

# Analisis Kinerja TabNet dan Random forest dalam Mengklasifikasikan Prevalensi Stunting Kabupaten/Kota di Indonesia = Analysis of TabNet and Random forest Performance in Classification of District/City Stunting Prevalence in Indonesia

Anatasya Oktaviani Handriati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551196&lokasi=lokal>

---

## Abstrak

Stunting merupakan kondisi yang terjadi pada balita yang mengakibatkan gagal tumbuh akibat dari kurangnya pemenuhan gizi yang diterima di masa kandungan dan setelah bayi lahir. Prevalensi stunting adalah persentase balita yang mengalami kasus stunting dibandingkan dengan jumlah anak balita di daerah tersebut. WHO mengategorikan prevalensi stunting menjadi 4 kategori yaitu low untuk prevalensi stunting < 20, medium untuk 20 prevalensi stunting < 30, high untuk 30 prevalensi stunting < 40 dan very high, untuk prevalensi stunting 40. Klasifikasi prevalensi stunting dilakukan untuk melihat faktor-faktor apa saja yang mempengaruhi tinggi rendahnya prevalensi stunting pada daerah tersebut. Random forest dan TabNet digunakan sebagai model klasifikasi. Sebelum melakukan klasifikasi, data dinormalisasikan terlebih dahulu jika menggunakan model random forest dan tidak dinormalisasikan jika menggunakan model TabNet. Kemudian pada data training dilakukan SMOTE dampak dari imbalance data. Hyperparameter pada random forest menggunakan Optuna. Model yang sudah dibuat, dievaluasi menggunakan confusion matrix dengan melihat nilai precision, recall, F1- Score, dan accuracy. Model dibandingkan dengan nilai rata-rata akurasi berdasarkan pembagian data training dan data testing di 70:30, 80:20 dan 90:10. Selain itu, dibandingkan hasil klasifikasi antara model random forest dan TabNet. Nilai rata-rata akurasi yang paling tinggi untuk random forest adalah 51% dan TabNet adalah 44,7% saat kedua model di proporsi data 70:30. Model yang paling terbaik adalah TabNet dengan akurasi 53% dibandingkan dengan random forest ketika random state sama dengan nol. Model TabNet memiliki bekerja optimal pada precision, recall dan F1-Score untuk target kelas low. Sedangkan model random forest bekerja optimal pada recall dan F1-Score untuk target kelas medium. Kedua model tidak bekerja dengan optimal di kelas target high dan very high.

.....Stunting is a condition that occurs in toddlers, resulting in failure to grow due to insufficient nutrition during pregnancy and after birth. The prevalence of stunting is the percentage of children under five who experience stunting compared to the total number of children in that area. The World Health Organization (WHO) categorizes the prevalence of stunting into four categories: low for prevalence < 20, medium for 20 prevalence < 30, high for 30 prevalence < 40, and very high for prevalence 40. The classification of stunting prevalence is done to identify factors that influence the high or low prevalence of stunting. Random forest and TabNet are used as classification models. Before performing the classification, the data is normalized, and SMOTE is applied to the training data. The hyperparameters of models are tuned using Optuna is used. The models are evaluated using a confusion matrix by examining precision, recall, F1-Score, and accuracy. The model are compared with the average accuracy value based on the distribution of training data and testing data at 70:30, 80:20 and 90:10. In addition, classification results will be compared between random forest and TabNet models. The highest average accuracy value for random forest is 51% and TabNet is 44.7% when the two models are in the proportion of data 70:30. The best model is TabNet with 53% accuracy compared to the random forest when the random state is zero. The TabNet model has

optimal work on precision, recall and F1-Score for low class targets. While the random forest model works optimally on recall and F1-Score for medium class targets. The two models do not work optimally in the high and very high target classes.