

Loan Approval Prediction System

Submitted in partial fulfillment of the requirements of the degree of

BACHELOR OF COMPUTER ENGINEERING

by

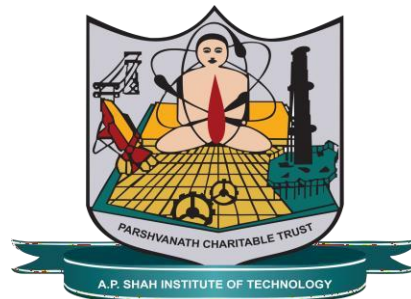
Moksha Shinde, 19102064

Karan Shah, 19102038

Kanchi Patil, 19102013

Guide:

Prof. Sofiya Mujawar



Department of Computer Engineering

A. P. SHAH INSTITUTE OF TECHNOLOGY, THANE

(2021-2022)



A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the Mini Project 2B entitled “**Loan Approval Prediction System**” is a bonafide work of “**Moksha Shinde (19102064), Karan Shah (19102038), and Kanchi Patil (19102013)**” submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Engineering in Computer Engineering**

Prof. Sofiya Mujawar
Guide

Prof. Sachin H Malave
Head of Department



A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval for TE

This Mini project report entitled ***Loan Approval Prediction System*** by ***Moksha Shinde, Karan Shah, and Kanchi Patil*** is approved for the degree of ***Bachelor of Engineering in Computer Engineering, 2021-22.***

Examiner Name Signature 1. _____
 2. _____

Date:

Place: A. P. Shah Institute of Technology, Thane

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Moksha Shinde, 19102064

(Signature)

Karan Shah, 19102038

(Signature)

Kanchi Patil, 19102013

Date:

Abstract

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. Though a lot of people are applying for loans, it's difficult to select the genuine applicant, who will repay the loan. While doing the process manually, a lot of misconceptions may happen when selecting the genuine applicant. Therefore we are developing a loan prediction system using machine learning, so the system automatically selects the eligible candidates. This is helpful to both bank staff and applicants. The time period for the sanction of the loan will be drastically reduced. In this project, we are predicting the loan data by using machine learning algorithms that are: Decision Tree, Naïve Baye's, and Random forest; and comparing these to receive the most efficient and accurate way of loan prediction.

The loan prediction machine learning system can be used to assess a customer's loan status and build strategies. This system extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status). These reports make a bank manager's job simpler and quicker, while also reducing the errors caused by manual calculation methods. Everyone that requires a loan may not always be informed about the requirements and necessary eligibility criteria.

CONTENTS

1 Introduction	08
2 Literature Survey.....	09
3 Problem Statement	12
4 Objective and Scope.....	12
5 Experimental Setup	13
5.1 Hardware Requirement	
5.2 Software Requirement	
6 System Design	14
6.1 UML Diagram	
6.2 Algorithm/Process	
7 Implementation	20
7.1 Code	
7.2 Screenshots	
8 Result.....	22
9 Conclusion	23
References	23
Annexure 1 - Project Planning (Using Gantt chart)	23

LIST OF FIGURES

1.1 UML Diagram.....	14
1.2 System Algorithm	14
1.3 Loan against Property Algorithm.....	15
1.4 Gold Loan Algorithm	16
1.5 Home Loan Algorithm.....	17
1.6 Business Loan Algorithm.....	18
1.7 Education Loan Algorithm.....	19
2.1 Age attribute graph 1	20
2.2 Age attribute graph 2	21
2.3 Age attribute graph 3	21
3.1 Credit Score attribute graph 1.....	21
3.2 Credit Score attribute graph 2.....	22
4.1 Comparing Different Algorithm Results	22

Introduction

A Prediction Model uses data mining, statistics, and probability to forecast an outcome. Every model has some variables known as predictors that are likely to influence future results. With the data that was collected from various resources then a statistical model is made. It can use a simple linear equation or a sophisticated neural network mapped using complex software. As more data becomes available the model becomes more refined and the error decreases meaning then it'll be able to predict with the least risk and consume as less time as it can. The Prediction Model helps the banks by minimizing the risk associated with the loan approval system and helps the applicant by decreasing the time taken in the process.

Banks make loans to customers in exchange for the guarantee of repayment. Some would default on their debts, unable to repay them for several reasons. The bank retains insurance to minimize the possibility of failure in the case of a default. The insured sum can cover the whole loan amount or just a portion of it. Banking processes use manual procedures to determine whether or not a borrower is suitable for a loan based on results. Manual procedures were most effective, but they were insufficient when there were a large number of loan applications. At that time, making a decision would take a long time. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies. This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status). These reports make a bank manager's job simpler and quicker.

48% to 77% percent of applied loans get rejected every year.

That is why you should be prepared before applying for a loan.

Loans are of several types: Home Loan, Business Loan, Educational Loan, etc. Each of these different types of loans has different factors on which their approval is determined. We have used various algorithms on several types of loans, keeping in regard the eligibility conditions of each loan, and found the algorithm with the highest accuracy rate to predict the loan approval and make this tedious task easier, while eliminating the plausible manual errors in the calculations.

The loan Approval Prediction System helps you predict your loan eligibility and lets you find your shortcomings so you can be prepared while applying for a loan.

This model intends to make the loan prediction process easier for both- the customer applying for the loan, as well as the bank sanctioning the loan.

Literature Survey

Paper Name: An Approach for Prediction of Loan Approval using Machine Learning Algorithm

Author: Mohammad Ahmad Sheikh, Amit Kumar Goel, Tapas Kumar

In our banking system, banks have many products to sell however the major source of earnings of any bank is its credit line. So, they could earn from the interest of these loans which they credit. A financial institution's profit or loss relies a massive extent upon loans i.e., whether the clients are paying back the mortgage or defaulting. By predicting the loan defaulters, the bank can lessen its Non-Performing Assets. This makes the study of this phenomenon very critical. Previous studies in this period have proven that there are a lot of strategies to study the hassle of controlling loan default. But because the proper predictions are very crucial for the maximization of profits, it is vital to study the nature of the distinct techniques and their comparison. This paper has taken the data of previous customers of various banks to whom on a set of parameters loans were approved. So the machine learning model is trained on that record to get accurate results. The main objective of this research is to predict the safety of loans. To predict loan safety, the logistic regression algorithm is used. First, the data is cleaned so as to avoid the missing values in the data set. To train this model, data set of 1500 cases and 10 numerical and 8 categorical attributes has been taken. To credit a loan to customer various parameters like CIBIL Score (Credit History), Business Value, Assets of Customer, etc. have been considered.

A very critical method in the predictive analysis is used to study the problem of predicting loan defaulters: The Logistic regression model. It is used for the description of data and uses to explain the relationship between a single binary variable and single or multiple nominal, ordinal, and ratio level variables that are independent in nature. The facts are accrued from the Kaggle for analysis and prediction. Logistic Regression models have been performed and the different measures of performance are computed. The models are compared on the premise of the overall performance measures together with sensitivity and specificity. The outcomes have proven that the model produces different results. The model is marginally better as it consists of variables (personal attributes of the client like age, purpose, credit score history, credit score amount, credit score duration, and so on.) apart from checking account information (which suggests the wealth of a client) that ought to be taken under consideration to calculate the probability of default on mortgage correctly. Therefore, through the use of a logistic regression approach, the right clients to be targeted for granting loans may be easily detected by evaluating their probability of default on loans. The model concludes that a bank should not only target wealthy customers for granting loans however it should determine the other attributes of a client as well which play a very crucial part in credit granting decisions and predicting the loan defaulters.

Paper name – Prediction for Loan Approval using Machine Learning Algorithm

Author – Ashwini S. Kadam , Shraddha R. Nikam , Ankita A. Aher , Gayatri V. Shelke , Amar S. Chandgude

A loan is when one or more people, organizations, or other entities lend money to other people, organizations, or entities. The recipient incurs a debt for which he or she is generally responsible for paying interest until the loan is repaid. The project's objective is to ensure that a person, institution, or organization applying for a loan is verified thoroughly before sanctioning them a loan. Several criteria like gender, education, number of dependents, to name a few, have to be taken into consideration before approving the loan.

“Prediction for Loan Approval using Machine Learning Algorithm” Author Ashwini S. Kadam, Shraddha R. Nikam , Ankita A. Aher , Gayatri V. Shelke , Amar S. Chandgude Year-2021. The main objective of this

project is to predict the safety of loan using Machine learning (SVM and Naïve Bayer's algorithm). This paper is divided into following sections:

- i. Data Collection
- ii. Data Pre-processing (Data Cleaning)
- iii. Model Selection and Evaluation
- iv. Classification
- v. Result (output)

This project has taken the data of previous customers of various banks to who on a set of parameters loan were approved. So the machine learning model is trained on that record to get accurate results. Our main objective of this project is to predict the safety of loan, by using the SVM and Naïve Bayer's algorithm. It was concluded in the research paper that, the Naïve Bayes model is extremely efficient and gives a better result when compared to other models (SVM algorithm). It works correctly and fulfils all requirements of bankers. It predicts the loan is approved or reject to loan applicant or customer.

Paper Name: Loan Approval Prediction Model a Comparative Analysis

Authors: Afrah Khan, Eakansh Bhadola, Abhishek Kumar, and Nidhi Singh.

In today's time the capital to require for any individual to purchase any asset is very high as the cost of assets is increasing day by day. It is not logical and practical to purchasing assets out of just savings. The easiest way an individual can get funds is to apply for a loan. There are different types of loans available for different purposes such as Property loan for buying property, Gold Loan for buying gold, Student loans for educational purposes, etc. But getting a loan is not easy, for any application to get approved it has to go through number of stages and verification and it's still not necessary that the application will be approved. To decrease the number of steps for an application to get approval and save time and also to calculate the risk factor associated with any application with the loan many loan predictions models were introduced. The main purpose of this project is to select the prediction model with least error and more accuracy and can be used in real world to predict if the loan is approvable or not taking the risk factor in mind. The algorithms that were considered for testing were: Random Forest, Naïve Bayes, KNearest Neighbors (KNN), Logistic Regression, and Decision Tree. Several Factors / Parameters were included such as type of loan, Credit Score rating, Assets and much more while applying the algorithms. The data was included in the training and testing set in the CSV format. The process of the project was divided into several steps such as Data loading, data exploring, data cleaning, data distribution, model selection and accuracy calculation.

For understanding the distribution of the data different kinds of plots were used such as histogram, box plot. After comparing and analyzing the models, they found out that the prediction model based on the Random Forest proved to be the most accurate and fitting of them all. This model can be useful in reducing the time and manpower required to approve loans and filter out the perfect candidate for providing loans.

Paper Name: Loan Prediction by using Machine Learning Models(Mar_April 2019)

Authors: - Pidikiti Supriya , Myneedi Pavani , Nagarapu Saisushma , Namburi Vimala Kumari , K Vikas

" Loan Prediction by using Machine Learning Models" Author- Pidikiti Supriya , Myneedi Pavani , Nagarapu Saisushma , Namburi Vimala Kumari , K Vikas Year- Mar_April 2019. With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets.

Decision tree algorithm in machine learning methods which efficiently performs both classification and regression tasks. Decision trees are widely used in the banking industry due to their high accuracy and ability to formulate a statistical model in plain language. In Decision tree each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continues value).

This problem is solved by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which gives the most accurate result. The dataset collected for predicting loan default customers is predicted into Training set and testing set. Generally, 80:20 ratio is applied to split the training set and testing set. The data model which was created using Decision tree is applied on the training set and based on the test result accuracy, Test set prediction is done. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

Problem Statement

Automate the loan eligibility process using various algorithms, based on customer information.

Objective and Scope

The main objective of the project is to compare the Loan Prediction Models, implemented by using various algorithms and choosing the best one out of them that can shorten the loan approval time and decrease the risk associated with it. It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

In the present scenario, a loan needs to be approved manually by a representative of the bank which means that person will be responsible for whether the person is eligible for the loan or not and also calculating the risk associated with it. As it is done by a human it is a time-consuming process and is susceptible to errors. If the loan is not repaid, then it accounts as a loss to the bank and banks earn most of their profits by the interest paid to them. If the banks lose too much money, then it will result in a banking crisis. This banking crisis affects the economy of the country. So it is very important that the loan should be approved with the least amount of error in risk calculation while taking up as the least time possible. So a loan prediction model is required that can predict quickly whether the loan can be passed or not with the least amount of risk possible. This model can be further developed by adding a user-friendly GUI, making the process simpler and the task less tedious.

Experimental Setup

Software requirements

1) Python 3.10.01

Modules Used: - NumPy

Pandas

matplotlib.pyplot

seaborn

scipy

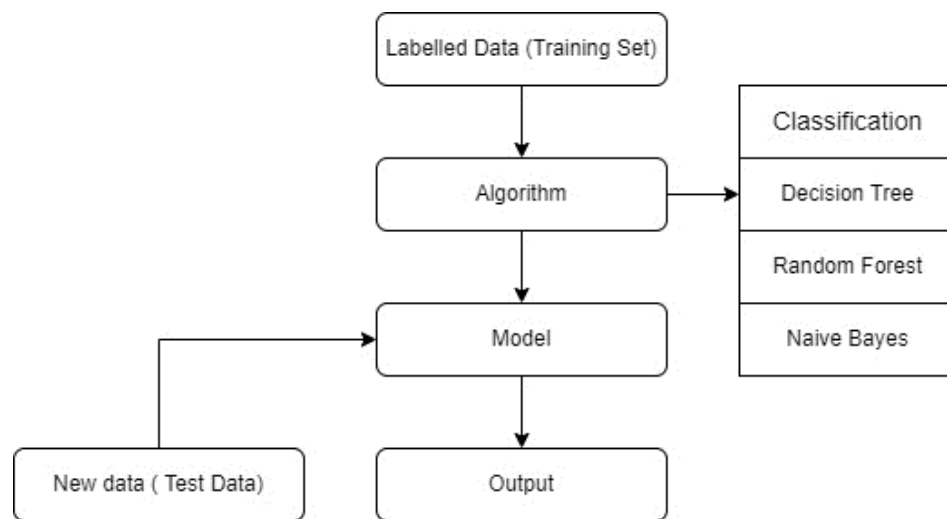
2) Google Colab

Hardware Requirements:-

1) i3, 8GB Ram.

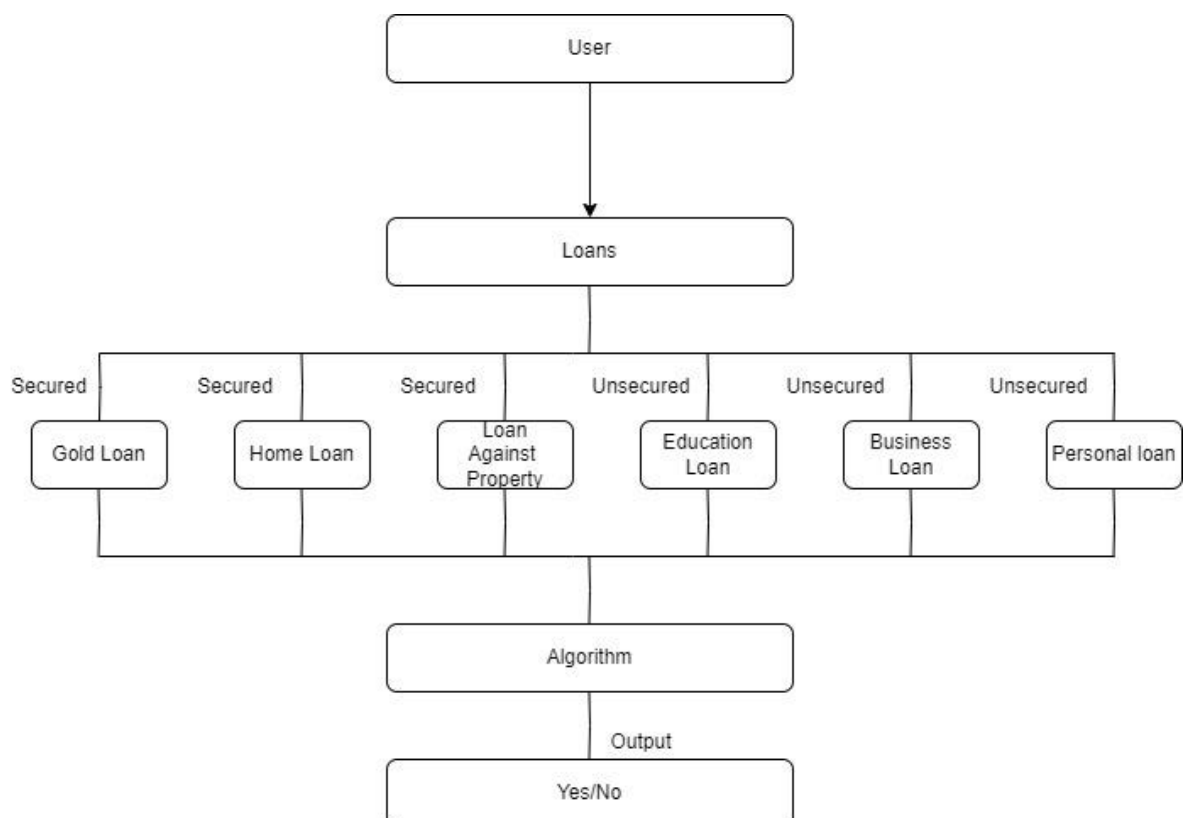
System Design

- UML Diagram

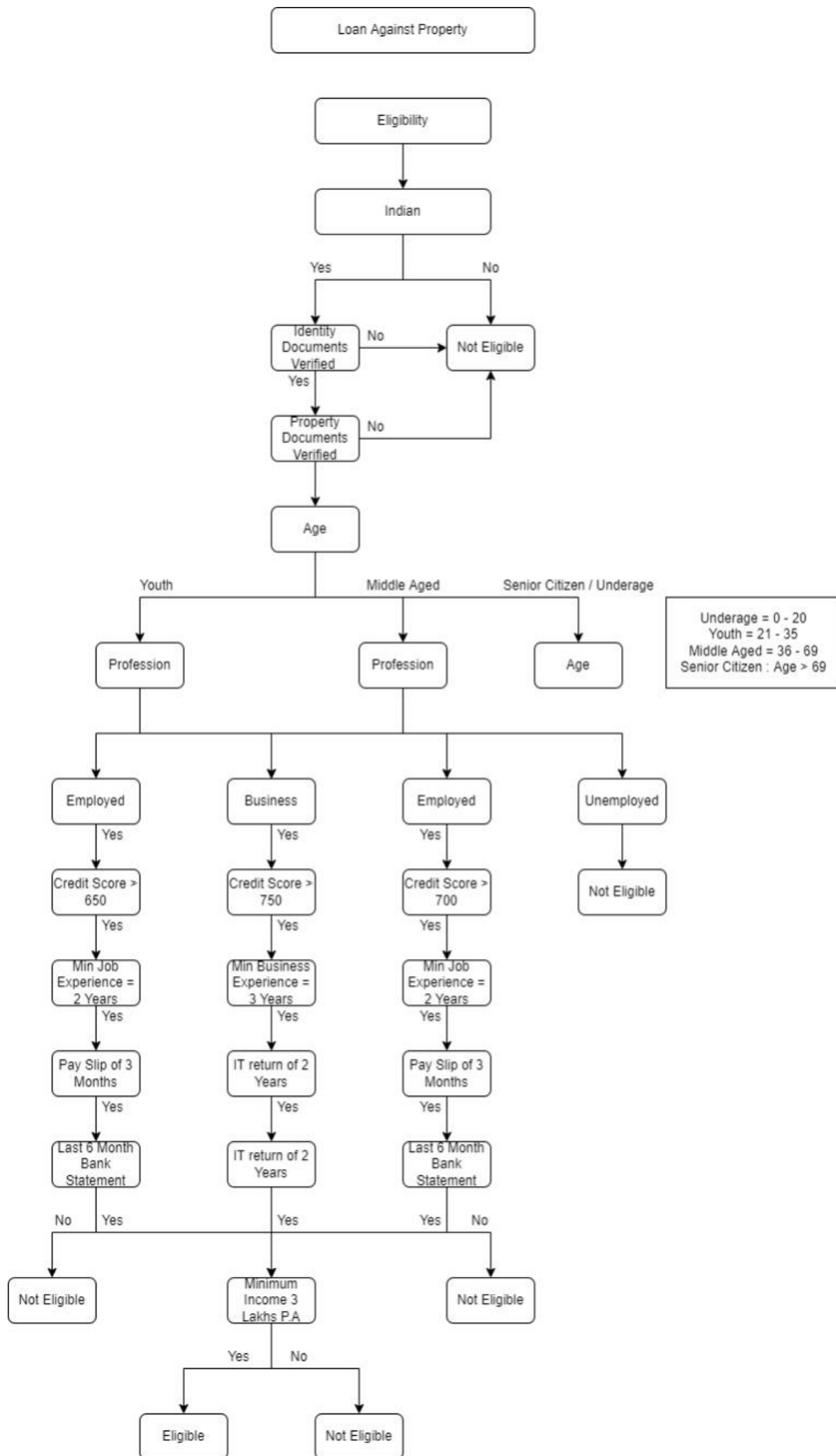


(1.1) UML Diagram

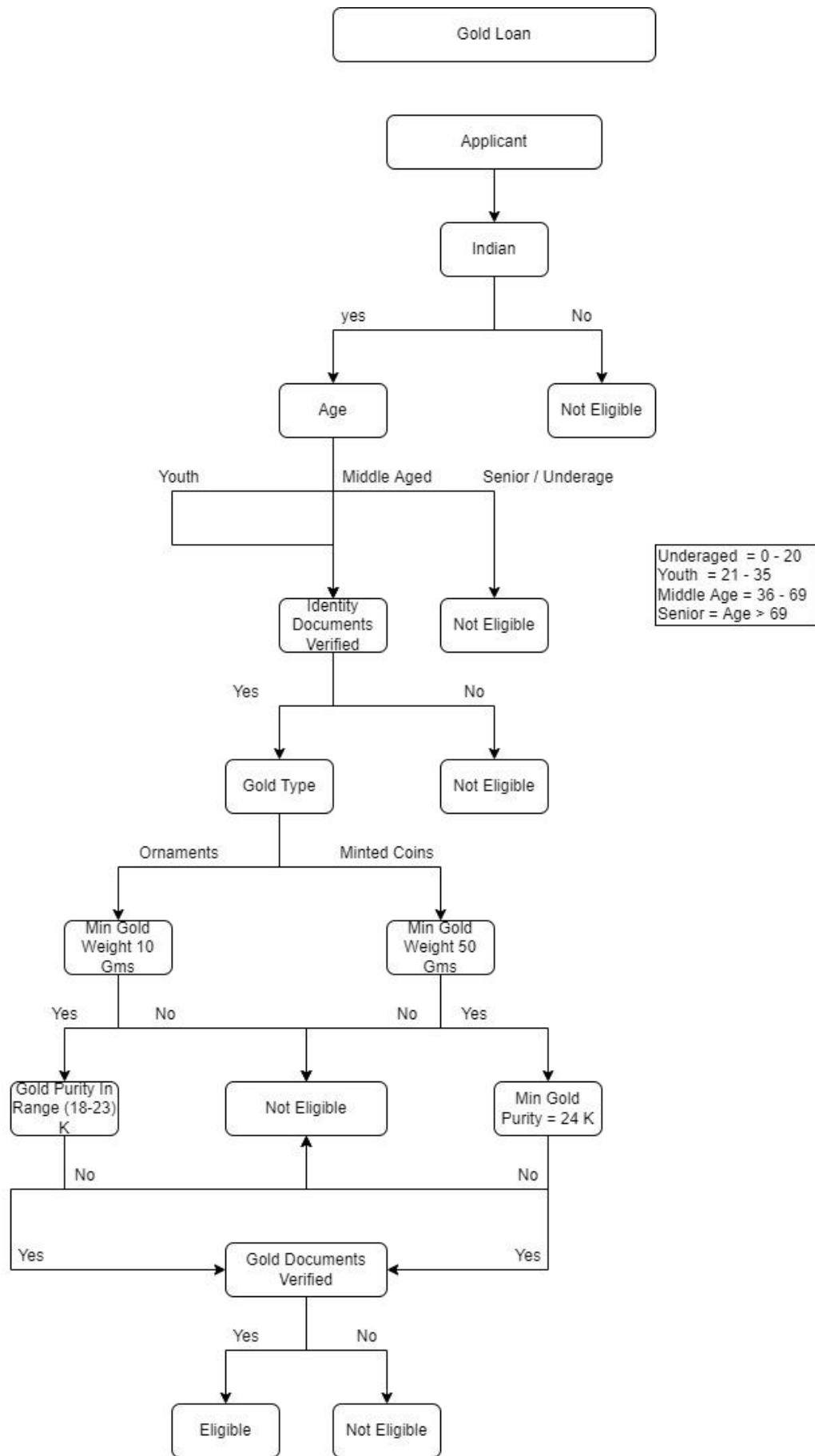
- Algorithm



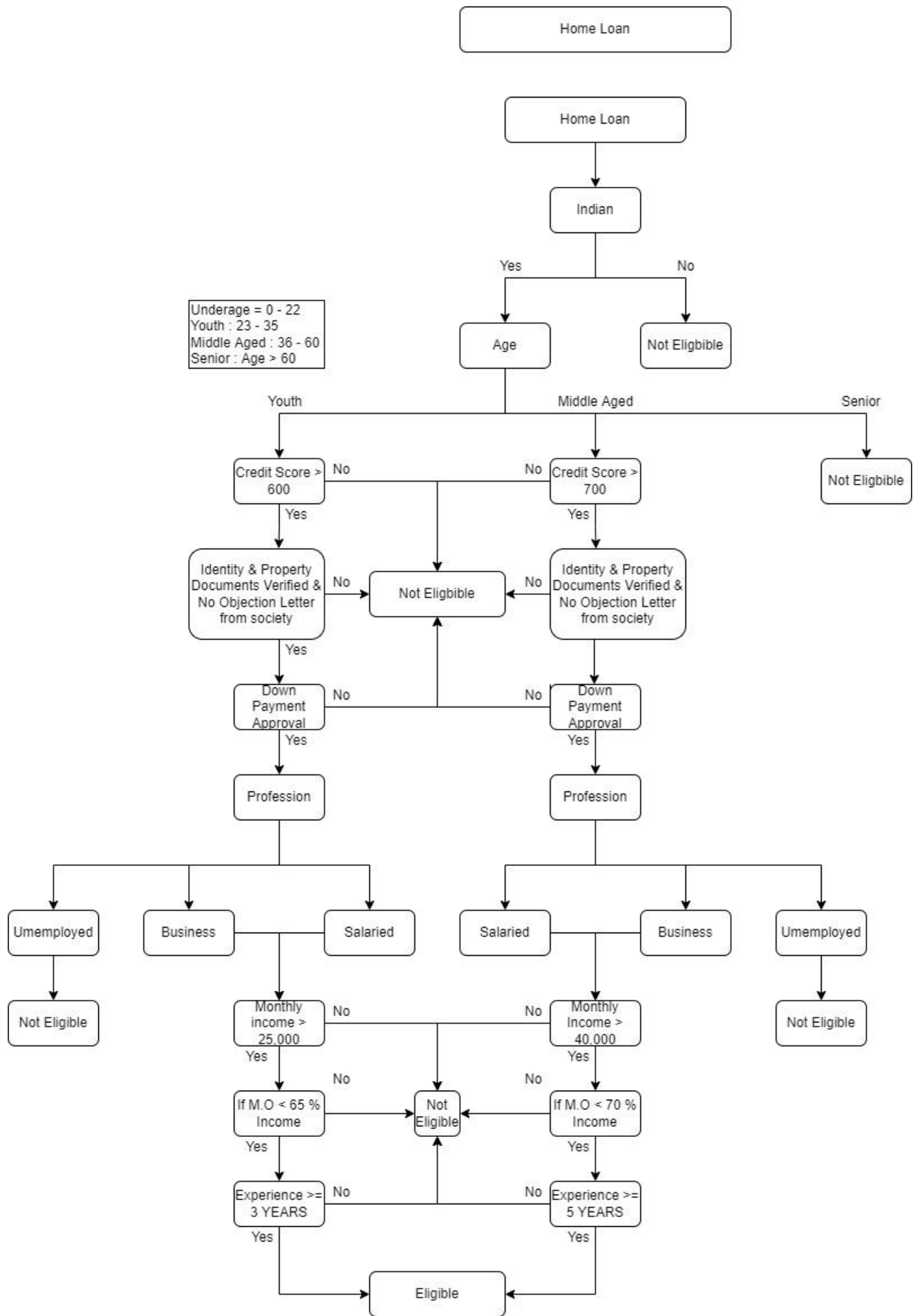
(1.2) System Algorithm



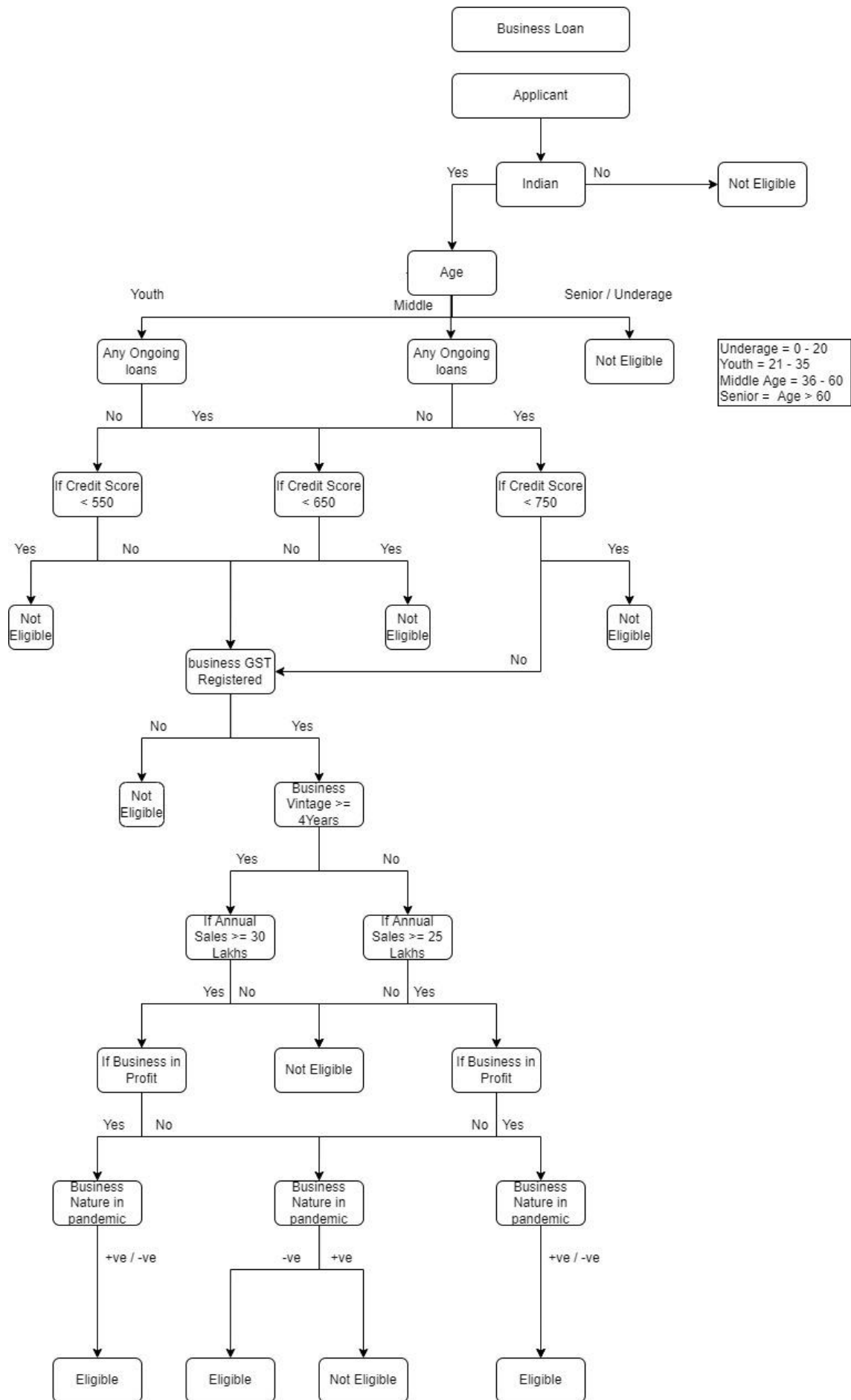
(1.3) Loan against Property Algorithm



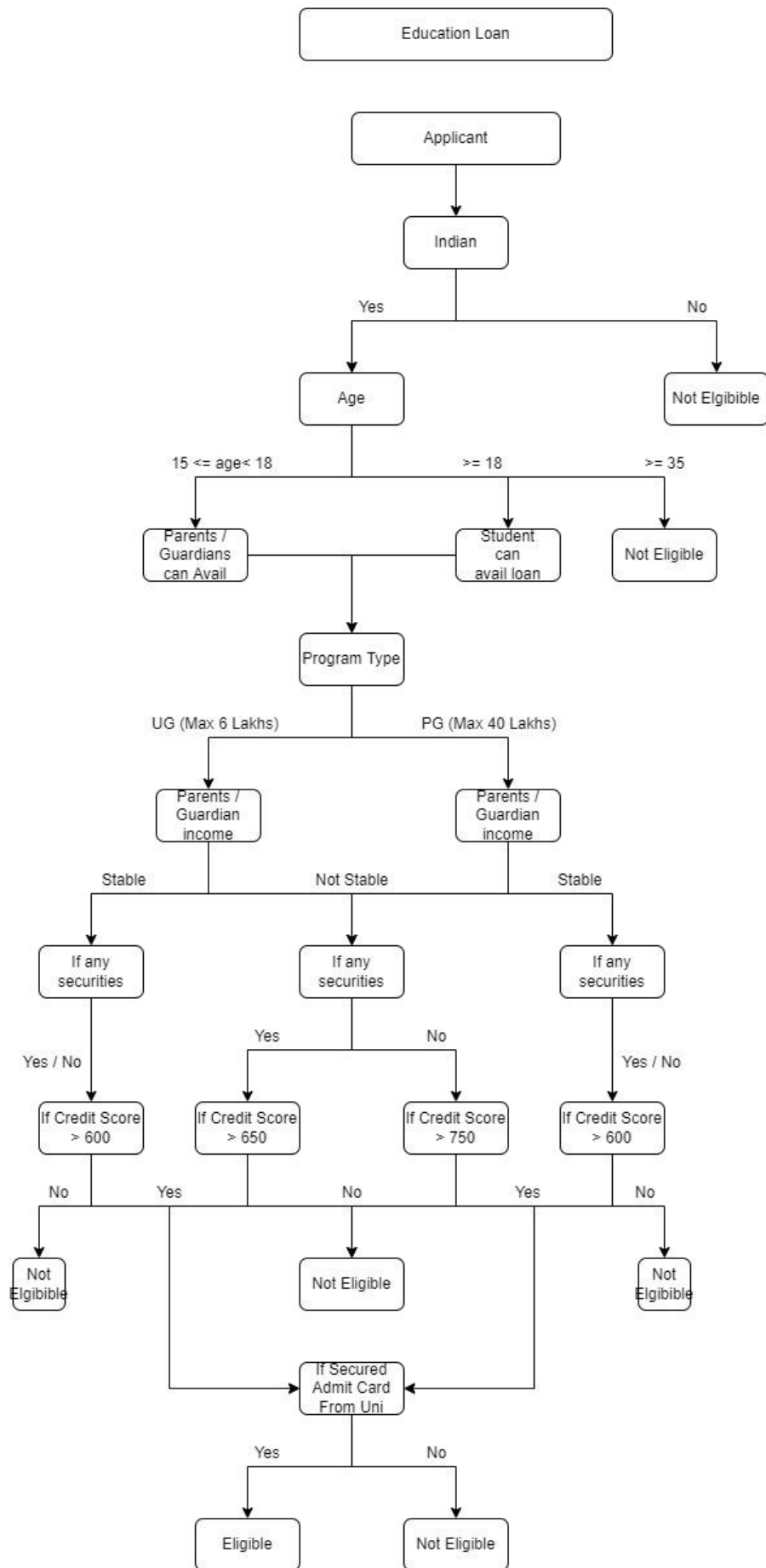
(1.4) Gold Loan Algorithm



(1.5) Home Loan Algorithm



(1.6) Business Loan Algorithm



(1.7) Education Loan Algorithm

Implementation

- **Code:**

<https://github.com/KaranShah-py/Loan-Prediction-Model-System>



- **Screenshots:**

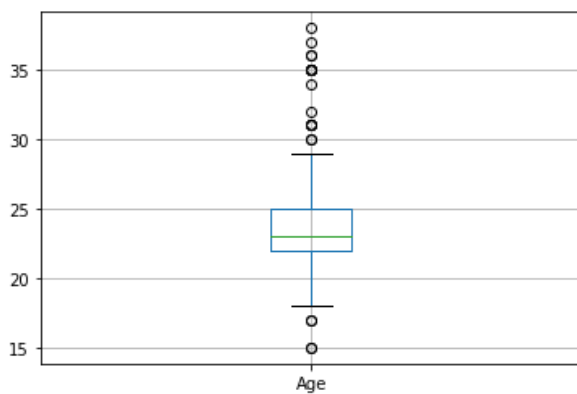
	Name	Nationality	Age	Program_Type	Total_income_P_G_S	Any_Securities	Credit Score	Secured Admit Card	Eligibility
0	Karan Shah	Indian	23.0	UG	Stable	Yes	660.0	Yes	Yes
1	Moksha Shinde	Indian	22.0	UG	Stable	No	558.0	Yes	No
2	Nikita Joshi	Indian	19.0	UG	Stable	No	525.0	Yes	No
3	Rahil Porwal	Indian	34.0	PG	Not Stable	No	529.0	Yes	No
4	Meet Shah	Indian	20.0	UG	Stable	No	764.0	Yes	Yes

As we can see that we have 6 independent and 1 dependent column in the dataset. Let us first check if there are any null values present in our dataset or not.

First let us check with the age attribute using a boxplot

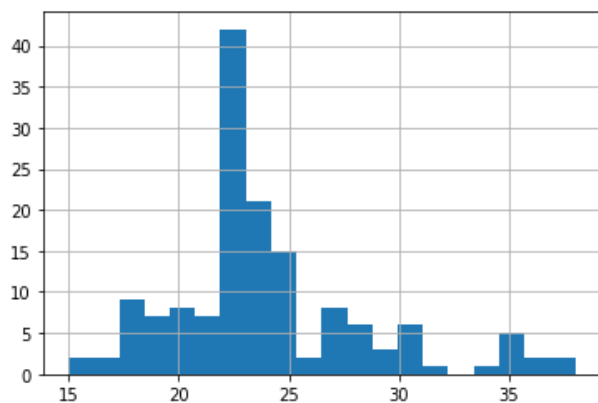
```
dataframe.boxplot(column="Age")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9daff59790>
```



(2.1) Age attribute graph 1

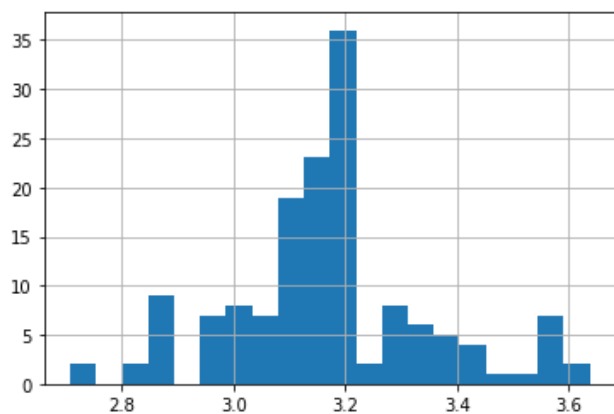
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9d9fe29590>
```



As we can see that the above graph is little left skewed let us not normalize the age data and create a new column of it named age_log.

(2.2) Age attribute graph 2

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9daf980f90>
```

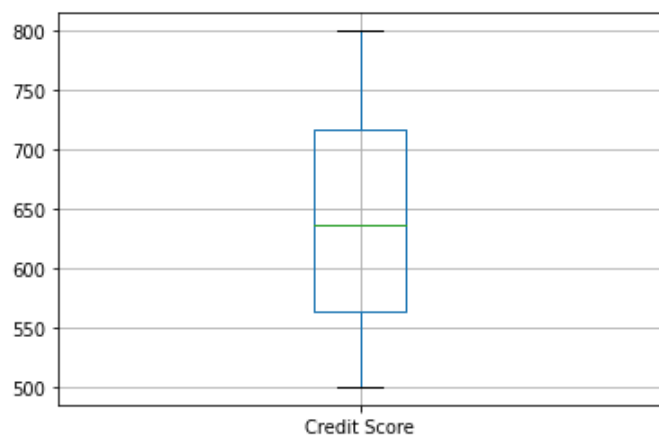


(2.3) Age attribute graph 3

Similarly let us check for credit score as well

```
dataframe.boxplot(column='Credit Score')
```

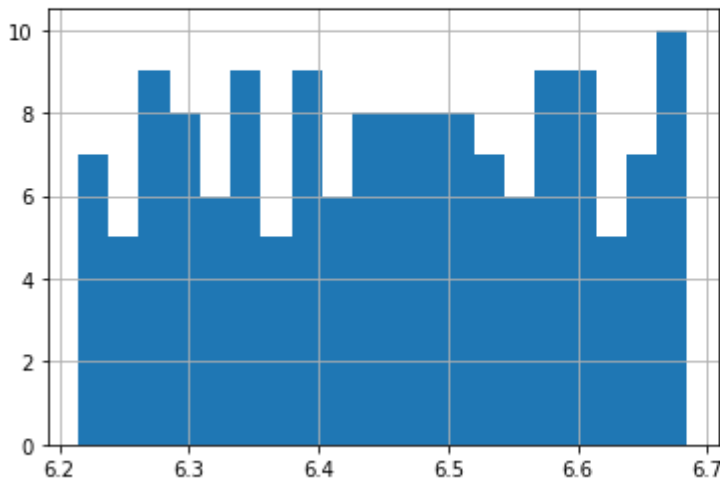
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f9daf868fd0>
```



(3.1) Credit Score attribute graph 1

```
dataframe['Credit_score_log'] = np.log(dataframe['Credit Score'])
dataframe['Credit_score_log'].hist(bins=20)
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f9daf868390>



since there are no much outliers present in the credit score column we do not need to normalize it.

(3.2) Credit Score attribute graph 2

RESULT:

Using Decision Tree Classifier Algorithm:

```
from sklearn import metrics
print('The Accuracy of our model using Decision Tree Algorithm is : ', metrics.accuracy_score(Y_pred_using_decision_tree,Y_test))
```

The Accuracy of our model using Decision Tree Algorithm is : 0.9333333333333333

Using Random Forest Classifier Algorithm:

```
print("The accuracy of our model using Random Forest Algorithm is : ", metrics.accuracy_score(Y_pred,Y_test))
```

The accuracy of our model using Random Forest Algorithm is : 0.9

Using Naive Bayes Classifier Algorithm:

```
print("The accuracy of our model using Naive Bayes Algorithm is : ", metrics.accuracy_score(Y_pred,Y_test))
```

The accuracy of our model using Naive Bayes Algorithm is : 0.4666666666666667

After the predicting of values, there comes the stage to check how accurate the model is by comparing the Predicted values with the test values. For model evolution using Sklearn's metrics model:

From sklearn import metrics,

The accuracy of our model using Decision Tree Classifier Algorithm is: **93 %**

The accuracy of our model using Random Forest Classifier Algorithm is: **90 %**

The accuracy of our model using Naive Bayes Classifier Algorithm is: **47 %**

Loan and Algorithm Accuracy		
Algorithm / Loan Type	Gold Loan	Education Loan
Decision Tree Algorithm	95%	93%
Random Forest Algorithm	76%	90%
Naive Bayes Algorithm	67%	47%

(4.1) Comparing Different Algorithm Results

Conclusion

After comparison and evaluation of all the chosen algorithms, the predictive models based on Decision Tree, Naïve Baye's, and Random Forest, give the accuracy as **93%**, **90%**, and **47%** for Education Loan and **95%**, **76%**, and **67%** for Gold Loan. It can be concluded with confidence that the Decision Tree Classifier Algorithm is extremely efficient and gives better results. It works correctly and fulfills all requirements of bankers. This system properly and accurately calculates the result. It predicts whether the loan is approved or rejected to the loan applicant or customer with very high accuracy.

References

- Research Methodologies for Student Performance Evaluation Using Educational Analytics Tools and Approaches. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019
- A Critical Review of Research on Student Self-Assessment. Educational Psychology and Methodology, University at Albany, Albany, NY, United States
- An Approach for Prediction of Loan Approval using Machine Learning Algorithm – IEEE. | IEEE Conference Publication |
- Loan Credibility Prediction System Based on Decision Tree Algorithm (ijert.org)
- International Research Journal of Engineering and Technology (IRJET) ,Dept. Of Computer engg., SND College of Engineering and research center, Yeola, Maharashtra, India., April 2021.
- Customer Loan Prediction Using Supervised Learning Technique (ijsrp.org)

Annexure 1 - Project Planning (Using Gantt chart)

