

# **PREDICTION OF MODERNIZED LOAN APPROVAL SYSTEM BASED ON MACHINE LEARNING APPROACH**

Submitted in partial fulfillment of the requirements for the award  
of  
**Bachelor of Engineering degree in Computer Science and Engineering**

*By*

**SUGAN ATHISH K.C (Reg. No. 39110981)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
SCHOOL OF COMPUTING**

## **SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY  
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI - 600119**

**APRIL - 2023**



# **SATHYABAMA**

**INSTITUTE OF SCIENCE AND TECHNOLOGY  
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

[www.sathyabama.ac.in](http://www.sathyabama.ac.in)

---

## **DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

### **BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work of **SUGAN ATHISH K.C(Reg.No - 39110981)** who carried out the Project Phase-2 entitled **"PREDICTION OF MODERNIZED LOAN APPROVAL SYSTEM BASED ON MACHINE LEARNING APPROACH"** under my supervision from January 2023 to April 2023.

**Internal Guide**

**Dr. RAMYA.G.FRANKLIN M.E., Ph.D**

**Head of the Department**

**Dr. L. LAKSHMANAN, M.E., Ph.D.**



**Submitted for Viva voce Examination held on 20.04.2023**

---

**Internal Examiner**

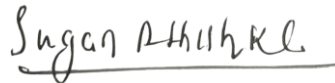
**External Examiner**

## DECLARATION

I, **SUGAN ATHISH K.C (Reg.No- 39110981)**, hereby declare that the Project Phase-2 Report entitled “**PREDICTION OF MODERNIZED LOAN APPROVAL SYSTEM BASED ON MACHINE LEARNING APPROACH**”

done by me under the guidance of **Dr.RAMYA.G.FRANKLIN, M.E.,Ph.D** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

**DATE: 20.04.2023**

A handwritten signature in black ink, reading "Sugan Athish K.C.", written over a horizontal line.

**PLACE: Chennai**

**SIGNATURE OF THE CANDIDATE**

## **ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph. D, Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. RAMYA G.FRANKLIN M.E., Ph.D**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my phase-2 project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**SUGAN ATHISH K.C**

## ABSTARCT

Technology has boosted the existence of human kind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set. 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.

## TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO.
	<b>ABSTRACT</b>	v
	<b>LIST OF FIGURES</b>	viii
1	<b>INTRODUCTION</b>	1
2	<b>LITERATURE SURVEY</b>	2
	2.1 Inferences from Literature Survey	2
	2.2 Open problems in Existing System	3
3	<b>REQUIREMENTS ANALYSIS</b>	4
	3.1 Feasibility Studies/Risk Analysis of the Project	4
	3.2 Software Requirements Specification Document	4
	3.3 System Use case	5
4	<b>DESCRIPTION OF PROPOSED SYSTEM</b>	27
	4.1 Proposed Work	27
	4.2 Methodology	27
	4.3 Architecture Design	28
	4.4 Sequence Diagram	30
	4.5 Collaboration Diagram	32
5	<b>IMPLEMENTATION DETAILS</b>	33
	5.1 Design and implementation constraints	33
	5.2 Other Nonfunctional Requirement	34

<b>CHAPTER NO</b>	<b>TITLE</b>		<b>PAGE NO.</b>
	5.3	MODULES	35
	5.4	CODING	35
	5.5	CODING STANDARD	36
	5.6	TEST PROCEDURE	37
	5.7	TEST DATA AND OUTPUT	38
	5.8	TESTING TECHNIQUES / TESTING STRATERGIES	40
<b>6</b>	<b>RESULT AND DISCUSION</b>		44
<b>7</b>	<b>CONCLUSION</b>		47
	<b>REFERENCES</b>		49
	<b>APPENDIX</b>		50
	<b>A. SOURCE CODE</b>		50
	<b>B. SCREENSHOTS</b>		55

## LIST OF FIGURE

FIGURE NO.	TITLE	PAGE NO.
4.1	Architecture Diagram	28
4.2	Sequence Diagram	29
4.3	Use case Diagram	30
4.4	Activity Diagram	31
4.5	Collabortion Diagram	32
B.1	Home Page	45
B2	Loan Approval System Recommendation Page	46



## LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
API	APPLICATION PROGRAMMING INTERFACE
CNN	CONVOLUTIONAL NEURAL NETWORK
CYP	CROP YEILD PREDICTION
CGI	COMMON GATEWAY INTERFACE
CLI	COMMAND-LINE INTERFACE
CSS	CASCADING STYLE SHEETS
HTML	HYPERTEXT MARKUP LANGUAGE
HTTP	HYPERTEXT TRANSFER PROTOCOL
JS	JAVA SCRIPT
ML	MACHINE LEARNING
MVC	MODEL VIEW CONTROLLER
PHP	HYPERTEXT PREPROCESSOR
SRS	SYSTEM REQUIREMENT SPECIFICATION
URD	USER REQUIREMENT DOCUMENT
UML	UNIFIED MODELLING LANGUAGE

## **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. One can predict whether or not a loan will be approved using a modernised loan approval system that uses machine learning techniques. We collect information about the customer using this way, including information on his monthly wage, marital status, loan amount, loan term, etc. In order to determine if the customer qualifies for the loan, the bank will then employ its criteria. Hence, there is a classification scheme. After building a model with a training set in this system, the class The classifier can provide the data items the right group. A sample dataset is created in order to train the data and deliver the accurate result, which the client's potential and capacity to repay the loan. A modernised loan approval procedure may be favourable for both banks and clients. The bank will be able to oversee the entire procedure when the client submits an application without interference from shareholders or the outside world

## **CHAPTER 2**

### **LITERATURE SURVEY**

J. contribute most to predicting a category or outcome using certain can tech of feature selection. we use the scikit-learn library which provides the SelectKBest class using the chi-squared (chi2) test to select a specific number of feature. we select seven relevant features from 11 different attributes of previous loan. finally the selected best features are fed into classification models for the classifying the outcome.

#### **2.1 INFERENCES FROM LITERATURE SURVEY**

Technology has boosted the existence of humankind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set. signals of the driver to classify the driver's state using logistic regression models.

## 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.

### **DIS Advantage**

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

## 2.2 PROPOSED 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 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 ANALYSIS**

#### **3.1 FEASIBILITY STUDIES**

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 for banks 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 party or 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 SOFTWARE REQUIREMENTS SPECIFICATION DOCUMENT**

##### **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 (Jupyter Note Book IDE)

### **3.3 SYSTEM USE CASE**

- Programming Language: Python.

#### **3.3.1 Introduction to Python**

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently.

It is used for:

- software development,
- mathematics,
- System scripting.

#### **What can Python do?**

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

#### **Why Python?**

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.

- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a functional way.

### **Good to know**

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
- Python 2.0 was released in 2000, and the 2.x versions were the prevalent releases until December 2008. At that time, the development team made the decision to release version 3.0, which contained a few relatively small but significant changes that were not backward compatible with the 2.x versions. Python 2 and 3 are very similar, and some features of Python 3 have been backported to Python 2. But in general, they remain not quite compatible.
- Both Python 2 and 3 have continued to be maintained and developed, with periodic release updates for both. As of this writing, the most recent versions available are 2.7.15 and 3.6.5. However, an official End Of Life date of January 1, 2020 has been established for Python 2, after which time it will no longer be maintained.
- Python is still maintained by a core development team at the Institute, and Guido is still in charge, having been given the title of BDFL (Benevolent Dictator For Life) by the Python community. The name Python, by the way, derives not from the snake, but from the British comedy troupe MontyPython's Flying Circus, of which Guido was, and presumably still is, a fan. It is common to find references to Monty Python sketches and movies scattered throughout the Python documentation.
- It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

## **Python Syntax compared to other programming languages**

- Python was designed for readability, and has some similarities to the English language with influence from mathematics.
- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.
- Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

## **Python is Interpreted**

- Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer's processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly.
- This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.
- One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting.
- In practice, however, for most programs, the difference in execution speed is measured in milliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications.
- For all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and object-oriented programming.
- Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation or GUI programming.



- Python accomplishes what many programming languages don't: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

### **3.3.2 Machine learning**

#### **Introduction:**

Machine learning (ML) is a means the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through learning. In its application across business problems, machine learning is also referred to as predictive analytics.

#### **Machine learning tasks:**

Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

In unsupervised learning, the algorithm builds a mathematical model from a set of data that contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data.

Active learning algorithms access the desired outputs (training labels) for a limited set of inputs based on a budget and optimize the choice of inputs for which it will acquire training labels. When used interactively, these can be presented to a human user for labeling. Reinforcement learning algorithms are given feedback in the form of positive or negative reinforcement in a dynamic environment and are used in autonomous vehicles or in learning to play a game against a human opponent. Other specialized algorithms in machine learning include topic modeling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Meta learning algorithms learn their own inductive bias based on previous experience. In developmental robotics, robot learning algorithms generate their own sequences of learning experiences, also known as a curriculum, to cumulatively acquire new skills through self-guided exploration and social interaction with humans. These robots use guidance mechanisms such as active learning, maturation, motor synergies, and imitation.

### **Types of learning algorithms:**

The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve.

### **Supervised learning:**

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs. The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and the desired output, also as a supervisory signal. In the mathematical model, each training example is represented by an array or vector, sometimes called a feature vector, and the training data is represented by a matrix.

Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs. An optimal function will allow the algorithm to correctly determine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its outputs or predictions over time is said to have learned to perform that task.

Supervised learning algorithms include classification and regression. Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. Similarity learning is an area of supervised machine learning closely related to regression and classification, but the goal is to learn from examples using a similarity function that measures how similar or related two objects are. It has applications in ranking, recommendation systems, visual identity tracking, face verification, and speaker verification.

### **Unsupervised learning:**

Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. A central application of unsupervised learning is in the field of density estimation in statistics, though unsupervised learning encompasses other domains involving summarizing and explaining data features.

Cluster analysis is the assignment of a set of observations into subsets (called *clusters*) so that observations within the same cluster are similar according to one or more pre designated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some *similarity metric* and evaluated, for example, by *internal compactness*, or the similarity between members of the same cluster, and *separation*, the difference between clusters. Other methods are based on *estimated density* and *graph connectivity*.

## Semi-supervised learning:

Semi-supervised learning falls between unsupervised learning (without any labeled training data) and supervised learning (with completely labeled training data). Many machine-learning researchers have found that unlabeled data, when used in conjunction with a small amount of labeled data, can produce a considerable improvement in learning accuracy.

## K-Nearest Neighbors

### Introduction

In four years of the analytics built more than 80% of classification models and just 15-20% regression models. These ratios can be more or less generalized throughout the industry. The reason of a bias towards classification models is that most analytical problem involves making a decision. For instance will a customer attrite or not, should we target customer X for digital campaigns, whether customer has a high potential or not etc. This analysis is more insightful and directly links to an implementation roadmap. In this article, we will talk about another widely used classification technique called K-nearest neighbors (KNN)

### KNN algorithm

KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. To evaluate any technique we generally look at 3 important aspects:

1. Ease to interpret output
2. Calculation time
3. Predictive Power

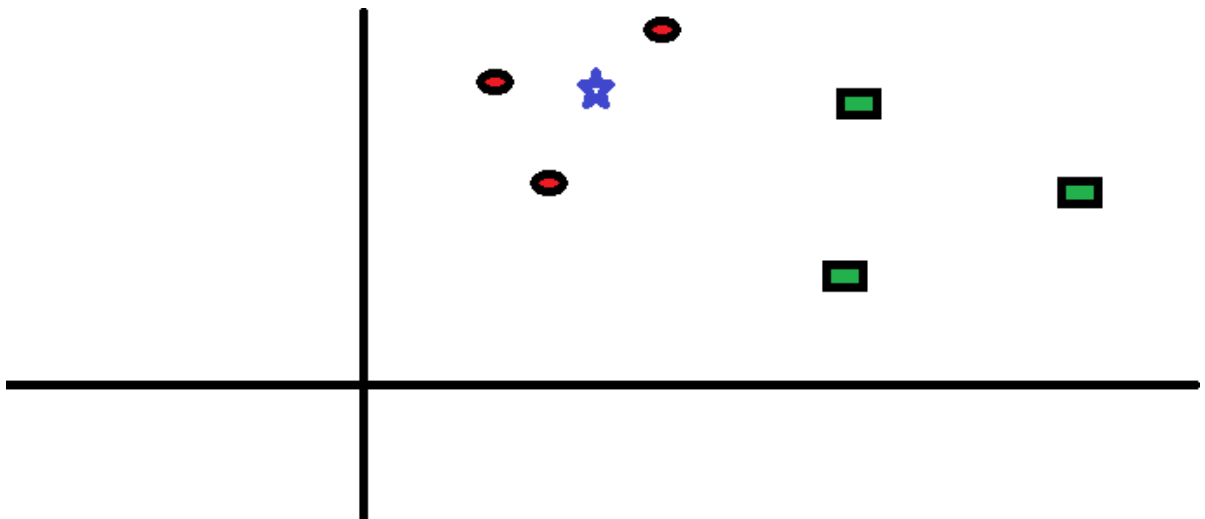
Let us take a few examples to place KNN in the scale:

	Logistic Regression	CART	Random Forest	KNN
1. Ease to interpret output	2	3	1	3
2. Calculation time	3	2	1	3
3. Predictive Power	2	2	3	2

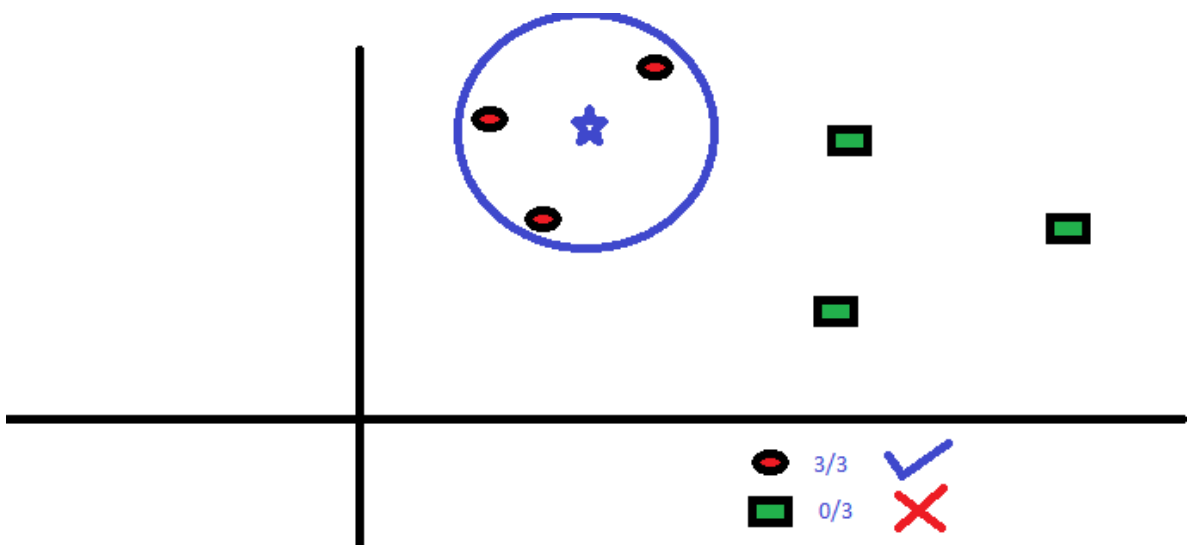
KNN algorithm works across all parameters of considerations. It is commonly used for its easy of interpretation and low calculation time.

### The KNN algorithm work

Let's take a simple case to understand this algorithm. Following is a spread of red circles (RC) and green squares (GS):



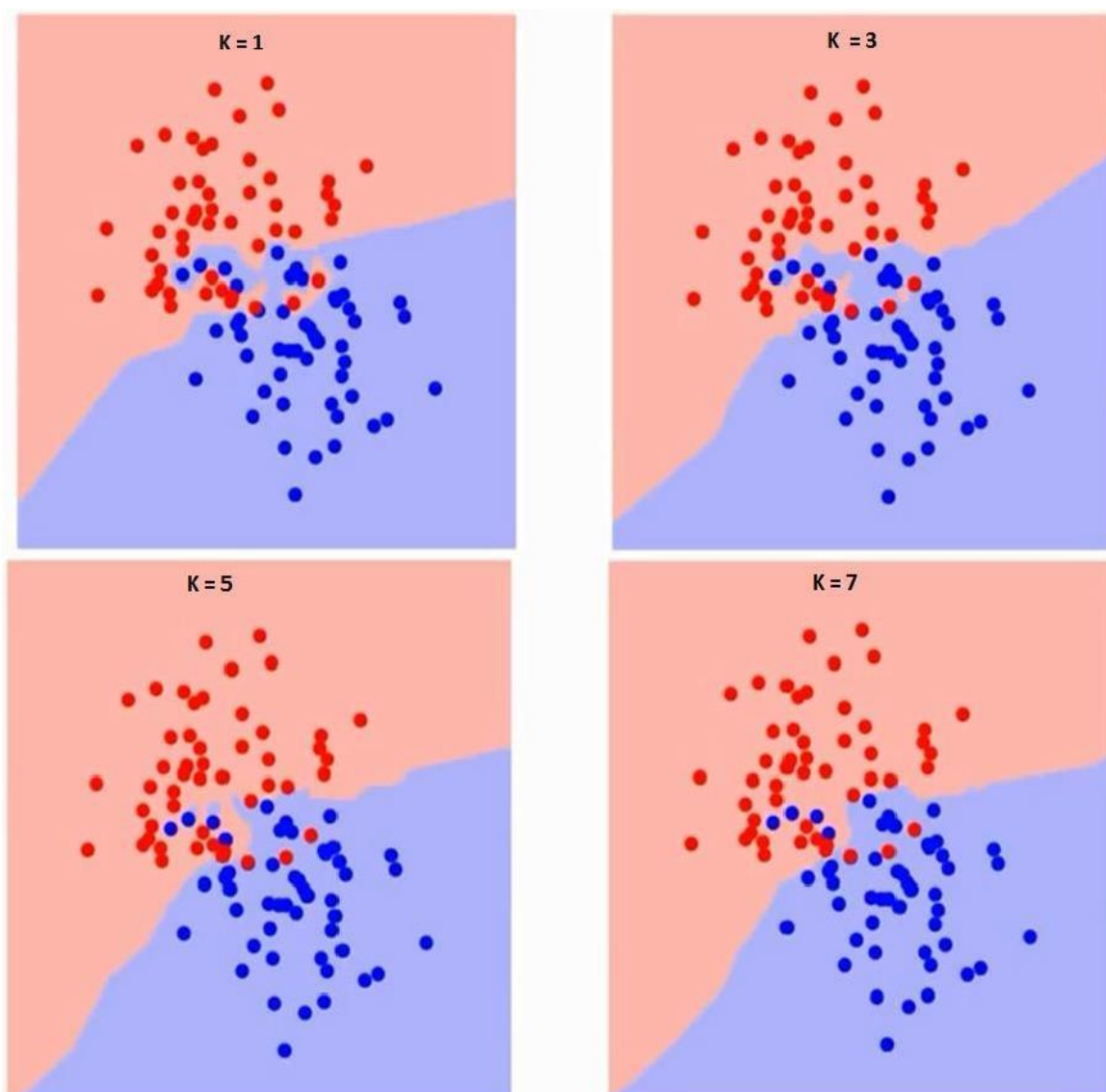
You intend to find out the class of the blue star (BS). BS can either be RC or GS and nothing else. The “K” in KNN algorithm is the nearest neighbors we wish to take vote from. Let's say  $K = 3$ . Hence, we will now make a circle with BS as center just as big as to enclose only three data points on the plane. Refer to following diagram for more details:



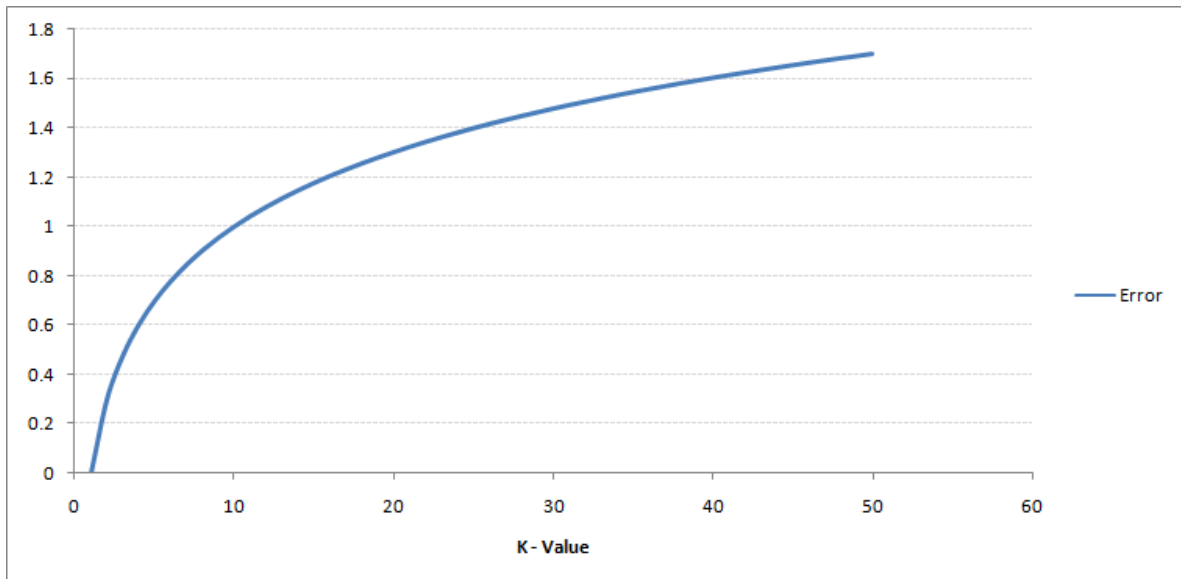
The three closest points to BS is all RC. Hence, with good confidence level we can say that the BS should belong to the class RC. Here, the choice became very obvious as all three votes from the closest neighbor went to RC. The choice of the parameter K is very crucial in this algorithm.

### How do we choose the factor K?

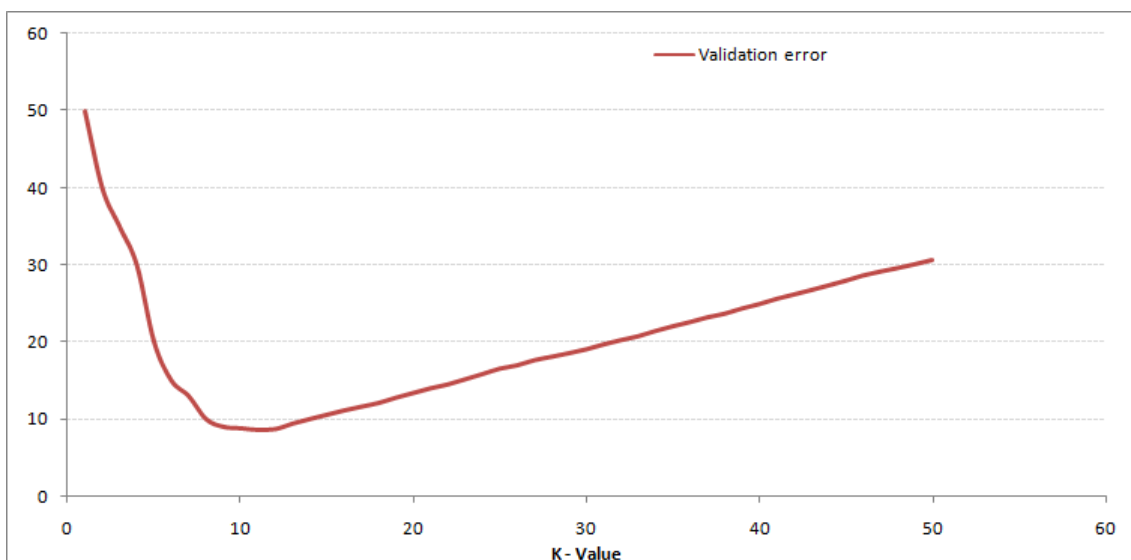
First let us try to understand what exactly does K influence in the algorithm. If we the last example, given that all the 6 training observation remain constant, with a given K value we can make boundaries of each class. These boundaries will the segregate RC from GS. The same way, let's try to see the effect of value "K" onthe class in boundaries. Following are the different boundaries separating the two classes with different values of K.



If you watch carefully, you can see that the boundary becomes smoother with increasing value of K. With K increasing to infinity it finally becomes all blue or all red depending on the total majority. The training error rate and the validation error rate are two parameters we need to access on different K-value. Following is the curve for the training error rate with varying value of K:



As you can see, the error rate at  $K=1$  is always zero for the training sample. This is because the closest point to any training data point is itself. Hence the prediction is always accurate with  $K=1$ . If validation error curve would have been similar, our choice of K would have been 1. Following is the validation error curve with varying value of K.



This makes the story more clear. At  $K=1$ , we were over fitting the boundaries. Hence, error rate initially decreases and reaches a minimal. After the minima point, it then increases with increasing  $K$ . To get the optimal value of  $K$ , you can segregate the training and validation from the initial dataset. Now plot the validation error curve to get the optimal value of  $K$ . This value of  $K$  should be used for all predictions.

### **Breaking it down – Pseudo Code of KNN**

We can implement a KNN model by following the below steps:

1. Load the data
2. Initialize the value of  $k$
3. For getting the predicted class, iterate from 1 to total number of training data points
  1. Calculate the distance between test data and each row of training data. Here we will use Euclidean distance as our distance metric since it's the most popular method. The other metrics that can be used are Chebyshev, cosine, etc.
  2. Sort the calculated distances in ascending order based on distance values
  3. Get top  $k$  rows from the sorted array
  4. Get the most frequent class of these rows
  5. Return the predicted class

### **Conclusion**

KNN algorithm is one of the simplest classification algorithms. Even with such simplicity, it can give highly competitive results. KNN algorithm can also be used for regression problems. The only difference from the discussed methodology will be using averages of nearest neighbors rather than voting from nearest neighbors.

### **Decision tree introduction**

In a decision tree, the algorithm starts with a root node of a tree then compares the value of different attributes and follows the next branch until it reaches the end leaf



node. It uses different algorithms to check about the split and variable that allow the best homogeneous sets of population.

Decision trees are considered to be widely used in [data science](#). It is a key proven tool for making decisions in complex scenarios. In Machine learning, ensemble methods like decision tree, [random forest](#) are widely used. Decision trees are a type of supervised learning algorithm where data will continuously be divided into different categories according to certain parameters.

So in this blog, I will explain the Decision tree algorithm. How is it used? How it functions will be covering everything that is related to the decision tree.

### **What is a Decision Tree?**

Decision tree as the name suggests it is a flow like a tree structure that works on the principle of conditions. It is efficient and has strong algorithms used for predictive analysis. It has mainly attributes that include internal nodes, branches and a terminal node.

Every internal node holds a “test” on an attribute, branches hold the conclusion of the test and every leaf node means the class label. This is the most used algorithm when it comes to [supervised learning](#) techniques.

It is used for both classifications as well as [regression](#). It is often termed as “CART” that means Classification and Regression Tree. Tree algorithms are always preferred due to stability and reliability.

### **How can an algorithm be used to represent a tree**

Let us see an example of a basic decision tree where it is to be decided in what conditions to play cricket and in what conditions not to play. You might have got a fair idea about the conditions on which decision trees work with the above example. Let us now see the common terms used in Decision Tree that is stated below:

**Branches** - Division of the whole tree is called branches.

**Root Node** - Represent the whole sample that is further divided.

**Splitting** - Division of nodes is called splitting.

**Terminal Node** - Node that does not split further is called a terminal node.

**Decision Node** - It is a node that also gets further divided into different sub-nodes being a sub node.

**Pruning** - Removal of subnodes from a decision node.

**Parent and Child Node** - When a node gets divided further then that node is termed as parent node whereas the divided nodes or the sub-nodes are termed as a child node of the parent node.

### **How Does Decision Tree Algorithm Work**

It works on both the type of input & output that is categorical and continuous. In classification problems, the decision tree asks questions, and based on their answers (yes/no) it splits data into further sub branches.

It can also be used as a [binary classification](#) problem like to predict whether a bank customer will churn or not, whether an individual who has requested a loan from the bank will default or not and can even work for multiclass classifications problems. But how does it do these tasks?

In a decision tree, the [algorithm](#) starts with a root node of a tree then compares the value of different attributes and follows the next branch until it reaches the end leaf node. It uses different algorithms to check about the split and variable that allow the best homogeneous sets of population.

**"Decision trees create a tree-like structure by computing the relationship between independent features and a target. This is done by making use of functions that are based on comparison operators on the independent features"**

### **Types of Decision Tree**

**Type of decision tree depends upon the type of input we have that is categorical or numerical :**

1. If the input is a categorical variable like whether the loan contender will defaulter or not, that is either yes/no. This type of decision tree is called a Categorical variable decision tree, also called classification trees.

2. If the input is numeric types and or is continuous in nature like when we have to predict a house price. Then the used decision tree is called a Continuous variable decision tree, also called Regression trees.

### **Lists of Algorithms**

**ID3 (Iterative Dichotomizer3)** – This DT algorithm was developed by Ross Quinlan that uses greedy algorithms to generate multiple branch trees. Trees extend to maximum size before pruning.

C4.5 flourished ID3 by overcoming restrictions of features that are required to be categorical. It effectively defines distinct attributes for numerical features. Using if-then condition it converts the trained trees.

C5.0 uses less space and creates smaller rulesets than C4.5.

The CART classification and regression tree are similar to C4.5 but it braces numerical target variables and does not calculate the rule sets. It generates a binary tree.

### **Why do we use Decision Trees?**

Decision trees provide an effective method of Decision Making because they: **Clearly lay out the problem so that all options can be challenged.** Allow us to analyze fully the possible consequences of a decision. Provide a framework to quantify the values of outcomes and the probabilities of achieving them.

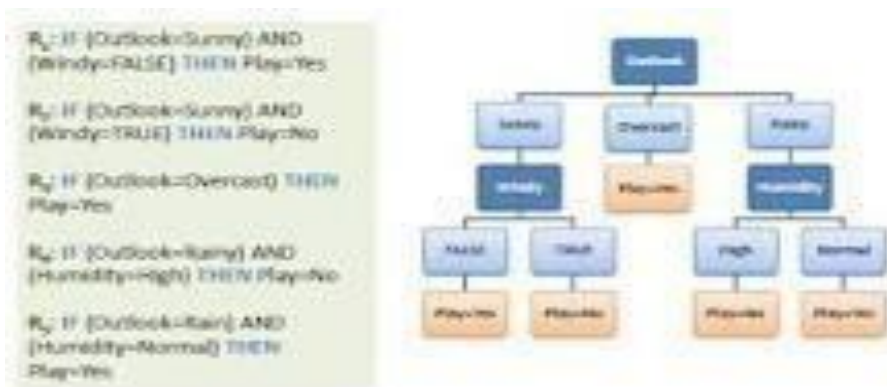
### **What is decision tree in interview explain?**

A Decision Tree is a supervised machine learning algorithm that can be used for both Regression and Classification problem statements. It divides the complete dataset into smaller subsets while at the same time an associated Decision Tree is incrementally developed.

### **Where are Decision Trees used?**

Decision trees are commonly used in **operations research**, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

### **What is a decision tree classification model?**



**Decision Tree – Classification:** Decision tree **builds classification or regression models in the form of a tree structure**. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Decision trees can handle both categorical and numerical data.

### What is the final objective of decision tree?

As the goal of a decision tree is that **it makes the optimal choice at the end of each node it needs an algorithm that is capable of doing just that**. That algorithm is known as Hunt's algorithm, which is both greedy, and recursive.

### Decision tree introduction

In a decision tree, the algorithm starts with a root node of a tree then compares the value of different attributes and follows the next branch until it reaches the end leaf node. It uses different algorithms to check about the split and variable that allow the best homogeneous sets of population. Decision trees are considered to be widely used in [data science](#). It is a key proven tool for making decisions in complex scenarios. In Machine learning, ensemble methods like decision tree, [random forest](#) are widely used. Decision trees are a type of supervised learning algorithm where data will continuously be divided into different categories according to certain parameters. So in this blog, I will explain the Decision tree algorithm. How is it used? How its functions will be covering everything that is related to the decision tree.

### What is a Decision Tree?

Decision tree as the name suggests it is a flow like a tree structure that works on the principle of conditions. It is efficient and has strong algorithms used for

predictive analysis. It has mainly attributes that include internal nodes, branches and a terminal node. Every internal node holds a “test” on an attribute, branches hold the conclusion of the test and every leaf node means the class label. This is the most used algorithm when it comes to [supervised learning](#) techniques. It is used for both classifications as well as [regression](#). It is often termed as “CART” that means Classification and Regression Tree. Tree algorithms are always preferred due to stability and reliability.

### **How can an algorithm be used to represent a tree**

Let us see an example of a basic decision tree where it is to be decided in what conditions to play cricket and in what conditions not to play. You might have got a fair idea about the conditions on which decision trees work with the above example. Let us now see the common terms used in Decision Tree that is stated below:

- **Branches** - Division of the whole tree is called branches.
- **Root Node** - Represent the whole sample that is further divided.
- **Splitting** - Division of nodes is called splitting.
- **Terminal Node** - Node that does not split further is called a terminal node.
- **Decision Node** - It is a node that also gets further divided into different sub-nodes being a sub node.
- **Pruning** - Removal of subnodes from a decision node.
- **Parent and Child Node** - When a node gets divided further then that node is termed as parent node whereas the divided nodes or the sub-nodes are termed as a child node of the parent node.

### **How Does Decision Tree Algorithm Work:**

It works on both the type of input & output that is categorical and continuous. In classification problems, the decision tree asks questions, and based on their answers (yes/no) it splits data into further sub branches.

It can also be used as a [binary classification](#) problem like to predict whether a bank customer will churn or not, whether an individual who has requested a loan from the bank will default or not and can even work for multiclass classifications problems. But how does it do these tasks?

In a decision tree, the [algorithm](#) starts with a root node of a tree then compares the value of different attributes and follows the next branch until it reaches the end leaf node. It uses different algorithms to check about the split and variable that allow the best homogeneous sets of population.

**"Decision trees create a tree-like structure by computing the relationship between independent features and a target. This is done by making use of functions that are based on comparison operators on the independent features"**

### **Types of Decision Tree:**

**Type of decision tree depends upon the type of input we have that is categorical or numerical :**

1. If the input is a categorical variable like whether the loan contender will defaulter or not, that is either yes/no. This type of decision tree is called a Categorical variable decision tree, also called classification trees.
2. If the input is numeric types and or is continuous in nature like when we have to predict a house price. Then the used decision tree is called a Continuous variable decision tree, also called Regression trees.

### *Lists of Algorithms*

**ID3 (Iterative Dichotomizer3)** – This DT algorithm was developed by Ross Quinlan that uses greedy algorithms to generate multiple branch trees. Trees extend to maximum size before pruning.

C4.5 flourished ID3 by overcoming restrictions of features that are required to be categorical. It effectively defines distinct attributes for numerical features. Using if-then condition it converts the trained trees.

C5.0 uses less space and creates smaller rulesets than C4.5.

The CART classification and regression tree are similar to C4.5 but it braces numerical target variables and does not calculate the rule sets. It generates a binary tree.

### Why do we use Decision Trees?

Decision trees provide an effective method of Decision Making because they: **Clearly lay out the problem so that all options can be challenged.** Allow us to analyze fully the possible consequences of a decision. Provide a framework to quantify the values of outcomes and the probabilities of achieving them.

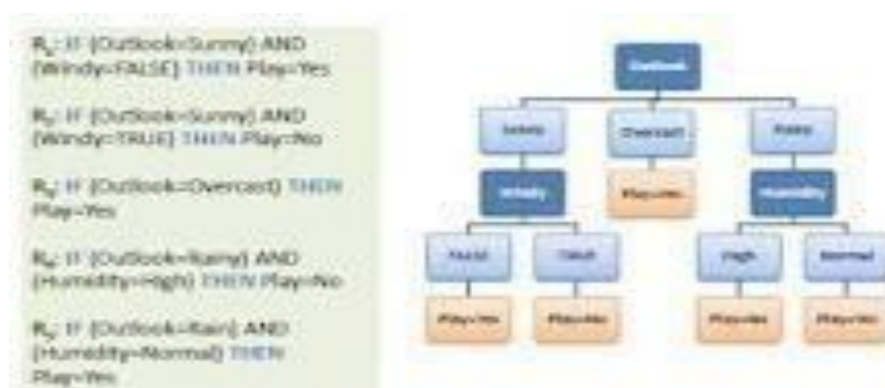
### What is decision tree in interview explain?

A Decision Tree is a supervised machine learning algorithm that can be used for both Regression and Classification problem statements. It divides the complete dataset into smaller subsets while at the same time an associated Decision.

### Where are Decision Trees used?

Decision trees are commonly used in **operations research**, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

### What is a decision tree classification model?



**DecisionTree-Classification:** Decision tree **builds classification or regression models in the form of a tree structure.** It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. . Decision trees can handle both categorical and numerical data.

## What is the final objective of decision tree?

As the goal of a decision tree is that **it makes the optimal choice at the end of each node it needs an algorithm that is capable of doing just that**. That algorithm is known as Hunt's algorithm, which is both greedy, and recursive

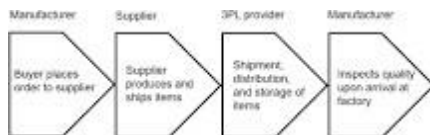
## Introduction to Logistics.:

LOGISTICS IS THE ART AND SCIENCE OF MANAGEMENT, ENGINEERING AND TECHNICAL ACTIVITIES CONCERNED WITH REQUIREMENTS, DESIGN AND SUPPLYING, MAINTAINING RESOURCES TO SUPPORT OBJECTIVES, PLANS AND OPERATION.

## How do you explain logistics?

Logistics refers to the **overall process of managing how resources are acquired, stored, and transported to their final destination**. Logistics management involves identifying prospective distributors and suppliers and determining their effectiveness and accessibility.

## What are the 3 types of logistics?



Logistics has three types; **inbound, outbound, and reverse logistics**.

## What are the 7 R's of logistics?

So, what are the 7 Rs? The Chartered Institute of Logistics & Transport UK (2019) defines them as: **Getting the Right product, in the Right quantity, in the Right condition, at the Right place, at the Right time, to the Right customer, at the Right price**.

## What are the importance of logistics?

Logistics is an important element of a successful supply chain that **helps increase the sales and profits of businesses that deal with the production, shipment, warehousing and delivery of products**. Moreover, a reliable logistics service can boost a business' value and help in maintaining a positive public image.



## **What is logistics in real life?**

Logistics is **the strategic vision of how you will create and deliver your product or service to your end customer**. If you take the city, town or village that you live in, you can see a very clear example of what the logistical strategy was when they were designing it.

## **What are the 3 main activities of logistics systems?**



## **Logistics activities or Functions of Logistics**

Order processing. The Logistics activities start from the order processing which might be the work of the commercial department in an organization. ...

- Materials handling. ...
- Warehousing. ...
- Inventory control. ...
- Transportation. ...
- Packaging.

## **What is 3PL and 4PL in logistics?**

A 3PL (third-party logistics) provider manages all aspects of fulfillment, from warehousing to shipping. A 4PL (fourth-party logistics) provider manages a 3PL on behalf of the customer and other aspects of the supply chain.

## **What are the five major components of logistics?**

**There are five elements of logistics:**

- Storage, warehousing and materials handling.

- Packaging and unitisation.
- Inventory.
- Transport.
- Information and control.

### **What is logistic cycle?**

Logistics management cycle includes key activities such as **product selection, quantification and procurement, inventory management, storage, and distribution**. Other activities that help drive the logistics cycle and are also at the heart of logistics are organisation and staffing, budget, supervision, and evaluation.

### **Why did you choose logistics?**

We choose logistics because **it is one of the most important career sectors in the globe and be more excited about it**. ... I prefer my profession to work in logistics and it can be a challenging field, and with working in it I want to make up an important level of satisfaction in their jobs.

### **What is logistics and SCM?**



The basic difference between Logistics and Supply Chain Management is that Logistics management is the process of integration and maintenance (flow and storage) of goods in an organization whereas Supply Chain Management is the coordination and management (movement) of supply chains of an organization

**Here are 6 steps logistics companies should follow to develop a sound logistics marketing plan.**

1. Define your service offer. ...

2. Determine your primary and secondary markets. ...
3. Identify your competition. ...
4. Articulate your value proposition. ...
5. Allocate a marketing budget. ...
6. Develop a tactical marketing plan

## CHAPTER 4

### DESCRIPTION OF PROPOSED SYSTEM

#### 4.1 PROPOSED 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 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.

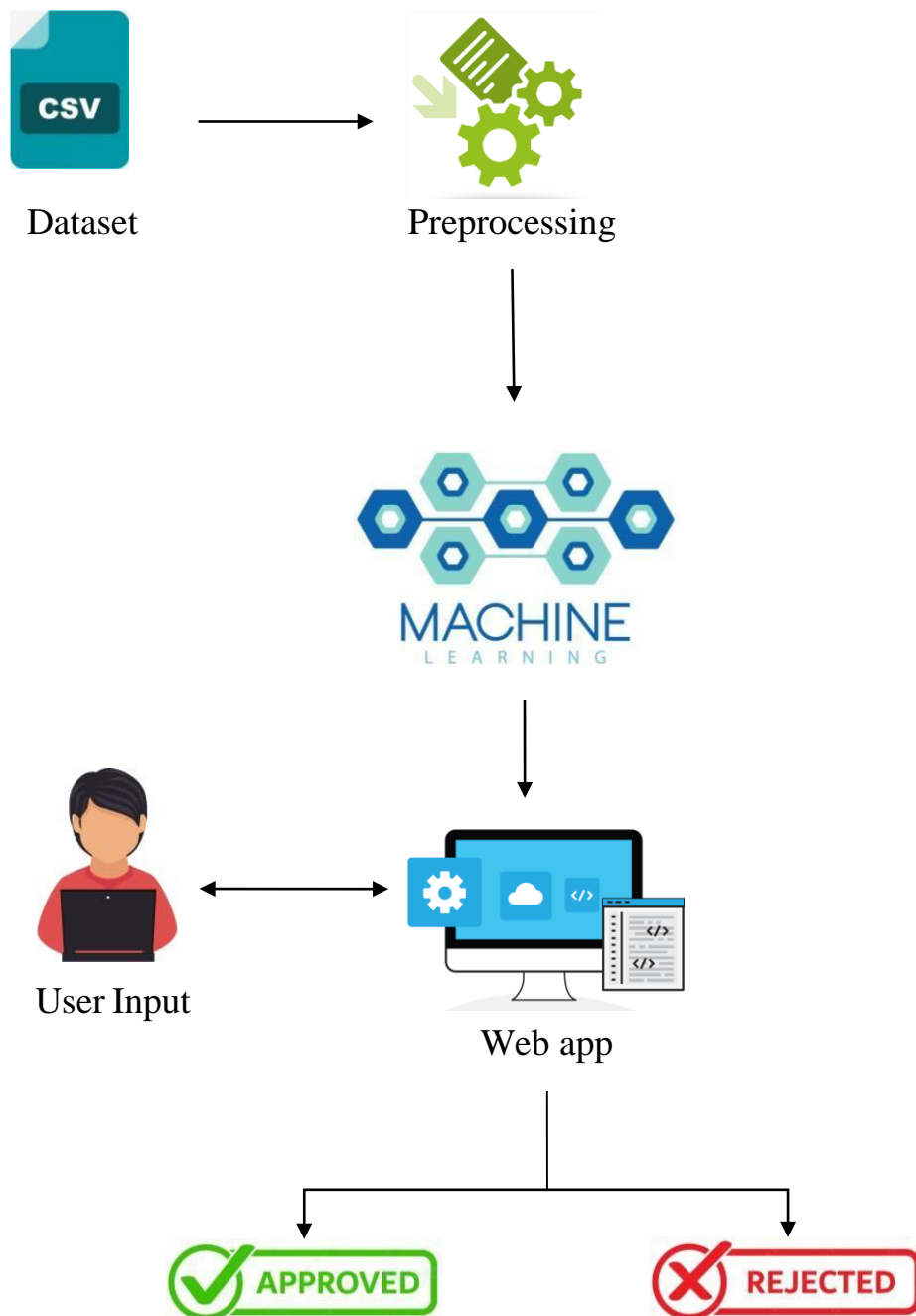
##### **Advantage**

- \*No need to go to bank We can do the transaction from house,
- \* we can consume the time doing from home.

#### 4.2 METHODOLOGY

The purpose of this chapter is to describe the methodology proposed for Estimating crop area and crop yield in the context of mixed and continuous cropping. Both list and area frames are considered. The sample selected for area estimation is to be used as the sampling Frame for selecting the sample for crop yield estimation. the sampling Design used to select the sample for crop area and yield estimation is explored briefly. The theory of domain estimation allows for the separate estimation, From a single sample of crop area and yield of different crops and of their mixtures. In addition, the criterion for determining sample size is also provided.

### 4.3 Architecture Diagram:

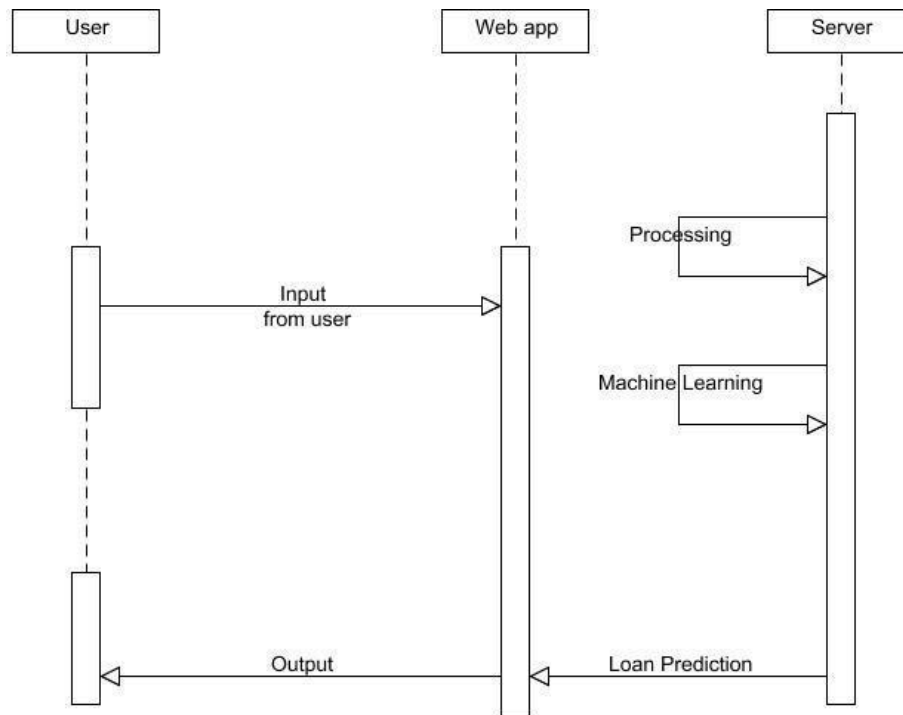


**Fig: 4.1 Architecture Diagram:**

### 4.2 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.



**FIG.4.2 ACTIVITY DIAGRAM OF THE SYSTEM**

### 4.3 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.

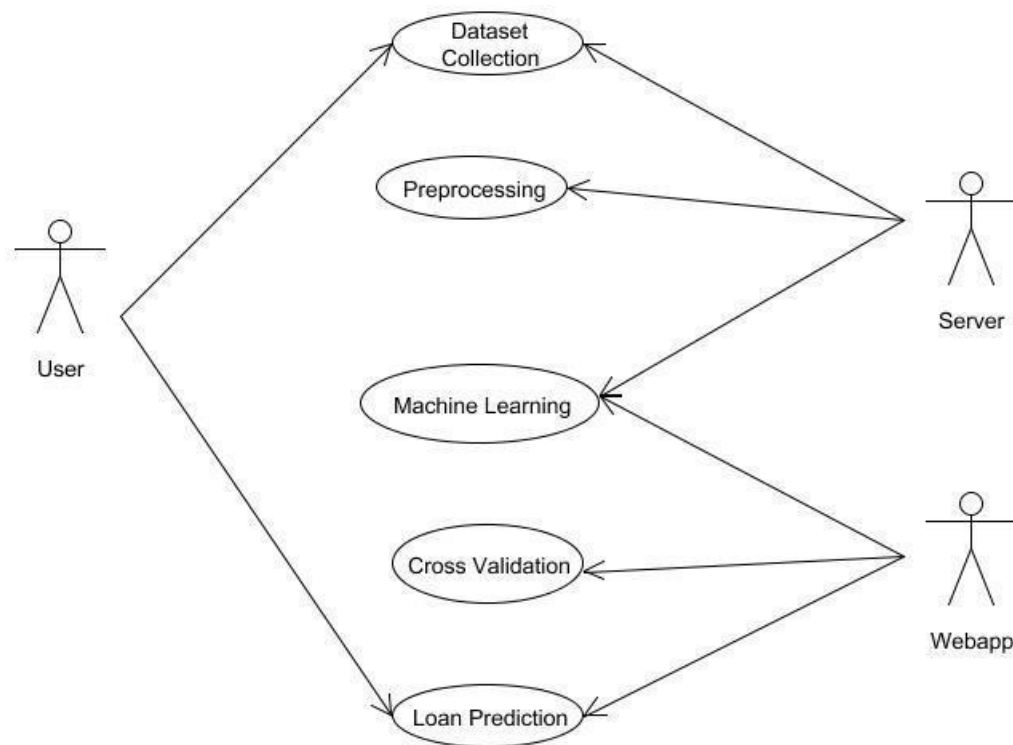
#### 4.3.1. USECASE DIAGRAM

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.



**FIG.4.3.1 UML USE CASE DIAGRAM OF THE SYSTEM**

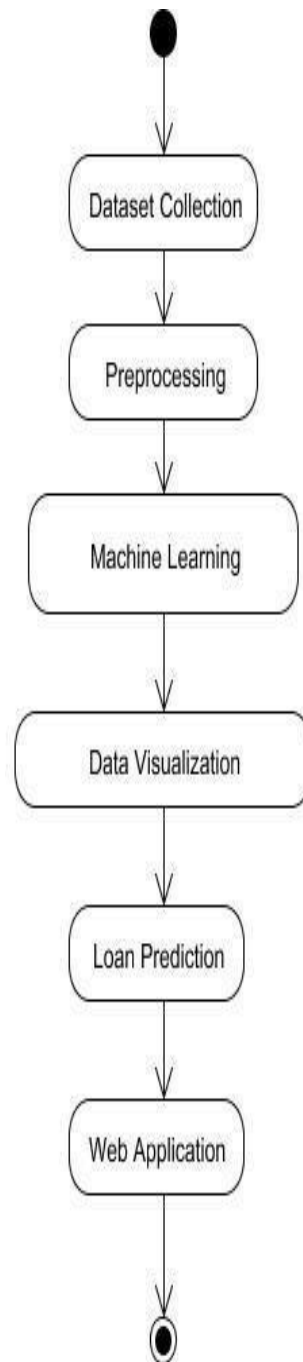
#### **4.4 Activity Diagram:**

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.

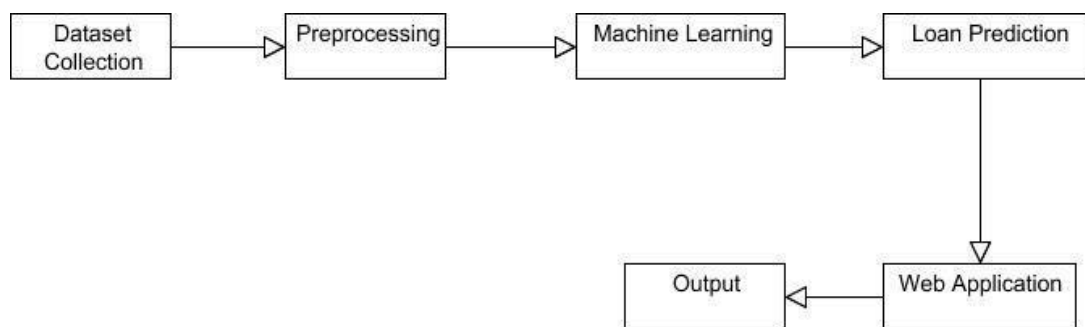


**FIG 4.4 ACTIVITY DIAGRAM OF THE SYSTEM**



#### 4.5 Collaboration Diagram:

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



**FIG 4.5 COLLABORATION DIAGRAM OF THE SYSTEM**

## **CHAPTER-5**

### **5.1 Design and Implementation Constraints**

#### **5.1.1 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

#### **5.2.2 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

#### **5.2.3 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 with the relation concept of entity-relationship modeling and many object oriented methods.

## **5.2 Other Nonfunctional Requirement**

### **5.2.1 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.

### **5.2.2 Safety Requirements**

1. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
2. 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.
3. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.
4. 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.
5. Systems with different requirements for safety levels must be separated.
6. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

## 5.3 MODULES

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

### 5.3 MODULE EXPLANATION:

#### 5.3.1 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 some accuracy 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.

## 5.4 CODING

Once the design aspect of the system is finalizes the system enters into the coding and testing phase. The coding phase brings the actual system into action

by converting the design of the system into the code in a given programming language. Therefore, a good coding style has to be taken whenever changes are required it easily screwed into the system.

## 5.5 CODING STANDARD

Coding standards are guidelines to programming that focuses on the physical structure and appearance of the program. They make the code easier to read, understand and maintain. This phase of the system actually implements the blueprint developed during the design phase. The coding specification should be in such a way that any programmer must be able to understand the code and can bring about changes whenever felt necessary. Some of the standard needed to achieve the above-mentioned objectives are as follows:

Program should be simple, clear and easy to understand.

Naming conventions

Value conventions

Script and comment procedure

Message box format

Exception and error handling

### 5.5.1 NAMING CONVENTIONS

Naming conventions of classes, data member, member functions, procedures etc., should be **self-descriptive**. One should even get the meaning and scope of the variable by its name. The conventions are adopted for **easy understanding** of the intended message by the user. So it is customary to follow the conventions. These conventions are as follows:

#### **Class names**

Class names are problem domain equivalence and begin with capital letter and have mixed cases.

#### **Member Function and Data Member name**

Member function and data member name begins with a lowercase letter

with each subsequent letters of the new words in uppercase and the rest of letters in lowercase.

### **5.5.2 VALUE CONVENTIONS**

Value conventions ensure values for variable at any point of time. This involves the following:

- Proper default values for the variables.
- Proper validation of values in the field.
- Proper documentation of flag values.

### **5.5.3 SCRIPT WRITING AND COMMENTING STANDARD**

Script writing is an art in which indentation is utmost important. Conditional and looping statements are to be properly aligned to facilitate easy understanding. Comments are included to minimize the number of surprises that could occur when going through the code.

### **5.5.4 MESSAGE BOX FORMAT**

When something has to be prompted to the user, he must be able to understand it properly. To achieve this, a specific format has been adopted in displaying messages to the user. They are as follows:

- X – User has performed illegal operation.
- ! – Information to the user.

## **5.6 TEST PROCEDURE**

### **SYSTEM TESTING**

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. The goal of the testing during phase is to verify that the specification has been accurately and completely incorporated into the design, as well as to ensure the correctness of the design itself. For example the design must not have any logic faults in the design is detected before coding commences, otherwise the cost of fixing the faults will be

considerably higher as reflected. Detection of design faults can be achieved by means of inspection as well as walkthrough.

Testing is one of the important steps in the software development phase. Testing checks for the errors, as a whole of the project testing involves the following test cases:

- Static analysis is used to investigate the structural properties of the Source code.
- Dynamic testing is used to investigate the behavior of the source code by executing the program on the test data.

## **5.7 TEST DATA AND OUTPUT**

### **5.7.1 UNIT TESTING**

Unit testing is conducted to verify the functional performance of each modular component of the software. Unit testing focuses on the smallest unit of the software design (i.e.), the module. The white-box testing techniques were heavily employed for unit testing.

### **5.7.2 FUNCTIONAL TESTS**

Functional test cases involved exercising the code with nominal input values for which the expected results are known, as well as boundary values and special values, such as logically related inputs, files of identical elements, and empty files. Three types of tests in Functional test:

- Performance Test
- Stress Test
- Structure Test

### **5.7.3 PERFORMANCE TEST**

It determines the amount of execution time spent in various parts of the unit, program throughput, and response time and device utilization by the program UNIT

#### **5.7.4 STRESS TEST**

Stress Test is those test designed to intentionally break the unit. A Great deal can be learned about the strength and limitations of a program by examining the manner in which a programmer in which a program unit breaks.

#### **5.7.5 STRUCTURED TEST**

Structure Tests are concerned with exercising the internal logic of a program and traversing particular execution paths. The way in which White-Box test strategy was employed to ensure that the test cases could Guarantee that all independent paths within a module have been have been exercised at least once.

- Exercise all logical decisions on their true or false sides.
- Execute all loops at their boundaries and within their operational bounds.
- Exercise internal data structures to assure their validity.
- Checking attributes for their correctness.
- Handling end of file condition, I/O errors, buffer problems and textual errors in output information

#### **5.7.6 INTEGRATION TESTING**

Integration testing is a systematic technique for construction the program structure while at the same time conducting tests to uncover errors associated with interfacing. i.e., integration testing is the complete testing of the set of modules which makes up the product. The objective is to take untested modules and build a program structure tester should identify critical modules. Critical modules should be tested as early as possible. One approach is to wait until all the units have passed testing, and then combine them and then tested. This approach is evolved from unstructured testing of small programs. Another strategy is to construct the product in increments of tested units. A small set of modules are integrated together and tested, to which another module is added and tested in combination. And so on. The advantages of this approach are that, interface dispenses can be easily found and corrected. The major error that was faced during the project is



linking error. When all the modules are combined the link is not set properly with all support files. Then we checked out for interconnection and the links. Errors are localized to the new module and its intercommunications. The product development can be staged, and modules integrated in as they complete unit testing. Testing is completed when the last module is integrated and tested.

## **5.8 TESTING TECHNIQUES / TESTING STRATEGIES**

### **5.8.1 TESTING**

Testing is a process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an as-yet –undiscovered error. A successful test is one that uncovers an as-yet- undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together. System testing requires a test consists of several key activities and steps for run program, string, system and is important in adopting a successful new system. This is the last chance to detect and correct errors before the system is installed for user acceptance testing.

The software testing process commences once the program is created and the documentation and related data structures are designed. Software testing is essential for correcting errors. Otherwise the program or the project is not said to be complete. Software testing is the critical element of software quality assurance and represents the ultimate the review of specification design and coding. Testing is the process of executing the program with the intent of finding the error. A good test case design is one that as a probability of finding an yet undiscovered error. A successful test is one that uncovers an yet undiscovered error. Any engineering product can be tested in one of the two ways:

#### **5.8.1 WHITE BOX TESTING**

This testing is also called as Glass box testing. In this testing, by knowing the specific functions that a product has been design to perform test can be conducted that demonstrate each function is fully operational at the same time searching for errors in each function. It is a test case design method that uses the control structure of the procedural design to derive test cases. Basis path testing is a white

box testing.

Basis path testing:

- Flow graph notation
- Cyclometric complexity
- Deriving test cases
- Graph matrices Control

### **5.8.2 BLACK BOX TESTING**

In this testing by knowing the internal operation of a product, test can be conducted to ensure that “all gears mesh”, that is the internal operation performs according to specification and all internal components have been adequately exercised. It fundamentally focuses on the functional requirements of the software.

The steps involved in black box test case design are:

- Graph based testing methods
- Equivalence partitioning
- Boundary value analysis
- Comparison testing

A software testing strategy provides a road map for the software developer. Testing is a set activity that can be planned in advance and conducted systematically. For this reason a template for software testing a set of steps into which we can place specific test case design methods should be strategy should have the following characteristics:

- Testing begins at the module level and works “outward” toward the integration of the entire computer based system.
- Different testing techniques are appropriate at different points in time.
- The developer of the software and an independent test group conducts testing.

- Testing and Debugging are different activities but debugging must be accommodated in any testing strategy.

### **5.8.2 INTEGRATION TESTING:**

Integration testing is a systematic technique for constructing the program structure while at the same time conducting tests to uncover errors associated with. Individual modules, which are highly prone to interface errors, should not be assumed to work instantly when we put them together. The problem of course, is “putting them together”- interfacing. There may be the chances of data lost across on another’s sub functions, when combined may not produce the desired major function; individually acceptable impression may be magnified to unacceptable levels; global data structures can present problems.

### **5.8.3 PROGRAM TESTING:**

The logical and syntax errors have been pointed out by program testing. A syntax error is an error in a program statement that in violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are common syntax error. These errors are shown through error messages generated by the computer. A logic error on the other hand deals with the incorrect data fields, out-off-range items and invalid combinations. Since the compiler s will not deduct logical error, the programmer must examine the output. Condition testing exercises the logical conditions contained in a module. The possible types of elements in a condition include a Boolean operator, Boolean variable, a pair of Boolean parentheses A relational operator or on arithmetic expression. Condition testing method focuses on testing each condition in the program the purpose of condition test is to deduct not only errors in the condition of a program but also other a errors in the program.

### **5.8.4 SECURITY TESTING:**

Security testing attempts to verify the protection mechanisms built in to a system well, in fact, protect it from improper penetration. The system security must be tested for invulnerability from frontal attack must also be tested for invulnerability from rear attack. During security, the tester places the role of individual who desires to penetrate system.

### **5.8.5 VALIDATION TESTING:**

At the culmination of integration testing, software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test-validation testing begins. Validation testing can be defined in many ways, but a simple definition is that validation succeeds when the software functions in manner that is reasonably expected by the customer. Software validation is achieved through a series of black box tests that demonstrate conformity with requirement. After validation test has been conducted, one of two conditions exists.

- \* The function or performance characteristics confirm to specifications and are accepted.
- \* A validation from specification is uncovered and a deficiency created.

Deviation or errors discovered at this step in this project is corrected prior to completion of the project with the help of the user by negotiating to establish a method for resolving deficiencies. Thus the proposed system under consideration has been tested by using validation testing and found to be working satisfactorily. Though there were deficiencies in the system they were not catastrophic.

### **5.8.6 USER ACCEPTANCE TESTING**

User acceptance of the system is key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system and user at the time of developing and making changes whenever required. This is done in regarding to the following points.

- Input screen design.
- Output screen design.

## **CHAPTER – 6**

### **RESULT AND DISCUSSION**

#### **LOAN APPROVAL RECOMMENDATION SECTION**

The modernization of loan approval systems using machine learning approaches is an active area of research and development. With the increasing availability of large amounts of data and advances in machine learning algorithms, it is now possible to build more accurate and efficient loan approval systems. One potential approach to modernizing loan approval systems is to use machine learning algorithms to analyze historical loan data and identify patterns and trends. This information can then be used to train predictive models that can estimate the likelihood of a loan being approved based on a set of applicant characteristics. These predictive models can also be used to generate loan approval recommendations for loan officers and other decision-makers. For example, a loan officer might input the relevant applicant data into the system, and the system would then generate a loan approval recommendation based on the predictive model. Another approach to modernizing loan approval systems is to use machine learning algorithms to automate the loan approval process. In this approach, the loan application data would be fed into the machine learning algorithm, which would then generate an approval or rejection decision based on the applicant's characteristics and other relevant factors. Overall, the use of machine learning algorithms has the potential to significantly improve the efficiency and accuracy of loan approval systems. However, it is important to ensure that these algorithms are retrained on high-quality data and are designed to avoid bias and discrimination. Additionally, human oversight is still necessary to ensure that loan approval decisions are fair and ethical.

Home Page

127.0.0.1:5000

# LOAN APPROVAL SYSTEM

Check your eligibility criteria to get loan

Home

ApplicantIncome: 100000

CoapplicantIncome: 100000

LoanAmount: 300000

Credit\_History: 0

Property\_Area: Urban

Predict

Activate Windows  
Go to Settings to activate Windows.

Type here to search

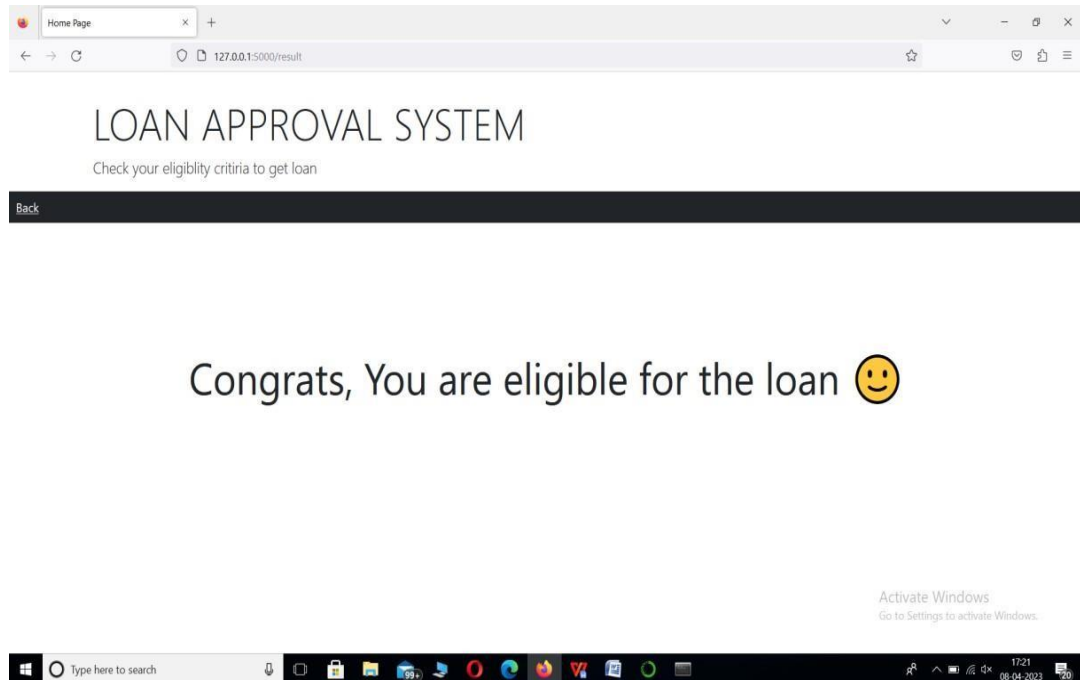
17:21 08-04-2023

**FIG 5.1 LOAN APPROVAL RECOMMENDATION SECTION**

## LOAN APPROVAL SYSTEM DETECTION SELECTION

The loan approval system can be modernized using machine learning approaches by incorporating detection and selection methods into the system. These methods can help to improve the accuracy and efficiency of the loan approval process by identifying potential risks and selecting the most appropriate loan products for each applicant. One detection method that can be used is fraud detection. Machine learning algorithms can analyze applicant data and identify patterns that may indicate fraudulent activity, such as inconsistencies in income or employment information. This information can then be used to flag potentially fraudulent applications for further review by loan officers. Another detection method is risk assessment. Machine learning algorithms can analyze applicant data and identify factors that may indicate a higher risk of default or late payments, such as high debt-to-income ratios or low credit scores. This information can then be used to assign a risk score to each application, which can be used to inform loan approval decisions. In addition to detection methods, selection methods can also be incorporated into the loan approval system. Machine learning algorithms can analyze applicant data **and identify** loan products that are best suited for each

applicant based on their financial history and current financial situation. This can help to ensure that each applicant is offered a loan product that is appropriate for their needs and reduces the risk of default. Overall, incorporating detection and selection methods into the loan approval system can help to improve the accuracy and efficiency of the loan approval process. However, it is important to ensure that these methods are designed to avoid bias and discrimination, and that human oversight is still necessary to ensure that loan approval decisions are fair and ethical.



**FIG .5.2. LOAN APPROVAL SYSTEM DETECTION SELECTION**

## **CHAPTER-7**

### **CONCLUSION**

A modernized loan approval system based on machine learning approach has the potential to significantly improve the efficiency and accuracy of loan approval processes. With the aid of machine learning algorithms, the system can quickly analyze a large amount of data from various sources, including credit scores, income levels, employment history, and other financial data, to make more accurate and objective decisions. Furthermore, a modernized loan approval system based on machine learning can reduce the potential for human error and bias, which may have previously led to incorrect decisions. This can lead to increased confidence in the accuracy of loan decisions and may reduce the risk of fraudulent loan applications being approved. In terms of project conclusion, a modernized loan approval system based on machine learning can help lenders to streamline the loan approval process, reducing the time required to make loan decisions, and enabling lenders to provide quicker responses to loan applicants. This can lead to increased customer satisfaction and may help to reduce the cost of lending. Overall, the implementation of a modernized loan approval system based on machine learning is likely to have a positive impact on the lending industry, and may help to improve the accuracy, efficiency, and speed of loan approvals, leading to increased customer satisfaction and a more streamlined lending process.

### **7.2 FUTURE ENHANCEMENT**

There are several ways in which a modernized loan approval system based on machine learning could be further enhanced. Here are some possibilities:

- Incorporating more data sources:** The loan approval system could be enhanced by incorporating more data sources such as social media data, transactional data, and other publicly available data sources. This would allow for a more comprehensive analysis of an applicant's financial behavior, creditworthiness, and risk profile.
- Improving accuracy:** One of the key goals of a loan approval system based on machine learning is to improve accuracy. This could be achieved by refining the algorithms used to analyze data, using more advanced machine learning techniques, and incorporating feedback from loan officers to improve the accuracy of the system.
- Streamlining the loan approval process:** A modernized loan



approval system based on machine learning could be further enhanced by streamlining the loan approval process. This could be done by automating more of the loan approval process, reducing the need for manual review, and speeding up the time it takes to process loan applications.

Enhancing transparency: Transparency is critical in any loan approval system. By providing more transparency into how loan decisions are made, borrowers can have greater confidence in the system and the decisions being made. This could be done by providing borrowers with more information about how their application was evaluated, what factors were considered, and what the decision-making process involved. Increasing flexibility: A modernized loan approval system based on machine learning could be enhanced by increasing flexibility. This could be done by allowing borrowers to customize their loan terms based on their unique needs and circumstances, and by providing more options for repayment and loan modification. Overall, there are many ways in which a modernized loan approval system based on machine learning could be further enhanced. By incorporating more data sources, improving accuracy, streamlining the loan approval process, enhancing transparency, and increasing flexibility, lenders can provide borrowers with a better experience and more personalized loan products.

## 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] Dhamodaran,s, Lakshmi ,M,(2019)"Ensampling data prediction using Sparse data in mobile intelligent system",International journal of interactive Mobile Technologies, 13(10) ,pp.106-119.
- [3] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
- [4] hakraborty, S., & Srivastava, S. (2021). Loan approval prediction using machine learning: A review. In Advances in Intelligent Systems and Computing (Vol. 1264, pp. 309-316). Springer.
- [5] Gajjar, D., & Vyas, A. (2020). Loan approval prediction using machine learning. In Advances in Intelligent Systems and Computing (Vol. 1176, pp. 421-427). Springer.
- [6] Jiang, S., Cao, J., & Xu, X. (2019). A machine learning-based approach for loan approval prediction. Journal of Ambient Intelligence and Humanized Computing, 10(9), 3609-3617.
- [7] Jiang, S., Cao, J., & Xu, X. (2020). A personalized loan approval system based on machine learning. Journal of Intelligent & Fuzzy Systems, 39(4), 5205-5217.
- [8] Exploratory data analysis  
[https://en.wikipedia.org/wiki/Exploratory\\_data\\_analysis](https://en.wikipedia.org/wiki/Exploratory_data_analysis).
- [9] Pandas Library <https://pandas.pydata.org/pandas-docs/stable/>
- [10] MeanDecreaseAccuracy  
<https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.h>

## APPENDIX

### Source Code:

```
import numpy as np
import pandas as pa
import matplotlib.pyplot as plt
import seaborn as sea
table = pa.read_csv('dataset/loan_data_set.csv')
ttab = table.copy()
ab = table.copy()
tab = table.copy()
Preprocessing
tab.shape
tab.info()
tab.isnull().sum()
tab.isna().sum()
np.sum(tab == 0)
# checking duplicates
tab[tab.Loan_ID.duplicated()]
tab = tab.drop(columns='Loan_ID').copy()
Missing value processing
tab['Self_Employed'].unique()
tab['Gender'] = tab['Gender'].fillna(value=tab['Gender'].dropna().mode()[0])
tab['Property_Area'] =
tab['Property_Area'].fillna(value=tab['Property_Area'].dropna().mode()[0])
tab['Dependents'] =
tab['Dependents'].fillna(value=tab['Dependents'].dropna().mode()[0])
tab['Married'] = tab['Married'].fillna(value=tab['Married'].dropna().mode()[0])
tab['Self_Employed'] =
tab['Self_Employed'].fillna(value=tab['Self_Employed'].dropna().mode()[0])
tab['Credit_History'] =
tab['Credit_History'].fillna(value=tab['Credit_History'].dropna().mode()[0])
tab['Loan_Status'] =
tab['Loan_Status'].fillna(value=tab['Loan_Status'].dropna().mode()[0])
tab['LoanAmount'] =
tab['LoanAmount'].fillna(value=np.round(tab['LoanAmount'].dropna().mean()))
tab['Loan_Amount_Term'] =
tab['Loan_Amount_Term'].fillna(value=tab['Loan_Amount_Term'].dropna().median(
))
tab.isna().sum()
Tab
tab.loc[tab['Dependents']=='3+', 'Dependents'] = 3
tab.loc[tab['Dependents']=='2', 'Dependents'] = 2
tab.loc[tab['Dependents']=='1', 'Dependents'] = 1
tab.loc[tab['Dependents']=='0', 'Dependents'] = 0
```

```

txt_cols=tab.columns.drop(['Dependents','ApplicantIncome','CoapplicantIncome','Loan_Amount_Term', 'Credit_History' ]).to_list()
for i in txt_cols:
    tab[i] = pa.factorize(tab[i])[0]
Tab
Remove outliers
txt_cols.append('Dependents')
txt_cols.append('Credit_History')
txt_cols
outtab = tab.drop(columns=txt_cols)
fig = plt.figure(figsize =(10, 7))
# Creating plot
plt.boxplot(outtab)
plt.show()
for i in outtab.columns.values:
    Q1,Q3 = np.percentile(outtab[i],[25,75])
    IQR = Q3 - Q1
    lf = Q1 - (1.5*IQR)
    hf = Q3 + (1.5*IQR)
    outtab.loc[outtab[i]>hf, i] = hf
    outtab.loc[outtab[i]<lf, i] = lf
Out tab
fin_tab = pa.concat([outtab,tab[txt_cols]], axis = 1).copy()
cls_0 = fin_tab[fin_tab['Loan_Status'] == 0]['Loan_Status']
cls_1 = fin_tab[fin_tab['Loan_Status'] == 1]['Loan_Status']
print("CLASS 0: ",cls_0.shape)
print("CLASS 1: ",cls_1.shape)
Status']X_train_res = fin_tab.drop(columns='Loan_Status')
Y_train_res = fin_tab['Loan
from imblearn.over_sampling import SMOTE
sm = SMOTE()
X_train_res,Y_train_res = sm.fit_resample(X_train_res,Y_train_res)
print("class 0: ",Y_train_res[Y_train_res == 0].shape)
pfinal = pa.concat([X_train_res, Y_train_res], axis = 1)
print("class 1: ",Y_train_res[Y_train_res == 1].shape)
final = pa.concat([X_train_res, Y_train_res], axis = 1)
Fin=final.loc[:,['ApplicantIncome','CoapplicantIncome','LoanAmount','Loan_Amount_Term','Dependents','Credit_History']]
outtab = fin.copy()
fig = plt.figure(figsize =(10, 7))
# Creating plot
plt.boxplot(outtab)
plt.show()
for i in outtab.columns.values:

```

```

Q1,Q3 = np.percentile(outtab[i],[25,75])
IQR = Q3 - Q1
lf = Q1 - (1.5*IQR)
hf = Q3 + (1.5*IQR)
outtab.loc[outtab[i]>hf, i] = hf
outtab.loc[outtab[i]<lf, i] = lf
for i in outtab.columns.values:
    Q1,Q3 = np.percentile(outtab[i],[25,75])
    IQR = Q3 - Q1
    lf = Q1 - (1.5*IQR)
    hf = Q3 + (1.5*IQR)
    outtab.loc[outtab[i]>hf, i] = hf
    outtab.loc[outtab[i]<lf, i] = lf
final.to_csv("preprocess_dataset/train_pre.csv", index = False)
na = pa.concat([outtab,final.drop(columns = outtab.columns)], axis = 1)
Start
Import numpy as np
Import pandas as pa
Import matplotlib.pyplot as plt
Import seaborn as sea
table = pa.read_csv('dataset/train_pre.CSV')
Table = table.copy()
Tab.head()
Feature selection
X = tab.drop(columns = 'Loan_Status')
Y = tab['Loan_Status']
from sklearn.feature_selection import SelectFromModel
from sklearn.linear_model import Lasso
# feature selection using lasso model
fsm = SelectFromModel(Lasso(alpha = 0.05, max_iter=3000000))
fsm.fit(X,Y)
selected_features = X.columns[fsm.get_support()]
fts = pa.concat((X[selected_features],Y), axis = 1)
print("Selected Features: ",selected_features)
Feature scaling
from sklearn.preprocessing import StandardScaler
X = table[selected_features]
sca = StandardScaler()
featuressca = sca.fit_transform(X)
featuressca = pa.DataFrame(featuressca, columns = X.columns)
featuressca.head()
X = featuressca.copy()
Y = table['Loan_Status']
Machine learning models
# Library
from sklearn.model_selection import train_test_split, cross_val_score

```

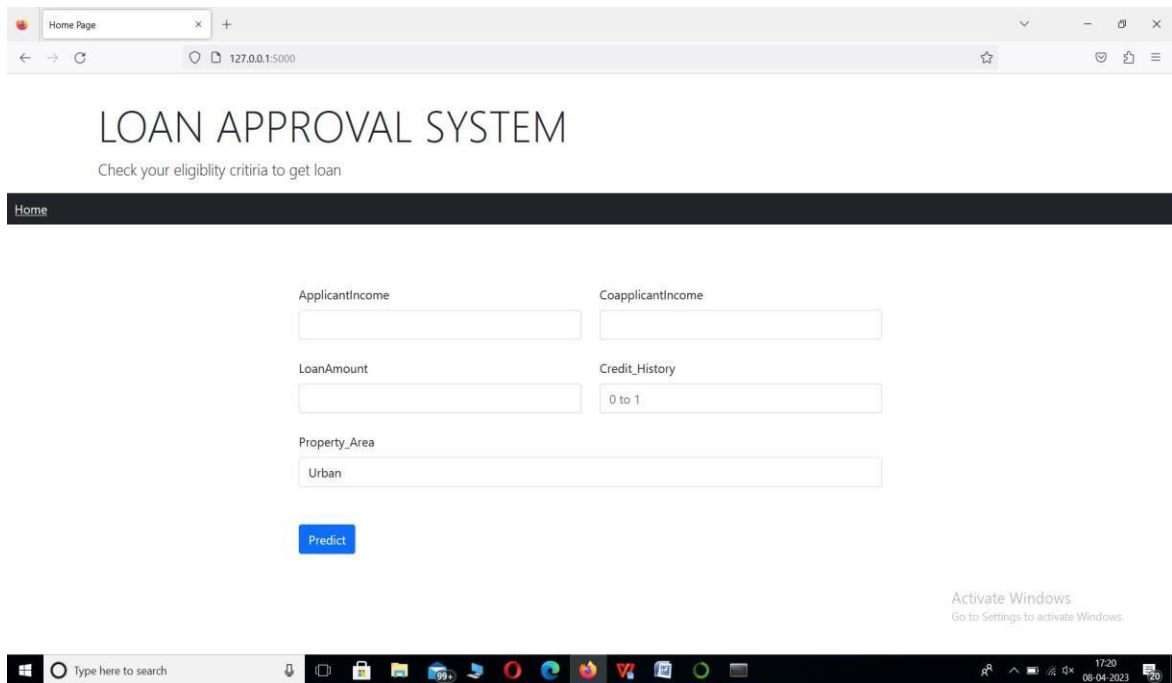
```

from sklearn.metrics import confusion_matrix, accuracy_score,
ConfusionMatrixDisplay
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
import warnings
warnings.filterwarnings(action='ignore')
KNN
X_train, X_test, y_train, y_test = train_test_split(X, Y, test_size = 0.3, random_state
=
27)
knn = KNeighborsClassifier(n_neighbors=20)
knn.fit(X_train, y_train)
test_accuracy = knn.score(X_test, y_test)
print('Score:{}'.format(test_accuracy))
y_pred = knn.predict(X_test)
confusion_mat = confusion_matrix(y_test,y_pred)
print("Confusion Matrix")
print(confusion_mat)
ksc = accuracy_score(y_test,y_pred)
cm_display= ConfusionMatrixDisplay(confusion_matrix = confusion_mat,
display_labels = [False, True]) cm_display.plot()
plt.show()
Model
## Writing to model file
import joblib
joblib.dump(rf, r'./randamfr.pkl')
Web_app
from flask import Flask,render_template,url_for,request
import pickle
import joblib
import numpy as np
app=Flask(__name__)
model_path = './Trained_Model/randamfr.pkl'
model = joblib.load(
open(model_path, 'rb'))
@app.route('/')
def home():
return render_template('home.html')
@app.route('/result',methods=['POST'])
def predict():
# Getting the data from the form
ApplicantIncome = int(request.form['ApplicantIncome'])
CoapplicantIncome = int(request.form['CoapplicantIncome'])
LoanAmount = int(request.form['LoanAmount'])

```

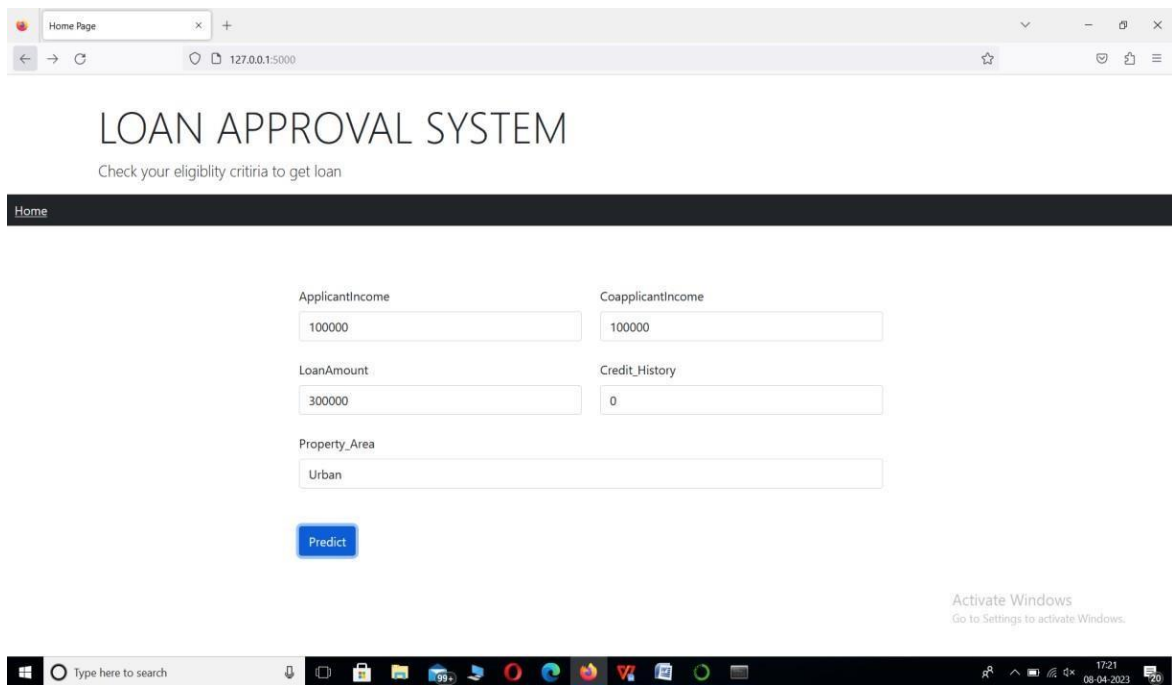
```
Credit_History = int(request.form['Credit_History'])
Property_Area = int(request.form['Property_Area'])
query=np.array([[ApplicantIncome,CoapplicantIncome,LoanAmount,Credit_History,
Property_Area]])
prediction = model.predict(query)
return render_template('result.html', prediction=prediction)
if __name__ == '__main__':
    app.run(debug=True)
```

## Screenshots



A screenshot of a web browser displaying the 'Home Page' of a 'LOAN APPROVAL SYSTEM'. The browser's address bar shows '127.0.0.1:5000'. The page has a dark header with the word 'Home' on the left. Below the header, the title 'LOAN APPROVAL SYSTEM' is centered, followed by the subtitle 'Check your eligibility criteria to get loan'. The main content area contains five input fields: 'ApplicantIncome', 'CoapplicantIncome', 'LoanAmount', 'Credit\_History', and 'Property\_Area'. The 'Credit\_History' field has a placeholder value '0 to 1'. The 'Property\_Area' field has a placeholder value 'Urban'. A blue 'Predict' button is located below the input fields. In the bottom right corner, there is a 'Activate Windows' watermark with the text 'Go to Settings to activate Windows.'.

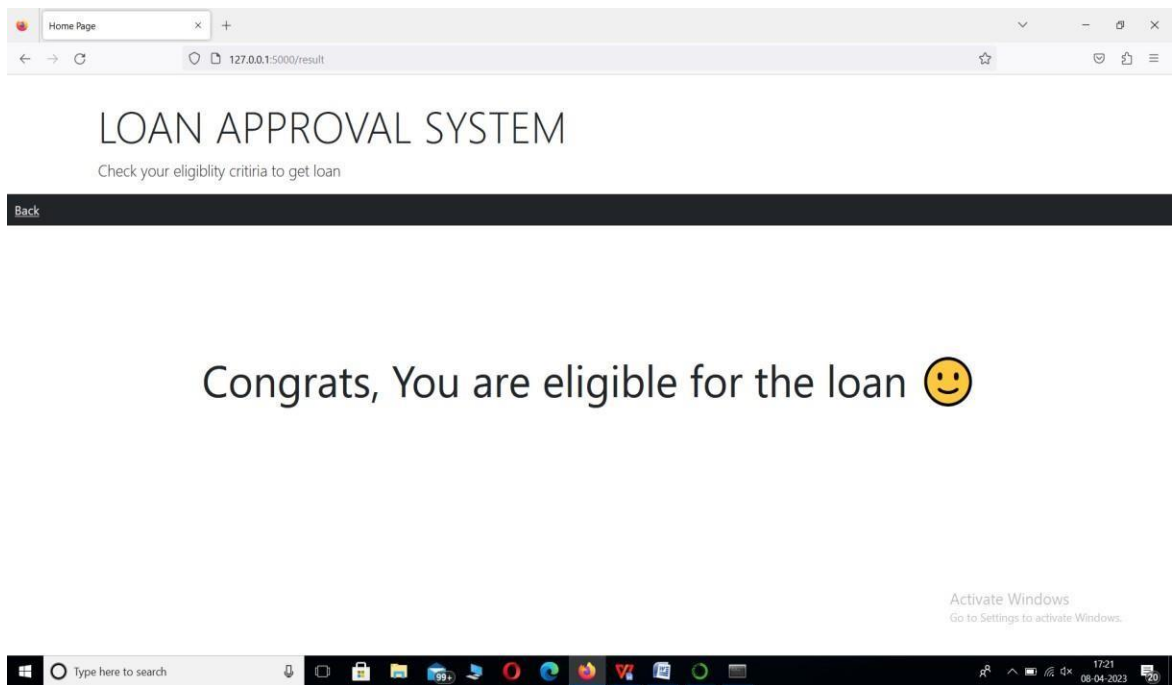
**FIG.B.1.HOME PAGE**



A screenshot of the same web browser displaying the 'Home Page' of the 'LOAN APPROVAL SYSTEM'. The browser's address bar shows '127.0.0.1:5000'. The page layout is identical to the previous screenshot, but the input fields now contain specific values: 'ApplicantIncome' is '100000', 'CoapplicantIncome' is '100000', 'LoanAmount' is '300000', 'Credit\_History' is '0', and 'Property\_Area' is 'Urban'. The blue 'Predict' button remains below the fields. The 'Activate Windows' watermark is still present in the bottom right corner.

**FIG.B.2.LOAN APPROVAL SYSTEM RECOMMENDATION SECTION**





**FIG.B.3 LOAN APPROVAL SYSTEM DETECTION SELECTION**

# Prediction of Modernized Loan Approval System Based on Machine Learning Approach

**Sugan Athish K.C**

**1**

Student

Department of Computer science And Engineering  
Sathyabama Institute of Science and Technology

Chennai,India

[Sugans1357@gmail.com](mailto:Sugans1357@gmail.com)

**DR.RAMYA G FRANKLIN**

**2**

Assistant professor

Department of Computer science And Engineer  
Sathyabama Institute of Science and Technology

Chennai,India

[ramyagfranklin@sathyabama.ac.in](mailto:ramyagfranklin@sathyabama.ac.in)

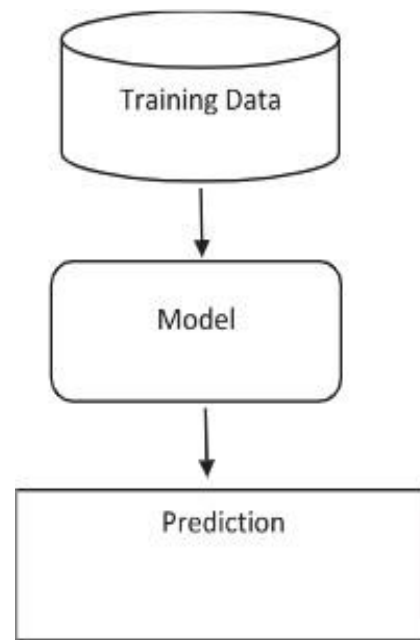
**Abstract:** Technology has improved both the quantity and The bank will next assess the candidate's merit utilising its quality of life for humans. We intend to provide something fresh and unique every day. In the banking industry, the candidate receives proofs/backup prior to approval of the loan amount..Machines make life easier and give us a sense of completion. The system's previous candidate data is what decides whether to approve the application. Several people ask for loans every day in the banking industry, yet the bank's resources are constrained. A classes-function algorithm would be very helpful in this situation if the proper prediction could be made. Examples include support vector machine, logistic regression, and random forest classifiers. A Bank's profit and loss vary according to the loan size, which also determines. If the borrower is making loan installments is unknown. The loan recovery represents the largest investment .The method of improvement is crucial in the banking industry A machine learning model was developed using the historical data of the candidates and a number of categorisation techniques.The main objective of this research is to forecast whether new loan applications will be granted or refused using machine learning models trained on past data.

**Keyword :** *Statistics, Credit, Instruction, Trial, divination, Machine Learning*

## 1. INTRODUCTION

One can predict whether or not a loan will be approved

using a modernised loan approval system that uses machine learning techniques. We collect information about the customer using this way, including information on his monthly wage, marital status, loan amount, loan term, etc. In order to determine if the customer qualifies for the loan, the bank will then employ its criteria. Hence, there is a classification scheme. After building a model with a training set in this system, the class The classifier can provide the data items the right group. A sample dataset is created in order to train the data and deliver the accurate result, which the client's potential and capacity to repay the loan. A modernised loan approval procedure may be favourable for both banks and clients. The bank will be able to oversee the entire procedure when the client submits an application without interference from shareholders or the outside world. Goal a machine learning mod.

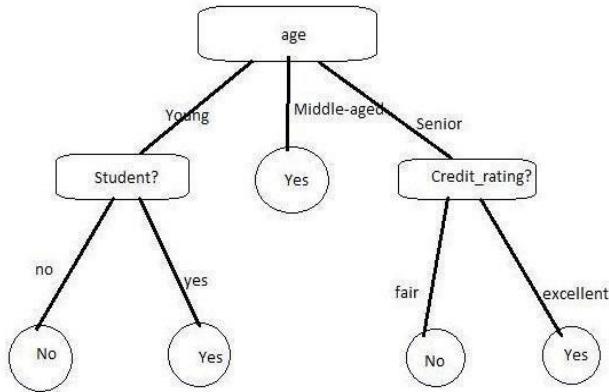


**Fig.1.Simple machine learning model**

## II ALGORITHMS FOR COMPUTER TECHNOLOGY

Three machine learning methods are employed in this study's research

- (a) **XGBoost** - XGBoost, a decision tree-based open source software tool. Machine learning methods are built using the gradient boosting framework. Linux and Windows are all supported.
- (b) **Random Forests** - This classifier produces a large number of prediction-accurate decision trees, each one superior than the others.
- (c) **Decision Tree** - A decision tree was used to divide the dataset into manageable chunks. then weigh all of your possibilities.



**Fig. 2. Decision Tree**

## FORMULATION OF THE PROBLEM

A significant issue is that many people are unable to repay bank debts. Also, banks are losing money. Banks get a lot of loan requests every day, but not all of them are approved. Most banks base their decisions on their own risk and credit scoring systems. It will become evident quickly why there is a debt issue. It is a serious problem that so many individuals struggle to pay their bills. Banks

are also in the red. Banks get a lot of loan requests every on several ground day, but not all of them are approved The answer to this question is that not everyone qualifies for loans because if the borrower is unable to repay the loan, either they

themselves or the business or bank that provided it would suffer a loss. The person giving the loan must first validate or establish certain criteria in order to evaluate if the person receiving it is capable of repaying it or not. Like banks, we provide the option of credit cards, but not everyone qualifies for one. An accessible credit score might be used to determine eligibility. For a loan, a person's creditworthiness is required. A NECESSARY TOOL To be accepted for a loan, a person has to have a good credit score. There are several prerequisites to getting a credit card, such as having a source of income. Banks provide loans on the borrower's behalf, who is required to present supporting papers and pass verification. like a company that cannot supplyIn this project, data processing techniques will be utilised to look for defaulters with similar characteristics in loan-approveddata. This will help banks in the future make better decisions. LS.

- Microsoft office
- Python
- Numpy+
- Dataset
- XGBoost
- Pandas

- Matplotlib
- Machine learning algorithms

## V. DATA ANALYSIS LOAN PREDICTION

The topic of how we evaluate whether to offer a loan to a customer or not arises. We have two target factors on which we base this decision. We must also verify all necessary paperwork, including confirmation of identity, address, and income. After that, we give the customer a loan that she might choose to return or not. As they need loans for both their enterprises and their children's education, middle-class parents have a considerable need for loans As a result, we must double-check everything because bank are not experiencing NPA loans.

- Higher possibilities of loan repayment are associated with better customers.
- Background checks should be thorough so that we can anticipate receiving the loan back at the ideal moment.

Thus, our objective variable is the result of our analysis

- A Data set

**TABLE I Default Data set**

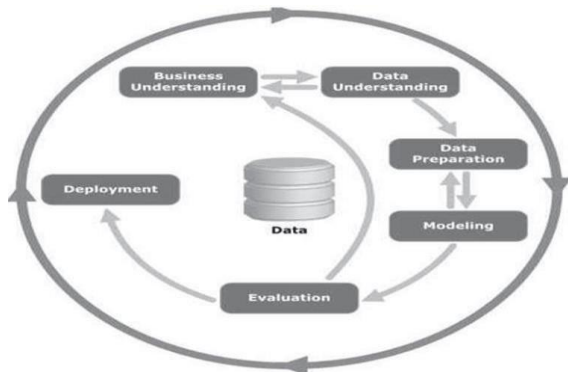
Fluctuating	Representation	kind
Credit	ParticularCredit	Integer
Gender	Mr/ Mrs	Person
Wedded	Possible	Person

**TABLE II. A DATA SET**

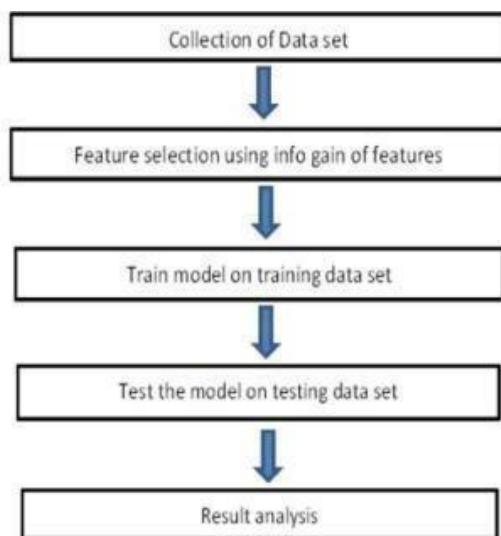
Fluctuating	Explanation	Kind
minor	Number of charger	integer
Intruccion	Degree noider/ Learner	string
freelance	Independent	character
Earnings	Applicant Salary	integer
Co –applicant income	Co-applicant remuneration	integer
Credit	Amount borrowed in thousands	integer
Loan_Amount_Ter m	Term of Loan in Months	integer
faith_History	credit history requirements ments	integer
Prossessions_Area	standards for credit history	string
credit_Status	(Y/N) Loan acceptance	character

## VI FORECAST TECHNIQUE FOR LOANS

This recommended method will use past behaviour while evaluating customer behaviour with these dataset, we can create a learning model that algorithms based on machine learning are used to determine if a client will be able to repay the loan.



**Fig.3.process Diagram**



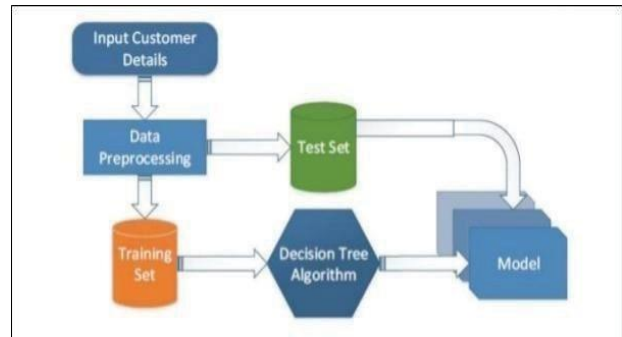
**Fig.4.Forecast Technique For Analysis**

## VII VALUE OF THE PROPOSED SYSTEM

The benefits of loan prediction will be covered in this section. With this technique, we may determine a loan applicant's capacity to pay back the loan. We forecast that the consumer would be given a loan if they are able to make the payments. In addition, we presume that the applicant won't be accepted if the application is denied. The advantage of this strategy is that we Gave some To establish a client's eligibility, we create the algorithms and simply consider the facts. This study assists banks in reducing potential losses and may boost the level of loans

## VIII DIAGRAM SHOWING THE METHODS

### ARCHITECTURE



**Fig.5.Architecture schematic**

## XI CONCLUSION:

This study claims that both datasets' prediction accuracy is excellent. When a client is experiencing a disaster, for example, the algorithm may not be able to forecast the right outcome. This study offers an accurate assessment of a borrower's capacity to repay a loan. Age, income, loan term, and loan amount are the primary determinants of qualifying for loans . whether the customer would have been). The two most important factors in establishing the loan applicant's category are both credit history

Submission message	Score	Code File	Solution File	Final Solution
XGBoost	0.777777777777778	---	<a href="#">Download</a>	<input type="radio"/>
Random Forest	0.763888888888889	---	<a href="#">Download</a>	<input type="radio"/>
Decision Tree	0.645833333333333	---	<a href="#">Download</a>	<input type="radio"/>
Third submission	0.777777777777778	---	<a href="#">Download</a>	<input type="radio"/>
Second submission	0.777777777777778	---	<a href="#">Download</a>	<input type="radio"/>
First submission	0.784722222222222	---	<a href="#">Download</a>	<input type="radio"/>

## REFERENCES

- [1] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
- [2] Anveshmi dumala and s.pallamsetty (2016) "Investigating then impact of simulation Time on Convergence Activity & Duration of EIGRP, OSPF Routing Protocols under Link Failure and Link Recovery in WAN Using OPNET Modeler", International Journal of Computer Science Trendsand Technology (IJCST) – Vol. 4 No. 5, pp. 38- 42
- [3] "Loan Approval Prediction Using Machine Learning Techniques" by R. S. Sambasivan, R. B. Lohani, and M. N. Bin Muhd Nor. This paper discusses the use of machine learning algorithms such as decision trees, support vector machines, and artificial neural networks for loan approval prediction.
- [4] "Loan Approval Prediction: A Deep Learning Approach" by D. Dheeraj, K. S. Arvind, and S. S. Suresh. This paper proposes a deep learning approach for loan approval prediction using a convolutional neural network (CNN) model.
- [5] "A Modern Approach to Loan Approval Prediction Using Machine Learning" by S. H. Kim, S. J. Choi, and Y. J. Lee. This paper proposes a modern approach to loan approval prediction using a hybrid machine learning model that combines random forest and gradient boosting algorithms.
- [6] "Credit Risk Assessment using Machine Learning Algorithms: A Systematic Review" by M. A. Firdaus and M. A. Islam. This article provides a systematic review of the use of machine learning algorithms for credit risk assessment, including loan approval prediction.
- [7] Lakshman Narayana Vejendla and A Peda Gopi, (2017)," Visual cryptography for gray scale images with enhanced security mechanisms", Traitement du Signal, Vol.35, No.3-4, pp.197-208. DOI: 10.3166/ts.34.197- 208.
- [8] Chaitanya,K.,andS.Venkateswarlu,(2016),"DETECT ION OF BLACKHOLE & GREYHOLE ATTACKS IN MANETs BASED ON ACKNOWLEDGEMENT BASED APPROACH." Journal of Theoretical and Applied Information Technology 89.1: 228
- [9] Patibandla R.S.M.L., Kurra S.S., Mundukur N.B (2012), "A Study on Scalability of Services and Privacy Issues in Cloud Computing". In:RamanujamR.,Ramaswamy S. (eds) Distributed Computing and Internet Technology. ICDCIT 2012. Lecture Notes in Computer Science, vol 7154. Springer, Berlin, Heidelberg
- [10] Patibandla R.S.M.L., Veeranjanyulu N. (2018), "Survey on Clustering Algorithms or Unstructured Data". In: Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (eds) Intelligent Engineering Informatics. Advances in Intelligent Systems and Computing, vol 695. Springer, Singap.Venkata Rao Maddumala, R. Arunkumar, and S. Arivalagan (2018)"An Empirica review. Science and Technology Vol.7 No.S1, pp. 96- 100. Review on Data Feature Selection and Big Data Clustering" Asian Journal of Computer.
- [11] Dhamodaran,s, Lakshmi ,M,(2019)"Ensampling data prediction using Sparse data in mobile intelligment system",International journal of interactive Mobile Technologies, 13(10) ,pp.106-119