**Logistic Regression on Bank Loan Data Set**

**Loading Data and Data Treatment:**

import pandas as pd

loan\_data = pd.read\_excel("Bank\_Personal\_Loan\_Modelling.xlsx", sheet\_name= "Data")

loan\_data.head(2)

Out[83]:

ID Age Experience ... CD Account Online CreditCard

0 1 25 1 ... 0 0 0

1 2 45 19 ... 0 0 0

[2 rows x 14 columns]

loan\_data.isnull().sum()

Out[85]:

ID 0

Age 0

Experience 0

Income 0

ZIP Code 0

Family 0

CCAvg 0

Education 0

Mortgage 0

Personal Loan 0

Securities Account 0

CD Account 0

Online 0

CreditCard 0

dtype: int64

loan\_data.columns

Out[87]:

Index(['ID', 'Age', 'Experience', 'Income', 'ZIP Code', 'Family', 'CCAvg',

'Education', 'Mortgage', 'Personal Loan', 'Securities Account',

'CD Account', 'Online', 'CreditCard'],

dtype='object')

**Model Generation:**

Y = loan\_data[['Personal Loan']]

X = loan\_data[['Age', 'Experience', 'Income', 'Family', 'CCAvg',

'Education', 'Mortgage', 'Securities Account',

'CD Account', 'Online', 'CreditCard']]

Corr\_matrix = X.corr()

import statsmodels.api as sm

X1 = sm.add\_constant(X)

Logistic = sm.Logit(Y, X1)

result = Logistic.fit()

Optimization terminated successfully.

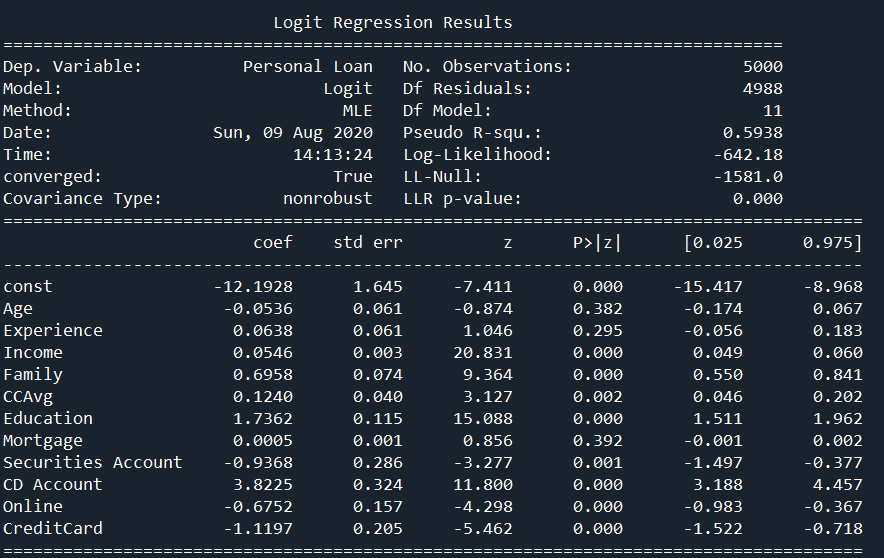
Current function value: 0.128435

Iterations 9

result.summary()

Out[95]:

<class 'statsmodels.iolib.summary.Summary'>



**Inference:**

Features Income, Family, CCAvg, Education, Securities Account, CD Account, Online, Credit are significantly important in getting the loan.

The probability for getting loan P(Y/N) is calculated as follows

P(Y/N) = 1/(1+e^(-k))

Where k = -12.1928 + (-0.0536)(Age) + (0.0638)( 'Experience') + (0.0546)( 'Income') + (0.6958)( 'Family') + (0.1240)( 'CCAvg') + (1.7362)( 'Education') + (0.0005)( 'Mortgage') + (-0.9368)( 'Securities Account') + (3.8225)( 'CD Account') + (-0.6752)( 'Online') + (-1.1197)( 'CreditCard')