

ABSTRAK

Diabetes merupakan penyakit yang terus meningkat dan semakin tinggi kasus kematian yang memakan korban, Penyakit kronis serius ini disebabkan oleh gangguan metabolik yang terjadi karena pankreas tidak dapat menghasilkan atau memproduksi cukup Insulin (hormon yang mengatur glukosa). Menurut *Internasional of Diabetic Federation (IDF)* tingkat prevalensi global penderita diabetes terus meningkat setiap tahunnya. Diabetes merupakan salah satu penyakit paling umum dan menjadi penyebab kematian terbesar di dunia. Pendeteksian penyakit diabetes dapat dilakukan dengan teknik *data mining*. *Data Mining* merupakan suatu proses pengumpulan informasi penting dari sebuah data yang besar pada suatu keahlian yang berkaitan dengan informatika. *Data Mining* juga dapat digunakan pada penelitian yang bergerak di aspek lainnya, salah satunya pada bagian kesehatan untuk melakukan prediksi penyakit Diabetes pada suatu kelompok individu dengan metode klasifikasi. Pada penelitian ini, *dataset* yang digunakan berasal dari *Pima Indians Diabetes Databases (PPID)*. Penelitian ini bertujuan untuk melakukan perbandingan performa klasifikasi dari Algoritma *Supervised Learning*, yaitu *Naïve Bayes* dan *XGBoost*. Penelitian ini juga akan melakukan penanganan *missing value* terhadap *dataset* dan membahas mengenai metode *Grid Search* sebagai optimisasi berdasarkan kinerja akurasi klasifikasi penyakit diabetes pada Algoritma *Naïve Bayes* dan *XGBoost*. Hasil akurasi dievaluasi dengan menggunakan *confusion matrix* serta menghitung nilai AUC. Sehingga dari hasil klasifikasi, didapat model klasifikasi Algoritma *Naïve Bayes* dengan nilai hasil akurasi model sebesar 79.68% dan Algoritma *XGBoost* memiliki performa yang lebih baik dengan nilai hasil akurasi yang didapat sebesar 90.10%.

Kata kunci— *Data Mining*, klasifikasi, *Diabetes*, *Naive Bayes*, *XGBoost*