

We are going to predict the probability of person accepting a bank loan according to the given dataset and the predictive model.

```
In [15]: import pandas as pd
dataset = pd.read_excel('dataset/Bank_Personal_Loan_Modelling.xlsx', sheet_name='Data')

In [12]: df = dataset[['Age', 'Experience', 'Income', 'Family', 'CCAvg', 'Education', 'Mortgage', 'Personal Loan', 'Securities Account', 'CD Account', 'Online', 'CreditCard']]
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 4999
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Age                    5000 non-null   int64
1   Experience              5000 non-null   int64
2   Income                  5000 non-null   int64
3   Family                  5000 non-null   int64
4   CCAvg                   5000 non-null   float64
5   Education               5000 non-null   int64
6   Mortgage                5000 non-null   int64
7   Personal Loan           5000 non-null   int64
8   Securities Account       5000 non-null   int64
9   CD Account              5000 non-null   int64
10  Online                  5000 non-null   int64
11  CreditCard              5000 non-null   int64
dtypes: float64(1), int64(11)
memory usage: 468.9 KB
```

The Dependent variable (y) is "Personal Loan"

and the rest of them are independent variables, contained in (x)

```
In [18]: y = df['Personal Loan']

x = df[['Age', 'Experience', 'Income', 'Family', 'CCAvg', 'Education', 'Mortgage', 'Securities Account', 'CD Account', 'Online', 'CreditCard']]
```

Now Performing Logistic Regression training

```
In [21]: import statsmodels.api as sm

x1 = sm.add_constant(x)

logistic = sm.Logit(y,x1)

result = logistic.fit()

print(result.summary())
```

```
Optimization terminated successfully.
Current function value: 0.128435
Iterations 9

Logit Regression Results
=====
Dep. Variable:          Personal Loan    No. Observations:          5000
Model:                  Logit           Df Residuals:              4988
Method:                  MLE            Df Model:                  11
Date:                   Mon, 10 Aug 2020 Pseudo R-squ.:              0.5938
Time:                   02:35:32         Log-Likelihood:            -642.18
converged:               True            LL-Null:                  -1581.0
Covariance Type:         nonrobust        LLR p-value:              0.000
=====

```

	coef	std err	z	P> z	[0.025	0.975]
const	-12.1928	1.645	-7.411	0.000	-15.417	-8.968
Age	-0.0536	0.061	-0.874	0.382	-0.174	0.067
Experience	0.0638	0.061	1.046	0.295	-0.056	0.183
Income	0.0546	0.003	20.831	0.000	0.049	0.060
Family	0.6958	0.074	9.364	0.000	0.550	0.841
CCAvg	0.1240	0.040	3.127	0.002	0.046	0.202
Education	1.7362	0.115	15.088	0.000	1.511	1.962
Mortgage	0.0005	0.001	0.856	0.392	-0.001	0.002
Securities Account	-0.9368	0.286	-3.277	0.001	-1.497	-0.377
CD Account	3.8225	0.324	11.800	0.000	3.188	4.457
Online	-0.6752	0.157	-4.298	0.000	-0.983	-0.367
CreditCard	-1.1197	0.205	-5.462	0.000	-1.522	-0.718

```
=====
```

Here the p-value of Age, Experience, Mortgage is more than 0.05, hence except these variables rest of them are significant in predicting the "Personal Loan" status

Now Creating the model

```
In [26]: #Coefficient values
B0      = -12.1928
ageX    = -0.0536
experienceX = 0.0638
incomeX = 0.0546
familyX = 0.6958
ccAvgX  = 0.1240
educationX = 1.7362
mortgageX = 0.0005
securitiesAccountX = -0.9368
cdAccountX = 3.8225
onlineX  = -0.6752
creditCardX = -1.1197

#Sample values for probability prediction
age      = 42
experience = 11
income   = 150
family   = 4
ccAvg    = 4.0
education = 2
mortgage = 100
securitiesAccount = 1
cdAccount = 1
online    = 1
creditCard = 1

import math

p = 1/(1+math.exp(-(B0+(age*ageX)+(experience*experienceX)+(income*incomeX)+(family*familyX)
+(ccAvg*ccAvgX)+(education*educationX)+(mortgage*mortgageX)+(securitiesAccount*securitiesAccountX)+(cdAccount*cdAccountX)+(online*onlineX)+(creditCard*creditCardX))))

In [28]: print("Probability of loan acceptance is ", p)
```

Probability of loan acceptance is 0.9121521126190076

After providing values to the model, it has generated a p-value of 0.912 which is greater than 0.5, so there is a very high chance of Personal loan acceptance as per the details provided by the person.