**Final Project - SCIPY**

Topic : Classification

Dataset : Bank marketing campaigns dataset analysis

(https://www.kaggle.com/volodymyrgavrysh/bank-marketing-campaigns-dataset)

----------------------------------------------------------------------------------------------------------------

1. Abstrak

* Dataset ini menjelaskan hasil dari marketing campaign suatu Bank di portugal. Target dari proses campaign ini adalah apakah seorang nasabah akan membuka deposto (“Yes”) atau tidak (“No”.)

1. Objective

* Tujuan dari penelitian ini adalah mengidentifikasi customer yang akan membuka deposito agar user marketing menjadi lebih efisien dalam proses customer approachment.

1. Metodologi Penelitian
   1. Melakukan eksplorasi terhadap dataset (EDA)
   2. Menentukan feature yang akan digunakan untuk membangun model
   3. Preprocessing Data
   4. Modelling menggunakan metode Logistic regression, KNN, Decision Tree, Random Forest dan XGBoost
   5. Tuning Model menggunakan random search
   6. Mencari nilai probabilistic 0 dan 1 untuk menentukan baseline/threshold value.
   7. Kesimpulan
2. Deskripsi Column dari Dataset
   1. Bank client data:
      * age (numeric)
      * Job (categorical): Tipe pekerjaan nasabah (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
      * Marital : marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
      * Education:(categorical:basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')
      * Default : Credit (categorical: 'no','yes','unknown')
      * Housing: housing loan (categorical: 'no','yes','unknown')
      * Loan: personal loan (categorical: 'no','yes','unknown')
   2. Related with the last contact of the current campaign:
      * Contact: Metode komunikasi yang digunakan (categorical: 'cellular','telephone')
      * Month: Month of last contact (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
      * day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
      * duration: last contact duration, in seconds (numeric). Important note: feature ini sangat mempengaruhi target. Contohnya if duration=0 then y='no'. Namun, duration ini baru dapat diketahui setelah nasabah dihubungi. Oleh karena itu sebaiknya feature ini tidak digunakan apabila ingin memiliki predictive model yang lebih realistis.
   3. other attributes:
      * campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
      * pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
      * previous: number of contacts performed before this campaign and for this client (numeric)
      * poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
   4. social and economic context attributes
      * emp.var.rate: employment variation rate - quarterly indicator (numeric)
      * cons.price.idx: consumer price index - monthly indicator (numeric)
      * cons.conf.idx: consumer confidence index - monthly indicator (numeric)
      * euribor3m: euribor 3 month rate - daily indicator (numeric)
      * nr.employed: number of employees - quarterly indicator (numeric)
   5. Output variable (desired target):
      * y – apakah customer membuka deposito? (binary: 'yes','no')