

new applicant granted the loan or not using machine learning models trained on the historical data set.

CONTENT		
Chapter	List of content	Page no.
I	Abstract	
1	Introduction	
2	Literature Survey	
3	Existing System	
3.1	Disadvantages of Existing System	
3.2	Proposed System	
3.3	Advantages of Proposed System	
3.4	System Architecture	
3.5	System Requirements	
3.6	Hardware Requirements	
3.7	Software Requirements	
4	Software Environment	
	Python	
4.1	History of Python	
4.2	Python Features	
4.3	Interactive Mode Programming	
4.4	Script Mode Programming	
4.5	Flask Frame Work	
5	Modules	

	5.1	Module Description	
	5.2	Dataset	
5.3		Data Preparation	
6		Model Selection	
	6.1	Analyze and Prediction	
	6.2	Requirement Analysis	
7		Functional Requirements	
	7.1	Non Functional Requirements	
	8	System Design and Testing Plan	
	8.1	System Design and Output Design	
	8.2	System Study	
	8.3	Feasibility Study	
	8.4	Economical Feasability	
	8.5	Social Feasability	
	8.6	Data Flow Diagram	
	9	UML Diagrams	
10		System testing	
	10.1	Unit Testing	
	10.2	Integration testing	
	10.3	Functional testing	
	10.4	White Box Testing	
	10.5	Black Box testing	
	10.6	Acceptance testing	
11		Conclusion	
12		References	

1 INTRODUCTION

As the data are increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Artificial intelligence (AI), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilizing AI calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high. Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications. One mistake can make a massive loss to a bank. Loan distribution is the primary business of almost every bank. This project aims to provide a loan to a deserving applicant out of all applicants. An efficient and non-biased system that reduces the bank's time employs checking every applicant on a priority basis. The bank authorities complete all other customer's other formalities on time, which positively impacts the customers. The best part is that it is efficient for both banks and applicants. This system allows jumping on particular applications that deserve to be approved on a priority basis. There are some features for the prediction like- 'Gender', 'Married', 'Dependents', 'Education', 'Self_Employed', 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Loan_Amount_Term', 'Credit_History', 'Property_Area', 'Loan_Status'.

2 LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

Loan Approval Prediction based on Machine Learning Approach

The main objective of this paper is to predict whether assigning the loan to particular person will be safe or not. 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.

Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process

Extending credits to corporates and individuals for the smooth functioning of growing economies like India is inevitable. As increasing number of customers apply for loans in the banks and non-banking financial companies (NBFC), it is really challenging for banks and NBFCs with limited capital to devise a standard resolution and safe procedure to lend money to its borrowers for their financial needs. In addition, in recent times NBFC inventories have suffered a significant downfall in terms of the stock price. It has contributed to a contagion that has also spread to other financial stocks, adversely affecting the benchmark in recent times. In this paper, an attempt is made to condense the risk involved in selecting the suitable person who could repay the loan on time thereby keeping the bank's nonperforming assets (NPA) on the hold. This is achieved by feeding the past records of the customer who acquired loans from the bank into a trained machine learning model which could yield an accurate result. The prime focus of the paper is to determine whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting.

Loan Prediction using machine learning model

whether or not it will be safe to allocate the loan to a particular person. This paper has the following sections (i) Collection of Data, (ii) Data Cleaning and (iii) Performance Evaluation. Experimental tests found that the Naïve Bayes model has better performance than other models in terms of loan forecasting. 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. This is done 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 give the most accurate result. The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. 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. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logic regression, Decision Tree and gradient boosting.

Loan Prediction using Decision Tree and Random Forest

In India the number of people or organization applying for loan gets increased every year. The bank have to put in a lot of work to analyse or predict whether the customer can pay back the loan amount or not (defaulter or non-defaulter) in the given time. The aim of this paper is to find the nature or background or credibility of client that is applying for the loan. We use exploratory data analysis technique to deal with problem of approving or rejecting the loan request or in short loan prediction. The main focus of this paper is to determine whether the loan given to a particular person or an organization shall be approved or not.

Prediction of Loan Risk using NB and Support Vector Machine

In recent years, banks and creditors face lots of challenges associated with the bank loan. In addition, banking transactions are growing rapidly and large data

sets that represent customer behaviour and risks around debt burdens have increased. So, it became mandatory to know the risks related to bank loans. Each consumer knows what they want and what kind of business they would like to choose. Customer can simply find another bank if administrators are not ready to pay attention. Many data analysis technique exists that focus on loan risk. These techniques are oriented toward extraction of quantity and statistical data properties. They facilitate useful data explanations and help to gain better insights in the process behind data. While there can be traditional data analysis techniques lead to intuition knowledge, it is still made by human researchers. Banks and many investment companies are pioneers using data mining. The proposed work in data mining focused on using data from banking sector to predict the status of loans. Two classification algorithms Naïve Bayes and Support Vector Machine are discussed here. The results have been proven that the speed and accuracy has been improved when combined these two techniques.

3 EXISITNG SYSTEM

- Y. Shi and P. Song proposed a method for evaluating project loans using risk analysis. The method evaluate the risk involved in loans of commercial banks.

- R. Zhang and D. Li used machine learning approaches in prediction systems. The machine learning approach was used for assessment of water quality. The paper concluded that machine learning is a very unimportant tool in prediction systems.
- C. Frank et al. used machine learning in prediction of smoking status. Different machine learning approaches were applied and investigated for finding the smoking status. From the results it was ensured that the logistic algorithm performs better.
- R. Lopes et al. applied machine learning approach for the prediction of credit recovery. Credit recovery is a very important issue for banking system. The prediction of credit recovery is a challenging task. Different machine learning approaches were applied to predict the credit recovery and gradient boosting algorithms (GBM) outperformed the other machine learning approaches.

3.1 DISADVANTAGES OF EXISTING SYSTEM

- Existing systems are often relatively inaccurate.
- Calculations can get very complex, particularly if many values are uncertain and/or if many outcomes are linked.

- Existing system often involves higher time to train the model.
- Existing system is relatively expensive as the complexity and time has taken are more.
- Existing system generally leads to the overfitting of data.
- Little bit of noise can make it unstable which leads to wrong predictions.

3.2 PROPOSED SYSTEM

- This proposed model will characterize the behavior of customers on the Basis of their record. These records are taken from the customers, and

create a data set. With the help of these data sets and training machine learning model, we predict that the customer's loan will pass or not.

- The aim of this Paper is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight. A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not.
- Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Paper is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing.

3.3 ADVANTAGES OF PROPOSED SYSTEM

- The advantage of this system is that we provided some conditions by setting the algorithms and just by evaluating the details, we get to know eligibility criteria that client is eligible or not.
- The proposed system also scales relatively well to high dimensional data.
- The proposed system is relatively memory efficient.
- The risk of over-fitting is less in our proposed system.
- A small change to the data does not greatly affect the hyperplane

3.4 SYSTEM ARCHITECTURE

3.6 HARDWARE REQUIREMENTS:

- System : Pentium Dual Core.
- Hard Disk : 120 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 4 GB.

3.7 SOFTWARE REQUIREMENTS:

- Operating system: Windows 7/10.
- Coding Language :Python

4 SOFTWARE ENVIRONMENT

Python: