

INTRODUCTION

1.1 Overview

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.

1.2 Purpose

The main aim of this use-case is to build a predictive model to predict if an applicant is able to repay the lending company or not.

LITERATURE SURVEY

2.1 Existing problem

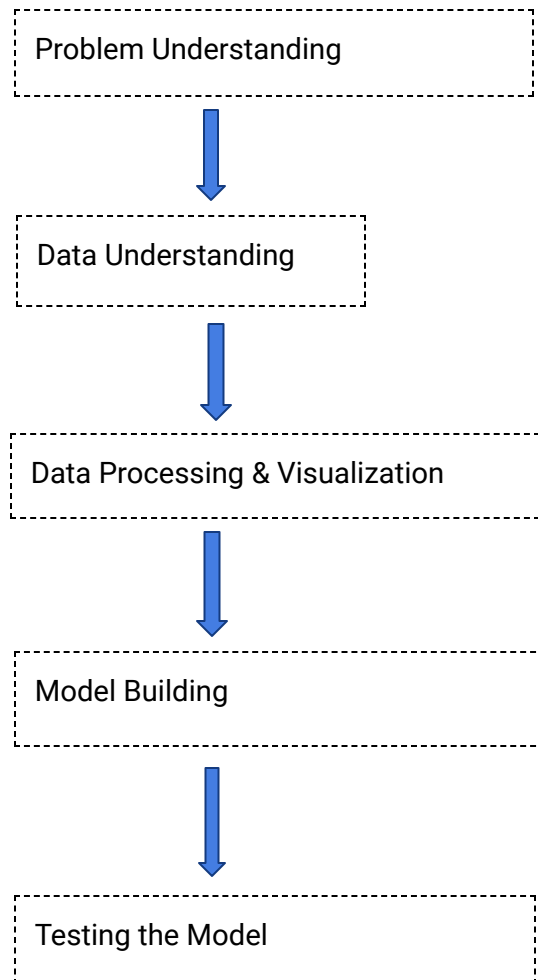
The Existing approach to solve this problem is to employ a person and do the background checks needed to make sure that the customer is eligible for the loan based on customer details provided while filling online application form.

2.2 Proposed Solution

It's a classification problem, given information about the application we have to predict whether they all be able to pay the loan or not. Details are Gender, Marital Status, Education, Number of dependents, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segment, those are eligible for loan amount so that they can specifically target these customers.

THEORITICAL ANALYSIS

3.1 Block Diagram



3.2 Hardware/Software Designing

Software Required:

- Weka software
- Eclipse IDE
- Install tablesaw and weka packages

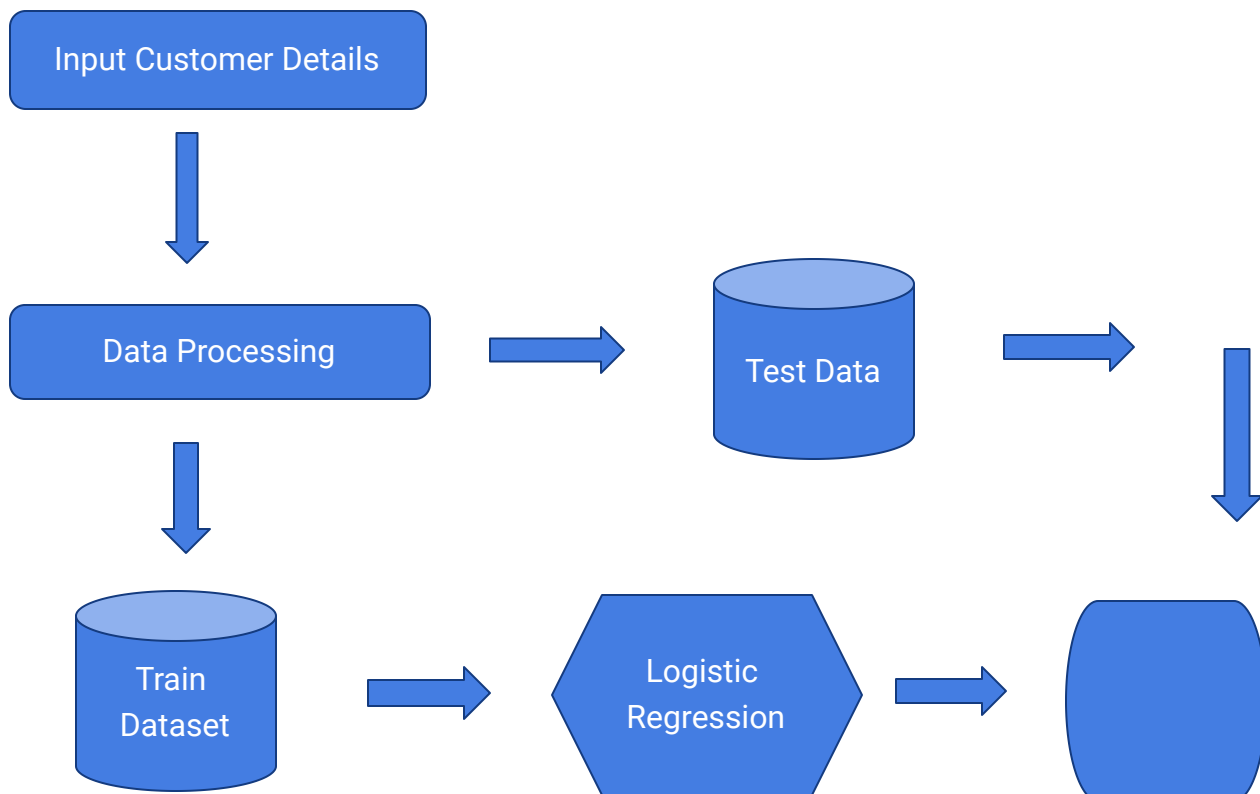
-

EXTERNAL INVESTIGATIONS

We have 12 independent variables and 1 target variable, i.e, Loan_status in the training dataset. We can see there are 2 formats of datatypes, i.e , Nominal and Numerical. The loan of 422(around 69%)people out of 614 were approved.

- It seems people with a credit history as 1 are more likely to get their loans approved.
- The proportion of loans getting approved in the semi-urban area is higher as compared to that in rural or urban areas.
- Application income does not affect the chances of loan approval.
- It shows that if co-applicants income is less the chances of loan approval are high. But this does not look right. The possible reason behind this may be that most of the applicants don't have any co-applicant so the co-applicant income for such applicants is 0 and hence the loan approval is not dependent on it.
- we can see that Proportion of loans getting approved for applicants having low Total_Income is very less compared to that of applicants with Average, High & very High Income.
- It can be seen that the proportion of approved loans is higher for Low and Average Loan Amount as compared to that of High Loan Amount.

Flow Chart



Result

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose None Apply Stop

Current relation: Relation: train_u6lujuX_CVtuZ9l Instances: 614 Attributes: 13 Sum of weights: 614

Attributes: All None Invert Pattern

No.	Name
1	Loan_ID
2	Gender
3	Married
4	Dependents
5	Education
6	Self_Employed
7	ApplicantIncome
8	CoapplicantIncome
9	LoanAmount
10	Loan_Amount_Term
11	Credit_History
12	Property_Area
13	Loan_Status

Remove

Selected attribute: Name: Loan_ID Missing: 0 (0%) Distinct: 614 Type: Nominal Unique: 614 (100%)

No.	Label	Count	Weight
1	LP001002	1	1.0
2	LP001003	1	1.0
3	LP001005	1	1.0
4	LP001006	1	1.0
5	LP001008	1	1.0
6	LP001011	1	1.0
7	LP001013	1	1.0

Class: Loan_Status (Nom) Visualize All

Too many values to display.

Status: OK Log x 0

Java desktop - ml/src/main/java/ml/LoanPredictionProject.java - Eclipse IDE

File Edit Source Refactor Navigate Search Project Run Window Help

Package Explorer: EmpDetails, external, ml, src/main/java, ml, LoanPredictionProject.java, src/main/resources, src/test/java, src/test/resources, JRE System Library [J2SE-1.5], Maven Dependencies, src, target, pom.xml, sample, StudentDetails

LoanPredictionProject.java

```
28 dataset = source.getDataSet();
29 dataset.setClassIndex(dataset.numAttributes()-1);
30
31
32 } catch (Exception e) {
33 // TODO Auto-generated catch block
34 e.printStackTrace();
35 }
```

Problems Javadoc Declaration Console

<terminated> LoanPrediction [Java Application] C:\Program Files\Java\jdk-15.0.1\bin\javaw.exe (06-May-2021, 12:21:00 pm)

614 rows X 13 cols

Structure of train_u6lujuX_CVtuZ9l.csv

Index	Column Name	Column Type
0	Loan_ID	STRING
1	Gender	STRING
2	Married	STRING
3	Dependents	STRING
4	Education	STRING
5	Self_Employed	STRING
6	ApplicantIncome	INTEGER
7	CoapplicantIncome	DOUBLE
8	LoanAmount	INTEGER
9	Loan_Amount_Term	INTEGER
10	Credit_History	INTEGER
11	Property_Area	STRING
12	Loan_Status	BOOLEAN

train_u6lujuX_CVtuZ9l.csv

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount
LP001002	Male	No	0	Graduate	No	5849	0	128
LP001003	Male	Yes	1	Graduate	No	4583	1508	66
LP001005	Male	Yes	0	Graduate	Yes	3000	0	120
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	141
LP001008	Male	No	0	Graduate	No	6000	0	

614

208M of 337M

19:51 08-05-2021

Advantages & Disadvantages

Advantages:

- The Proposed Model has automated the Loan Eligibility Process of the customer.
- The Model showed 84% percent accuracy
- Logistics Regression proved to be the best algorithm to built the model,since all other classification algorithms showed only 70-80%accuracy.
- A Loan company can use this prediction model without any concern.

Disadvntges:

- The loan companies grant a loan after an intensive proess of verification and validation. However,They still dont have assurance if the applicant is able to repay the laon with no difficulties.

Applications

The main application of the model are in Banking sector and other loan lending companies.Loans are the core business of banks. The main profit comes directly from the loan's interest.

Conclusion

The conclusion that are can be derived from this project are:

- The Project is more usefull in predicting the loan eligibility status of a customer.
- The Accuracy is 84%.
- Banking sector and other loan lending companies cen be benifited by this prediction model.
- With just limited information we can do great predictions.

Future Scope

In Future we can improve the accuracy of the model by adding some more attributes like previous loans status,years of more services etc..

-

Bibliography

References:

<https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/>

<https://towardsdatascience.com/predict-loan-eligibility-using-machine-learning-models-7a14ef904057>

Kaggle for datasets.

<https://towardsdatascience.com/introduction-to-logistic-regression-66248243c148>