Model Explainability in Healthcare Risk Prediction using SHAP .

Abstract

Machine Learning models are increasingly used for healthcare risk prediction; however, their lack of interpretability poses challenges for clinical trust and adoption. This project investigates model explainability in healthcare risk prediction using Shapley Addictive explanations (SHAP). Supervised learning models were trained on structured clinical datasets to predict disease risk, and SHAP values were used to quantify the contribution of individual features to model predictions at both global and local levels. The analysis revealed clinically meaningful feature importance patterns patterns and highlighted differences between population-level trend and individual-level explanations. Experimental results demonstrate that SHAP-based interpretability improves transparency without compromising predictive performance, enabling better understanding of model behavior. This work underscores the role of explainable AI techniques in supporting trustworthy and accountable machine learning systems for heathcare decision-Maker.