Data Science-Based Key Factor Analysis and Risk Prediction of Diabetic

Authors : Fei Gao, Rodolfo C. Raga Jr.

Abstract : This research proposal will ascertain the major risk factors for diabetes and to design a predictive model for risk assessment. The project aims to improve diabetes early detection and management by utilizing data science techniques, which may improve patient outcomes and healthcare efficiency. The phase relation values of each attribute were used to analyze and choose the attributes that might influence the examiner's survival probability using Diabetes Health Indicators Dataset from Kaggle's data as the research data. We compare and evaluate eight machine learning algorithms. Our investigation begins with comprehensive data preprocessing, including feature engineering and dimensionality reduction, aimed at enhancing data quality. The dataset, comprising health indicators and medical data, serves as a foundation for training and testing these algorithms. A rigorous cross-validation process is applied, and we assess their performance using five key metrics like accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). After analyzing the data characteristics, investigate their impact on the likelihood of diabetes and develop corresponding risk indicators.

Keywords : diabetes, risk factors, predictive model, risk assessment, data science techniques, early detection, data analysis, Kaggle

Conference Title : ICBDMLH 2023 : International Conference on Big Data and Machine Learning for Healthcare

Conference Location : Jeddah, Saudi Arabia

Conference Dates : November 20-21, 2023