

Laporan Homework

Setelah melakukan preprocessing, pada tahap pemodelan ini kami membuat beberapa model klasifikasi karena memiliki target nilai kategorikal. Beberapa model yang sudah dilakukan yaitu sebagai berikut:

- Model Random Forest
- Model Light GBM
- Model Neural Network
- Model XGBoost
- Model Gradient Boosting
- Model Linear Regression
- Model Linear SVC
- Model Decision Tree
- Model Naive Baiyes

Kami mengevaluasi model machine learning menggunakan metrics Recall karena fokus untuk mengurangi nilai false negative. Setelah melakukan pemodelan dan evaluasi beberapa algoritma didapatkan bahwa:

- Model Random Forest mampu menghasilkan nilai recall sebesar 88.18%.
- Model Light GBM mampu mneghasilkan nilai recall sebesar 83.60%.
- Model Neural Network mampu menghasilkan nilai recall sebesar 67.46%.
- Model XGBoost mampu menghasilkan nilai recall sebesar 71.17%.
- Model Gradient Boosting mampu menghasilkan nilai recall sebesar 71.68%.
- Model Linear Regression mampu menghasilkan recall sebesar 67.64%.
- Model Linear SVC mampu menghasilkan nilai recall sebesar 66.38%.
- Model Decision Tree mampu menghasilkan nilai recall sebesar 61.27%.
- Model Naive Baiyes mampu menghasilkan nilai recall sebesar 32.17%.

Dari info di atas dapat disimpulkan Model Random Forest merupakan base model terbaik, namun kita masih harus melakukan tuning lanjutan agar tidak terjadi overfitting. Kami melakukan Hyperparameter tuning dengan RandomizedSearchCV. Setelah dilakukan hyperparameter tuning berhasil untuk meningkatkan Recall dan mengurangi overfitting dan Model Random Forest mampu menghasilkan Recall sebesar 92.43%.