**Task 1 Report**

**Students Names:**

**Team ID : SC\_H8**

|  |  |  |  |
| --- | --- | --- | --- |
| Student IDs | Department | Name | No |
| 20191700245 | Scientific Computing | Rehab Hosam | 1 |
| 20191700484 | Scientific Computing | Maivel Maher | 2 |
| 20191700407 | Scientific Computing | Omar Khaled | 3 |
| 20191700347 | Scientific Computing | Abdelrahman Mohamed | 4 |
| 20191700007 | Scientific Computing | Ibrahim Youssef | 5 |

**Visualizations :**

Chart, scatter chart

Description automatically generatedChart, bar chart

Description automatically generatedChart

Description automatically generated with medium confidenceChart, scatter chart

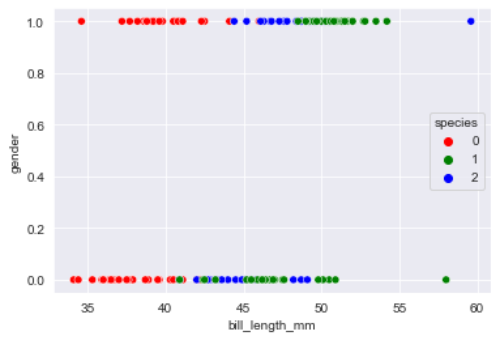
Description automatically generatedChart, scatter chart

Description automatically generatedChart

Description automatically generatedChart, scatter chart

Description automatically generated

**Observations :**

* There are 7 visulization plots only instead of 10 and here is why :
  + The col. Name “Gender” is defined by only 0 or 1 which is Male and female encoded.
  + And thus its scatter visualization with any features will be as follows :
  + ****
  + As you can see here its difficult to be separated and not visualized correctly so the best describe to it was the box plot between the feature itself (gender) and the frequency of class “species” only in 1 plot instead of 3.
* Some plots are not a scatter depending on what describes best if it’s a linear feature or not.
* Displots are seen better than scatter in some and we can separate linearly separable displots by a horizontal line but those features were not the best ones.
* We did the shuffling between training and testing sets manually and have chosen as required to be total of 60 training examples and 40 test between 2 different equally separated classes
* You can find all operations in file operations and all algorithms of single perceptron in the other file and of course the GUI in main.. hope you enjoy the background 😊

**Conclusions :**

* As you can see from the plotting here the best described which can be separated linearly easily is the plot between X1 and X2 (Bill shape length and Bill shape depth)
* Data contains 6 Null values in col. Type(male/female) and ZERO duplicated values.
  + Null values are could be replaced with bell shape or bell depth as these 2 feautures greatly distinguish between if the null value will be male(1) or female (0)
* All String values are encoded
* Y label col. Is encoded as 1 or -1 due to the signum nature
* Best accuracies got was between bill\_shape, bill\_depth and was 0.975 and bill\_shape and gender.