# Loan Eligibility Prediction

## Introduction Overview

Banks fundamental business model rely on financial intermediation by raising finance and lending (mortgage, real estate, consumer and companies loans). The latter is the major source of credit risk composed from 2 main points loan approval and fraud. Loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don't have assurance if the applicant is able to repay the loan with no difficulties.

**Purpose**

The main aim of use-case is to build the predictive model to predict if a applicant is able to repay the loan or not

# LITERATURE SURVEY

- **Existing problem**
- The existing approach to solve this problem is to employ a person and do the background checks needed to make sure that the customer is eligible for the loan based on customer detail provided while filling online application form

- **Proposed Solution**
- Logistic regression and decision trees are both popular classification techniques (supervised learning) used to build behavioral scorecards, they are statistical methods that analyse a dataset to bring out the relationship between "predictors" (or explanatories) that are independent variables and a "response" (or Outcome variable) that is a dependent variable. In our case we try to estimate the probability of granting a loan .

# THEORITICAL ANALYSIS

Block diagram of loan eligibility prediction
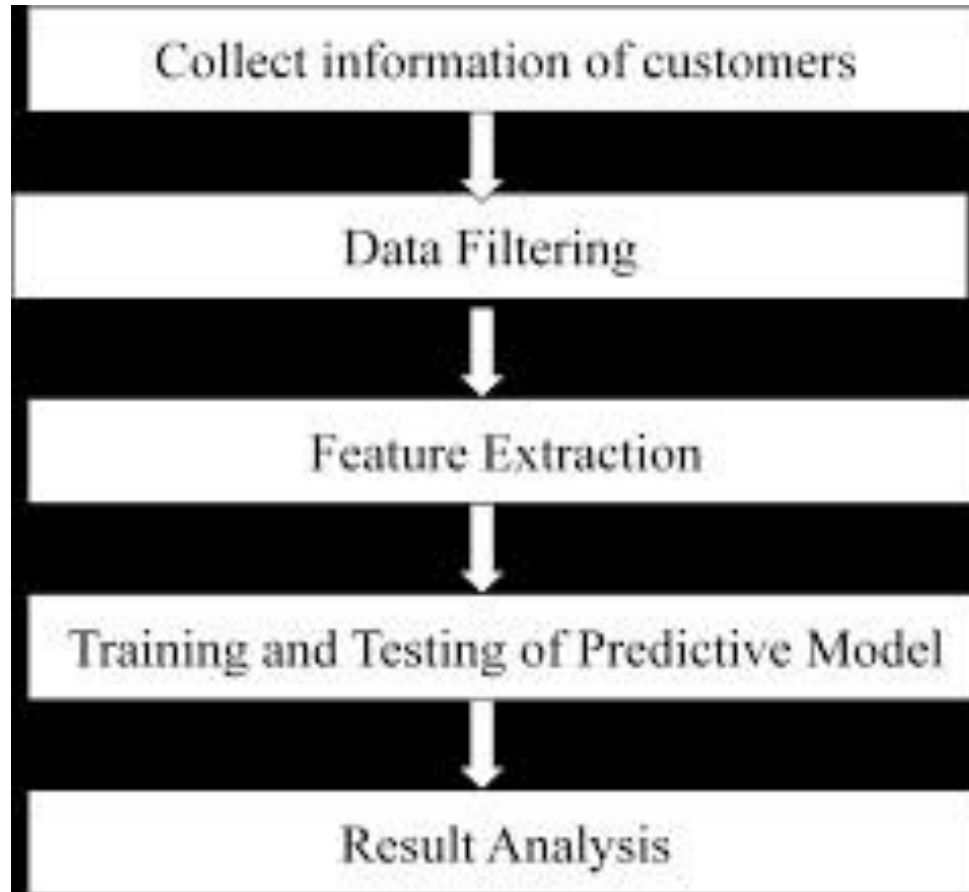
# Hardware/software requirements

**Software Requirements**

• Java jdk 10

• Weka

• Eclipse IDE

**Hardware Requirements**

• Processor-i5

• Min hardDisk-4GB

• Min Memory-4GBRAM

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

**Classifier**

Choose | **Logistic** -R 1.0E-8 -M -1 -num-decimal-places 4

**Test options**

- ○ Use training set
- ○ Supplied test set    Set...
- ● Cross-validation   Folds  10
- ○ Percentage split    %   66

More options...

(Nom) Loan_Status

Start | Stop

**Result list (right-click for options)**

16:09:00 - functions.SMO
16:17:42 - functions.SMO
16:18:10 - functions.Logistic
16:38:59 - functions.Logistic
18:00:40 - functions.Logistic

**Classifier output**

```
                          Class
Variable                      Y
=================================
Gender=Female            1.0245
Married=Yes              1.8335
Education=Not Graduate    0.678
Self_Employed=Yes        0.9811
ApplicantIncome               1
CoapplicantIncome             1
LoanAmount               0.9982
Loan_Amount_Term         0.9989
Credit_History          47.8801
Property_Area=Urban       0.839
Property_Area=Rural      0.6911
Property_Area=Semiurban  1.6302


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         497               80.9446 %
Incorrectly Classified Instances       117               19.0554 %
Kappa statistic                          0.4825
Mean absolute error                      0.2966
Root mean squared error                  0.3901
Relative absolute error                 68.9518 %
Root relative squared error             84.1492 %
Total Number of Instances              614

=== Detailed Accuracy By Class ===
```

**Status**

OK                                                                     Log

Type here to search

7:42 PM
5/8/2021

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

**Classifier**

Choose | Logistic -R 1.0E-8 -M -1 -num-decimal-places 4

**Test options**

○ Use training set
○ Supplied test set    Set...
● Cross-validation  Folds  10
○ Percentage split    %  66

More options...

(Nom) Loan_Status

Start | Stop

**Result list (right-click for options)**

16:09:00 - functions.SMO
16:17:42 - functions.SMO
16:18:10 - functions.Logistic
16:38:59 - functions.Logistic
18:00:40 - functions.Logistic

**Classifier output**

```
Property_Area=Urban           0.839
Property_Area=Rural           0.6911
Property_Area=Semiurban       1.6302


Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         497               80.9446 %
Incorrectly Classified Instances       117               19.0554 %
Kappa statistic                          0.4825
Mean absolute error                      0.2966
Root mean squared error                  0.3901
Relative absolute error                 68.9518 %
Root relative squared error             84.1492 %
Total Number of Instances              614

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                 0.981    0.568    0.792      0.981   0.876      0.539    0.755     0.843     Y
                 0.432    0.019    0.912      0.432   0.587      0.539    0.755     0.670     N
Weighted Avg.    0.809    0.396    0.829      0.809   0.786      0.539    0.755     0.789

=== Confusion Matrix ===

   a    b   <-- classified as
 414    8 |   a = Y
 109   83 |   b = N
```

**Status**

OK

Log    x 0

Type here to search    |    Distribution ...    train.csv Pro...    smartinternz...    Weka GUI C...    Weka Explor...    1299 - org...    ENG    6:25 PM    5/8/2021

```
12
13⊖        public static Instances getInstances (String filename)
14         {
```

Problems  @ Javadoc  Declaration  Console ✕

<terminated> LogReg [Java Application] C:\Program Files\Java\jdk-16.0.1\bin\javaw.exe  (04-May-2021, 4:25:57 pm – 4:26:10 pm)

```
614
** Logistic Regression Evaluation with Datasets **

Correctly Classified Instances        182                84.2593 %
Incorrectly Classified Instances       34                15.7407 %
Kappa statistic                         0.6034
Mean absolute error                     0.3106
Root mean squared error                 0.3824
Relative absolute error                71.2395 %
Root relative squared error            81.3534 %
Total Number of Instances              216

Confusion matrix:
[142.0, 3.0]
[31.0, 40.0]
--------------------
Area under the curve
0.7565808644973288
--------------------
[Correct, Incorrect, Kappa, Total cost, Average cost, KB relative, KB information, Correlation, Complexity 0, Complexity s
Recall :0.56
Precision:0.93
F1 score:0.7
Accuracy:0.84
--------------------
```

# ADVANTAGES AND DISADVANTAGES

Advantages

- The proposed model has automated the loan eligibility process of the customer

- Logistic regression proved to the best algorithm to build the model.

A loan company can use this algorithm without any concerns.

Disadvantages

• The loan companies grant the loan after an intensive process of verification and validation.however they still don't have assurance if the applicant is able to repay the loan with no difficulties.

# **Conclusion**

The conclusion that can be dervided from the project are

• The project is more useful in  predicting the loan eligibility status of customer.

• The accuracy is 84%

- with just limited data we can predict status of any customer