Final project

**The report**

Name: **Mohamed Ali idris**

ID: **234306905**

**Git-hub Link** : https://github.com/Mohamed8560/uskudar-final

# Output Data preprocessing

| **label** | **feature\_1** | **feature\_2** | **feature\_3** | **...** | **feature\_25** | **feature\_26** | **feature\_27** | **feature\_28** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0.332782 | -0.720114 | -0.001521 | ... | -0.799281 | 2.214463 | 1.950241 | 1.218493 |
| 1 | -1.137488 | 1.007146 | -1.039270 | ... | 2.385355 | -1.024235 | 0.884599 | 2.203402 |
| 2 | 0.330583 | 0.323197 | 1.448780 | ... | -0.437876 | -1.286548 | -0.992809 | -1.038670 |
| 3 | 0.817007 | -0.589700 | -0.718352 | ... | -0.364194 | -1.048436 | -1.180356 | -0.215117 |
| 4 | -0.019848 | 1.181997 | -1.150103 | ... | 0.635942 | -0.029522 | 0.324787 | 0.223970 |

# Feature Selection

**the top 15 features**

['feature\_4', 'feature\_6', 'feature\_7', 'feature\_9', 'feature\_13', 'feature\_14',

'feature\_17', 'feature\_18', 'feature\_20', 'feature\_21', 'feature\_23', 'feature\_25',

'feature\_26', 'feature\_27', 'feature\_28']

# 3. Interpretation

**Best Model:**

* Identify the best-performing model (eg, XGBoost)
* Justify based on AUC/F1 and generalization stability

**Feature Representation:**

* Mutual Information + Standard Scaling
* Discuss why it performed well with the chosen model