

# Implementasi K-Prototypes dengan Seleksi Fitur dalam Clustering Pasien Kanker Serviks berdasarkan Faktor Risiko = Implementation of K-Prototypes with Feature Selection in Clustering Cervical Cancer Patients based on Risk Factors

Wanda Puspita Hati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553557&lokasi=lokal>

---

## Abstrak

Kanker menjadi penyebab utama kematian global yang menyebabkan hampir 10 juta kematian atau hampir seperenam kasus kematian pada tahun 2020. Tindakan pencegahan primer yang efektif dapat mencegah setidaknya 40% kasus kanker. Tingkat kematian akibat kanker di negara berkembang lebih tinggi dibandingkan dengan negara maju, mencerminkan kesenjangan dalam menangani faktor risiko, keberhasilan deteksi, dan pengobatan. Wanita di negara berkembang paling sering menderita kanker serviks. Masyarakat terutama wanita memerlukan pengetahuan mengenai faktor risiko kanker serviks. Salah satu solusi potensial untuk masalah ini peran machine learning dalam mempelajari data pasien kanker serviks. Penelitian ini menggunakan algoritma clustering K-Prototypes, yang dapat mengelompokkan data campuran, baik numerik maupun kategorik. Data faktor risiko kanker serviks dari pasien di RSUPN X digunakan dalam penelitian ini. Seleksi fitur dilakukan untuk meningkatkan kinerja algoritma K-Prototypes, dengan membandingkan seleksi fitur menggunakan Variance Threshold dan Correlation Coefficient. Kinerja algoritma K-Prototypes terbaik didapatkan dengan menggunakan Correlation Coefficient yang ditinjau berdasarkan Silhouette Coefficient sebesar 0,6; Davies-Bouldin Index sebesar 0,6; dan Callinzki-Harabasz Index sebesar 1.080. Interpretasi cluster yang terbentuk dari simulasi menghasilkan perbedaan utama karakteristik faktor risiko dari dua cluster, yaitu umur, menopause, dan kondisi kesehatan seperti keputihan, pendarahan, nyeri perut bawah, dan penurunan nafsu makan. Sementara, faktor terkait riwayat terdahulu, kesehatan reproduksi, dan masalah gizi tidak menunjukkan perbedaan yang signifikan. Algoritma K-Prototypes diharapkan dapat menjadi solusi dalam mengidentifikasi kelompok berdasarkan faktor risiko kanker serviks untuk membantu tenaga medis dalam mengambil keputusan dan tindakan selanjutnya, serta pengetahuan bagi masyarakat.

.....Cancer is a leading cause of death worldwide, resulting in nearly 10 million deaths, or almost one-sixth of all deaths, in 2020. Effective primary prevention measures can prevent at least 40% of cancer cases. Cancer mortality rates are higher in developing countries compared to developed countries, reflecting disparities in addressing risk factors, detection success, and available treatments. Women in developing countries most frequently suffer from cervical cancer. It is crucial for communities, especially women, to have knowledge about the risk factors for cervical cancer. One potential solution to this issue is the role of machine learning in analyzing cervical cancer patient data. This study uses the K-Prototypes clustering algorithm, which can cluster mixed data, both numerical and categorical. Cervical cancer risk factor data from patients at X National General Hospital were used in this research. Feature selection was performed to improve the performance of the K-Prototypes algorithm, comparing feature selection using Variance Threshold and Correlation Coefficient. The best performance of the K-Prototypes algorithm was obtained using the Correlation Coefficient, as reviewed based on a Silhouette Coefficient of 0,6; a Davies-Bouldin Index of 0,6; and a Callinzki-Harabasz Index of 1.080. Interpretation of the clusters formed from the

simulation revealed major differences in the characteristics of risk factors between two clusters, namely age, menopause, and health conditions such as leukorrhea, bleeding, lower abdominal pain, and loss of appetite. Meanwhile, factors related to previous history, reproductive health, and nutritional issues did not show significant differences. The K-Prototypes algorithm is expected to be a solution in identifying groups based on cervical cancer risk factors to assist medical professionals in decision-making and subsequent actions, as well as to provide knowledge to the public.