* **Team Members**

|  |  |
| --- | --- |
| **Names (Arabic)** | **IDs** |
| 1- جون جورج عدلي اسكندر | 20171338375 |
| 2- فادي ماجد محارب فهيم | 20171456150 |
| 3- يمنى اسامه جابر حسن | 20171349957 |

* **Dataset Description** (You can download datasets from [Kaggle](https://www.kaggle.com/datasets), [UCI datasets](https://archive.ics.uci.edu/ml/datasets.php), [Google Datasets](https://toolbox.google.com/datasetsearch))

|  |  |
| --- | --- |
| **Dataset download link** | <https://archive.ics.uci.edu/ml/datasets/HTRU2> |
| **Number of samples** | 17898 |
| **Number of classes** | 2 |
| **Number of samples per class** | {0: 16259, 1: 1639} |
| **Names of classes** | [0, 1] |
| **Dimensionality** | 8 |
| **Names of features** | [Profile\_mean, Profile\_stdev, Profile\_skewness, Profile\_kurtosis, DM\_mean, DM\_stdev, DM\_skewness, DM\_kurtosis] |

* **Results** (Apply all of the following classifiers on your datasets. Don’t forget to use train\_test\_split function)

|  |  |  |
| --- | --- | --- |
| **Classifier name** | **Parameters used** | **Accuracy score** |
| **Support Vector Machine** | SVC(kernel=’linear’) | 0.983240223464 |
| **Naïve Bayes** | GaussianNB() | 0.946145251397 |
| **Decision Tree** | DecisionTreeClassifier(random\_state=4) | 0.967374301675 |
| **Neural Networks with at least 1 hidden layer** | MLPClassifier(hidden\_layer\_sizes=(20,15,10,5)) | 0.983016759776 |

* **Which algorithm is the most suitable for your dataset?**

**Answer: Support Vector Machine**

* **Python code + Printed report:** 10 Marks.
* **Bonus:** GUI that contains buttons for showing data, selecting each algorithm, and viewing results.
* **Discussion duration:** 15 minutes per team.
* **Discussion dates:**
  + Sunday, Dec 15, 2019 (1 PM:4 PM) – Lab C
  + Wednesday, Dec 18, 2019 (10 AM:2 PM) – Lab D
  + Thursday, Dec 19, 2019 (3 PM:6 PM) – Lab D
* **Note that:** All team members must attend during discussion and no late projects are accepted.