davies_bouldin_score() pada Scikit-learn. Semakin kecil nilai DBI, semakin baik kualitas Cluster yang dihasilkan, yaitu semakin kompak anggota dalam satu Cluster dan semakin terpisah antar-Cluster [7]. Hasil pengelompokan yang optimal selanjutnya diinterpretasikan untuk menyusun Kalender Pola Cuaca Kabupaten Sambas dan disajikan melalui sistem berbasis Web yang dibangun menggunakan framework Flask dengan basis data MySQL.

3. HASIL DAN PEMBAHASAN

Penentuan Jumlah Cluster Optimal

Implementasi Elbow Method dilakukan dengan menguji jumlah Cluster sebanyak 2 hingga 6. Hasil visualisasi menunjukkan bahwa penurunan nilai WCSS dari K = 2 ke K = 3 cukup signifikan, sedangkan setelah K = 3 penurunan nilai WCSS mulai melandai, sehingga K = 3 teridentifikasi sebagai titik siku (elbow point). Berdasarkan evaluasi Davies-Bouldin Index, nilai DBI terkecil secara matematis diperoleh pada K = 2 sebesar 1,0653. Meskipun demikian, penelitian ini tetap menggunakan K = 3 pada tahap implementasi dengan mempertimbangkan hasil Elbow Method yang menunjukkan titik siku pada K = 3, serta tujuan penelitian untuk mengelompokkan pola cuaca ke dalam tiga kategori, yaitu basah, normal, dan kering, sehingga hasil pengelompokan lebih mudah diinterpretasikan dan lebih sesuai dengan kebutuhan penyusunan kalender pola cuaca.

Karakteristik Hasil Clustering

Algoritma K-Means Clustering berhasil mengelompokkan data cuaca harian Kabupaten Sambas menjadi tiga Cluster berdasarkan tingkat kemiripan karakteristik enam parameter cuaca yang digunakan. Karakteristik rata-rata (Centroid) setiap Cluster disajikan pada Tabel 2.

Tabel 2. Profil Rata-Rata Centroid Tiap Cluster

Cluster	TX (°C)	TAVG (°C)	RHAVG (%)	RR (mm)	SS (jam)	Kategori
0	30,9984	26,6016	90,0215	10,7371	7,3032	Normal/ Lembap
1	33,4045	28,3174	83,1208	1,6064	7,3623	Kering
2	30,8637	26,7021	90,5556	15,4705	1,6034	Basah

Cluster 0 memiliki kelembapan udara tinggi (90,02%), curah hujan sedang (10,74 mm), dan suhu rata-rata 26,60°C, sehingga merepresentasikan kondisi cuaca normal atau lembap. Cluster 1 memiliki suhu rata-rata tertinggi (28,32°C), kelembapan udara lebih rendah (83,12%), dan curah hujan paling rendah (1,61 mm), sehingga menggambarkan kondisi cuaca kering. Cluster 2 memiliki kelembapan udara tertinggi (90,56%), curah hujan paling tinggi (15,47 mm), serta lama penyinaran matahari paling rendah (1,60 jam), sehingga merepresentasikan kondisi cuaca basah. Hasil ini menunjukkan bahwa keenam parameter cuaca yang digunakan mampu membedakan karakteristik ketiga kondisi cuaca secara jelas, sejalan dengan temuan penelitian sejenis pada wilayah lain yang juga menggunakan parameter suhu, kelembapan, dan curah hujan sebagai pembeda utama antar-Cluster [2], [3].

Distribusi Cluster Berdasarkan Bulan

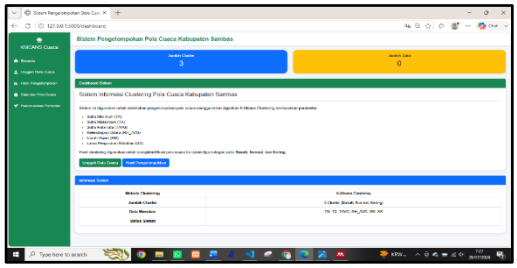
Analisis distribusi kemunculan Cluster pada setiap bulan menunjukkan pola musiman yang konsisten selama periode pengamatan. Bulan Januari dan Februari didominasi oleh Cluster 2 (basah), yaitu masing-masing sebanyak 34 dan 37 kejadian, yang menunjukkan bahwa awal tahun di Kabupaten Sambas cenderung berada pada kondisi basah dengan intensitas curah hujan relatif tinggi. Memasuki bulan April hingga September, Cluster 1 (kering) menjadi Cluster yang paling dominan, dengan puncak kemunculan pada bulan Juli sebanyak 44 kejadian, menandakan periode pertengahan tahun didominasi oleh cuaca kering dengan curah hujan rendah dan lama penyinaran matahari yang lebih tinggi. Sementara itu, bulan Maret dan Oktober didominasi oleh Cluster 0 (normal), yang menggambarkan masa transisi antara musim hujan dan musim kemarau. Bulan November dan Desember kembali didominasi oleh Cluster 2, sehingga menutup pola musiman tahunan dengan kondisi basah.

Berdasarkan pola dominasi Cluster tersebut, disusun Matriks Kalender Pola Cuaca Kabupaten Sambas yang menyederhanakan hasil Clustering ke dalam bentuk yang lebih mudah dipahami. Matriks ini memberikan ringkasan kecenderungan kondisi cuaca setiap bulan sehingga dapat dimanfaatkan sebagai bahan pertimbangan dalam perencanaan kegiatan yang dipengaruhi oleh kondisi cuaca, khususnya pada sektor pertanian.

Implementasi Sistem Berbasis Web

Hasil pengelompokan pola cuaca diimplementasikan ke dalam sistem berbasis Web yang dibangun menggunakan bahasa pemrograman Python dengan framework Flask dan basis data MySQL. Sistem tidak melakukan proses Clustering secara langsung, melainkan menampilkan hasil pengolahan data yang telah diperoleh sebelumnya. Sistem terdiri atas halaman Login, Dashboard, Unggah Data Cuaca, Hasil Clustering, Kalender Pola Cuaca, dan Rekomendasi Pertanian.

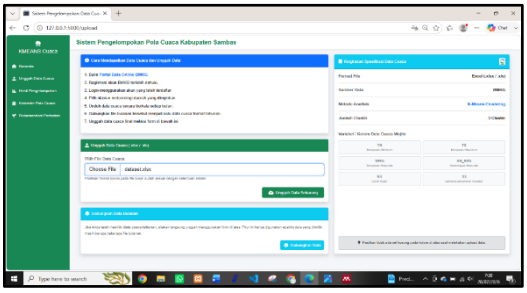
1. Halaman Dashboard



Gambar 1. Halaman Dashboard

Halaman Dashboard merupakan halaman utama sistem yang digunakan sebagai pusat informasi dan navigasi. Halaman ini menyediakan akses menuju fitur-fitur yang tersedia pada sistem.

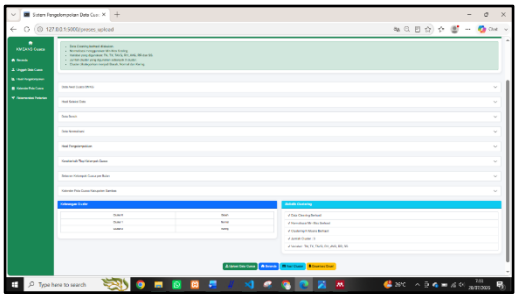
2. Halaman Unggah Data Cuaca



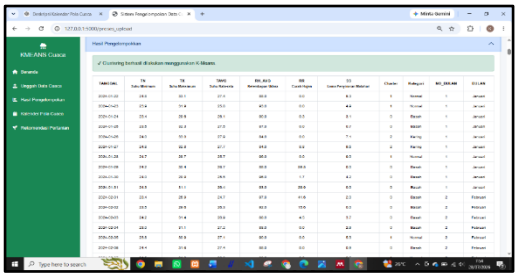
Gambar 2. Halaman Unggah Data Cuaca

Pada halaman Unggah Data Cuaca, pengguna dapat mengunggah dataset berformat Microsoft Excel (.xlsx/.xls), termasuk fitur penggabungan data bulanan menjadi satu dataset tahunan

3. Halaman Hasil Clustering



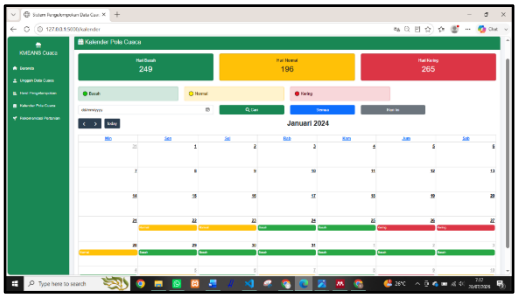
Gambar 3. Halaman Proses Hasil Clustering



Gambar 4. Halaman Hasil Clustering

Halaman Hasil Clustering menyajikan setiap tahapan pengolahan data, yaitu Data Awal, Hasil Seleksi, Data Bersih, Data Normalisasi, Hasil Pengelompokan, karakteristik tiap Cluster, distribusi Cluster per bulan, serta Kalender Pola Cuaca, dan menyediakan fitur unduh hasil dalam format Excel.

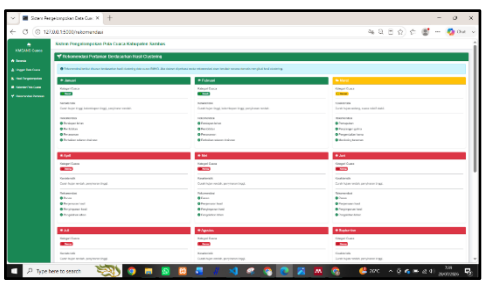
4. Halaman Kalender Pola Cuaca



Gambar 5. Halaman Kalender Pola Cuaca

Halaman Kalender Pola Cuaca digunakan untuk menyajikan hasil pengelompokan dalam bentuk kalender. Tampilan ini memudahkan pengguna dalam melihat kecenderungan kondisi cuaca berdasarkan tanggal dan hasil Cluster.

5. Halaman Rekomendasi Pertanian



Gambar 6. Halaman Rekomendasi Pertanian

Halaman Rekomendasi Pertanian digunakan untuk menampilkan rekomendasi berdasarkan kategori pola cuaca yang diperoleh dari hasil clustering. Informasi rekomendasi disajikan berdasarkan bulan sehingga dapat menjadi informasi pendukung dalam perencanaan kegiatan pertanian.

Dengan adanya sistem ini, informasi pola cuaca hasil Clustering dapat disajikan secara lebih terstruktur, interaktif, dan mudah diakses oleh masyarakat maupun petani di Kabupaten Sambas.

4. KESIMPULAN

Algoritma K-Means Clustering berhasil diimplementasikan untuk mengelompokkan data cuaca harian Kabupaten Sambas periode 2024–2025 menjadi tiga Cluster optimal melalui tahapan Knowledge Discovery in Databases (KDD) dan normalisasi Min-Max Scaling, yaitu Cluster 0 sebagai kondisi normal/lembap, Cluster 1 sebagai kondisi kering, dan Cluster 2 sebagai kondisi basah. Distribusi Cluster pada setiap bulan menunjukkan pola musiman yang konsisten, dengan awal dan akhir tahun (Januari, Februari, November, dan Desember) didominasi kondisi basah, pertengahan tahun (April–September) didominasi kondisi kering, serta bulan Maret dan Oktober sebagai masa transisi dengan kondisi normal. Hasil pengelompokan tersebut berhasil disajikan secara interaktif melalui purwarupa sistem berbasis Web, yang terbukti efektif menyederhanakan data mentah menjadi informasi pola cuaca yang mudah dipahami oleh masyarakat maupun petani di Kabupaten Sambas.

DAFTAR PUSTAKA

[1] M. Firmandani Pramoedya, R. A. Ramadhani, and A. B. Setiawan, "Pemanfaatan K-Means Clustering dalam Pengelompokan Cuaca di Wilayah Kediri," INOTEK, vol. 9, 2025.

[2] G. Andrian, D. Arisandi, and T. Handhayani, "Clustering Data Meteorologi Wilayah Indonesia Timur dengan Metode K-Means dan Fuzzy C-Means," INTI Nusa Mandiri, vol. 18, no. 2, pp. 100–106, 2024, doi: 10.33480/inti.v18i2.5039.

[3] E. Manalu, E. Surbakti, and S. Pardingotan Sipayung, "Analisis Pola Cuaca di Provinsi Sumatera Utara Menggunakan Metode Clustering K-Means," Jurnal Ilmu Komputer dan Informatika, vol. 2, no. 3, 2026.

[4] A. S. Hasugian, J. A. Aziz, and N. A. Nasution, "Penerapan K-Means Clustering untuk Identifikasi Pola Iklim Stasiun Cuaca Indonesia Menggunakan Data NOAA Global Surface Summary of the Day (GSOD)," Jurnal Komputer Teknologi Informasi Sistem Komputer (JUKTISI), vol. 5, no. 1, pp. 685–693, 2026, doi: 10.62712/juktisi.v5i1.941.

[5] D. Athallah Ubaid and M. Ichsan Farel Rachmad, "Klasifikasi Risiko Strok Menggunakan Algoritma Random Forest dengan Teknik Knowledge Discovery in Database," Journal Computer Science, vol. 4, no. 1, 2025.

[6] A. Nur Khomarudin, "Teknik Data Mining: Algoritma K-Means Clustering," 2003. [Online]. Available: https://agusnkhom.wordpress.com

[7] Hendrik, Kusrini, and Kusnawi, "Optimasi Penentuan Sentroid Awal pada K-Means untuk Meningkatkan Hasil Evaluasi Davies-Bouldin Index," vol. 6, no. 1, 2024.

[8] N. A. Y. et al., "Analisis Perbandingan K-Means dan DBSCAN dalam Pengelompokan Data Travel Review Ratings Menggunakan Evaluasi Silhouette Index dan Davies-Bouldin Index," Jurnal Informatika dan Teknik Elektro Terapan, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.6884.

[9] N. A. Maori, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," Jurnal SIMETRIS, vol. 14, 2023.

[10] B. Hartono, S. Eniyati, and K. Hadiono, "Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering," Jurnal Sistem Komputer dan Informatika (JSON), vol. 4, no. 3, p. 503, 2023, doi: 10.30865/json.v4i3.6021.

[11] M. Benri, H. Metisen, and S. Latipa, "Analisis Clustering Menggunakan Metode K-Means dalam Pengelompokkan Penjualan Produk pada Swalayan Fadhila," Jurnal Media Infotama, vol. 11, no. 2, 2015.

[12] Miftahuddin, "Analisis Unsur-unsur Cuaca dan Iklim Melalui Uji Mann-Kendall Multivariat," vol. 13, no. 1, 2016.

[13] A. Andi Rusdin, P. Palloan, and A. Prasetiyo, "Uji Akurasi Ambang Batas Indeks Stabilitas Atmosfer terhadap Pembentukan Thunderstorm dan Awan Cumulonimbus di Stasiun Meteorologi Kelas I Sultan Hasanuddin," vol. 12, no. 2, pp. 269–275, 2023, doi: 10.25077/jfu.12.2.269-275.2023.

[14] Badan Meteorologi, Klimatologi, dan Geofisika, "Profil BMKG," 2024. [Online]. Available: https://www.bmkg.go.id/

[15] Badan Meteorologi, Klimatologi, dan Geofisika, "Data Online," 2025.

[16] Badan Meteorologi, Klimatologi, dan Geofisika, "Suhu Udara," 2025. [Online]. Available: https://gaw-bariri.bmkg.go.id/publikasi/gawsarium/258-suhu-udara

[17] A. Nelvi, R. A. Nata, et al., "Sunshine Duration and Diurnal Temperature Range and Its Relation to

Climate Change in Pontianak." [Online]. Available: www.ogimet.com

[18] A. Srirahayu and L. Setya Pribadie, "Review Paper Data Mining Klasifikasi Data Mining," Jurnal Ilmiah Informatika Global, doi: 10.36982/jiig.v13i2.2307.

[19] Y. S. Swarinoto, "Pemanfaatan Suhu Udara dan Kelembapan Udara dalam Persamaan Regresi untuk Simulasi Prediksi Total Hujan Bulanan di Bandar Lampung."