Modelling Stroke Risk Based on Age, Objective Factors, and Food Consumption Using Nonparametric Binary Logistic Regression with Least Square Spline Estimator

Nur Chamidah

Faculty of Science and Technology, Universitas Airlangga, Indonesia

Marisa Rifada

Faculty of Science and Technology, Universitas Airlangga, Indonesia

I Nyoman Budiantara

Faculy of Science and Data Analysis, Institut Teknologi Sepuluh Nopember, Indonesia

Muhammad Anshari

The Institute of Policy Studies, Universiti Brunei Darussalam, Brunei Darussalam

Naufal Ramadhan Al Akhwal Siregar

Faculty of Science and Technology, Universitas Airlangga, Indonesia

Renatalia Fika

Faculty of Science and Technology, Universitas Airlangga, Indonesia

Abstract

Stroke is a leading global cause of mortality and disability, which is categorized into ischemic and hemorrhagic stroke. This study develops a nonparametric binary logistic regression model based on the least square spline estimator (LSS-NBLR) to assess the risk of ischemic and hemorrhagic stroke. The model estimates the odds ratio for each risk factor, allowing for flexible analysis of how variables impact stroke risk. Risk factors analyzed include age, LDL cholesterol, uric acid, triglycerides, blood glucose and sodium levels. Data were sourced from Dr. Drs. M. Hatta Brain Hospital, Bukittinggi, Indonesia, in 2023. The results indicate that age ≥ 50 years, LDL ≥ 125 mg/dL, triglycerides ≥ 125 mg/dL, and blood glucose ≥ 150 mg/dL increase the odds of ischemic stroke. In contrast, uric acid and sodium levels consistently showed odds ratios below 1 across all spline segments, indicating a decreasing tendency in the odds. The LSS-NBLR model achieved 87.14% accuracy for in-sample data and 93.3% for out-of-sample data. These findings highlight the model's flexibility and accuracy in stroke risk prediction. Furthermore, this study contributes to SDGs by supporting early detection and prevention strategies for stroke through data-driven biostatistical modeling.