

ABSTRAK

Diabetes merupakan salah satu penyakit kronis yang semakin meningkat prevalensinya di seluruh dunia. Pencegahan dan edukasi mengenai diabetes menjadi prioritas penting dalam agenda kesehatan global. Implementasi teknologi seperti *data mining* dan prediksi berbasis *machine learning* dalam diagnosis dini dan pengelolaan diabetes juga merupakan langkah signifikan untuk mengurangi dampak dari penyakit ini. *Data mining* membangun model yang dapat memprediksi apakah seseorang terkena penyakit diabetes atau tidak. Baik dan tidaknya model dalam melakukan klasifikasi dapat ditentukan dari nilai akurasi. Hasil dari akurasi pemodelan klasifikasi tidak menentu hasilnya selalu baik atau tinggi, peningkatan nilai akurasi dapat dilakukan dengan melakukan seleksi fitur. Seleksi fitur berperan penting dalam mengurangi kompleksitas model dengan mengeliminasi fitur-fitur yang tidak relevan atau memiliki korelasi yang tinggi satu sama lain. Data yang digunakan dalam penelitian ini adalah data diabetes yang terdiri dari beberapa fitur yang berpotensi mempengaruhi diagnosis diabetes. Implementasi dilakukan menggunakan bahasa pemrograman *Python 3* di *Google Colaboratory*. Setelah dilakukan seleksi fitur, model Regresi Logistik dibangun dan dievaluasi dengan menggunakan matriks akurasi, presisi, *recall*, dan skor-F1. Hasil penelitian menunjukkan bahwa seleksi fitur menggunakan *Pearson Correlation* berhasil meningkatkan performa akurasi model prediksi diabetes sebesar 1,03% dari akurasi tanpa seleksi fitur 89,74% meningkat menjadi 90,77% setelah dilakukan seleksi fitur.

Kata Kunci: Seleksi fitur, *Pearson Correlation*, Regresi Logistik, Prediksi Diabetes