**PENDAFTARAN PRA-PROPOSAL TUGAS AKHIR**

Nama : Reo Sahobby

NIM : 123170067

Program Studi : Informatika

Keminatan : Sistem Cerdas

Bidang Tugas Akhir : Machine Learning

Bahasa Pemrograman : Python, R, C++, HTML, PHP, CSS, JavaScript, Java

yang dikuasai

Rencana Judul Tugas Akhir : Penerapan *SMOTE* Pada Adult Census Income Datasets Menggunakan

Algoritma XgBoost dan Randomized Search Optimizer.

Calon Dosen Pembimbing : Mangaras Yanu F, S.T., M.Eng.

DESKRIPSI SINGKAT RENCANA TUGAS AKHIR

Sesuai dengan rencana judul tersebut, nantinya saya ingin melakukan penelitian tentang penerapan teknik SMOTE yang dikombinasikan dengan algoritma XgBoost yang menggunakan optimizer RandomizedSearch untuk mencari parameter terbaik.

Datasest yang saya gunakan adalah datasets Adults Census Income, datasets yang dikeluarkan oleh UCI Machine Learning. Datasets tersebut berisi data diri yang digunakan untuk menentukan pendapatan, terdapat dua kelas label target di dalam datasets tersebut, label target yang ada dirasa tidak seimbang karena terdapat label <=50k sebanyak 76% dan label >50k sebanyak 24% dari total keseluruhan data.

Teknik SMOTE banyak digunakan untuk menangani permasalahan *imbalance* data (data tidak seimbang), dengan cara kerja yaitu membuat data sintetik pada kelas minor (label >50k) sebanyak kelas major (label <=50k). Hasil dari SMOTE akan membuat semua kelas target berjumlah sama. Diharapkan adanya peningkatan akurasi model, apabila *preprocessing* data dilakukan menggunakan teknik SMOTE.

Algoritma klasifikasi yang saya gunakan adalah XgBoost karena memiliki performa yang baik dalam melakukan klasifikasi, dan memiliki *regularization* sehingga menghindari model yang overfit.

Rencana yang akan saya lakukan adalah:

Saya ingin membandingkan hasil performa dari model algoritma XgBoost yang menggunakan SMOTE, dan algoritma XgBoost tanpa menggunakan SMOTE. Parameter hasil yang akan diukur meliputi nilai akurasi, presisi, recall, ROC, AUC, dan F1 score.

Nantinya model akan ditraining menggunakan optimizer Randomized Search untuk mencari gradient descent terbaik dan menemukan parameter terbaik, dan menggunakan K-Fold Cross Validation untuk memvalidasi model. Dari hasil perbandingan performa tersebut, nantinya dapat digunakan untuk mengetahui model terbaik, dan pengaruh SMOTE dalam meningkatkan akurasi model dalam menangani *imbalance* *data* (data tidak seimbang).

Rencana rumusan masalah:

a. Bagaimana pengaruh SMOTE dalam menangani *imbalance data* (data tidak seimbang) pada model XgBoost yang dibuat?

b. Bagaimana perbandingan hasil dari model XgBoost yang dikombinasikan dengan SMOTE dan model XgBoost tanpa SMOTE, manakah dari kedua model tersebut yang memiliki performa yang lebih baik?

Rencana tahapan:

1. Preprocessing Data. Meliputi data *cleaning*, *handle outlier* data, *scalling*, *normalization*.

2. Modeling. Pembuatan model dan proses training.

3. Testing. pengujian performa model.

4. Validasi hasil.

Berdasarkan keputusan Koordinator Tugas akhir,

Status Pra-Proposal : DITERIMA / TIDAK DITERIMA \*)

Usulan Calon Dosen Pembimbing : …………………………………………………………………………………………….

Catatan :

Yogyakarta, ……………..…………………………

Koordinator Tugas Akhir Prodi …………………..,

(…………………………………………………….)

NIP/ NIK.

Menyetujui,

Calon Dosen Pembimbing

(…………………………………………………….)

NIP/ NIK.

\*) *Coret yang tidak perlu*