Intelligent Loan Assistant using Machine Learning and Data Science

Ashwini Shinde[1], Mitali Yadav[2], Pranoti Deshmukh[3], Mrs. Asmeeta Mali[4]

[1-3]Student, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

ABSTRACT

In this technology era there are so many enhancements in the banking sector as well. Nowadays people are preferring taking loans to buy a home, a new car, pursuing education (Degree). So the number of loan applications are increasing day by day. Loan approval is very difficult as the bank wants to invest their asset in safe hands and needs to analyse various aspects of approving the loan. As this process is difficult it needs to be automated. To automate this process we have used machine learning technique which will predict if the person is eligible to take a loan or not. We have considered various parameters for prediction. With this implementation, the user can quickly check his/her loan status.

Keywords: Machine Learning, Prediction, Loan Prediction, Loan Dataset, Logistic Regression, Random Forest, SVM, Decision Tree

INTRODUCTION

Machine Learning is an AI technique which is very useful in prediction systems. Machine learning is a type of artificial intelligence that allows a software application to learn from the data & become more accurate in predicting outcomes without human intervention. Loan Prediction is very helpful for employees of banks as well as for the applicant also. It is very difficult to predict the possibility of payment of loan by the customer. Using Machine learning we predict loan approval.

There are some features for the prediction like- Gender, Age, Married Status, Dependents, Education, Applicant Income, Co Applicant Income, Loan Amount, Credit History, Down Payment, Property Area.

PROBLEM DEFINITION

The banking sector deals with various types of loan daily. It exists in Rural, Semi Urban and Urban areas. After applying for a loan by a customer, the bank needs to validate the eligibility of customers to get the loan or not. So the customer will fill an online loan application form. This form consists of details like Gender, Marital Status, qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others. Also to get information about a specific loan and its eligibility criteria customers need to visit the bank's branch. So instead of it, customers can get information about a particular loan and its eligibility criteria in the same app and also can check the status whether he/she is capable of taking a loan or not. The loan approval process is automated, it can save a lot of man hours and improve the speed of service to the customers and provide some additional services.

THEORY

MACHINE LEARNING

Machine learning is a type of artificial intelligence. It allows a software application to learn from the data & become more accurate in predicting outcomes without human intervention.

Supervised Learning

Supervised learning is one of the most basic types of machine learning. Supervised learning is the point at which the model is getting prepared on a labelled dataset.It allows collecting data and produces data output from previous experiences.
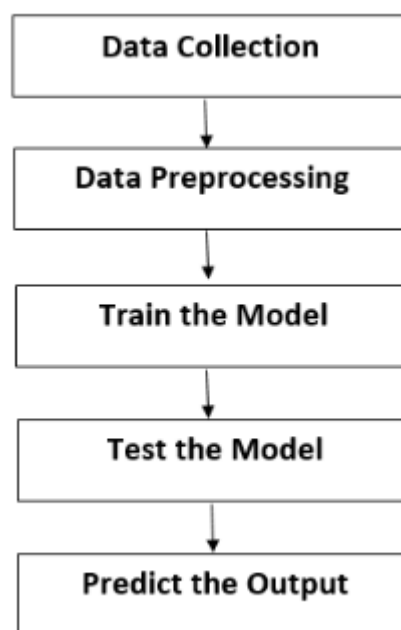
Types of Supervised Learning:

1. Classification
2. Regression

Unsupervised Learning

In unsupervised learning, the input data are not labelled dataset. It used to identify patterns in data sets containing data points that are neither classified nor labelled.

PROPOSED METHODOLOGY

PROPOSED SYSTEM

The proposed system automates the process of determining loan eligibility status. A data set contains the details of the loan applicant.

The data set is classified into two categories:

Train Model - We have used the major dataset to train the model. It consists of the sample output data

and the corresponding sets of input data. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output.

Test Model - We have used the minor dataset to test the model. It consists of only input data and it will predict the output variable.

SYSTEM DESIGN AND IMPLEMENTATION

To implement our project we have used Python language. MySQL is used to store the data and HTML and CSS is used for GUI.

Data description :

Data has been collected from GitHub and some of the attributes which are required to test eligibility are added by us. Dataset is consist of two datasets which are the train dataset and test dataset. Major dataset is used to train the model and minor dataset is used to test the model.

ApplicantIncom	Applicant's Income
CoapplicantInc me	Co Applicant Income
LoanAmount	Loan Amount
Loan_Amount_ em	Terms of loan months
Credit_History	Meets Guidelines Not
Property_Area	Urban/Semi-Urban/Rural
Prev_Loan	Yes/No
Down_Paymen	Down payment
Emplyement_S tus	Applicant Employe duration in months
Age	Applicant Age
Loan_Status	Loan Status(Y/N)

Dataset Attributes

Attribute	Description
Gender	Female/Male
Married	Yes/No
Dependents	Number of dependan
Education	Graduate/Not Graduate
Self_Employed	Yes/No

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	
LP001002	Male	No	0	Graduate	No	5849	0	2000	
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120	
LP001008	Male	No	0	Graduate	No	6000	0	141	
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	

Sample Dataset

	ApplicantIncome	CoapplicantIncome	LoanAmount
count	981.000000	981.000000	955.000000
mean	5179.795107	1601.916330	144.456545
std	5695.104533	2718.772806	97.983082
min	0.000000	0.000000	9.000000
25%	2875.000000	0.000000	100.000000
50%	3800.000000	1110.000000	126.000000
75%	5516.000000	2365.000000	162.000000
max	81000.000000	41667.000000	2000.000000

Analysis of Data

Data processing :

Data preprocessing is a data mining technique which is used to transform the raw data into useful and efficient data.

It is pre-processing of the data to prepare an ideal data set for modelling.

The data can have irrelevant and missing value. This will be handled by data cleaning. It involves handling of the noisy data, missing data, null value, etc. There are various methods to fill missing values.

1) Handling missing values:

We have filled the missing value manually by attribute median value method and for some attributes we have used a random number within the required range using randint() method.

```python
data.Credit_History.fillna(np.random.randint(0,2),inplace=True)

data.LoanAmount.fillna(data.LoanAmount.median(),inplace=True)
```

2) Data transformation :

In data transformation we have performed data mapping for some of the attributes to map the values.

```python
data.Education=data.Education.map({'Graduate':1,'Not Graduate':0})

data.Property_Area=data.Property_Area.map({'Urban':2,'Rural':0,'Semiurban':1}
```

3) Fitting data into model:

To train a model we need to fit data into the model. In our project we have used a supervised learning model. We didn't consider all the attributes otherwise it might lead to the problem of overfitting. Model will fail to give correct results for real time cases if we take more numbers of attributes as it will result in the model learning specific to our dataset only.

```python
LR=LogisticRegression()

LR.fit(train_X,train_y)
```

LOGISTIC REGRESSION

Logistic regression is a Machine Learning algorithm. It comes under the Supervised Learning technique. Logistic regression predicts the output of a categorical dependent variable. Logistic regression is used for solving classification problems.

Logistic Regression Types -

Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable.

Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables.

In our application, we used a supervised learning approach. In that, we used Binomial Logistic Regression.

**Logistic Function (Sigmoid Function):**
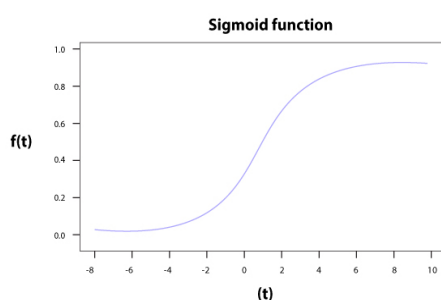
The sigmoid function is a mathematical function used to map the predicted values to probabilities.

In order to map predicted values to probabilities, we use the Sigmoid function.

It maps any value within a range of 0 and 1.

The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form.
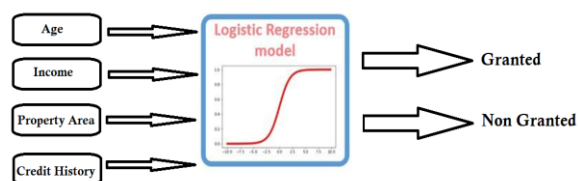
FEATURES OF OUR SOLUTION

This application predicts how much loan users can take.

This application provides various sections for different types of Loan like Home Loan, Car Loan, Personal Loan, etc.

There is one section which suggests which period of time is suitable to take the Loan.

Rate of Interest varies from bank to bank, so it helps users to choose feasible bank loan.

If a user searches for a particular loan many times then there is one option which is 'notify me' in which the user will get a notification about the loan.

CONCLUSION

The IT industry needs to update old technology with new technology, having an efficient model with additional services. Our application is used by the user to check his/her loan application status as well as will get some information about the particular loan. From the analysis of the data, it is very clear that it reduces all the frauds done at the time of loan approval.

REFERENCES

1. Gupta, Anshika, et al. "Bank Loan Prediction System using Machine Learning." 2020 9th International Conference System Modelling and Advancement in Research Trends (SMART). IEEE, 2020.

2. Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, 2020.

3. Kumar, R., et al. "Prediction of loan approval using machine learning." Int J Adv Sci Technol 28 (2019): 455-460.

4. Ndayisenga, Theoneste. Bank Loan Approval Prediction Using Machine Learning Techniques. Diss. 2021.

5. Arun, Kumar, Garg Ishan, and Kaur Sanmeet. "Loan approval prediction based on machine learning approach." IOSR J. Comput. Eng 18.3 (2016): 18-21

6. Sreesouthry, S., et al. "Loan Prediction Using Logistic Regression in Machine Learning." Annals of the Romanian Society for Cell Biology (2021): 2790-2794.

7. Vaidya, Ashlesha. "Predictive and probabilistic approach using logistic regression: application to prediction of loan approval." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.