أكاديمية سدايا
SDAIA Academy

Abstract

This selected data from the famous data site kaggle represents one of the important data in the detection of financial fraud, and to clarify harmful behavior in evaluating the performance of financial fraud detection methods .

Design

This dataset shows transactions that took place within two days, the dataset is very unbalanced, the positive category (fraud) represents 0.172% of all transactions, we will prepare a model that shows us whether these transactions are fraudulent or legitimate .

Data

We have 492 scams out of 284807 transactions, we will read the data and link the columns together and understand the relationships between them .

Algorithm

- Import data and then read it to know the number of rows and columns. etc.
- It does not contain null values.
- Does not contain duplicate values.
- All questions answered in graphs.
- The data is broken down to create a model that determines whether these transactions are fraudulent or legitimate
- We used five of the models and the best result was voting

Tools

- Numpy
- Pandas
- Matplotlib
- Seaborn
- Sklearn
- Patsy

Communication

Presentation , Visualaization and Choice best model .