

## INTISARI

### Perbandingan Metode *Preprocessing* Terhadap Model Artificial Neural Network (ANN) dan Random Forest (RF) dalam Klasifikasi Penyakit

Oleh

Dian Novitaningrum

21/483928/PPA/06177

Permasalahan yang timbul pada data seringkali dipengaruhi oleh tidak bersihnya sebuah data. Data *preprocessing* dapat menjadi rekomendasi untuk penanganan dari permasalahan tersebut. Beberapa teknik *preprocessing* dan *normalization* yang diusulkan digunakan untuk dapat menghasilkan model yang dianggap efisien. Model yang dipilih adalah Algoritma Artificial Neural Network (ANN) dan Random Forest (RF).

Penggunaan *dataset* terdiri dari enam buah *dataset* yaitu *Diabetes Mellitus* penggunaan sebanyak 947 data, *Cardiovascular* penggunaan sebanyak 1000 data, *Heart Disease* sebanyak 1025 data, *Hyphothyroid* sebanyak 1000 data, *Hepatitis* 155 dan *Chronic Kidney Disease* sebanyak 400 data. *Dataset* tersebut dimanipulasi hingga memiliki prosentase missing values sebanyak 10%, 20%, dan 30%. Sedangkan performa model menggunakan *accuracy*, *sensitivity*, *spesificity* dan *f1-score*. Tujuan dengan adanya penelitian ini digunakan untuk mendapatkan rekomendasi dari model penggunaan data *preprocessing* dan mendapat hasil yang optimal untuk penanganan setiap *dataset* yang tersedia. Sehingga penelitian ini diharapkan mampu menjadi rekomendasi yang tepat untuk mengolah data *preprocessing* yang tepat untuk setiap *dataset*nya dan menganalisis perbandingan kinerja dari setiap *dataset* yang dipergunakan.

Hasil penelitian ini teknik MeanMinmax mampu menjadi teknik terbaik pada algoritma ANN, sedangkan untuk algoritma RF teknik terbaik yang didapatkan adalah MeanZscore. Kemudian, disimpulkan bahwa semakin besar jumlah prosentase dalam *generate missing values* sangat menentukan dalam meningkatkan nilai *mean rank* yang dihasilkan. Parameter terbaik dari algoritma ANN adalah fungsi aktivasi yaitu Relu, sedangkan untuk algoritma RF memiliki *criterion* terbaik adalah Log\_loss.

**Kata Kunci :** Klasifikasi Penyakit, Data *Preprocessing*, Friedman, ANN, RF

## **ABSTRACT**

### **Comparison of Preprocessing Methods on Artificial Neural Network (ANN) and Random Forest (RF) Models in Disease Classification**

By

Dian Novitaningrum

21/483928/PPA/06177

Unclean data often influence problems that arise with data. Data preprocessing can be a recommendation for handling this problem. Several proposed preprocessing and normalization techniques are used to produce a model that is considered efficient. The models chosen are the Artificial Neural Network (ANN) and Random Forest (RF) algorithms.

The dataset used consists of six datasets, namely Diabetes Mellitus with 947 data, Cardiovascular with 1000 data, Heart Disease with 1025 data, Hypothyroid with 1000 data, Hepatitis 155, and Chronic Kidney Disease with 400 data. The datasets are manipulated to have 10%, 20%, and 30% missing values. In comparison, the model performance uses accuracy, sensitivity, specificity, and f1-score. This research aims to obtain recommendations for data preprocessing usage models and obtain optimal results for handling each available dataset. It is hoped that this research will provide appropriate recommendations for processing appropriate data preprocessing for each dataset and analyzing the performance of each dataset used.

The results of this research mean that the MeanMinmax technique is able to become the best technique for the ANN algorithm, while for the RF algorithm the best technique obtained is MeanZscore. Then, it was concluded that the greater the percentage of missing values that was crucial in increasing the resulting mean rank value. The best parameter of the ANN algorithm is the activation function, namely Relu, while for the RF algorithm the best criterion is Log\_loss.

**Keywords : Disease Classification, Data Preprocessing, Friedman, ANN, RF**