Prediction of Modernized Loan Approval System Based on Machine Learning Approach

SYNOPSIS

At

Guru Tegh Bahadur Institute of Technology

(Deemed To be University)

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering Degree in Computer Science and Engineering

BY

Pradhumn Gupta (09113202720)

Shekhar Gupta (10513202720)

Prabhneet Singh (08813202720)

Prabhjot Singh (0101327221)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERINGSCHOOL OF COMPUTING
Guru Tegh Bahadur Institute of Technology

Guru Gobind Singh Indraprastha University
Dwarka, New Delhi
Year 2020-2024

CHAPTER 1

INTRODUCTION

Aim: To determine the loan approval system using machine learning algorithms.

.

Synopsis:

Loan approval is a very important process for banking organizations. The systems approved or reject the loan applications. Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers worked on loan approval prediction systems. Machine Learning (ML) techniques are very useful in predicting outcomes for large amount of data. In this paper different machine learning algorithms are applied to predict the loan approval of customers..In this paper, various machine learning algorithms that have been used in past are discussed and their accuracy is evaluated. 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.

CHAPTER 2

SYSTEM ANALYSIS

2.1 EXISTING SYSTEM

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. In existing process, they are use RF algorithm in loan approval system. But the efficiency and accuracy was pretty low. Already banks are provide online transaction system, online bank account opening system, etc,. But there is no loan approval system in the banking sector. Then now we create a new system for loan approval. So now we move on to the proposed system.

DisAdvantage

*****To apply the loan we need to go to bank to apply it

2.2PROPOSED SYSTEM

The proposed model focuses on predicting the credibility of customers for loan repayment by analyzing their details. The input to the model is the customer details collected. On the output from the classifier, decision on whether to approve or reject the customer request can be made. Using different data analytics tools loan prediction and there severity can be forecasted. In this process it is required to train the data using different algorithms and then compare user data with trained data to predict the nature of loan. The training data set is now supplied to machine learning model; on the basis of this data set the model is trained. Every new applicant details filled at the time of application form acts as a test data set. After the operation of testing,

model predict whether the new applicant is a fit case for approval of the loan or not based upon the inference it conclude on the basis of the training data sets. By providing real time input on the web app. In our project, Logistic Regression gives high accuracy level compared with other algorithms. Finally, we are predicting the result via data visualization and display the predicted output using web app using flask.

Advantage

*No need to go to bank We can do the transaction from house,

***** we can consume the time doing from home

CHAPTER 3

REQUIREMENT SPECIFICATIONS

3.1 INTRODUCTION

Prediction of modernized loan approval system based on machine learning approach is a loan approval system from where we can know whether the loan will pass or not. In this system, we take some data from the user like his monthly income, marriage status, loan amount, loan duration, etc. Then the bank will decide according to its parameters whether the client will get the loan or not. So there is a classification system, in this system, a training set is employed to make the model and the classifier may classify the data items into their appropriate class. A test dataset is created that trains the data and gives the appropriate result that, is the client potential and can repay the loan. Prediction of a modernized loan approval system is incredibly helpful forbanks and also the clients. This system checks the candidate on his priority basis. Customer can submit his application directly to the bank so the bank will do the whole process, no third partyor stockholder will interfere in it. And finally, the bank will decide that the candidate is deserving or not on its priority basis. The only object of this research paper is that the deserving candidate gets straight forward and quick results.

3.2 HARDWARE AND SOFTWARE SPECIFICATION

3.2.1 HARDWARE REQUIREMENTS

- Hard disk : 500 GB and above.
- Processor : i3 and above.
- Ram : 4GB and above.

3.2.2 SOFTWARE REQUIREMENTS

- Operating System : Windows 10
- Software : python
- Tools : Anaconda (Jupiter Note Book IDE)

CHAPTER 4

4.1 Design and Implementation Constraints

4.5.1 Constraints in Analysis

- ♦ Constraints as Informal Text
- ♦ Constraints as Operational Restrictions
- ♦ Constraints Integrated in Existing Model Concepts
- ♦ Constraints as a Separate Concept
- ♦ Constraints Implied by the Model Structure

4.1.2 Constraints in Design

- ♦ Determination of the Involved Classes
- ♦ Determination of the Involved Objects
- ♦ Determination of the Involved Actions
- ♦ Determination of the Require Clauses
- ♦ Global actions and Constraint Realization

4.1.3 Constraints in Implementation

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds withthe relation concept of entity-relationship modeling and many object oriented methods.

4.2 Other Nonfunctional Requirements

4.2.1 Performance Requirements

The application at this side controls and communicates with the following three main general components.

- embedded browser in charge of the navigation and accessing to the web service;
- Server Tier: The server side contains the main parts of the functionality of the proposed architecture. The components at this tier are the following.

Web Server, Security Module, Server-Side Capturing Engine, Preprocessing Engine, Database System, Verification Engine, Output Module.

4.2.2 Safety Requirements

1. The software may be safety-critical. If so, there are issues associated with its integrity level
2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
3. If a system must be of a high integrity level and if the software is shown to be of that

integrity level, then the hardware must be at least of the same integrity level.

4. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.

5. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.

6. Systems with different requirements for safety levels must be separated.

7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

CHAPTER 5

5.1 Architecture Diagram:

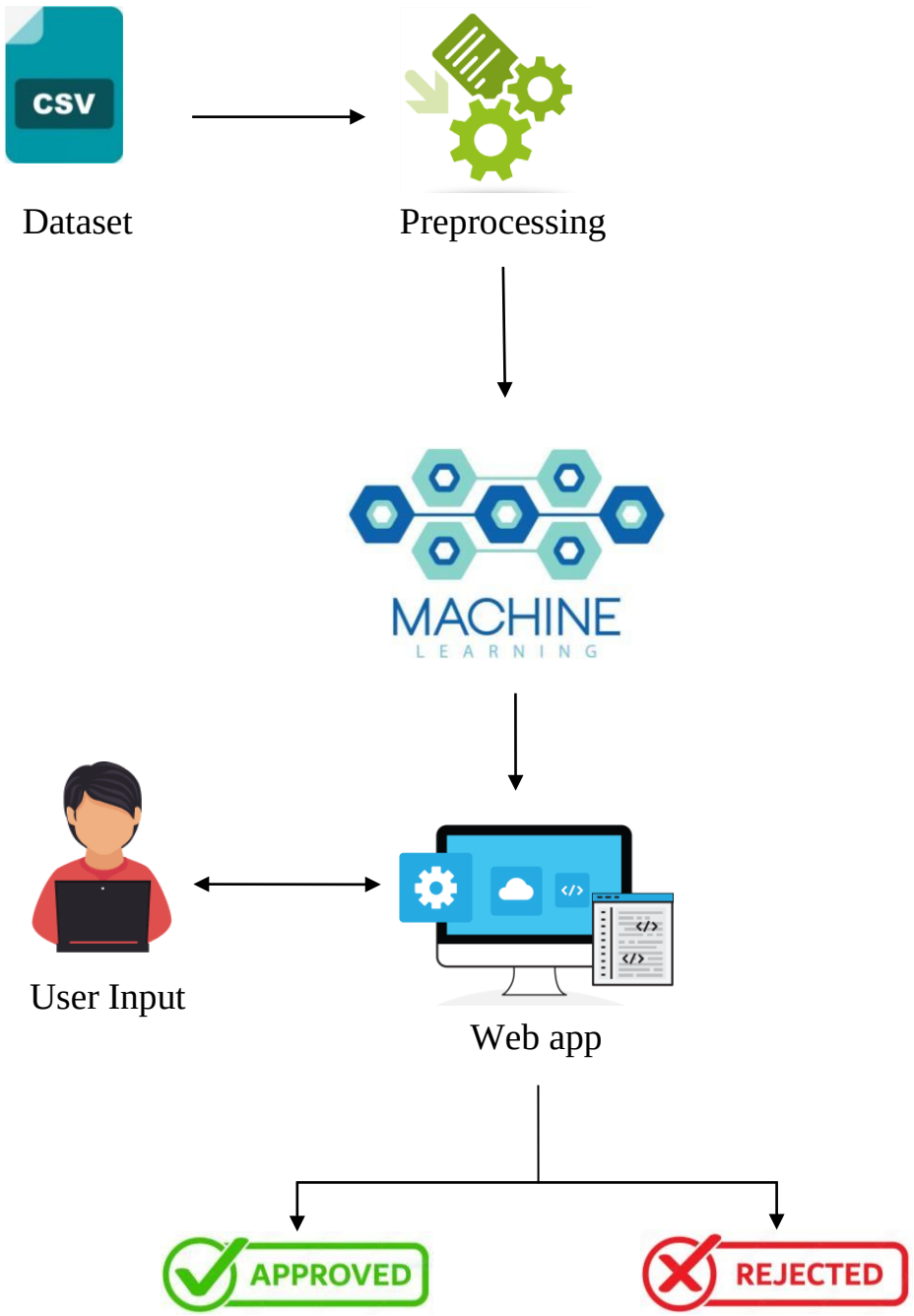
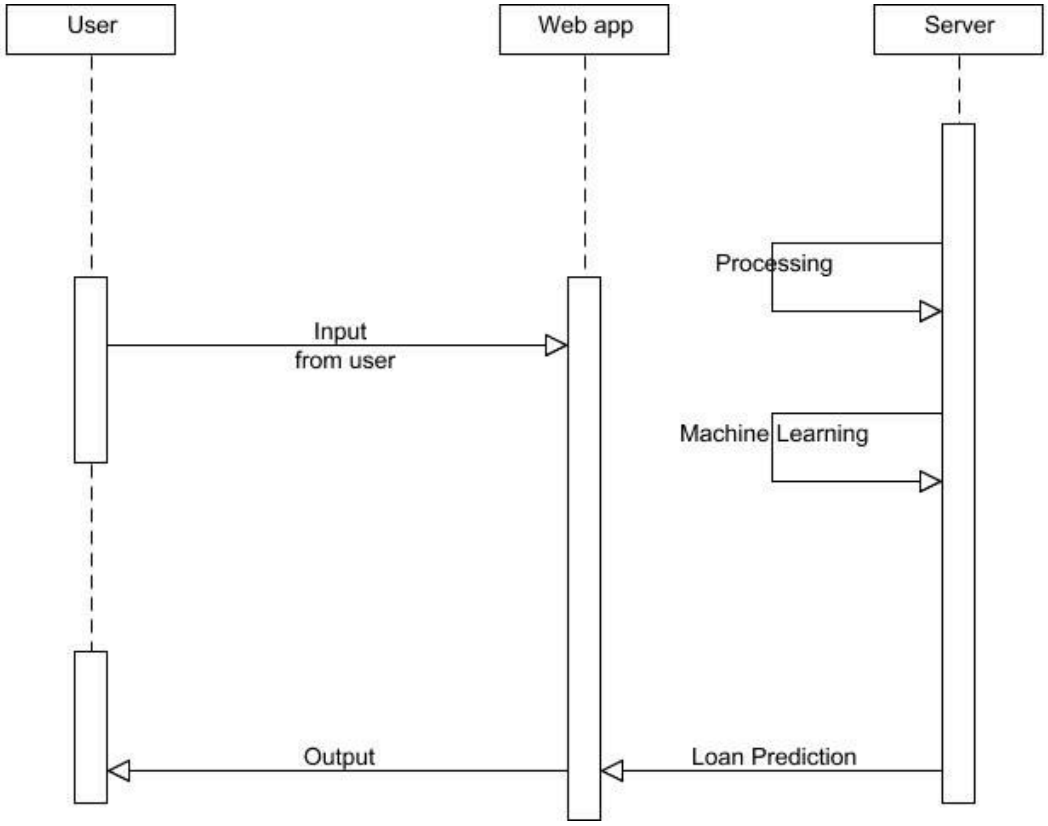


Fig: 5.1

5.2 Sequence Diagram:

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of Message Sequence diagrams are sometimes called event diagrams, event sceneries and timing diagram.



5.3 Use Case Diagram:

Unified Modeling Language (UML) is a standardized general-purpose modeling language in the field of software engineering. The standard is managed and was created by the Object Management Group. UML includes a set of graphic notation techniques to create visual models of software intensive systems. This language is used to specify, visualize, modify,construct and document the artifacts of an object oriented software intensive system under development.

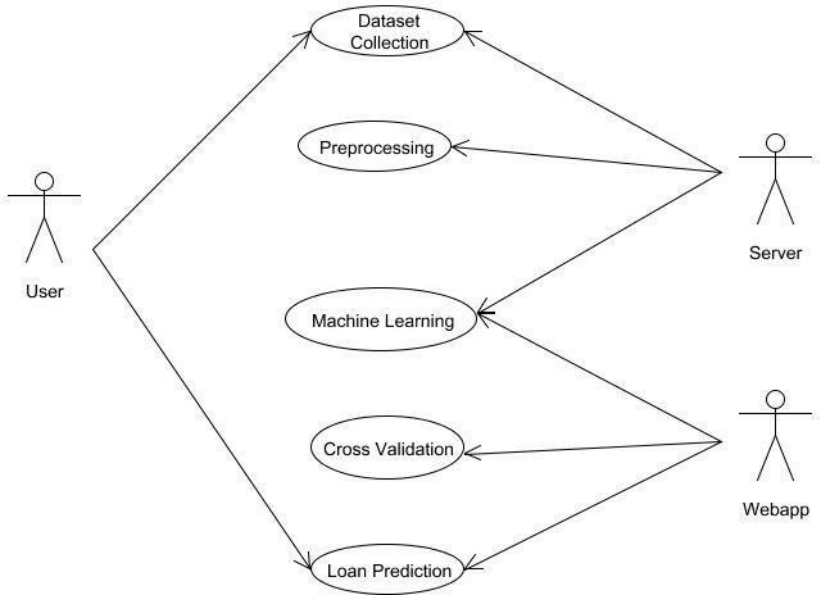
5.3.1. USECASE DIAGRAM

A Use case Diagram is used to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those use cases.

Use case diagram consists of two parts:

Use case: A use case describes a sequence of actions that provided something of measurable value to an actor and is drawn as a horizontal ellipse.

Actor: An actor is a person, organization or external system that plays a role in one or more interaction with the system.

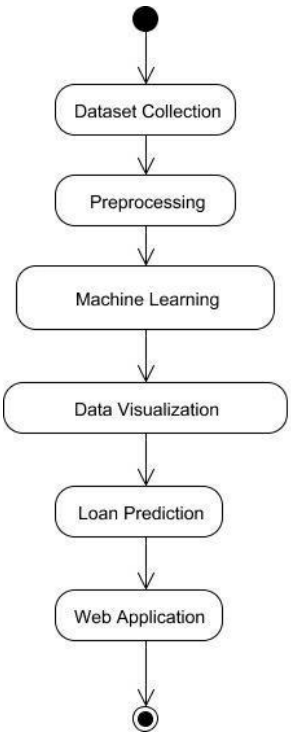


5.4 Activity Diagram:

Activity diagram is a graphical representation of workflows of stepwise activities and actions with support for choice, iteration and concurrency. An activity diagram shows the overall flow of control.

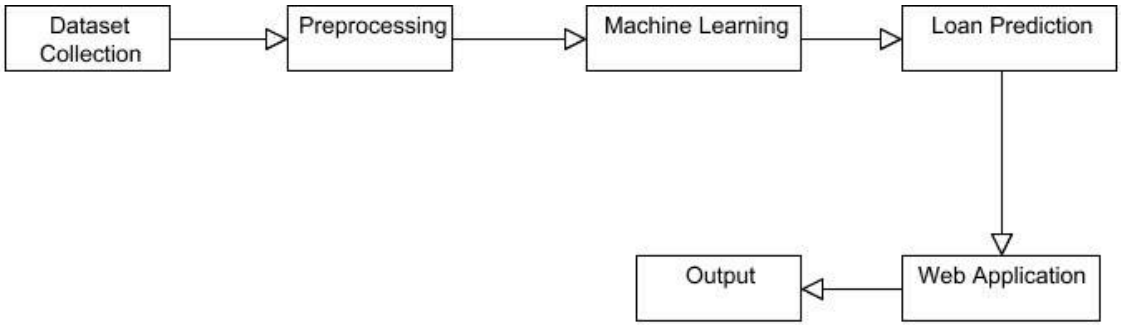
The most important shape types:

- Rounded rectangles represent activities.
- Diamonds represent decisions.
- Bars represent the start or end of concurrent activities.
- A black circle represents the start of the workflow.
- An encircled circle represents the end of the workflow.



5.5 Collaboration Diagram:

UML Collaboration Diagrams illustrate the relationship and interaction between software objects. They require use cases, system operation contracts and domain model toalready exist. The collaboration diagram illustrates messages being sent between classes and objects.



CHAPTER 6

6.1 MODULES

- **Dataset collection**
- **Machine Learning Algorithm**
- **Prediction**

6.2 MODULE EXPLANATION:

6.2.1 Dataset collection:

Dataset is collected from the kaggle.com. That dataset have some value like gender, marital status, self-employed or not, monthly income, etc,. Dataset has the information, whether the previous loan is approved or not depends up on the customer information. That data well be preprocessed and proceed to the next step.

Machine learning Algorithm:

In this stage, the collected data will be given to the machine algorithm for training process. We use multiple algorithms to get high accuracy range of prediction. A preprocessed dataset are processed in different machine learning algorithms. Each algorithm gives someaccuracy level. Each one is undergoes for the comparison.

- ✔ **Logistic Regression**
- ✔ **K-Nearest Neighbors**
- ✔ **Decision Tree Classifier**

Prediction:

Preprocessed data are trained and input given by the user goes to the trained dataset. The Logistic Regression trained model is used to predict and determine whether the loan given to a particular person shall be approved or not.

REFERENCES

1] Amruta S. Aphale and R. Prof. Dr. Sandeep. R Shinde, "Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval", International Journal of Engineering Trends and Applications (IJETA), vol. 9, issue 8, 2020)

[2] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016

[3] Exploratory data analysis https://en.wikipedia.org/wiki/Exploratory_data_analysis

[4] Pandas Library https://pandas.pydata.org/pandas-docs/stable/

[5] MeanDecreaseAccuracy https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.hml