

# 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.

### **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**

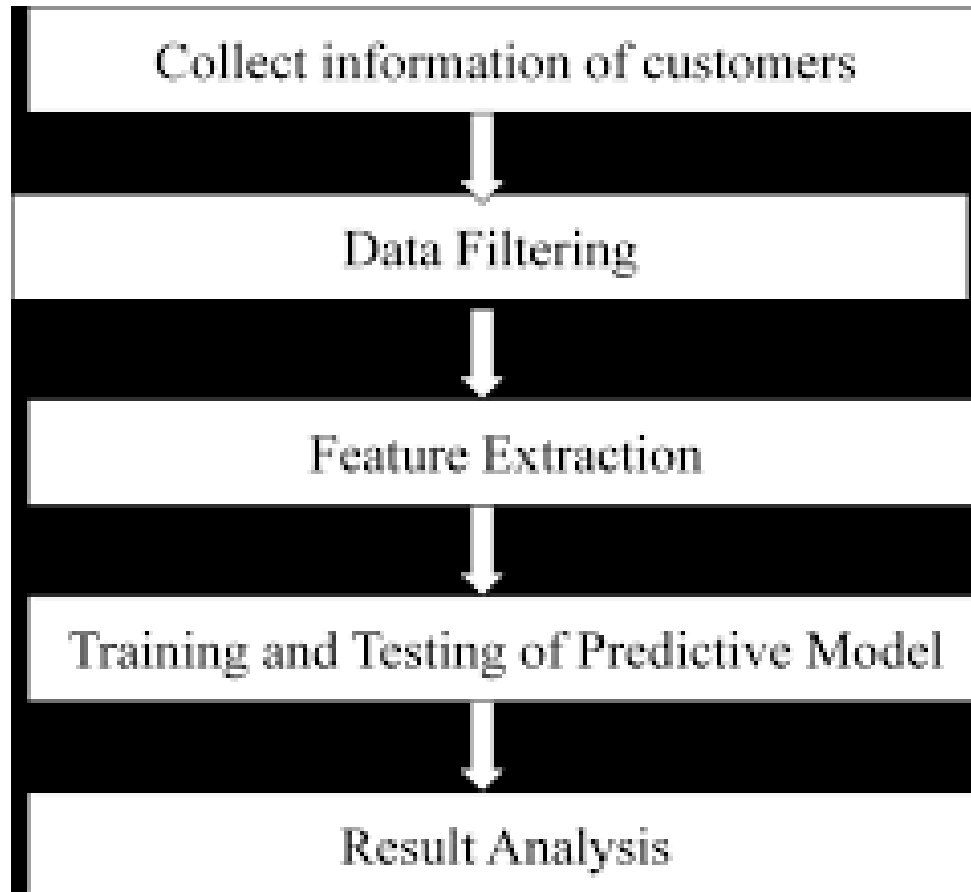
- The existing approach to solve this problem is to employ a person and do the background checks needed to make sure that the customer is eligible for the loan based on customer detail provided while filling online application form

- **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

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Classifier

ChooseLogistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Set...

Folds10

%88

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

VariableClass

Y

=====

Gender=Female1.0245

Married=Yes1.8335

Education=Not Graduate0.678

Self\_Employed=Yes0.9811

ApplicantIncome1

CoapplicantIncome1

LoanAmount0.9982

Loan\_Amount\_Term0.9989

Credit\_History47.8801

Property\_Area=Urban0.839

Property\_Area=Rural0.6911

Property\_Area=Semiurban1.6302

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances49780.9446 %

Incorrectly Classified Instances11719.0554 %

Kappa statistic0.4825

Mean absolute error0.2966

Root mean squared error0.3901

Relative absolute error68.9518 %

Root relative squared error84.1492 %

Total Number of Instances614

=== Detailed Accuracy By Class ===

Status

OK

Log

Type here to search

SPS-106...

Weka G...

Weka Ex...

1299 - o...

Presenta...

Screens...

7:42 PM5/8/2021

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Classifier

ChooseLogistic -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Set...

Folds10

%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=Urban0.839

Property\_Area=Rural0.6911

Property\_Area=Semiurban1.6302

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances49780.9446 %

Incorrectly Classified Instances11719.0554 %

Kappa statistic0.4825

Mean absolute error0.2966

Root mean squared error0.3901

Relative absolute error68.9518 %

Root relative squared error84.1492 %

Total Number of Instances614

=== 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

4148 | a = Y

10983 | b = N

Status

OK

Log

Type here to search

Distribution ...

train.csv Pro...

smartinternz...

Weka GUI C...

Weka Explor...

1299 - org...

6:25 PM5/8/2021

Weka Explorer

PreprocessClassifyClusterAssociateSelect attributesVisualize

Open file...Open URL...Open DB...Generate...UndoEdit...Save...

Filter

ChooseReplaceMissingValuesApplyStop

Current relation

Relation: train (1)-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.RepInstances: 614Attributes: 11Sum of weights: 614

Attributes

AllNoneInvertPattern

No.	Name
1	<input type="checkbox"/> Gender
2	<input type="checkbox"/> Married
3	<input type="checkbox"/> Education
4	<input type="checkbox"/> Self_Employed
5	<input type="checkbox"/> ApplicantIncome
6	<input type="checkbox"/> CoapplicantIncome
7	<input type="checkbox"/> LoanAmount
8	<input type="checkbox"/> Loan_Amount_Term
9	<input type="checkbox"/> Credit_History
10	<input type="checkbox"/> Property_Area
11	<input checked="" type="checkbox"/> Loan_Status

Remove

Selected attribute

Name: Loan\_StatusType: NominalMissing: 0 (0%)Distinct: 2Unique: 0 (0%)

No.	Label	Count	Weight
1	Y	422	422.0
2	N	192	192.0

Class: Loan\_Status (Nom)Visualize All

Label	Count
Y	422
N	192

Status

OKLog

Type here to search

(1) WhatsApp...

Weka GUI Ch...

Weka Explorer

1299 - org.m...

Presentation...

7:19 PM5/8/2021

```
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 be 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 derived 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