

ABSTRAK

Latar Belakang: Penyakit Tidak Menular (PTM) adalah penyebab utama kesakitan dan kematian di seluruh dunia dengan jumlah kasus yang terus meningkat. Faktor risiko PTM, baik perilaku maupun fisiologis, memiliki pengaruh signifikan terhadap peningkatan ini. Melalui manajemen data dan informasi, salah satu cara untuk meningkatkan pengendalian PTM secara efisien dan terfokus adalah dengan memanfaatkan data faktor risiko PTM yang ada untuk dikelompokkan dan dianalisis faktor risiko PTM di setiap daerahnya, serta divisualisasikan ke dalam bentuk pemetaan.

Tujuan: Menganalisis data faktor risiko PTM dengan menggunakan teknik *data mining clustering* dengan algoritma K-Means dan memvisualisasikan data spasial berdasarkan faktor risiko penyakit tidak menular di DIY.

Metode: Jenis penelitian ini adalah deskriptif kuantitatif dengan rancangan *cross-sectional*. Data yang digunakan adalah data sekunder faktor risiko perilaku dan fisiologis PTM dari 78 kecamatan di DIY tahun 2022-2023. Penelitian dilaksanakan di Dinas Kesehatan DIY pada bulan April – Mei 2024. *Clustering* K-Means dilakukan dengan Python dan visualisasi data spasial menggunakan Tableau Public.

Hasil: Penerapan *clustering* K-Means berhasil mengidentifikasi empat *cluster* optimal. *Cluster* dengan banyaknya variabel faktor risiko perilaku PTM yang tinggi sebagian besar tersebar di kecamatan di Kabupaten Gunungkidul, yaitu 10 dari 14 kecamatan yang tergolong *cluster* tersebut, sementara untuk faktor risiko fisiologis PTM hanya ditemukan di Kecamatan Kraton, Kota Yogyakarta.

Kesimpulan: *Data mining clustering* telah berhasil diterapkan dan menghasilkan pola persebaran *cluster* yang jelas di antara 78 kecamatan di DIY. Hasil ini dapat memberikan wawasan penting untuk merencanakan tindakan preventif sesuai dengan karakteristik setiap *cluster*.

Kata Kunci: *data mining*, *clustering*, K-Means, Penyakit Tidak Menular, visualisasi data spasial

ABSTRACT

Background: *Non-communicable diseases (NCDs) are the leading causes of morbidity and mortality globally, with an increasing number of cases. Risk factors for NCDs, such as behavioral and physiological risk factors, play a significant role in this increase. Through data and information management, one way to improve NCD control efficiently and purposefully is by utilizing existing risk factor data for clustering, analyzing NCD risk factors in each region, and visualizing the results as maps.*

Objective: *Analyzed NCD risk factor data using data mining clustering techniques with the K-Means algorithm and visualized spatial data based on NCD risk factors in the Special Region of Yogyakarta.*

Method: *This study was a quantitative descriptive research with a cross-sectional design. The data were secondary data on behavioral and physiological risk factors for NCDs from 78 sub-districts in the Special Region of Yogyakarta for 2022-2023. The research was conducted at the Dinas Kesehatan DIY from April to May 2024. K-Means clustering was performed using Python and spatial data visualization was done using Tableau Public.*

Output: *The implementation of K-Means clustering successfully identified four optimal clusters. The cluster with the highest number of behavioral risk factors was mostly found in Gunungkidul regency, with 10 out of 14 sub-districts classified in this cluster, whereas physiological risk factors were only found in the Kraton sub-district of Yogyakarta city.*

Conclusion: *Data mining clustering has been successfully applied, revealing distinct cluster distribution patterns among 78 sub-districts in the Special Region of Yogyakarta. This provides valuable insights for planning preventive actions according to the characteristics of each cluster.*

Keywords: *data mining, clustering, K-Means, Non-communicable Disease, data spatial visualization*