Machine *Learning*
Fraud with KNN

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)

Fraud with KNN

- We worked on Fraud detection with learning-based algorithms such as LR, NN and RF
- In this one, we will try to use KNN algorithm for the classification
- Your goals:
  - Find the highest f1-score you can (aim for 80% and higher on val set)
    - f1_score(y_test, y_pred)
  - Your inference time for the val/test set should be within 4 seconds
    - Remember KNN uses all the data. But this is so slow for large datasets
  - In addition, consider also using PCA. Does it help? Make it worth? With/Without?

Code

- Like the previous project
- 1) develop a train script that utilizes the train.csv and val.csv to build a KNN model
  - Save results in a pkl file
- 2) develop a test script that reads the model and test.csv and prints the f1-score
- Upload your code and pkl on drive and share with me

Report

- Prepare a report explaining your approach
- Discuss your findings

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."