

4. ABSTRAK

Diabetes Mellitus adalah salah satu penyakit yang memiliki dampak signifikan pada penduduk Indonesia, sekitar 88,51% penduduk Indonesia yang berusia 44-94 tahun terdiagnosa Diabetes Militus, seperti yang diungkapkan oleh program Asuransi Kesehatan Indonesia. Menurut data kesehatan oleh *International Diabetes Foundation* (IDF): satu orang meninggal akibat Diabetes Mellitus setiap lima detik, sehingga penanganan terhadap penyakit diabetes militus perlu segera dihadapi agar dapat mengurangi risiko yang terjadi. Penelitian ini berfokus pada prediksi *Length of Stay* (LOS) pasien Diabetes Mellitus, menggunakan 3 metode machine learning : Logistic Regression, Random Forest dan XGBoost. Secara bersamaan, Process Mining digunakan dalam penelitian ini dengan mencakup Process Discovery dan Conformance Checking, untuk meningkatkan kualitas model. Faktor kunci yang memengaruhi LOS pasien, antara lain lokasi fasilitas kesehatan, usia pasien, dan waktu kedatangan. Random Forest mendapatkan nilai akurasi 0,88 dengan *F1 score* sebanyak 0.82, *Precision* 0.82 and *Recall* 0.81 and *Time Prediction* 0,1027 detik. Akurasi ini lebih besar dibandingkan dengan Logistic Regressi hanya mendapatkan akurasi 0,76, *F1 score* 0,63 , *Precision* 0,57, *Recall* 0,6 dan *Time Prediction* 0,00062. Sedangkan XGBoost mendapatkan akurasi 0,86, *F1 score* 0,79, *Precision* 0,79, *Recall* 0.79 dan *Prediction Time* 0,06499 detik. Prosedur medis tertentu, seperti yang melibatkan '*Diabetes & Nutritional/Metabolic Disorders*' merupakan procedure treatment dengan sojourn time terlama dan case yang paling banyak terjadi pada semua kategori LOS. Prosedur yang diberikan secara berulang dapat berpengaruh pada LOS (Length of Stay) pasien, terutama untuk prosedur Perawatan Rawat Inap dengan Gangguan Pembuluh Darah Perifer tingkat ringan, sedang, dan berat dan Perawatan Rawat Inap di Ruang Umum. Sedangkan pasien yang melakukan kunjungan Puskesmas sebagai fasilitas Kesehatan memiliki LOS yang lebih lama dibandingkan dengan Dokter umum dan Klinik Pratama. Temuan dari penelitian ini menekankan peran penting pemilihan model dan distribusi kelas dalam prediksi LOS. Random Forest menjadi model yang lebih unggul dibandingkan dengan Regresi Logistik dan XGBoost.

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