

ANALISIS *TIME SERIES* TEMPERATUR DI KOTA KUPANG MENGUNAKAN METODE *K-MEANS CLUSTERING*

Frengky Fernandez

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S. Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: *Frengky.535190007@stu.untar.ac.id*

ABSTRAK

Cuaca merupakan bagian penting dari aktivitas sehari-hari manusia. Oleh karena itu, banyak orang yang membutuhkan informasi kondisi atmosfer (cuaca) lebih cepat, lengkap, dan akurat. Prediksi cuaca dapat digunakan untuk memecahkan masalah yang timbul akibat pengaruh cuaca seperti deteksi kekeringan, cuaca buruk, panen dan produksi, industri perencanaan energi, penerbangan, komunikasi dan lain-lain. Metode Neural Network lebih efisien dalam perhitungan yang cepat dan dapat menangani data tidak stabil dalam kasus data prakiraan cuaca biasa. Untuk Prediksi Cuaca dengan input data sinoptik adalah datanya. Beberapa percobaan dilakukan untuk mendapatkan arsitektur yang optimal dan menghasilkan prediksi yang akurat.

Kata kunci: Peramalan, Time Series, Cuaca, K-Means, Clustering.

ABSTRACT

Weather is an important part of human daily activities. Therefore, many people need faster, more complete, and more accurate information about atmospheric conditions (weather). Weather forecasts can be used to solve problems caused by weather, such as drought detection, bad weather, harvesting and production, energy planning, aviation, communications, and others. The Neural Network method is more efficient in fast calculations and is able to handle unstable data in the case of ordinary weather forecast data. For Weather Prediction with synoptic data input, the data is used. Several experiments were conducted to obtain the optimal architecture and produce accurate predictions..

Keywords: Forecasting, Time Series, Weather, K-Means, Clustering.

1. PENDAHULUAN

Topik yang dipilih dalam penelitian ini adalah analisis dan menghitung nilai *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) pada Temperatur dan Temperatur rata-rata Kota Kupang 2019 dan 2021. Selain itu kami memilih topik ini juga untuk mengetahui apakah algoritma *K-means Clustering* ini dapat cocok dipakai untuk memprediksi cuaca. pada Kota Kupang. Tujuan dari penelitian ini untuk memprediksi temperatur maksimum musiman dan temperatur rata-rata dari Kota Kupang yang akan terjadi pada kemudian hari menggunakan metode klasterisasi dan model dari algoritma *K-Means*.

2. METODE PENELITIAN

Metode K-Means adalah suatu metode analisis data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya [1].

Metode ini merupakan metode evaluasi cluster yang menggabungkan metode *cohesion* dan *separation*. Cohesion diukur dengan menghitung seluruh objek yang terdapat dalam sebuah cluster

dan *separation* diukur dengan menghitung jarak rata-rata setiap objek dalam sebuah cluster dengan cluster terdekatnya [2] [3] [4]. Jarak antara data dihitung dengan menggunakan rumus *euclidean distance*. Untuk menyediakan informasi tentang kualitas hasil clustering pada proses clustering, dapat dihitung *silhouette* dari masing-masing cluster bahkan keseluruhan cluster dari hasil kerja suatu algoritma clustering.

Nilai *silhouette* untuk keseluruhan data dengan jumlah cluster k , dapat didefinisikan sebagai $s(i, k)$ yang dihitung dengan persamaan (3) yakni rata-rata *silhouette value* untuk semua cluster. Terdapat beberapa langkah algoritma *K-Means* adalah dengan memilih secara acak k buah sebagai pusat klaster. Lalu jarak antara data dan pusat klaster dihitung menggunakan *euclidean distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori dan persamaan jarak *euclidean* yang dirumuskan dengan:

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2},$$

dimana, $d(i, j)$ adalah jarak data ke- i ke pusat klaster j , X_{ki} adalah data ke- i pada atribut data ke- k , X_{kj} . Data ditempatkan ke dalam klaster yang terdekat yang di hitung dari tengah klaster. Pusat klaster baru akan ditentukan bila semua data telah ditetapkan dalam klaster terdekat. Proses penentuan pusat klaster dan penempatan data dalam klaster diulangi sampai nilai *centroid* tidak berubah lagi.

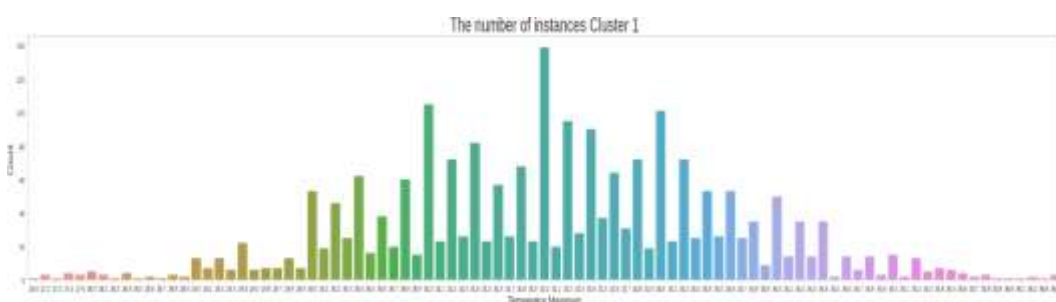
3.1 Dataset

Dataset yang akan dianalisis di ambil dari data iklim pada website Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) yaitu dari pengamatan sinoptik Stasiun Meteorologi Tardamu Kupang. Dataset yang digunakan yaitu dari periode Januari 2010 sampai Oktober 2022. Data tersebut terdiri dari atribut variabel Temperature Minimum (Tn), Temperatur Maksimum (Tx), Temperatur Rata-rata (Tavg), Kelembapan Rata-rata (RH_avg), Curah Hujan (RR), Kecepatan Angin Maksimum (ff_x), Arah Angin Saat Kecepatan Maksimum (ddd_x), Kecepatan Angin Rata-rata (ff_avg), dan Arah Angin Terbanyak (ddd_car).

Pada tahap pra-pemrosesan data, data akan digabungkan dan dari data tersebut akan dihitung jumlah *missing values*-nya. Pra-pemrosesan bertujuan untuk menghilangkan noise, missing value, dan melakukan normalisasi pada data. Nilai yang tidak terukur seperti 8888 dan 9999 diubah menjadi 0 bertujuan untuk mengisi nilai yang kosong.

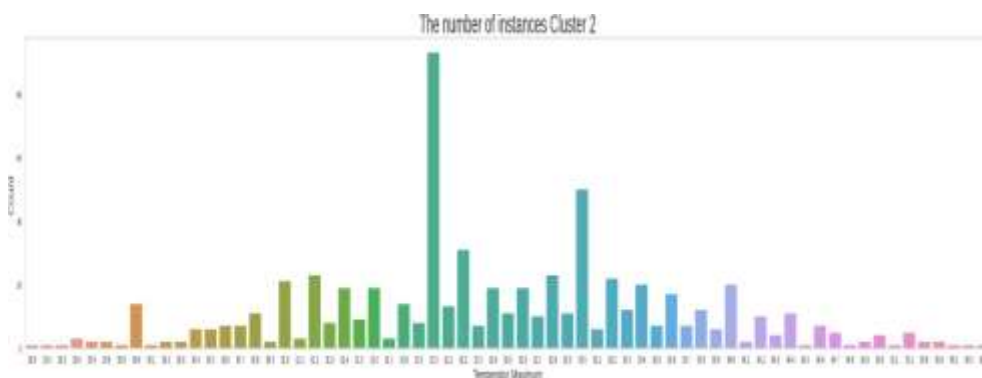
3. HASIL DAN PEMBAHASAN

Pada eksperimen ini, data yang digunakan adalah data cuaca dari BMKG Kupang Stasiun Meteorologi Tardamu. Eksperimen yang dilakukan memiliki tujuan, yaitu untuk memprediksi temperatur maksimum dan temperatur rata-rata tahun dengan menggunakan metode K-means Clustering. Hasil visualisasi prediksi data aktual Cluster 1 Temperatur maksimum dapat dilihat pada Gambar 1 berikut.



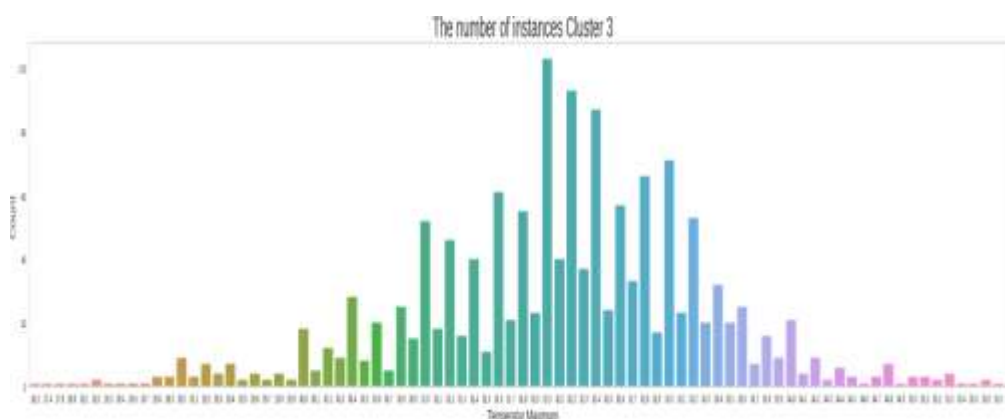
Gambar 1. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum Cluster 1.

Hasil visualisasi prediksi data aktual Cluster 2 Temperatur maksimum dapat dilihat pada Gambar 2 berikut.



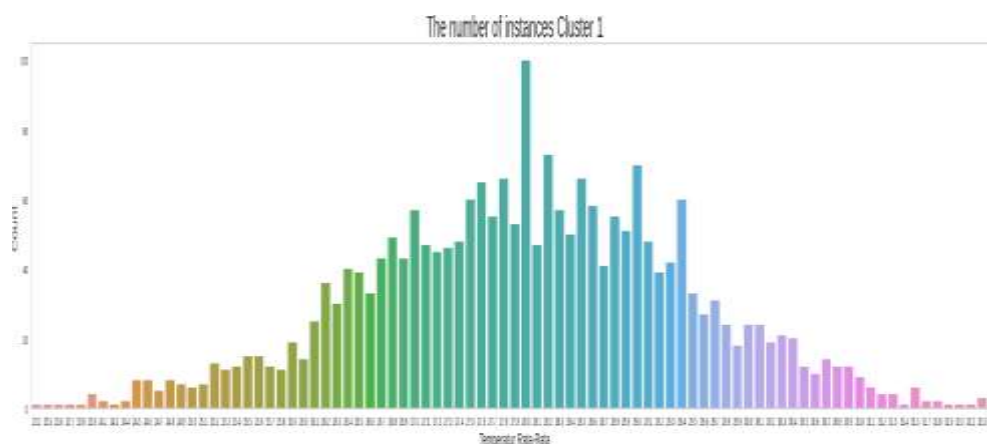
Gambar 2. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum *Cluster 2*

Hasil visualisasi prediksi data aktual Cluster 3 Temperatur maksimum dapat dilihat pada Gambar 3 berikut.



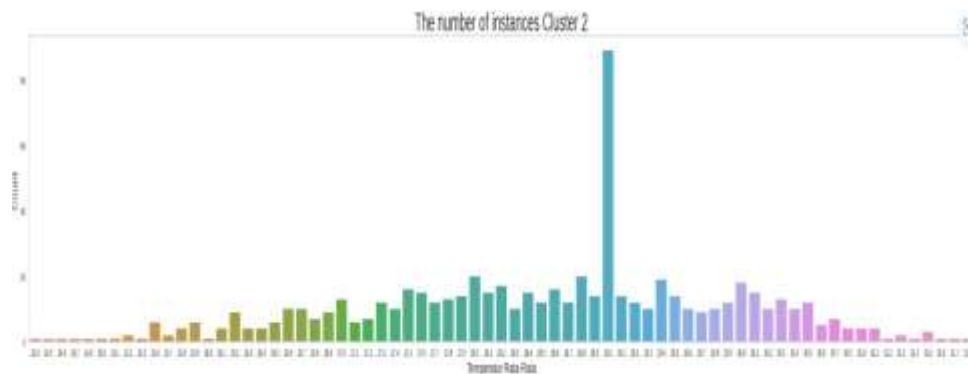
Gambar 3. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum *Cluster 3*

Hasil visualisasi prediksi data aktual Cluster 4 Temperatur maksimum dapat dilihat pada Gambar 4 berikut.



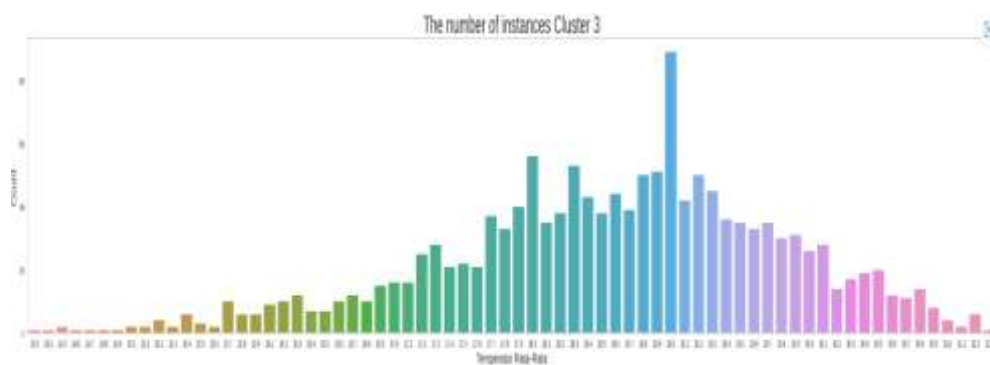
Gambar 4. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum *Cluster 4*

Hasil visualisasi prediksi data aktual Cluster 5 Temperatur maksimum dapat dilihat pada Gambar 5 berikut.



Gambar 5. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum Cluster 5

Hasil visualisasi prediksi data aktual Cluster 6 Temperatur maksimum dapat dilihat pada Gambar 6 berikut.



Gambar 8. Hasil Visualisasi Prediksi Data Aktual Temperatur Maksimum Cluster 6

Dari eksperimen ini untuk memprediksi Temperatur maksimum dan Temperatur rata-rata dengan menggunakan metode *K-means Clustering* menggunakan data latih tahun 2012 sampai dengan 2020. eksperimen ini menghasilkan silhouette average score dengan nilai 0.66. Dimana nilai yang tinggi menunjukkan bahwa objek tersebut cocok dengan *cluster*-nya sendiri dan kurang cocok dengan cluster yang berdekatan. Jika sebagian besar objek memiliki nilai tinggi, maka konfigurasi pengelompokan sesuai. Jika banyak titik memiliki nilai rendah atau negatif, maka konfigurasi pengelompokan mungkin memiliki terlalu banyak atau terlalu sedikit kluster.

4. KESIMPULAN

Dari hasil eksperimen latih data BMKG kota Kupang dari tahun 2012 sampai tahun 2018 dengan tujuan memprediksi nilai pada tahun 2019 mendapatkan hasil *silhouette average score* dengan nilai 0.66, sehingga nilai yang tinggi menunjukkan bahwa objek tersebut cocok dengan cluster-nya sendiri dan kurang cocok dengan cluster yang berdekatan. Jika sebagian besar objek memiliki nilai tinggi, maka konfigurasi pengelompokan sesuai.

Jika banyak titik memiliki nilai rendah atau negatif, maka konfigurasi pengelompokan mungkin memiliki terlalu banyak atau terlalu sedikit kluster.. Untuk menemukan jumlah cluster yang optimal. Analisis *silhouette* dapat digunakan untuk mempelajari jarak pemisahan antara cluster yang dihasilkan dan dapat dianggap sebagai metode yang lebih baik. Analisis Silhouette juga memiliki keuntungan tambahan untuk menemukan penciran jika ada dalam sebuah cluster.

DAFTAR PUSAKA

- [1] Agusta, Y. 2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol.3 , 47- 60.
- [2] E. Rendón, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus External cluster validation indexes," vol. 5, no. 1, 2011.
- [3] Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.Y..
- [4] Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. Rainstek: Jurnal Terapan Sains & Teknologi, 1(3), 53-68.
- [5] Daniel Riano Kaparang dan Eko Sedyono. (2013). "Penentuan Alih Fungsi Lahan Marginal Menjadi Lahan Pangan Berbasis Algoritma K-means di Wilayah Kabupaten Boyolali." JdC. 2. 20.