

ABSTRACT

Machine Learning and Pareto Analysis for Institutional Healthcare Fraud Detection

by

A. A. S. Wicaksono 19/444302/PA/19364

Institutional healthcare fraud poses a complex and significant threat, leading to substantial financial losses and compromising the integrity of healthcare systems. Traditional fraud detection approaches often overlook institutional actors. This research addresses that gap by developing and evaluating a robust machine learning framework specifically designed to identify fraudulent activities originating from healthcare institutions, with the goal of enhancing detection capabilities.

The study utilizes relevant healthcare datasets and conducts a comparative analysis of various machine learning algorithms, selecting the most effective models based on Area Under the Curve (AUC) score. Given the complexity of the data, Pareto Analysis (80/20 principle) is applied for feature selection to improve the model efficiency and interpretability. This research aims to deliver a validated and optimized methodology, offering insights into effective computational techniques and key data features for combating institutional healthcare fraud, ultimately benefiting regulatory bodies and healthcare integrity programs.

ABSTRAK

Pembelajaran Mesin dan Analisis Pareto untuk Pendeteksian Kecurangan Kesehatan Institusional

oleh

A. A. S. Wicaksono 19/444302/PA/19364

Penipuan institusional dalam sektor layanan kesehatan merupakan ancaman yang kompleks dan signifikan, menyebabkan kerugian finansial yang besar serta melemahkan integritas sistem kesehatan. Deteksi penipuan konvensional sering kali mengabaikan aktor-aktor institusional. Penelitian ini bertujuan untuk mengatasi kesenjangan tersebut dengan mengembangkan dan mengevaluasi kerangka kerja pembelajaran mesin yang tangguh, yang secara khusus dirancang untuk mengidentifikasi aktivitas penipuan yang berasal dari institusi layanan kesehatan, dengan tujuan meningkatkan kapabilitas deteksi.

Penelitian ini memanfaatkan dataset layanan kesehatan yang relevan serta melakukan analisis komparatif terhadap berbagai algoritma pembelajaran mesin, dengan memilih model paling efektif berdasarkan skor *Area Under Curve*. Mengingat kompleksitas data yang tinggi, analisis Pareto (prinsip 80/20) diterapkan dalam seleksi fitur untuk meningkatkan efisiensi dan interpretabilitas model. Studi ini bertujuan menghasilkan metodologi yang tervalidasi dan teroptimasi, serta memberikan wawasan mengenai teknik komputasi yang efektif dan seleksi fitur dalam upaya pemberantasan penipuan institusional di sektor layanan kesehatan.