Faculty Of Computers And Artificial Intelligence

Helwan University

Course: Selected Topics In CS -1

CARD CUSTOMERS PREDICTION

Logistic Regression and SVM Models implemented on numerical dataset

A) General information on both Numerical Datasets (LG and SVM):

Name of dataset used: BankChurners

Number of classes: 23 Class **Labels**

of classes:

```
CLIENTNUM
Attrition_Flag
Customer_Age
Gender
Dependent_count
Education_Level
Marital_Status
Income_Category
Card_Category
Months_on_book
Total_Relationship_Count
Months_Inactive_12_mon
Contacts_Count_12_mon
Credit_Limit
Total_Revolving_Bal
Avg_Open_To_Buy
Total_Amt_Chng_Q4_Q1
Total_Trans_Amt
Total_Trans_Ct
Total_Ct_Chng_Q4_Q1
Avg_Utilization_Ratio
Naive_Bayes_Classifier
classification
```

Total number of samples: 2999 sample Number

of samples used in:

Training: 80%

Testing: 20%

B) Implementation details of LG numerical dataset:

-no feature extraction was done on the numerical dataset

-No cross validation was used

Before we optimized the accuracy :

-**hyperparameters used:**

- penalty='l2',
- solver='newton-cg',
- C=1.0,
- dual=False,
- tol=0.0001,
- class_weight='balanced',
- max_iter=100,
- l1_ratio=None,
- multi_class='auto',
- verbose=0,
- warm_start=False,
- n_jobs=None

c) Results details of LG numerical dataset:

Loss Values:

[0.4534928627676029,

0.22374015385866114,

0.15966835234232155,

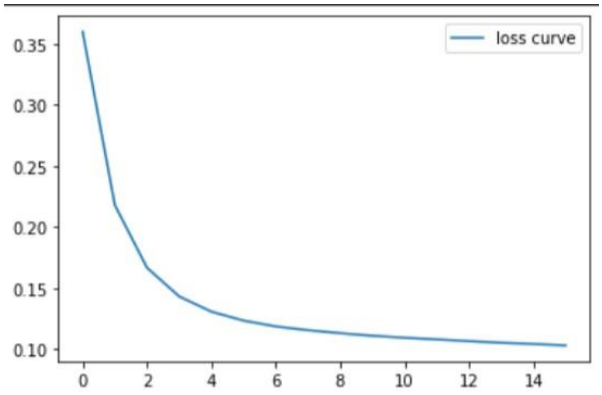
…..

, 0.0956188666041105,

0.09518745927467737,

0.09465393900011664]

Loss curve:



F1-Score: 0.97

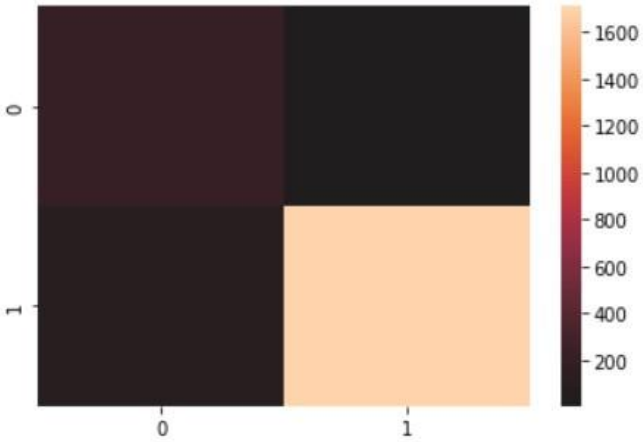
Accuracy: 0.95

Confusion matrix:

Array ([[227 , 3],

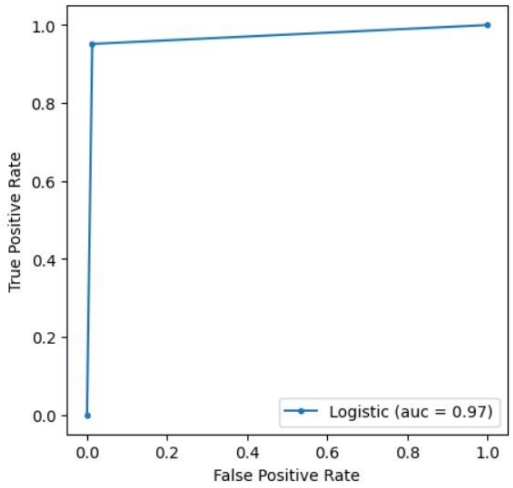
[87, 1709]])

Confusion matrix plot:



AUC value: 0.97

ROC curve:



Recall Score: 0.95

Precision Score: 0.99

After we optimized the accuracy

-**hyperparameters used:**

- penalty='l1',
- solver='saga',
- C=0.2,
- dual=False,
- tol=0.0001,
- class_weight='None',
- multi_class='auto',
- verbose=0,
- warm_start=False,
- n_jobs=None
- fit_intercept=True
- intercept_scaling=1
- random_state=None

c) Results details of LG numerical dataset:

Loss Values:

[0.8487668012625855,

0.6109143726566947,

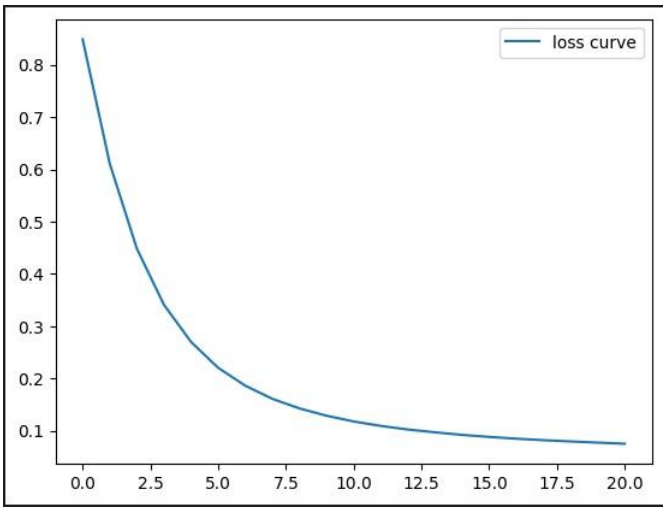
0.44837926853632665,

…..

0.07960229362475647,

0.07736735448715581,

0.07513018510509227] **Loss curve:**



F1-Score: 0.9857142857142858

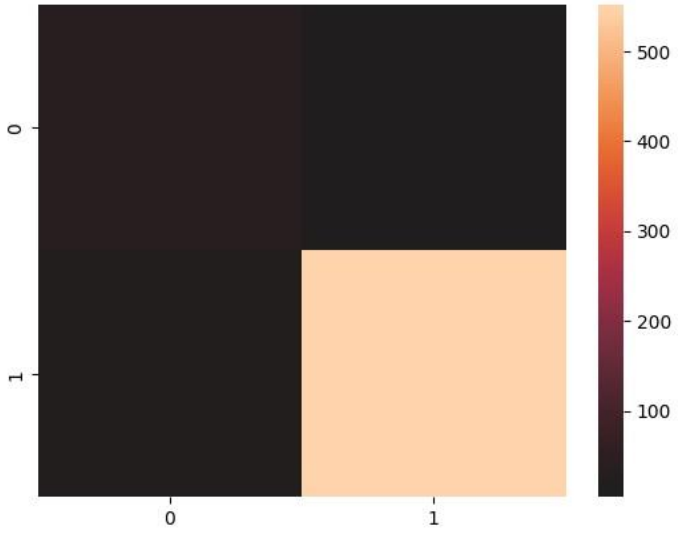
Accuracy: 0.9733333333333334

Confusion matrix: array([[

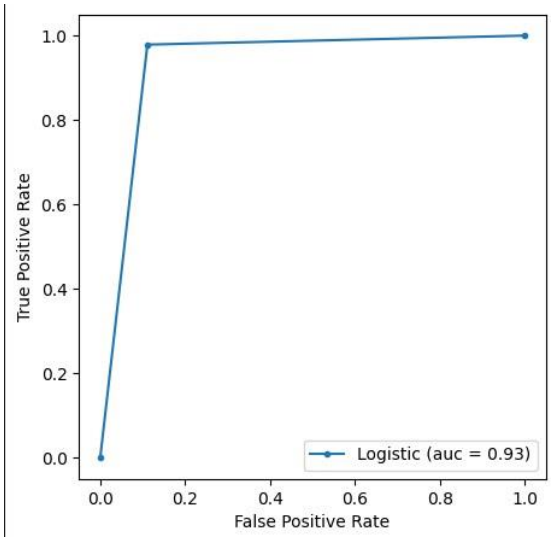
32, 4],

[12, 552]])

Confusion matrix plot:



AUC value: 0.9338061465721039 **ROC curve:**



Recall Score: 0.9787234042553191

Precision Score: 0.9928057553956835

B) Implementation details of SVM numerical dataset:

-no feature extraction was done on the numerical dataset

-no cross validation was used

Before we optimized the accuracy :

-hyperparameters used:

- C=1, ○ kernel='poly', ○ degree=33, ○ gamma='auto', ○ shrinking=False, ○ probability=True, ○ tol=0.1, ○ class_weight=None, ○ verbose=False, ○ max_iter=-1,
- random_state =4

c) Results details of SVM numerical dataset:

Loss values:

[0.4579088608604389,

0.24674559998127407,

0.1775408559076878,

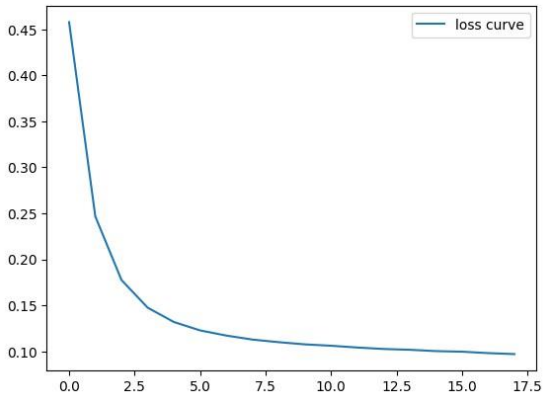
……

0.09962811908373023,

0.09808817212813536,

0.09701903103707903]

Loss curve:



F1-score: 0.973

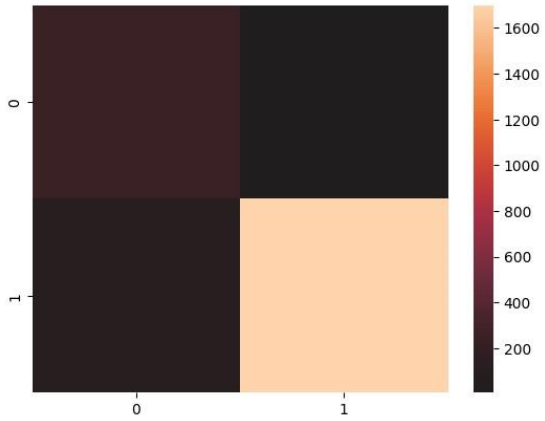
Accuracy Score: 0.954

Confusion matrix:

([[233 5]

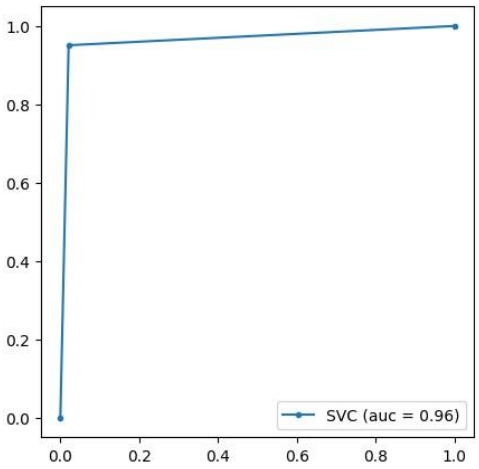
[88 1700]])

Confusion matrix plot:



AUC Value: 0.964 **ROC**

Curve:



Recall Score: 0.950

Precision Score: 0.997

After we optimized the accuracy :

Hyperparameters <u>used</u> :

- C=0.1,
- kernel='linear', ○ degree=3, ○ gamma='auto', ○ shrinking=False, ○ probability=True, ○ tol=0.001, ○ cache_size=200, ○ class_weight=None, ○ verbose=False, ○ max_iter=-1, ○ random_state =40

<u>c) Results details of SVM numerical dataset:</u>

Loss values:

[0.5054872927985871,

0.3745298811345925,

0.28772872069848326

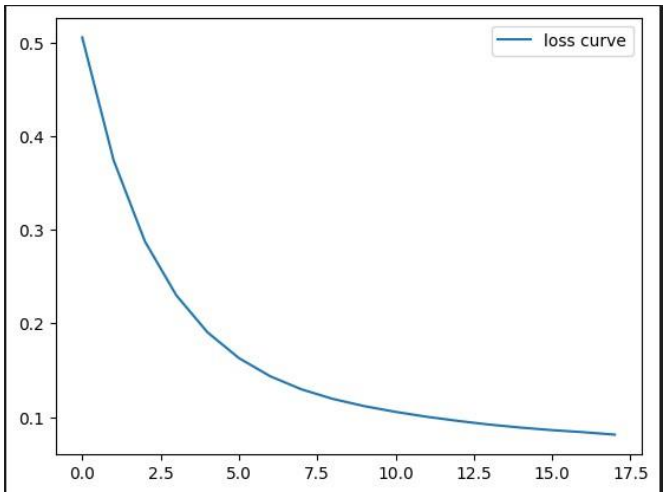
……

0.08614844011702817,

0.08394865461430984,

0.08134572714116146]

Loss curve:



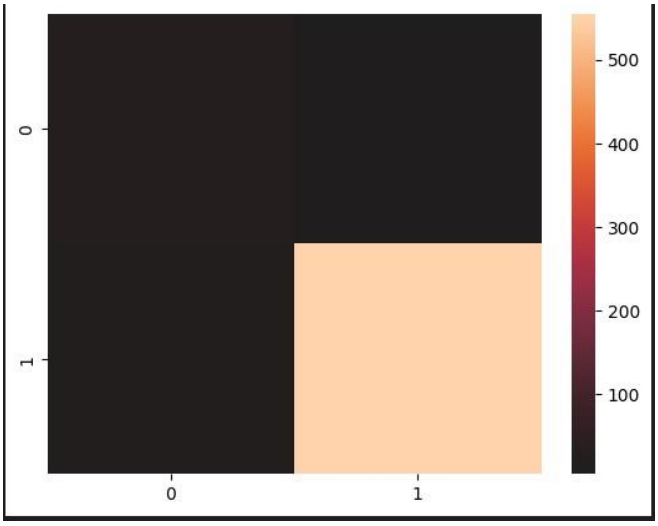
F1-score: 0.9849157054125999

Accuracy Score: 0.975 **Confusion matrix**:

[[28 5]

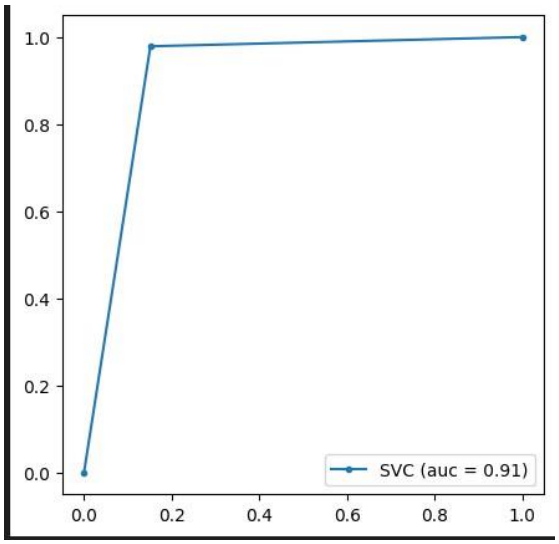
[12 555]]

Confusion matrix plot:



AUC Value: 0.9136604136604137 **ROC**

Curve:



Recall Score: 0.9788359788359788

Precision Score: 0.9910714285714286