

## ABSTRAK

Penyakit diabetes adalah penyakit kronis yang berupa gangguan metabolisme yang ditandai dengan kadar gula darah yang meningkat lebih dari batas normal. Jumlah pasien diabetes selal yang Setelah dilakukan pelatihan dan pengujian data dengan model klasifikasi *Random Forest* dengan beberapa rasio *split*, model yang memiliki nilai akurasi tertinggi adalah model yang menggunakan rasio *split* 70:30 dengan nilai akurasi 72,89%. Waktu yang digunakan untuk melatih dan menguji dengan model ini juga relatif rendah dengan waktu latih 21,23 detik dan waktu uji 0,76 detik. Karena dataset yang digunakan memiliki data label 0 (tidak terdiagnosis) yang jauh lebih banyak dibandingkan data label 1 (terdiagnosis), hasil yang diperoleh model-model tersebut tidak seakurat mungkin.

**Kata Kunci:** Diabetes, Data Mining, Klasifikasi

## **ABSTRACT**

Diabetes is a chronic disease in the form of metabolic disorders that are indicated by the increase of glucose in a patient's blood to an abnormal level. The amount of diabetes patients continues to increase all around the world. The diagnoses of diabetes on a patient can be predicted with the Random Forest classification algorithm. After training and testing data with the Random Forest classification model with different split ratios, the model with the highest accuracy is the model that uses the split ratio of 70:30 with 72,89% accuracy. The time taken to learn and test with this model is also relatively short with 21,23 seconds of training time and 0,76 seconds of testing time. The result may not be as accurate as possible due to the fact that the amount of data with the label 0 (undiagnosed) is far higher than the amount of data with the label 1 (diagnosed).

**Keywords:** Diabetes, Data Mining, Classification