

## ABSTRAK

Penyakit diabetes merupakan penyakit yang cukup berbahaya yang ditandai oleh tingginya gula darah yang melebihi batas normal. Ada beberapa faktor yang menjadi penyebab penyakit diabetes diantaranya jenis kelamin, usia, hipertensi, penyakit jantung, riwayat merokok, index masa tubuh (imb), hemoglobin dan tingkat glukosa darah. Menurut *International Diabetes Federation* yang mengidap penyakit diabetes pada tahun 2015 sebanyak 415 juta jiwa, dan diperkirakan akan terus meningkat setiap tahunnya. Upaya yang dapat dilakukan untuk mencapai tujuan tersebut adalah dengan dikembangkannya suatu sistem untuk memprediksi penyakit diabetes dengan prinsip klasifikasi untuk memprediksi apakah seseorang terkena diabetes atau tidak. Dataset yang digunakan pada penelitian ini diambil dari website kaggle, yaitu *diabetes prediction dataset*. Metode klasifikasi yang digunakan yaitu *Naive Bayes* dan *Support Vector Machine (SVM)*. Hasil dari penelitian ini disimpulkan bahwa dari ketiga pengujian tersebut dengan *percentage split* 60%, 70% dan 80% menunjukkan bahwa algoritma *Support Vector Machine (SVM)* mendapatkan nilai persentase yang lebih tinggi dari algoritma *Naive Bayes* kecuali pada pengujian kedua dengan *percentage split* 70% yaitu pada parameter *recall* nya, algoritma *Naive Bayes* lebih unggul daripada *Support Vector Machine (SVM)*. Hal itu menunjukkan bahwa data gagal diklasifikasi dengan baik. Pengujian pertama *Support Vector Machine (SVM)* pada *percentage split* 60% dengan parameter *accuracy* mendapatkan nilai 96,40%, *precision* 99,78% dan *recall* 96,39%. Kemudian pengujian kedua *Support Vector Machine (SVM)* pada *percentage split* 70% dengan parameter *accuracy* mendapatkan nilai 97,07%, *precision* 100% dan *recall* 96,90% dan terakhir pada pengujian ketiga *Support Vector Machine (SVM)* dengan *percentage split* 80% pada parameter *accuracy* mendapatkan nilai 98%, *precision* 100% dan *recall* 97,90%.

Kata kunci : diabetes, klasifikasi, *Naive Bayes* dan *Support Vector Machine (SVM)*

## **ABSTRACT**

*Diabetes is a disease that is quite dangerous which is characterized by high blood sugar that exceeds normal limits. There are several factors that cause diabetes including gender, age, hypertension, heart disease, smoking history, body mass index (IMB), hemoglobin and blood glucose level. According to the International Diabetes Federation, there were 415 million people with diabetes in 2015, and it is estimated that this will continue to increase every year. Efforts that can be made to achieve this goal is to develop a system to predict diabetes with the principle of classification to predict whether a person has diabetes or not. The dataset used in this study was taken from the kaggle website, namely the diabetes prediction dataset. The classification method used is Naive Bayes and Support Vector Machine (SVM). The results of this study concluded that the three tests with a percentage split of 60%, 70% and 80% show that the Support Vector Machine (SVM) algorithm gets a higher percentage value than the Naive Bayes algorithm except in the second test with a percentage split of 70%, namely on the recall parameter, the Naive Bayes algorithm is superior to the Support Vector Machine (SVM). This shows that the data failed to be classified properly. The first test of the Support Vector Machine (SVM) at a percentage split of 60% with accuracy parameters obtained a value of 96.40%, precision of 99.78% and recall of 96.39%. Then the second test of Support Vector Machine (SVM) on a percentage split of 70% with accuracy parameters getting a value of 97.07%, precision of 100% and recall of 96.90% and finally on the third test of Support Vector Machine (SVM) with a percentage split of 80% on the accuracy parameters get a value of 98%, precision 100% and recall 97.90%.*

*Keywords : diabetes, classification, Naive Bayes and Support Vector Machine (SVM)*