ABSTRACT

Cancer remains a significant public health issue in Indonesia and worldwide. Central Java Province is one of the regions with the highest cancer rates, at 1.7 per 1000 population. Identifying the most influential cancer risk factors in each region is essential for more selective prevention strategies. This study aims to determine the geographic distribution of cancer risk factors in Central Java Province using the Random Forest method to identify global feature importance, and SHAP to analyze the contribution of risk factors at the local level. The results indicate that the presence of sugary drink outlets is the most influential factor in 28.6% of regions in Central Java. Environmental factors, such as air pollution (SO_2, NO_2) and slum areas, affect 22.9%, 11.4%, and 5.7% of the regions. Understanding the geographic characteristics of socio-economic, environmental, and dietary factors related to cancer cases is crucial for the design of more targeted policies and decision-making.

Keywords*: cancer, risk factors, machine learning, Random Forest, SHAP, Central Java, geographic distribution.*