

**Faculty of Computers and Artificial Intelligence, Helwan University**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Course Name : Selected CS-1**

**Team ID : 12**

**TA: Eng\ Salma**

| **Name** | **ID** | **Notes** |
| --- | --- | --- |
| **حسن جمعه حسن** | **201900265** |  |
| **عبدالله اسماعيل محمد محمد** | **2019004545** |  |
| **فادي هاني فهيم جيد** | **202000623** |  |
| **على خالد على** | **202000566** |  |
| **محمد اسامة سيد حجازى** | **201900629** |  |
| **محمود عبدالحميد عبدالجواد عبدالحميد** | **201900775** |  |

**Code And DataSet Project on the GitHub**

**Link:**

[**https://github.com/hassangomaa/Selected\_CS\_1\_Project\_SVM\_ANN\_logistic\_regression.git**](https://github.com/hassangomaa/Selected_CS_1_Project_SVM_ANN_logistic_regression.git)

**QR Code :**

****

**Logistic Regression Documintation**

**Context**

**About The Data**

**The following fruits and are included:**

**Apples (different varieties: Crimson Snow, Golden, Golden-Red, Granny Smith, Pink Lady, Red, Red Delicious), Apricot, Avocado, Avocado ripe, Banana (Yellow, Red, Lady Finger), Beetroot Red, Blueberry, Cactus fruit, Cantaloupe (2 varieties), Carambula, Cauliflower, Cherry (different varieties, Rainier), Cherry Wax (Yellow, Red, Black), Chestnut, Clementine, Cocos, Corn (with husk), Cucumber (ripened), Dates, Eggplant, Fig, Ginger Root, Granadilla, Grape (Blue, Pink, White (different varieties)), Grapefruit (Pink, White), Guava, Hazelnut, Huckleberry, Kiwi, Kaki, Kohlrabi, Kumsquats, Lemon (normal, Meyer), Lime, Lychee, Mandarine, Mango (Green, Red), Mangostan, Maracuja, Melon Piel de Sapo, Mulberry, Nectarine (Regular, Flat), Nut (Forest, Pecan), Onion (Red, White), Orange, Papaya, Passion fruit, Peach (different varieties), Pepino, Pear (different varieties, Abate, Forelle, Kaiser, Monster, Red, Stone, Williams), Pepper (Red, Green, Orange, Yellow), Physalis (normal, with Husk), Pineapple (normal, Mini), Pitahaya Red, Plum (different varieties), Pomegranate, Pomelo Sweetie, Potato (Red, Sweet, White), Quince, Rambutan, Raspberry, Redcurrant, Salak, Strawberry (normal, Wedge), Tamarillo, Tangelo, Tomato (different varieties, Maroon, C**

**herry Red, Yellow, not ripened, Heart), Walnut, Watermelon.**

**Labeling Data :**

**Selecting only 2 classes of fruits (banana-mandrine).**

**Data Size:**

**980 as a training size**

**332 as testing size**

**Tools and IDE used:**

**IDE: Kaggle notebook.**

**Programming Language: python**

**Packages used:**

* **Numpy**
* **Pandas**
* **Matplotlib**
* **Seaborn**
* **Keras**
* **Sklearn**
* **cv2**
* **os**

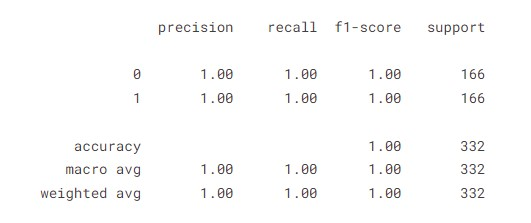
**Data Preparation:**

1. **Resizing image to 100\*100**
2. **Normalizing pictures (dividing pictures pixels by 255)**
3. **Transforming images to array**
4. **Flattening the array**

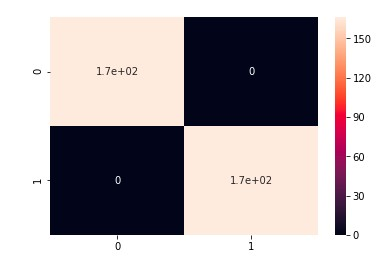
**Model:**

**Logistic regression with five cross validation splits.**

**Test and Validation (Result details):**

****

**Confusion\_Matrix:**

****

**Logistic Regression Documintation (titanic data set)**

**Context:**

**About The Data..**

**This is the legendary Titanic ML competition – the best, first challenge for you to dive into ML competitions and familiarize yourself with how the Kaggle platform works.**

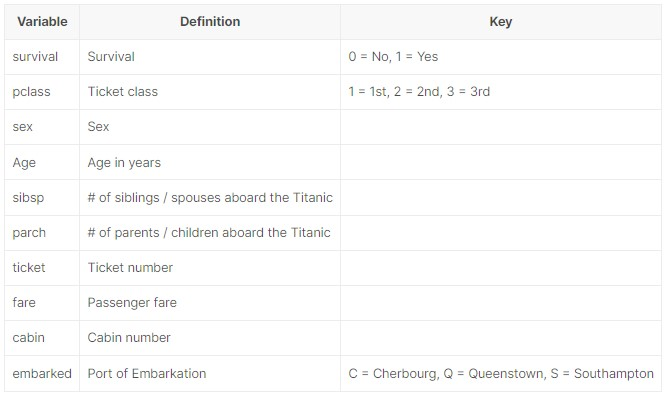
**The competition is simple: use machine learning to create a model that predicts which passengers survived the Titanic shipwreck.**

**Labeling Data :**

**For dataset total shape is 891 Example.**

**Testing size 178.**

**Training size 713.**

****

**Tools and IDE used:**

**IDE: Kaggle notebook.**

**Programming Language: python**

**Packages used:**

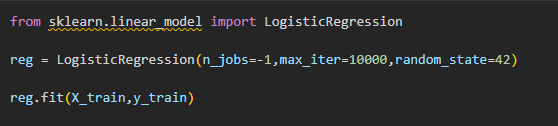
* **Numpy**
* **Pandas**
* **Matplotlib**
* **Seaborn**
* **Keras**
* **Sklearn**
* **cv2**
* **os**

**Data Preparation:**

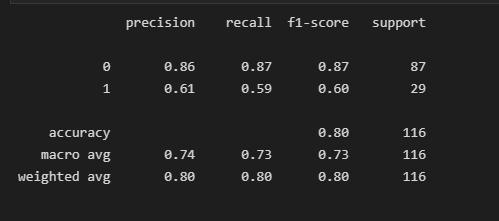
1. **preprossesing names to get the martial status**
2. **filling nan values**
3. **ranking embarked feature**
4. **removing outliers using IQR**
5. **one hot encoding**
6. **using standard scaler to scale the data**

**Model:**

**Logistic regression.**

****

**Test and Validation (Result details):**

****

**ANN**

## **a-General Information on dataset:**

**The name of dataset used: Malaria Cell Images Dataset**

**Number of classes and their labels: \_**

**The total number of samples in dataset: 27558 image**

**The size of each (in case of images):** 8 kb

**The number of samples used in training: 20,668 image(75%) The number of samples used in validation: 20,668 image(75%) The number of samples used in testing: 6890 image(25%)**

## **b- Implementation details:**

### **At feature extraction phase**

**How many features were extracted: many features**

**Their names:** \_

**The dimension of resulted features:** \_

### **Is cross-validation**

**Used in any of the implemented models? If yes, specify the number of fold and ratio of training/validation.**

**Number of fold:**

**Ratio of validation**

### **Hyperparameters used in your model**

**, as initial learning rate, optimizer, regularization, batch size, no. of epochs, etc…**

# 

## **C- Results details:**

**loss curve:**

**accuracy:**

**training 97%**

**test 71%**

**confusion matrix:**

**ROC curve**:

# 

# 

# 

# 

# 

# 

# **SVM (Numerical)**

## **a-General Information on dataset:**

**The name of dataset used: Used-cars-catalog**

**Number of classes and their labels: 24 classe and their labels(manufacturer\_name - model\_name - transmission – color – odometer\_ value – year\_produced – engine fuel – engine\_has\_gas – engine\_type – engine\_capacity – body\_type – has\_warranty - state - drive**

**\_train – price\_USD – is\_exchangable – number of photos – up\_counter – feature\_0 - feature\_1**

**- feature\_2 - feature\_3)**

***After drop:(* transmission– odometer\_ value – year\_produced – engine fuel – engine\_has\_gas – engine\_type – engine\_capacity – body\_type – has\_warranty - state - drive \_train – price\_USD )**

**The total number of samples in dataset: 4999 sample**

**The number of samples used in training:** :(75%) **The number of samples used in validation**:(75%) **The number of samples used in testing:** (25%)

## **b- Implementation details:**

### **At feature extraction phase**

**How many features were extracted:**

**Their names:**

**The dimension of resulted features:**

### **Is cross-validation**

**Used in any of the implemented models? If yes, specify the number of fold and ratio of training/validation.**

**Number of fold:** n

**Ratio of validation:**

**Hyperparameters used in your model**

**, as initial learning rate, optimizer, regularization, batch size, no. of epochs, etc…**

## **C- Results details:**

**accuracy: training 97% test 71% confusion matrix:**

**ROC curve:**

# **SVM(Images)**

## **a-General Information on dataset:**

**The name of dataset used: Malaria Cell Images Dataset**

**Number of classes and their labels: \_**

**The total number of samples in dataset: 27558 image**

**The size of each (in case of images):** 8 kb

**The number of samples used in training: 13779 image(50%) The number of samples used in validation: 13779 image(50%) The number of samples used in testing: 13779 image(50%)**

## **b- Implementation details:**

### **At feature extraction phase**

**How many features were extracted: many features**

**Their names:** \_

**The dimension of resulted features:** \_

### **Is cross-validation**

**Used in any of the implemented models? If yes, specify the number of fold and ratio of training/validation.**

**Number of fold:**

**Ratio of validation:**

**Hyperparameters used in your model**

**, as initial learning rate, optimizer, regularization, batch size, no. of epochs, etc…**

## **5**

**C- Results details:**

**loss curve:**

**accuracy:**

**confusion matrix:**

**ROC curve:**