

Kaggle Competition 2

Instruction & Guidelines

- You are allowed to use any library, built-in function or code base for this assignment.
- You are only allowed to use classical ML models or their ensemble, and your pipeline must not contain Deep Learning (DL) models. For example, use of the AutoGrad library from Pytorch is allowed but not a CNN/MLP. Similarly, generating code from GenAI/code-LLMs is allowed but generating directly predictions from these is not allowed.
- This is a competition, so each is competing against the others in the class. The higher your rank on the leaderboard, the higher the marks. Currently, there is no lower bar on performance, but after a week we will reassess if students are pushing the performance or not.
- You are given Train and Test splits. You are supposed to use only the Train set for model training. Groundtruth for Test split is not available. Sample submission is available in the competition link.

Dataset Description

This dataset is used to predict customer churn, where "churn" refers to customers ending their relationship with the company. The dataset includes information about each customer's demographics, usage patterns, subscription details, and customer service interactions.

Table 1: Dataset Features Description

Features	Unit	Description
Customer ID	int	ID of the Customer
Credit Score	float	Credit Score of customer
Country	string	Customer residence country
Gender	string	Customer Gender
Age	int	Customer Age
Tenure	int	Years customer is having bank account
Balance	float	Account balance of customer
Products Number	int	Number of products from bank
Credit Card	bool	Customer having credit card or not
Active Member	bool	Is customer active member of bank
Estimated Salary	float	Estimated salary of customer

Task

Your task is to use any kind of classification model to predict the churn of customer.

Competition Link

Kaggle Competition 2 Link