Name : Siddhant M Mangale

Class : B.Tech (A)

Roll No. : 31

Experiment No. : 10

In [ ]:

*# Building a predictive ML model- Logistic Regression*

In [1]:

**using** Pkg

In [2]:

Pkg**.**add("Pandas")

**Updating** registry at `C:\Users\aniket\.julia\registries\Gen eral.toml`

**Resolving** package versions...

**Installed** TableTraitsUtils ─ v1.0.2 **Installed** Lazy ───────────── v0.15.1 **Installed** Pandas ─────────── v1.6.1

**Updating** `C:\Users\aniket\.julia\environments\v1.8\Project. toml`

[eadc2687] + Pandas v1.6.1

**Updating** `C:\Users\aniket\.julia\environments\v1.8\Manifes t.toml`

[50d2b5c4] + Lazy v0.15.1 [eadc2687] + Pandas v1.6.1

[382cd787] + TableTraitsUtils v1.0.2

**Precompiling** project...

* TableTraitsUtils
* Lazy
* Pandas

3 dependencies successfully precompiled in 47 seconds. 213 already precomp iled. 4 skipped during auto due to previous errors.

**using** Pandas

┌ Info: Installing pandas via the Conda pandas package...

└ @ PyCall C:\Users\aniket\.julia\packages\PyCall\ygXW2\src\PyC all.jl:719

┌ Info: Running `conda install -y pandas` in root environment

└ @ Conda C:\Users\aniket\.julia\packages\Conda\x2UxR\src\Cond a.jl:127

Collecting package metadata (current\_repodata.json): ...working... done Solving environment: ...working... done

## Package Plan ##

environment location: C:\Users\aniket\.julia\conda\3 added / updated specs:

- pandas

The following packages will be downloaded:

package | build

---------------------------|-----------------

pandas-1.5.2 | py310h1c4a608\_0 9.1 MB conda-forg

e

pytz-2022.6 | pyhd8ed1ab\_0 235 KB conda-forg

e

Total: 9.4 MB The following NEW packages will be INSTALLED:

pandas conda-forge/win-64::pandas-1.5.2-py310h1c4a608\_0 None pytz conda-forge/noarch::pytz-2022.6-pyhd8ed1ab\_0 None

Downloading and Extracting Packages

pytz-2022.6 | 235 KB | ########## | 100% pandas-1.5.2 | 9.1 MB | ########## | 100%

Preparing transaction: ...working... done Verifying transaction: ...working... done Executing transaction: ...working... done Retrieving notices: ...working... done

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| df\_train **=** read\_csv("ProcessedData.csv") | | | | | | | | | |
| **117** | LP001405 | 1 | 1 | 1 | 0 | 0 | 2214 | 1398.000 |  |
| **118** | LP001421 | 1 | 1 | 0 | 0 | 0 | 5568 | 2142.000 |
| **119** | LP001422 | 0 | 0 | 0 | 0 | 0 | 10408 | 0.000 |
| **120** | LP001426 | 1 | 1 | 0 | 0 | 0 | 5667 | 2667.000 |
| **121** | LP001430 | 0 | 0 | 0 | 0 | 0 | 4166 | 0.000 |
| **122** | LP001431 | 0 | 0 | 0 | 0 | 0 | 2137 | 8980.000 |
| **123** | LP001432 | 1 | 1 | 2 | 0 | 0 | 2957 | 0.000 |
| **124** | LP001439 | 1 | 1 | 0 | 1 | 0 | 4300 | 2014.000 |
| **125** | LP001443 | 0 | 0 | 0 | 0 | 0 | 3692 | 0.000 |
| **126** | LP001448 | 1 | 1 | 3 | 0 | 0 | 23803 | 0.000 |
| **127** | LP001449 | 1 | 0 | 0 | 0 | 0 | 3865 | 1640.000 |
|  |  |  |  |  |  |  |  |  |  |

In [5]:

x**=**df\_train[["Credit\_History","Education","Married","Self\_Employed","Property\_Area"]] y **=** df\_train["Loan\_Status"]

Out[5]:

0 1 1 0 2 1 3 1 4 1 .. 609 1 610 1 611 1 612 1 613 0 Name: Loan\_Status, Length: 614, dtype: int64

In [6]:

**using** ScikitLearn**:** fit!, predict, @sk\_import, fit\_transform!

In [7]:

@sk\_import model\_selection**:** train\_test\_split

┌ Warning: Module model\_selection has been ported to Julia - try `import Sci kitLearn: CrossValidation` instead

└ @ ScikitLearn.Skcore C:\Users\aniket\.julia\packages\ScikitLe arn\7Csem\src\Skcore.jl:193

Out[7]:

PyObject <function train\_test\_split at 0x00000000553EE200>

In [8]:

x\_train, x\_test, y\_train, y\_test **=** train\_test\_split(x, y, test\_size**=**0.3, random\_state**=**42)

Out[8]:

4-element Vector{PyCall.PyObject}:

PyObject Credit\_History Education Married Self\_Employed Propert y\_Area

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 553 | 0 | 1 | 1 | 0 | 2 |
| 601 | 1 | 1 | 1 | 0 | 0 |
| 261 | 1 | 0 | 0 | 0 | 1 |
| 496 | 1 | 1 | 1 | 0 | 0 |
| 529 | 1 | 1 | 0 | 0 | 1 |
| .. | ... | ... | ... | ... | ... |
| 71 | 1 | 1 | 1 | 1 | 1 |
| 106 | 1 | 0 | 1 | 0 | 2 |
| 270 | 1 | 0 | 0 | 0 | 2 |
| 435 | 1 | 0 | 1 | 0 | 1 |
| 102 | 1 | 0 | 1 | 0 | 2 |

[429 rows x 5 columns]

PyObject Credit\_History Education Married Self\_Employed Propert y\_Area

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 350 | 1 | 0 | 1 | 0 | 1 |
| 377 | 1 | 0 | 1 | 0 | 1 |
| 163 | 1 | 0 | 1 | 0 | 0 |
| 609 | 1 | 0 | 0 | 0 | 0 |
| 132 | 1 | 0 | 0 | 0 | 1 |
| .. | ... | ... | ... | ... | ... |
| 380 | 1 | 0 | 1 | 0 | 1 |
| 446 | 1 | 1 | 1 | 0 | 0 |
| 144 | 1 | 0 | 1 | 0 | 2 |
| 290 | 1 | 0 | 1 | 0 | 0 |
| 353 | 0 | 0 | 1 | 1 | 0 |

[185 rows x 5 columns]

PyObject 553 0

|  |  |  |  |
| --- | --- | --- | --- |
| 601 1 | |  | |
| 261 1 | |
| 496 1 | |
| 529 1 | |
| .. | |
| 71 1 | |
| 106 1 | |
| 270 1 | |
| 435 1 | |
| 102 1 | |
| Name: Loan\_Status, Length: | | 429, | dtype: int64 |
| PyObject 350 1 | |  |  |
| 377 | 1 | | |
| 163 | 1 | | |
| 609 | 1 | | |
| 132 | 1 | | |
|  | .. | | |
| 380 | 1 | | |
| 446 | 1 | | |
| 144 | 1 | | |
| 290 | 1 | | |

353 0

Name: Loan\_Status, Length: 185, dtype: int64

In [9]:

@sk\_import linear\_model**:** LogisticRegression model**=**LogisticRegression()

model**.**fit(x\_train,y\_train)

println("Predicted Accuracy of the Model : ",model**.**score(x\_test,y\_test)) println("Actual Accuracy of the Model : ",model**.**score(x\_train,y\_train))

Predicted Accuracy of the Model : 0.7837837837837838 Actual Accuracy of the Model : 0.8205128205128205

In [10]:

@sk\_import model\_selection**:** cross\_val\_score

┌ Warning: Module model\_selection has been ported to Julia - try `import Sci kitLearn: CrossValidation` instead

└ @ ScikitLearn.Skcore C:\Users\aniket\.julia\packages\ScikitLe arn\7Csem\src\Skcore.jl:193

Out[10]:

PyObject <function cross\_val\_score at 0x000000005554E170>

In [11]:

cross\_score **=** cross\_val\_score(model, x, y, cv**=**5)

println("cross\_validation\_score: ", mean(cross\_score))

cross\_validation\_score: 0.809462881514061