Abstract

Income is money that a person or a business receives in return for working, providing a product or service, or investing capital. A person ' s income may also derive from a pension, a government benefit, building a model to predict whether an individual's income will be greater than $50,000 per year based on several attributes. Helping Governments for income tax or any finance company.

Design

The US Adult Census dataset is a repository of 48,842 entries extracted from the 1994 US Census database.

Data

We will be using (income classification) Dataset from Kaggle. The dataset contains information about people and their income, 15 columns and 32,561 rows.

Algorithm

Problem understanding, data description, data visualization, Experiments.

Tools

Numpy, Pandas, Seaborn, matplot, sklearn, xgboost, imblearn, mlxtend.

Communication

ROC curve

- tune_lr_RUS (AUC = 0.68)
- tune_dt (AUC = 0.89)
- tune_rf (AUC = 0.91)
- tune_knn (AUC = 0.63)
- tune_xgb (AUC = 0.93)
- voting_classifer1 (AUC = 0.91)

Observation

This plot shows the roc curve for all models.