

3. ABSTRACT

Diabetes Mellitus has a substantial impact on Indonesians, affecting 88.51% of those aged 44-94, as revealed by the Indonesia Health Insurance program. The urgency of this health concern is underscored by the International Diabetes Federation's alarming statistic: one person succumbs to Diabetes Mellitus every five seconds, demanding immediate attention to diagnosis and risk reduction. This study focuses on predicting the Length of Stay (LOS) of Diabetes Mellitus patients using three machine learning methods: Logistic Regression, Random Forest, and XGBoost. Concurrently, Process Mining, encompassing Process Discovery and Conformance Checking, elevates model quality. Key factors influencing patient LOS, including the location of healthcare facility, patient age, and arrival time, are unveiled. Random Forest achieves an accuracy score 0.88 with an F1 score of 0.82, Precision of 0.82, and Recall of 0.81, with a Time Prediction of 0.1027 seconds. This accuracy is higher compared to Logistic Regression with accuracy 0.76, F1 score 0.63, Precision 0.57, Recall 0.6 and time prediction 0.00062. While XGBoost gets an accuracy of 0.86, F1 score 0.79, Precision 0.79, Recall 0.79 and time prediction 0.06499. Certain medical procedures, such as those involving 'Diabetes & Nutritional/Metabolic Disorders,' represent treatments with the longest sojourn time and the highest frequency of cases across all LOS categories. Giving procedures multiple times can affect the patient's LOS, especially for procedure Inpatient with Peripheral Vascular Disorders with mild, moderate, and severe level procedures and Inpatient Care in the General Ward. Facility usage like is Puskesmas being the primary destination for diabetes patients and has longer lengths of stay compared to general practitioners (Dokter Umum) or primary clinics (Klinik Pratama). The finding of this study underscores the pivotal role of model selection and class distribution in LOS prediction. Random Forests get superior performance, than Logistic Regression and XGBoost.

Keywords: Machine Learning, Process Mining, Length of Stay, Logistic Regression, Random Forest XGBoost, BPJS