AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM

A PROJECT REPORT

Submitted by

P.S.A.BHASKAR REDDY [Reg No:RA1911003010072]

M.GNANESWAR REDDY [Reg No:RA1911003010113]

Under the guidance of

Dr. R. S. PONMAGAL

Associate Professor,Department of Computing Technologies

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

DEPARTMENT OF COMPUTING TECHNOLOGIES

COLLEGE OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR – 603 203

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that this B.Tech project report titled **"AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM"** is the bonafide work of **P.S.A.BHASKAR REDDY [RA1911003010072]** and **M.GNANESWAR REDDY [RA1911003010113]** who carried out the projectwork under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion for this or any other candidate.

Dr. R. S. PONMAGAL
SUPERVISOR
Associate Professor
Department of Computing Technologies

Dr. R. S. PONMAGAL
PANEL HEAD
Associate Professor
Department of Computing Technologies

M. Pushpalatha

Department of Computing Technology
Kattankulathur
603 203
SRM IST

Dr. M. PUSHPALATHA
HEAD OF THE DEPARTMENT
Department of Computing Technologies

INTERNAL EXAMINER

15/5/23

EXTERNAL EXAMINER

Department of Computing Technologies

SRM Institute of Science and Technology

Own Work Declaration Form

Degree/ Course : B.Tech in Computer Science and Engineering

Student Names : P.S.A.Bhaskar Reddy, M.Gnaneswar Reddy

Registration Number: RA1911003010072, RA1911003010113

Title of Work : An Approach For Prediction Of Loan Approval Using ML Algorithm

We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism, as listed in the UniversityWebsite, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that we have met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web,etc.)
- Given the sources of all pictures, data etc. that are not my own

- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook /University website

I understand that any false claim for this work will be penalized in accordance with theUniversity policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.
Student 1 Signature: P.S.A. BHASKAR REDDY **Student 2 Signature:** H. Gyan Swar Reddy **Date:** 09/05/2023
If you are working in a group, please write your registration numbers and sign with the date for every student in your group.

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr. T. V. Gopal**, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the projectwork.

We are incredibly grateful to our Head of the Department, **Dr. M. Pusphalatha ,** Professor, Department of Computing Technologies , SRM Institute of Science and Technology,for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators, **Dr. M. Kanchana**, **Dr. G. Usha, Dr. R. Yamini and Dr. K. Geetha,**Panel Head, **Dr. R. S. Ponmagal,** Associate Professor and Panel Members, **Dr. S. Nagadevi,** Assistant Professor, **Dr. M. Gayathri,** Assistant Professor and **Dr. Kishore Anthuvan Sahayaraj,** Assistant Professor, Department of Computing Technologies , SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Dr. S. BABU**, Associate Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide **Dr.R.S.PONMAGAL**, Associate Professor, Department of Computing Technologies, SRM Institute of Science and Technology,for providing us with an opportunity to pursue our project under her mentorship. She provided us with the freedom and support to explore the research topics of our interest.Her passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank the Computing Technologies Department staff and students, SRM Institute of Science and Technology, for their help duringour project. Finally, we would like to thank parents, family members, and friends for their unconditional love, constant support, and encouragement.

P.S.A.BHASKAR REDDY

[RA1911003010072]

M.GNANESWAR REDDY

[RA1911003010113]

TABLE OF CONTENTS

ABSTRACT

Numerous people are seeking for bank advances as a result of the financial sector's improvement, but the bank only has a limited amount of resources that it can allocate to certain people. Therefore, it is a typical interaction to find out who can be granted credit in order to make the bank feel more comfortable.Therefore, in this activity, we work to lessen the risk associated with selecting the covered person in order to conserve a significant amount of bank resources and efforts. This is accomplished by extracting information from historical records of those to whom advances have previously been granted, and based on these records and interactions, a machine was built using the model that produces the most accurate results. Four areas make up this project: (i)Data Collection (ii)Comparison of ML models on gathered information (iii)Training of framework on most encouraging model (iv)Testing.

LIST OF FIGURES

LIST OF ABBREVIATIONS

HTML	Hyper Text Markup Language
CSS	Cascading Style Sheet
JS	Javascript
CV	Computer Vision
DB	Data Base
UI	User Interface
ML	Machine Learning
SVM	Support Vector Machine
KNN	K-Nearest Neighbour

CHAPTER 1
INTRODUCTION

1.1 Introductory Part

A bank is essential to the market economy. Assessing credit risk is critical to the success or failure of an organization. Before granting credit, banks evaluate borrowers to determine if they are good (non-defaulters) or bad (defaulters). However, For any business or bank, projecting whether a borrower will fail in the future is a difficult undertaking. Essentially, identifying the loan defaulters involves a binary classification problem that can be difficult to solve. Loan amount; customer history determines his capacity to obtain credit. Sorting out whether a borrower has defaulted or not is the issue. However, because of Creating a model to assess credit risk is a challenging task, particularly given the increasing demand for loans. One potential source of information for such a model could be the data on past customers from various banks who were granted credit was used for this task. The AI model is built using that data in order to generate reliable results.This investigation's main goal is to foresee the security of credit. The strategic relapse calculation is used to forecast future health. In order to fill in any gaps in the informational collection, the data must first be cleaned 1000 examples, 10 mathematical formulas, and 8 blatantly obvious characteristics was used to create our model. Different criteria, such as CIBIL Score (Credit History), Business Value, are used to credit an advance to a client.Banks today must deal with the critical issue of credit risk assessment, which allows them to determine whether or not to approve a loan application based on the likelihood that the borrower would default in the future. This can enhance the volume of credits while also assisting banks in reducing potential losses.The information of previous clients of several institutions, to whom a number of international loan applications were submitted accepted, was used for this task.

1.2 Motivation

The loan approval process is critical for financial institutions. The system either approves or denies loan applications, and loan recovery plays a big role in a bank's financial health performance. Predicting whether a borrower will repay a debt is an extremely challenging task for financial companies. The loan approval is predicted using machine learning.

1.3 Objective

The software programme is used to carry out the goal. This suggested model delivers crucial data with the maximum level of precision. In commercial banks, it is used with a machine learning classifier to forecast loan status.

1.4 Existing System

A decision tree technique of loan forecasting had been proposed by Abhilash. The decision tree and the highly practical Regression algorithm are two of the most used models. Decision trees make use of a number of different techniques to determine whether dividing a knot into two or more smaller knots. It becomes simpler to create future subknots. Therefore, we can state that as the desired variable increases, the bump's chasteness increases. Cons of Current System include

(1).Less efficiency

(2).Less accuracy

(3).Issue With Prediction

(4).It is not a web-based programme.

1.5 Proposed System

In this study, we provide a logistic regression analysis algorithm for loan approval that is machine learning-based. Logistic regression is one of the most popular and successful classification-based techniques. Because Logistic Regression employs the concept of predictive analysis, which was enough for summarising the data, it serves a function or has significance.The benefits of the proposed system are

(1).Great accuracy

(2).Time efficient

(3).Simple UI

(4).It’s a web based application

(5).It gives perfect result.

1.6 System Requirements

- Anaconda (Spyder)
- Python
- OS : Windows, Linux (environment)
- Processor : Intel Dual Core(Minimum)
- RAM : 4GB(Minimum)

CHAPTER 2

LITERATURE SURVEY

2.1 Review of Literature Survey

1. PAPER NAME: An Approach for Prediction of Loan Approval using Machine Learning Algorithm

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

ABSTRACT: It is true that banks rely heavily on their credit lines as a primary source of income, and loan default has a significant impact on their profitability. Therefore, predicting loan defaulters is a crucial aspect of managing risk and maximizing earnings. In this regard, the logistic regression model has been widely used in predictive analytics to forecast defaulters on loans. The logistic regression model considers a number of variables, including the customer's individual characteristics (age, purpose, credit history, credit amount, and credit term) and checking account information (wealth), to accurately calculate the probability of loan default. By using this approach, banks can identify the ideal customers to target for loan approval and manage their risk effectively.

Therefore, it is important for banks to consider all relevant customer characteristics when making credit decisions and identifying loan defaulters.In conclusion, the logistic regression model is a powerful tool for predicting loan defaulters and managing risk in the banking industry. By accurately assessing the probability of loan default, banks can make informed credit decisions and target the right customers for loan approval. This approach can ultimately lead to improved profitability and reduced Non-Performing Assets for banks.

Keywords: outlier, component, over-fitting, transform, and loan.

2. PAPER NAME: Loan Default Forecasting using Data Mining

AUTHOR: Bhoomi Patel1, Harshal Patil2, Jovita Hembram3, Shree Jaswal

ABSTRACT: It is indeed crucial for banks and other financial organizations to accurately estimate the likelihood of default on a debt, in order to minimize their losses and increase the amount of credit they can offer. Building a model that considers various features of an applicant is one way to achieve this goal.With the increasing use of technology in criminal activities, there is a rising problem ofdistressed banks and bad debts. Employing data mining techniques to identify from datasets containing information on mortgage loan applications, likely defaulters can be an effective way to help banks make better judgments in the future.Data mining involves analyzing large amounts of data to uncover patterns and relationships that can be used to make predictions or inform decision-making.

By applying data mining techniques to datasets containing information about loan applicants, banks can determine the patterns and traits that are connected to default and use this knowledge to improve their assessment of loan applications.There are many different data mining techniques that can be used for this purpose, including decision trees, neural networks, and logistic regression. These techniques involve analyzing various features of loan applicants, such as their credit history, income, and employment status, and using this information to make predictions about their likelihood of default.Overall, the use of data mining techniques can be a powerful tool for banks and other financial organizations to improve their assessment of loan applications and reduce their risk of losses from default.

Keywords: credit, loan, forecasting, and data analysis.

3. PAPER NAME: Prediction of Loan Status in Commercial Bank using Machine Learning Classifier

AUTHOR: G. Arutjothi,Dr. C. Senthamara

ABSTRACT: Great! Developing a credit scoring model using the use of machine learning techniques is a promising option to increase the precision of identifying credit defaulters and legitimate clients in the banking industry. It's critical to remember that the calibre and quantityof the data affect the model's accuracy credit data used in the analysis. Therefore, it's crucial to ensure that the data is clean, relevant, and sufficient for the purpose of credit scoring.The Min-Max normalization technique is an excellent way to standardize the data for use in the machine learning classifier. It rescales the data to a fixed range of values, typically between 0 and 1, which makes it easier to compare the variables and analyze the data.Using the K-NN classifier is also a suitable approach for this problem. It is a non-parametric classification algorithm that assigns a new observation to a class determined by its k-nearest neighbours' majority vote in the feature space.

This method is suitable for small to medium-sized datasets and can handle both numerical and categorical variables.It's also essential to assess the credit scoring model's effectiveness and make sure it is providing the necessary degree of accuracy.Accuracy, precision, recall, and F1 score are a few examples of metrics that can be employed to assess the model's effectiveness.These measures aid in evaluating the model's capability to accurately identify credit defaulters and legitimate clients.Overall,developing a credit scoring model using machine learning classifiers is a promising approach for the banking sector to forecast the status of loan defaulters and enhance the loan-lending procedure.

Keywords: Loan status, the loan-lending procedure, credit score, and min-max normalisation.

4.PAPER NAME: Overdue Prediction of Bank Loans Based on LSTM-SVM

AUTHOR: Xin Li, Xianzhong Long, Guozi Sun, Geng Yang, and Huakang Li

ABSTRACT: It sounds like you are discussing a research study that proposes a new approach to predicting loan default riskcombining various machine learning algorithms. Specifically, study suggests using the LSTM algorithm to analyze user dynamic behavior and SVM algorithm is used to analyse user static data.The static data, such as fundamental user data, bank records, internet habits, Credit card billing records and loan term details, are used as the input for the SVM model, while the LSTM model takes the In order to calculate the chance of behaviour, the user's most recent transaction type from browsing behaviour is used.

Finally, the study computes the average of the two algorithms to arrive at a final prediction of the user's loan default risk. According to the experimental findings, the proposed LSTM-SVM model outperforms more conventional techniques. It is important to note that this is just one study, and further research would be needed to validate these findings and determine the generalizability of the approach. Furthermore, it's crucial to take into account the ethical implications using personal data to make loan risk predictions and ensure that any such models are used in a fair and unbiased manner.

Keywords: LSTM, SVM, and Overdue Prediction for Bank Loans.

5. PAPER NAME: Prediction Defaults for Networked-guarantee Loans

AUTHOR: Dawei Cheng, Zhibin Niu†, Yi Tu and Liqing Zhang

ABSTRACT: It seems that the Chinese government and banks are facing a potential systemic risk due to networked-guaranteed loans. Such loans involve a commitment to pay back debt, and if one of the businesses within the assurance network experiences the financial difficulties,the chance of default may spread and cause a crisis.To address this challenge, your research proposes a model of uneven network risk diffusion that incorporates a data-driven default Distillation model to predict.In the near future, business default risk.In situations where there is no default contagion, your research suggests using a Algorithm for standalone positive weighted k-nearest neighbours (pwkNN) cases.

However, to further improve prediction accuracy, your approach integrates a diffusion model with a data-driven default.You conducted an empirical investigation based on a three-year identified in a loan record from a big commercial bank that your suggested strategy performs better than traditional credit danger methodologies according to AUC. Regulators and stakeholders may find your quantitative risk evaluation approach to have high real-world data prediction performance.Overall, it appears that your research provides valuable contribution to addressing the challenges posed by networked-guaranteed loans the potential for default contagion.

Keywords: Data mining, classification, and positive weighted k-nearest neighbours.

6. PAPER NAME: Personal Credit Rating Using Artificial Intelligence Technology for the National Student Loans

AUTHOR: Jian HU, Zibo, China

ABSTRACT: It is true that national student loans are a type of financial credit service offered by commercial banks, and they often have characteristics of commercial loans. However, traditional credit assessment systems may not be suitable for college students who lack a credit history.In order to address this issue, a backpropagation neural network was developed to assess a college student's credit rating.Among the machine learning algorithms, neural networks are able to learn and adapt to patterns in Consequently, they are suitable for tasks like credit assessment.The algorithm was trained and tested using a set of samples provided by a bank, and the results showed that the neural network was reasonably effective at assessing the personal credit condition of college students.

The maximum error between the network's prediction value and real value was only 2.92, indicating that the neural network was able to accurately predict credit ratings.Overall, the use of neural networks for credit assessment may be a promising approach for evaluating the creditworthiness of college students who lack a credit history. However, it is important to note that additional research and testing may be needed to fully evaluate the effectiveness of this approach.

Keywords: National Student Loans, Back Propagation Neural Networks, Artificial Intelligence, and Credit Rating.

7. PAPER NAME: Dynamic Loan Service Monitoring using Segmented Hidden Markov Models

AUTHOR: -Haengju Lee, Nathan Gnanasambandam, Raj Minhas, and Shi Zha

ABSTRACT: A statistical model called the HMM (Hidden Markov Model) is utilised it to analyze time series of data. In this context, the HMM is used to make the loan service monitoring procedure automated. By using historical payment information for the borrowers, the model is constructed to forecast the likelihood of defaulting soon. The model is dynamic, meaning it continuously changesas more recently realised data is combined with the earlier historical data.The composite data consisting the loan's current status and its days past due for each month. was used to produce Data from the time series sequence is used. Different HMMs are trained throughout the training stage, with one being a HMMs that were paid and those that were in default.

The segmentation of the defaulted data and training each segment independently allows for more accurate monitoring than training a single defaulted HMM.During the prediction stage, two steps are used for each active loan: categorization of the loan and prediction of the probability of a future default frame. If probability exceeds predetermined level,Signal is sent by the monitoring system.The selection of the best threshold level is investigated using precision and recall analysis. This allows for the determination of the best level to use when identifying loans that are at risk of defaulting.Overall, the use of HMMs allows for the automation of the loan service monitoring process, providing a more accurate and efficient way to monitor loans and identify those that are at risk of defaulting.

Keywords: Default monitoring systems, the Hidden Markov Model, and sequences and sequential data analysis are all used to anticipate loan default.

8. PAPER NAME: Azure ML Based Analysis and Prediction Loan Borrowers Creditworthy

AUTHOR: Khaldoon Alshouiliy, Ali Al Ghamdi, Dharma P Agrawal

ABSTRACT: It seems like you are describing a research study that aims to analyze Lending Club datasets and create a model using Azure ML to predict whether clients will repay their loans. The study utilizes the Two Jungle algorithm and Two Decision tree algorithms to evaluate The algorithms' effectiveness was assessed using the metrics Accuracy, Precision, Recall, F1, and AUC. The use of machine learning tools like Azure ML in the field of P2P lending and credit scoring is becoming increasingly popular due to the vast amount of data available.

By analyzing this data, lenders can make better-informed decisions regarding loan approvals and risk assessments.It's important to note that the study collects Lending Club datasets from 2007 to 2018, which may not represent current data. It's also unclear how the datasets were cleaned or preprocessed before uploading them to Azure ML for analysis. Overall, the study's findings can provide useful insights into the effectiveness of the Two Jungle algorithm and Two Decision tree algorithms for credit scoring in the context of P2P lending. However, it's essential to keep in mind the limitations of the data used and the methods employed.

Keywords: Data mining, classification, and positive weighted k-nearest neighbours.

9. PAPER NAME: Loan Prediction by Using Machine Learning Models

AUTHOR: P. Supriya, M. Pavani, N. Saisushma, N. Kumari, and K. Vikas

ABSTRACT: It appears that the study aims to explore the properties of Lending Club datasets and create a machine learning model on Microsoft Azure platform to predict loan repayment. The study used data from 2007 to 2018 and employed the Two Jungle algorithm and Two Decision tree algorithm. The algorithms' effectiveness was assessed using the metrics Accuracy, Precision, Recall, F1, and AUC metrics. Additionally, the study compared its results with those of other researchers in the field of P2P lending, credit scoring, big data, and data analytics. Machine learning algorithms are employed can be beneficial predicting loan defaults, and the Microsoft Azure platform provides a robust environment for developing and deploying machine learning models. The study's use of different algorithms and evaluation metrics indicates a thorough approach to model development and assessment. However, it is unclear how the datasets were cleaned, and whether any feature engineering was performed to improve model performance.

Furthermore, the study's reliance on historical data may limit the model's ability to predict future outcomes accurately, as lending patterns and economic conditions can change over time. Future research could explore the use of more recent data to improve model accuracy and assess the impact of external factors on loan repayment. Overall, the study provides insights into how machine learning is used Machine learning algorithms are employed in credit scoring and highlights the importance of rigorous model evaluation.

Keywords: Distance and attribute weighting using dynamic K-Nearest Neighbour.

10.PAPER NAME: Analysis of Feature Selection and Extraction Algorithm for Loan Data: A Big Data Approach

AUTHOR: Girija Attigeri, Manohara Pai M M*, Radhika M Pai

ABSTRACT: The passage highlights the importance of using algorithms for identifying fraudulent activities in financial institutions. It emphasizes that the quality of data supplied to these algorithms is crucial to their effectiveness. Financial data is gathered from a number of sources, in different formats, making it large and unstructured, and thus, Pre-processing that is parallel and distributed necessary to enhance its quality.The research aims to reduce the dimensions of big financial data while taking feature selection and extraction algorithms intoconsideration. The principal feature analysis-based feature extraction and transformation algorithm's impact on financial data is analyzed, and the reduction in dimension is examined in various classification algorithms for financial loans.

The study uses Spark notebook to implement the algorithm in a distributed and parallel manner on thea cloud computing platform from IBM. The findings indicate that reducing the features greatly enhances execution time without affectingaccuracy. The implication of this is that using algorithms that incorporate featureselection and extraction can help financial institutions identify fraudulent activities efficiently while minimizing execution time.

Keywords: Big financial data, feature extraction and selection, support vector, and classification.

11.PAPER NAME: Credit Collectibility Prediction of Debtor Candidate Using Dynamic K-Nearest Neighbor Algorithm and Distance and Attribute Weighted

AUTHOR: Tiara Fajrin, Ragil Saputra, Indra Waspada

ABSTRACT: It seems that Jepara Branch of BPR Bank is facing a bad loan problem when providing loans to MSME activists, and they need an application to help them anticipate the Receivability of the Debtor applicants to make the issue smaller To address this issue, a data mining categorization algorithm, specificallythe Algorithm dynamic K-Nearest Neighbour and attribute-weighted distance, was used in this study's application. This algorithm adds attribute and distance weights to the k-Nearest Neighbor algorithm to provide an output that can be used as a second opinion when deciding whether to approve or turn away a borrower.The deciding elements to predict the outcome of the loan application are the five Cs (character, capacity, capital, collateral,andeconomic condition), monthly income, other debt situations, dependents, age, commodity kind, and business status.

The study found that using attribute weighting in the algorithm improves its Compared to the one that doesn't use it, this one has higherprecision, recall, and accuracy.Correlation Attribute Evaluation Modifications-assessed attributes' relative relevance also increased the domain expert's recall value of 54.35 attributes. These results suggest that the application of the Algorithm using K-Nearest Neighbour, Distance, and Attribute Weights, along with that identified factors, can help BPR Bank Jepara Artha anticipate the collectibility of debtor applicants and reduce the bad loan problem.

Keywords: Cross-validation, attribute and distance weighting, dynamic K-Nearest Neighbour, classification, and data mining are all used in loan prediction.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

3.1 UML DIAGRAMS

3.1.1 Architecture Diagram

An architecture diagram is a graphical representation of the various components and interactions within a system or application. It can take many different forms depending on the system being modeled and the purpose of the diagram, but in general, it should provide a clear and concise overview of the system's structure and functionality.

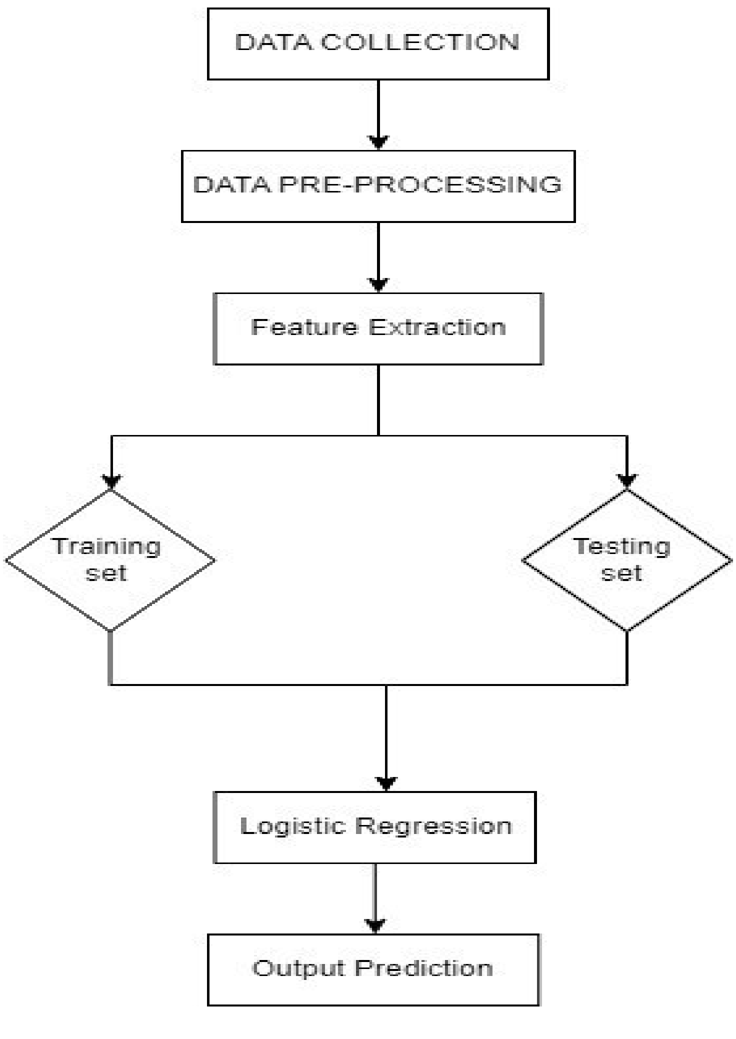


FIG - 3.1(Architecture Diagram)

3.1.2 Data Flow Diagram

The movement of data is shown in a data flow diagram (DFD) graphically inside from a system.It serves to model the processes, data stores, and flows of information inside of a system or organization.DFDs can be used to model any system or process that involves data, from a simple business process to a complex information system.They are useful for identifying inefficiencies,redundancies,and opportunities for improvement in a system, as well as for communicating the flow of data to stakeholders .

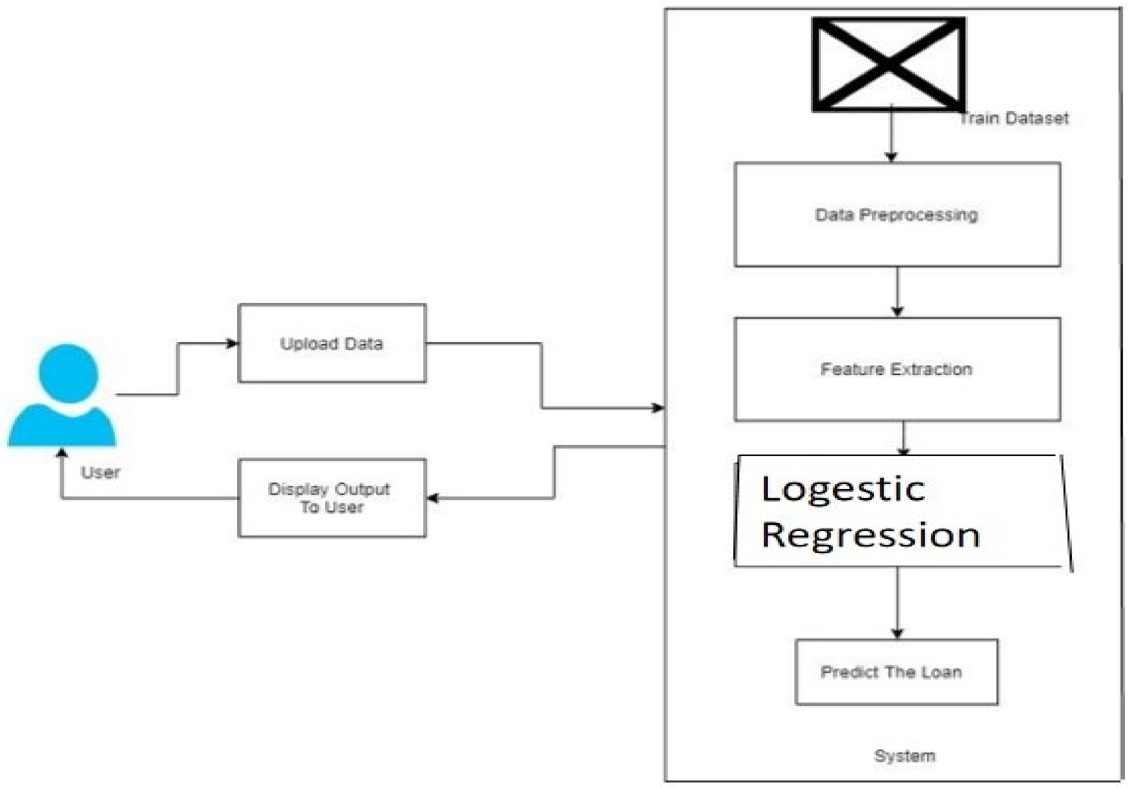


FIG - 3.2(Data Flow Diagram)

CHAPTER 4
METHODOLOGY

4.1 Proposed Modules

- Data Collection
- Data Pre-Processing
- Feature Extraction
- Spliting into training set
- Spliting into testing set
- Model Evaluation
- Prediction

4.1.1 Data Collection

Depending Depending on the project type we want to create, we can construct an IoT system to collect information from multiple sensors to develop a real-time machine learning project. The data set may come from many different sources, such as data from files, databases, sensors, and other sources can be used, but not immediately for analysis.since it may contain sizable volumes of data with errors, abnormally high values, disorganised language, or noisy data.Thus, To solve this issue, data preprocessing is performed.Additionally, we can benefit from some of the free internet data sets.The two The UCI Machine Learning Repository and Kaggle are the two most well-known repositories for building machine learning models. Kaggle is among the most well-known websites for training machine learning algorithms. They also hold contests where anyone may compete and demonstrate their machine learning abilities.

Data-collection is the process of gathering and acquiring data from various sources to be used for analysis or machine learning. The success of any data analysis or The effectiveness and volume of the data obtained are key factors in machine learning projects.

Here are some common methods of data collection:

1. Surveys: By posing questions to a certain set of people to learn their thoughts or behaviours, surveys are a popular method of gathering data.

2. Interviews: By posing open-ended or closed-ended questions to individuals or groups, interviews are one way to gather data.

3.Observations: By seeing and documenting actions, events, and behaviours, observers gather data.

4. Experiments: Experiments entail gathering data while modifying variables in a controlled environment to see how they affect the result.

5. Secondary sources: Secondary sources involve collecting data from existing sources such as databases, reports, or publications.

6. Sensors: Sensors are used to collect data from physical or digital environments, such as temperature, pressure, or location data.

7. Web scraping: Web scraping involves collecting data from websites by extracting relevant information using software tools.

The data collected may be either semi-structured, unstructured, or both. Detailed information is organized and defined by a set of rules, while unstructured data is not organized and is free-form. Semi-structured data has elements of both structured and unstructured data.

Making sure the information is accurate is crucial collected is accurate, relevant, and unbiased. The data collection process should be carefully planned and executed to avoid errors and inconsistencies that could affect the results of the analysis or machine learning models.

4.1.2 Data Pre-Processing

Among the most significant machine learning Data pre-processing is a procedure. It is the most important step to do in order to create machine learning models that are more accurate.The 80/20 principle is applicable to machine learning. For each data scientist, processing data should take up 80% of their work, and analysis should take up 20% of their time. Purifying unclean data or converting data gathered Data pre-processing is the conversion of raw data from the real world into clean data sets.Or, to put it another way, data is collected from many sources in a raw format, which hinders the analysis of data.The stage of the process where numerous measures are done to transform the data into a manageable, clean data collection is known as data pre-processing.Data pre-processing, as we all know, is the process of removing errors from raw data so that the model may be trained.Data pre-processing is thus certainly required if we want to achieve decent outcomes from the model that was used in deep learning and machine learning projects.

Data preparation which entails transforming unformatted data into a format that can be used to train computer models, is a vital stage in data analysis and machine learning. Making ensuring the data is correct, comprehensive, consistent, and pertinent for analysis or modelling is the aim of data pre-processing.

Here are a few typical methods used in data pre-processing:

1. Data cleaning: It entails locating and fixing data flaws or anomalies such missing values, duplicate entries, and outliers.

2. Data normalization: To prevent biases in the research or modelling, it entails scaling the data to a standard range, such as 0 to 1.

3. Data transformation: It involves transforming the data to a more suitable format, such as converting categorical variables to numerical variables, or applying log or square root transformations to skewed data.

4. Data integration: It involves combining multiple data sources into a single dataset for analysis or modeling.

5. Data reduction: It entails either applying Reduce the number of components using principal component analysis (PCA) variables or choosing a subset of pertinent features to minimise the dimensionality of the data.

6.Data splitting: It involves segmenting the data into testing, validation, and training sets to assess the effectiveness of the model on untested data.

7. Data imputation: It involves filling missing values using statistical methods such as mean or median imputation or regression-based imputation.

Depending on the nature and quality of the data, the goals of the analysis or modelling, and the computational resources available, these techniques are often applied iteratively and in a precise order. The quality and dependability of the outcomes from data analysis or machine learning models can be significantly impacted by how well the data is pre-processed.

4.1.3 Feature Extraction

The process of selecting and modifying a portion of the input data into features that are more useful and pertinent for a particular machine learning task is known as feature extraction. This method is frequently used to simplify the input data's dimensions and draw attention to the aspects of the data that are crucial to solving a particular issue.

In general, feature extraction involves the following steps:

1. Data Preprocessing: This entails preparing the data for analysis by cleaning it.This can entail filling in any missing numbers, normalising the data, and formatting the data appropriately.

2. Feature Selection: In order to do this, a subset of the features from the initial data must be chosen. Different methods, including feature ranking, backward elimination, and correlation analysis, can be used to accomplish this.

3. Feature Transformation: This entails repurposing the chosen features into a fresh set of features that are more instructive and applicable to the current task. the Principal component analysis (PCA), linear discriminant analysis (LDA), and non-linear feature extraction approaches can all be used for this.

The type of data and the particular machine learning problem determine the best feature extraction technique to use. While some tasks might only need a straightforward feature selection step, others might call for a more involved feature extraction strategy.Feature extraction generally aims to provide a set of features that can capture the key aspects of the data and enhance.The effectiveness of the computer learning paradigm.Our main goal is to train the most effective model using the pre-processed data.

4.1.4 Splitting into Training Set

We originally split a training model with three components: "Training data," "Validation data," & "Testing data".The "training data set" is utilised For the classifier's training, the "validation set"to adjust the settings, and the "test data set" to assess it's the classifier effectiveness. It's critical to remember that when the classifier is being trained, justThe validation and/or training set is accessible. The classifier's training process cannot make use of the test set of data. The test set won't be available prior to evaluating the classifier.The practise set is body of information that teaches the computer how to handle data. Machine learning performs the training step using algorithms.

A crucial stageThe information is split into training and testing sets in machine learning. It is essential to assess the model's performance using data that was not utilised for training.A portion of the dataset used to train the machine learning model is called the drill set. In order to learn how the input variables and the output variable are related, this subset is applied to fit the parameters of the model.A variety of methods, including random sampling, stratified sampling, and time-series splitting, can be used using the dataset, produce training and testing sets. The most popular method is using the dataset, produce training and testing sets. At random, with a usual divided into 30% for testing and 70% for training.It is crucial to make sure the split accurately reflects the overall properties of the dataset and that the model is not either over- or underfitting the data. Another method for ensuring that the model is not overfitting or underfitting the data is cross-validation. Cross-validation uses a dataset that is divided into several subsets, and each subset is used to train and test the model and assess its performance.

To summarise, it is essential for machine learning to divide the dataset into sets for testing and training. It is vital to ensure that the model can generalise to new data in an adequate manner and that it does not overfit or underfit the data.

4.1.5 Splitting into Testing Set

Cross-validation is a machine learning technique used in frequently used to evaluate a machine learning model's performance on untrained data. Using an unknown set of data from the training set, a classifier's parameters are changed. After the data has been divided into the three aforementioned categories, we can start the training process.a lesson set and a test (or validation) set are used in a data set to build a model and check the model's accuracy, respectively. None of the data points from the training set are present in the test (validation) set.For each iteration, a data set is frequently divided into a training set, a validation set, and a test set. (However, some people prefer to this as a "test set" instead).The model can employ whichever of the models we decide upon in step 3 point 3. After the model has been trained, we can use testing data to use it to forecast the future or unrecognised data. We can then generate a confusion matrix to show how well our model has been trained once this is done. A matrix of ambiguity comprises four columns dimensions: "True Positives," "True Negatives," "False Positives," and "False Negatives." We would like to collect more data for the True Positives and True Negatives to construct a more accurate model. The number of classes has a direct relationship with the size of the Confusion matrix.positive real-world examples These are instances that we correctly predicted that the result would be TRUE.On the other hand, the testing set is a portion of the dataset used to evaluate the machine learning model's performance performed. Data that it has never seen before, this subset is used to assess the performance of the model in terms of accuracy, precision, recall, and other metrics.

A critical step in assessing a Performance of a machine learning model involves creating a testing set from a dataset. Using the testing set, the model's generalizability to brand-new, untested data is assessed.Usually, two sets: a practise set and a test set are created from the dataset. In a common split ratio, the training set contains 70% or 80% of the data, while the testing set has the remaining 20% or less remaining 30% or 20%, respectively.

Examples in the testing set should be comparable to those in the training set and should be representative of the entire dataset. This is essential to ensure that the model's performance on the testing set is an accurate representation of its performance on brand-new, untested data.It's crucial to remember that the testing set should only be used to assess how well the final model performs. It shouldn't be utilised to train the model or select the features that should be included in the model.

In conclusion, dividing the dataset into a testing set is crucial for assessing performance of a machine learning model. The testing set should only be used to evaluate how well the final model performs, and it should be representative of the entire dataset.

Actual negatives : We anticipated FALSE, and the outcome is consistent what we expected.

Positive errors: The projected result is FALSE even though we predicted TRUE.

Untrue negatives : Although we expected FALSE, the actual outcome is TRUE.

The model's accuracy can be evaluated using the confusion matrix as well.

Accuracy equals (True Negatives plus True Positives) divided by the amount of classes overall.

4.1.6 Model Evaluation

A step in the model development process is model assessment. searching for the model that most closely matches Our data and future performance predictions of the model are helpful. The model's hyper-parameters could be changed to increase accuracy. The confusion matrix may also be examined in order tothe ratio of real positives to real negatives should be increased.A crucial phase in machine learning is model evaluation, which involves evaluating how well a trained model performs on a test dataset. In order to assess a model's generalizability to fresh, untested data and to spot any flaws, such as over- or underfitting, it must first be evaluated.

Depending on the issue's nature and its severity data's characteristics, a machine learning model's performance can be assessed using a variety of performance indicators.Several typical metrics include:

1.Accuracy: determines how many of the model's predictions were accurate.

2. Precision: the percentage of accurate positive predictions among all positive model predictions.

3. Recall: calculates the percentage of samples in the dataset that are actually positive that are also true positives.

4. F1-score: a precision and recall weighted harmonic mean that equally weights both parameters.

5. The ROC curve's area under it (AUC-ROC) measures how well the model can distinguish between positive and negative samples.

6. The average squared difference between the values is known as the mean squared error (MSE) that were anticipated and the actual ones.

7. R-squared (R2): the model's ability to forecast outcomes and how well the model fits the data.

The test dataset is used, and the selected metrics are computed, to assess the model's performance. The model is then modified in accordance with how these measures' values compare to the targeted performance thresholds. The best-performing model may be chosen for deployment after several models have been trained and tested.

In conclusion, evaluating thow well a model performs on a test dataset using a variety of performance indicators is a critical component of machine learning. The evaluation's findings are used to pinpoint any model flaws and make the necessary changes to enhance the model's performance.

4.1.7 Prediction

By implementing our model, we give it the ability to work in a practical environment, facilitate normal discourse, and be practical.Make sure your models are functioning as anticipated and delivering correct results by keeping an eye on them. To evaluate performance, it entails comparing model predictions to real data sets.Making predictions on fresh, unforeseen data involves employing a trained machine learning model. Prediction is the process of using the model to create precise predictions on previously unknown fresh data.The trained model is run over the input data to produce an output, which is then used to generate a forecast.

The quality and quantity of the training data, the model's complexity, and the hyperparameters used to train the model are some of the variables that affect how accurately the model predicts the future. The goal is to develop a model that can accurately and confidently anticipate new data.Prediction may occasionally be used to guide decisions or actions based on model output. For instance, the model's prediction in a medical diagnosis application can be used to suggest a course of therapy or send the patient to an expert.It is crucial to remember that prediction is only as accurate as the training data's quality and the model's capacity for generalisation to fresh data. As a result, it is crucial to assess the model's effectiveness on a test dataset before applying it to prediction.

In conclusion, prediction is the process of making predictions on fresh, unforeseen data using a trained machine learning model. the model's complexity, the calibre of data from training, and the hyperparameters used preparing the model all affect how accurate the predictions are before utilising the model for prediction,its performance should be assessed using a test dataset.

4.2 Logistic Regression Algorithm

A binary one or more independent variables and a dependent variable can be statistically modelled using logistic regression. In this specific kind of regression analysis, the data are modelled using a logistic function.The probability that the dependent variable will have a specific value (typically 0 or 1), based on the values of the independent variables.In other words, logistic regression is a method used to predict the likelihood that an event will occur based on the supplied variable values. In binary classification tasks, when attempting to predict which of two classes an observation belongs to, It is frequently used in statistical analysis and machine learning.A logistic curve—an S-shaped curve created using logistic regression—shows how the dependent and independent variables are related. Because the logistic function used in logistic regression ensures that the predicted probabilities are always between 0 and 1, it is excellent for binary classification tasks.

When attempting to forecast a binary outcome (0 or 1) for binary classification issues, a common statistical strategy is logistic regression. Because it employs a logistic function to simulate the probability of the binary outcome, the algorithm is known as logistic regression.Based on the values of the input features, the logistic regression method calculates the likelihood of the binary outcome. Any real-valued function can be transformed using the logistic function, also known as the sigmoid function quantity into a probability that can be expressed as a value between 0 and 1.

In the logistic function, z is the linear combination of the input characteristics and their respective weights, and (z) = 1 / (1 + e-z) is the definition of the logistic function.

z = w0 + w1x1 + w2x2 + ... + wnxn

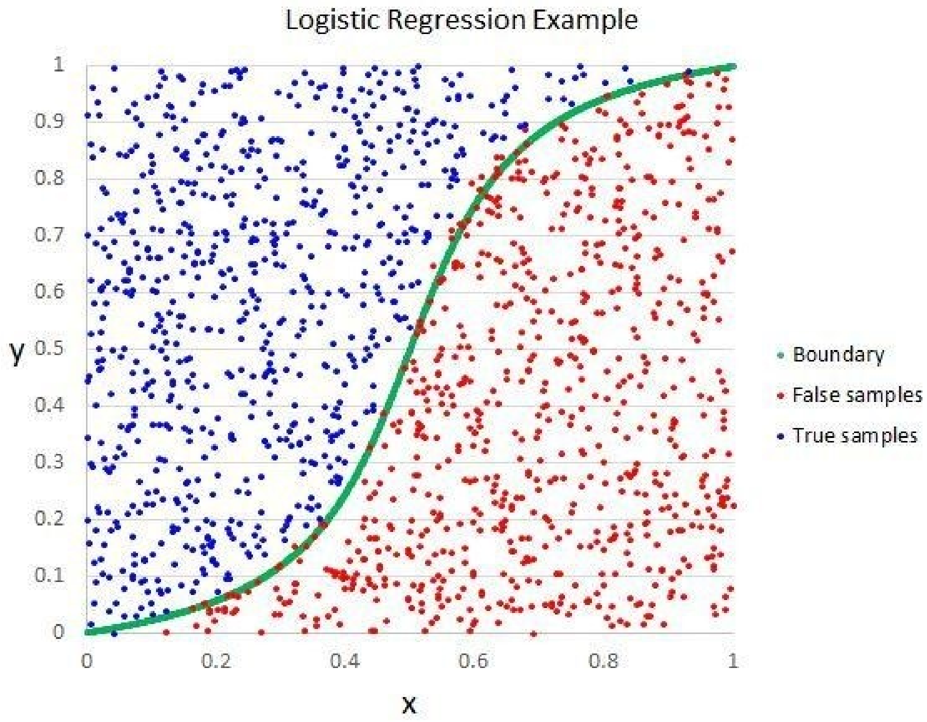


Fig - 4.2(Logistic Regression)

With the use of a maximum likelihood estimation technique, the weights are learned from the training data. The objective is to identify the weights that, given the input attributes, increase the likelihood of the observed outcomes.The logistic function can be used to forecast the likelihood of the binary outcome for brand-new, untested data once the weights have been learned. To transform the probability into a binary forecast, a threshold can be set.Numerous benefits of the logistic regression algorithm include its readability, interpretability, and capacity for handling both linearly and nonlinearly separable datasets. On datasets with high-dimensional or highly correlated features, it might not work as well.

In conclusion, logistic regression is a popular approach for issues involving binary categorization. A logistic function is used to estimate the binary outcome's probability, and maximum likelihood estimation is applied to the training data to learn the weights. The approach can handle datasets that can be separated linearly and nonlinearly, and it is easy to understand.

4.3 Implementation

There are two sections to the dataset: test and train. To improve accuracy, we used training and testing ratios: 70:30, and 80:20. Then in all of ratio the data goes into data preprocessing It addresses the data must be altered or encoded in order for a device to quickly parse it. The programme should immediately identify and comprehend the model must take into account the properties of the data in order to produce precise and accurate predictions. Unprocessed data are converted into manageable numerical features using the feature extraction procedure, which keeps the integrity of the original data set intact. It produces better outcomes than usingML on only raw data. The data is subsequently passed on to the ML algorithms that have been developed. They are trained first, then tested to determine the correctness of each method.

Here is a straightforward scikit-learn implementation of the logistic regression algorithm in Python:

Python: Import the necessary libraries.

import pandas as pd from sklearn import train_test_split import.model_selection

from SkLearn import.linear_model Logisticsklearn.import metrics accuracy_score regression

Dataset = "dataset.csv" read by pd # Activate the dataset.

Create training and test sets from the dataset.

X_train, X_test, Y_train, and Y_test (test) = Train_test_split_size=0.2, random_state=42, data.iloc[:,:-1], data.iloc[:, -1])

Construct the Using the Fit the training data model with a logistic regression model to it using the LogisticRegression() model.fitting X_train and Y_train)

Use the test data for forecasting (y_pred = model).predict(X_test)

Determine the model accuracy using the formula score of correctness (y_test, y_pred).

accuracy = print('Accuracy:', accuracy) '''

In this implementation, we first import the necessary libraries, including pandas for loading and handling the dataset, scikit-learn for putting the logistic regression technique into practise, and accuracy_score for assessing the model's performance.The dataset is then loaded and divided Using the scikit-learn 'train_test_split' tool, divide the data sets for testing and training.Next, apply the adjust method, we make a specimen of the LogisticRegression class and adjust it for training set of data.

The 'predict' method is then employed to produce forecasts based on the testing data, and the "accuracy"_score' function is used to determine the model's accuracy.It should be noted that the logistic regression algorithm can be tailored and tuned in a variety of ways to enhance performance on a particular dataset.

CHAPTER 5

5 .CODING AND TESTING

Fig.5.1 - Importing Libraries

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
#%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
#from google.colab import files
from sklearn.model_selection import StratifiedKFold
from sklearn import metrics
#from google.colab import files
from sklearn.tree import  DecisionTreeClassifier
from sklearn.metrics import accuracy_score, f1_score
import warnings
warnings.filterwarnings("ignore")
```

```python
#1. pip install markupsafe==2.0.1
#   pip install werkzeug==2.0.3
#2. pip install jinja2<3.1.0
#   from markupsafe import escape
#   pip install Flask==2.1.0


import flask
import pickle
import pandas as pd
import numpy as np
from markupsafe import escape


from sklearn.preprocessing import StandardScaler
```

Fig.5.2 - Checking Dataset Labels

```
loan_data  = pd.read_csv("Data/loan_train.csv",index_col=False)
#loan_data = loan_data.drop(['Unnamed: 0'], axis=1)
loan_data.head()

test_data = pd.read_csv('Data/loan_test.csv')
test_data.head()

loan_data.shape

test_data.shape

loan_data['Loan_Status'].value_counts()

loan_data['Loan_Status'].value_counts(normalize=True)

loan_data['Loan_Status'].value_counts().plot.bar()

loan_data['Gender'].value_counts(normalize=True).plot.bar(title='Gender')
plt.show()
loan_data['Married'].value_counts(normalize=True).plot.bar(title='Married')
plt.show()
loan_data['Self_Employed'].value_counts(normalize=True).plot.bar(title='Self_Employed')
plt.show()
loan_data['Credit_History'].value_counts(normalize=True).plot.bar(title='Credit_History')
plt.show()

loan_data['Dependents'].value_counts(normalize=True).plot.bar( title='Dependents')
plt.show()
loan_data['Education'].value_counts(normalize=True).plot.bar(title='Education')
plt.show()
loan_data['Property_Area'].value_counts(normalize=True).plot.bar(title='Property_Area')
plt.show()
```

Fig.5.3 - Checking The Data

```
loan_data['Gender'].fillna(loan_data['Gender'].mode()[0], inplace=True)
loan_data['Married'].fillna(loan_data['Married'].mode()[0], inplace=True)
loan_data['Dependents'].fillna(loan_data['Dependents'].mode()[0], inplace=True)
loan_data['Self_Employed'].fillna(loan_data['Self_Employed'].mode()[0], inplace=True)
loan_data['Credit_History'].fillna(loan_data['Credit_History'].mode()[0], inplace=True)

loan_data['Loan_Amount_Term'].value_counts()

loan_data['Loan_Amount_Term'].fillna(loan_data['Loan_Amount_Term'].mode()[0], inplace=True)

loan_data['LoanAmount'].fillna(loan_data['LoanAmount'].median(), inplace=True)

loan_data.isnull().sum()
```

Fig.5.4 - Loan Application Form

```html
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

    <!-- OG Meta Tags to improve the way the post looks when you share the page on LinkedIn, Facebook, Google+ -->
    <meta property="og:site_name" content="" /> <!-- website name -->
    <meta property="og:site" content="" /> <!-- website link -->
    <meta property="og:title" content="" /> <!-- title shown in the actual shared post -->
    <meta property="og:description" content="" /> <!-- description shown in the actual shared post -->
    <meta property="og:image" content="" /> <!-- image link, make sure it's jpg -->
    <meta property="og:url" content="" /> <!-- where do you want your post to link to -->
    <meta property="og:type" content="article" />

    <!-- Website Title -->
    <title>Loan Applicantion Form</title>
    <!-- Styles -->
    <link href="https://fonts.googleapis.com/css?family=Montserrat:500,700&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="https://fonts.googleapis.com/css?family=Open+Sans:400,400i,600&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="{{url_for('static', filename='css/styles.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/fontawesome-all.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/swiper.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/magnific-popup.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/bootstrap.css')}}" rel="stylesheet">


    <!-- Favicon  -->
    <link rel="icon" href="{{url_for('static', filename='images/logo.svg')}}">
</head>
<style>
    body {
        background-image: url({{ url_for('static', filename='loan_app.jpg')}});
        background-repeat: no-repeat;
        background-attachment: fixed;
        background-size: cover;
    }
<body data-spy="scroll" data-target=".fixed-top">

    <!-- Preloader -->
    <div class="spinner-wrapper">
        <div class="spinner">
            <div class="bounce1"></div>
            <div class="bounce2"></div>
            <div class="bounce3"></div>
        </div>
    </div>
    <!-- end of preloader -->


    <!-- Navbar -->
    <nav class="navbar navbar-expand-md navbar-dark navbar-custom fixed-top">
        <!-- Text Logo - Use this if you don't have a graphic logo -->
        <!-- <a class="navbar-brand logo-text page-scroll" href="index.html">Aria</a> -->

        <!-- Image Logo -->
        <a style="color:white;font-size:20px" href="/">Home</a>


        <!-- Mobile Menu Toggle Button -->
        <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarsExampleDefault"
            aria-controls="navbarsExampleDefault" aria-expanded="false" aria-label="Toggle navigation">
            <span class="navbar-toggler-awesome fas fa-bars"></span>
            <span class="navbar-toggler-awesome fas fa-times"></span>
        </button>
```

Fig.5.5 - Sytle.css

```css
body,
html {
    width: 100%;
    height: 100%;
}

body, p {
    color: #787976;
    font: 400 1rem/1.5625rem "Open Sans", sans-serif;
}

.p-large {
    font: 400 1.125rem/1.625rem "Open Sans", sans-serif;
}

.p-small {
    font: 400 0.875rem/1.375rem "Open Sans", sans-serif;
}

.p-heading {
    margin-bottom: 3.5rem;
    text-align: center;
}

.li-space-lg li {
    margin-bottom: 0.375rem;
}

.indent {
    padding-left: 1.25rem;
}

h1 {
    color: #484a46;
    font: 700 2.5rem/3rem "Montserrat", sans-serif;
}

h2 {
    color: #484a46;
    font: 700 1.75rem/2.125rem "Montserrat", sans-serif;
```

```css
.btn-outline-lg {
    display: inline-block;
    padding: 1.375rem 2.125rem 1.375rem 2.125rem;
    border: 0.125rem solid #787976;
    border-radius: 0.25rem;
    background-color: transparent;
    color: #787976;
    font: 700 0.75rem/0 "Montserrat", sans-serif;
    text-decoration: none;
    transition: all 0.2s;
}
```

Fig.5.6 - Training and Testing the Dataset

```
loan_data['LoanAmount_log']=np.log(loan_data['LoanAmount'])
loan_data['LoanAmount_log'].hist(bins=20)
test_data['LoanAmount_log']=np.log(test_data['LoanAmount'])

loan_data=loan_data.drop('Loan_ID',axis=1)
test_data=test_data.drop('Loan_ID',axis=1)

X = loan_data.drop('Loan_Status',axis=1)
y = loan_data.Loan_Status

train = loan_data.copy()
test = test_data.copy()

X = pd.get_dummies(X)
train=pd.get_dummies(train)
test=pd.get_dummies(test)

x_train, x_valid, y_train, y_valid = train_test_split(X,y, test_size=0.3)
```

Fig.5.7 - Logistic Regression Algorithm

```
X = loan_data.drop('Loan_Status',axis=1)
y = loan_data.Loan_Status

train = loan_data.copy()
test = test_data.copy()

X = pd.get_dummies(X)
train=pd.get_dummies(train)
test=pd.get_dummies(test)

x_train, x_valid, y_train, y_valid = train_test_split(X,y, test_size=0.3)
model = LogisticRegression()
model.fit(x_train, y_train)
```

Fig.5.8 - Accuracy Score

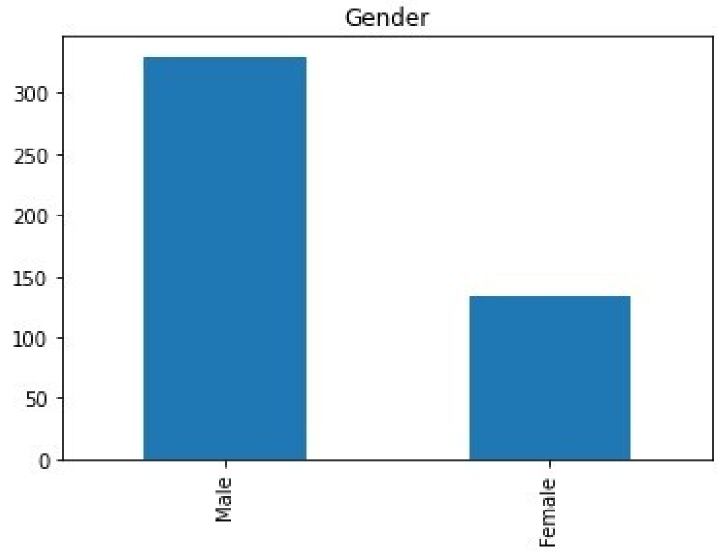
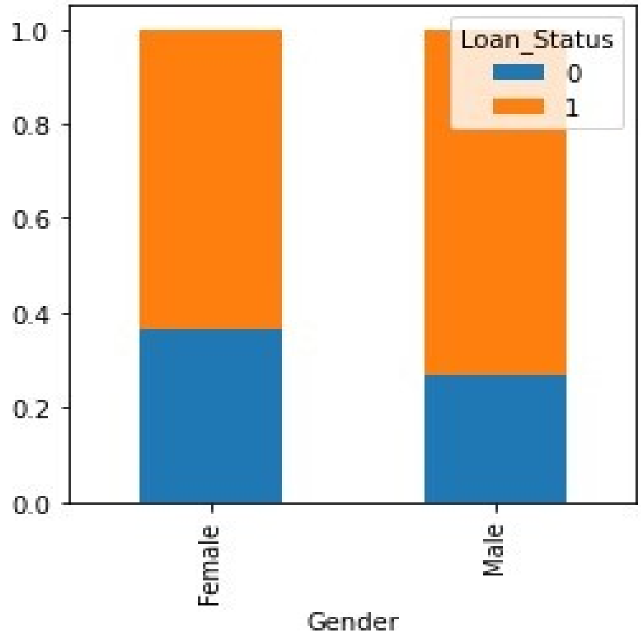
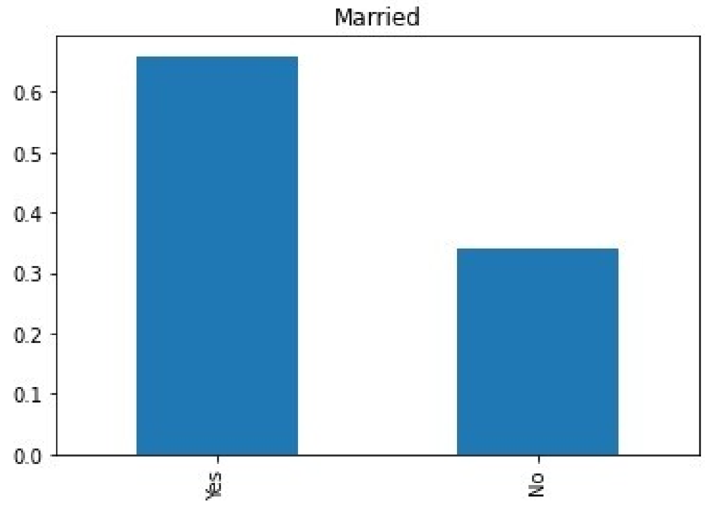
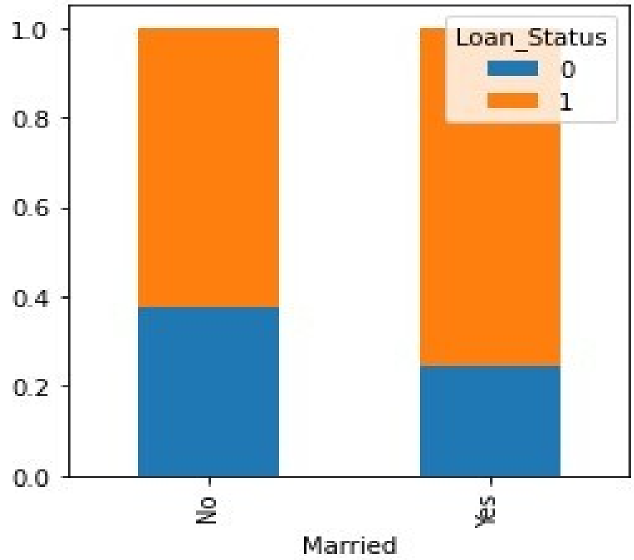
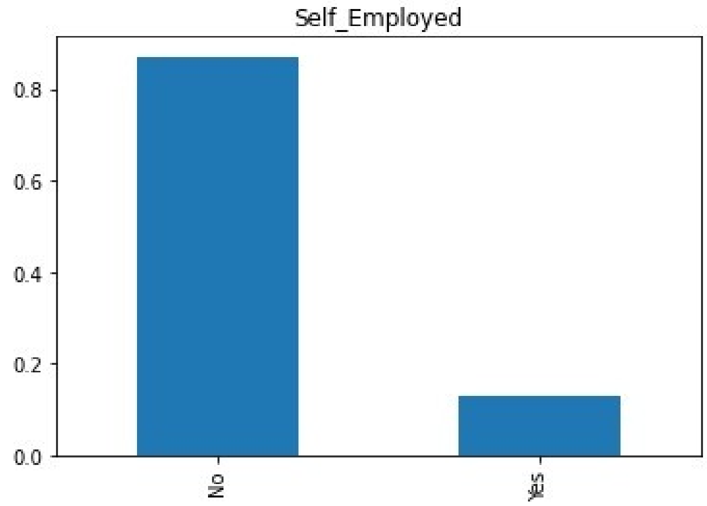
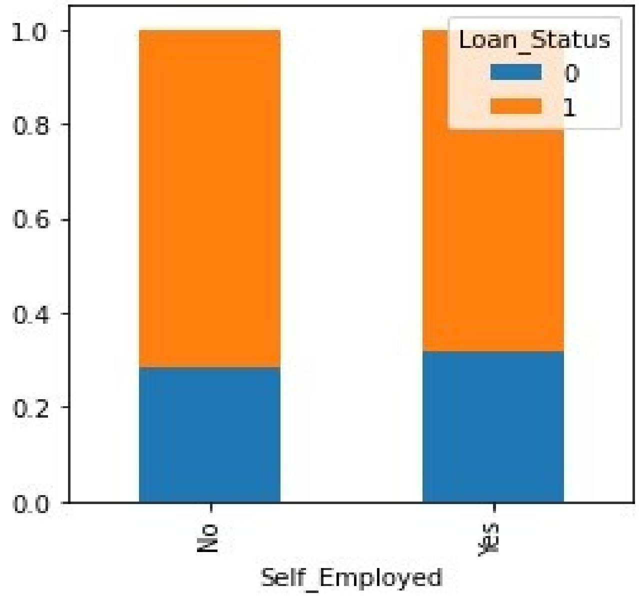
```
train = loan_data.copy()
test = test_data.copy()

X = pd.get_dummies(X)
train=pd.get_dummies(train)
test=pd.get_dummies(test)

x_train, x_valid, y_train, y_valid = train_test_split(X,y, test_size=0.3)
model = LogisticRegression()
model.fit(x_train, y_train)

pred_cv = model.predict(x_valid)
print('Model Accuracy = ', accuracy_score(y_valid,pred_cv))
print('Model F1-Score = ', f1_score(y_valid,pred_cv))
```

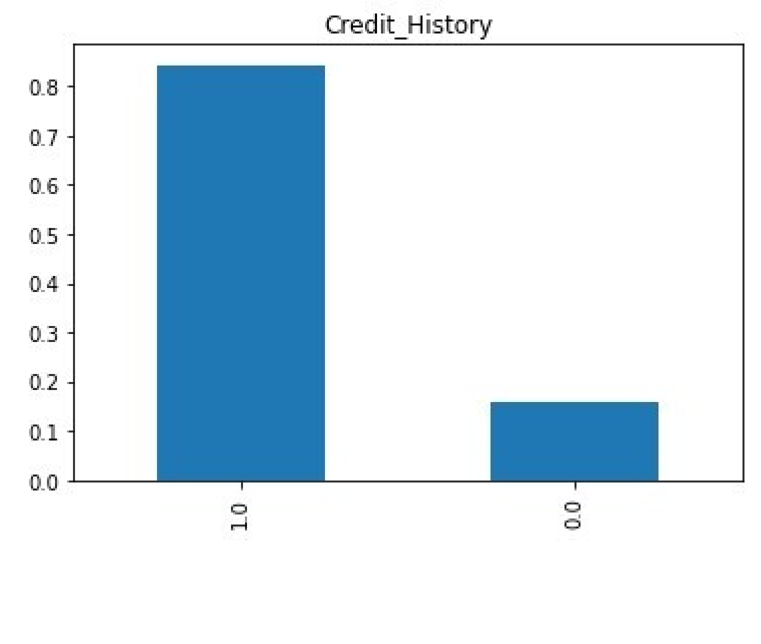
```
Python 3.9.13 (main, Aug 25 2022, 23:51:50) [MSC v.1916 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 7.31.1 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/user/OneDrive/Documents/Desktop/LOAN-APPROVAL-main/LOAN-APPROVAL-main/LOAN-APPROVAL-main/model.py',
wdir='C:/Users/user/OneDrive/Documents/Desktop/LOAN-APPROVAL-main/LOAN-APPROVAL-main/LOAN-APPROVAL-main')

                                   Warning
Figures now render in the Plots pane by default. To make them also appear inline in the Console, uncheck "Mute
Inline Plotting" under the Plots pane options menu.
 Model Accuracy =  0.8428571428571429
Model F1-Score =  0.9017857142857143

In [2]:
```

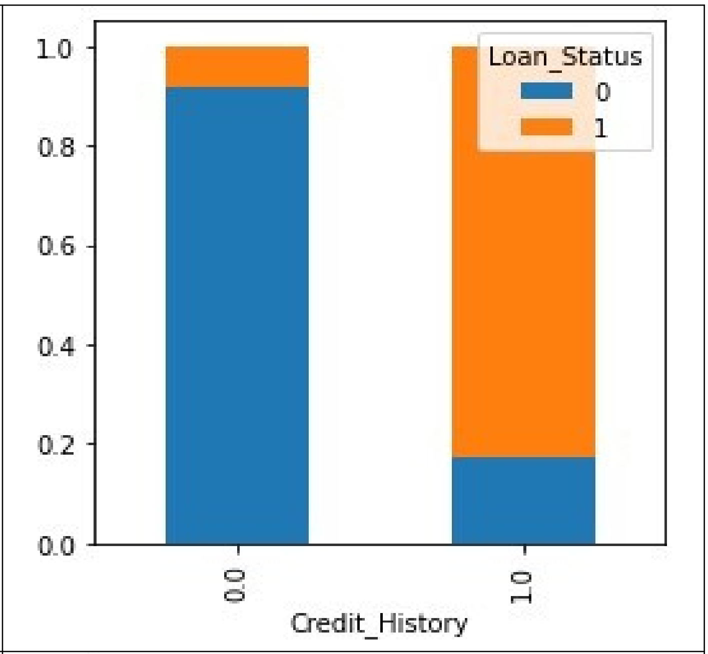
Fig.5.9 - Python to Frontend Linking

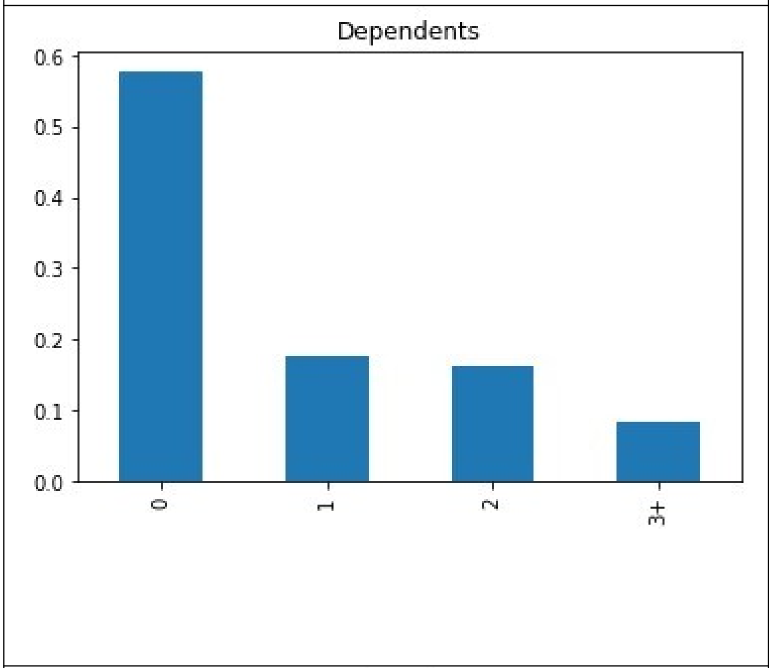
```python
app = flask.Flask(__name__, template_folder='templates')
@app.route('/')
def main():
    return (flask.render_template('index.html'))

@app.route('/report')
def report():
    return (flask.render_template('report.html'))

@app.route('/jointreport')
def jointreport():
    return (flask.render_template('jointreport.html'))


@app.route("/Loan_Application", methods=['GET', 'POST'])
def Loan_Application():

    if flask.request.method == 'GET':
        return (flask.render_template('Loan_Application.html'))

    if flask.request.method =='POST':

        #get input
        #gender as string
        genders_type = flask.request.form['genders_type']
        #marriage status as boolean YES: 1 , NO: 0
        marital_status = flask.request.form['marital_status']
        #Dependents: No. of people dependent on the applicant (0,1,2,3+)
        dependents = flask.request.form['dependents']

        #dependents = dependents_to_int[dependents.upper()]

        #education status as boolean Graduated, Not graduated.
        education_status = flask.request.form['education_status']
        #Self_Employed: If the applicant is self-employed or not (Yes, No)
        self_employment = flask.request.form['self_employment']
        #Applicant Income
        applicantIncome = float(flask.request.form['applicantIncome'])
        #Co-Applicant Income
        coapplicantIncome = float(flask.request.form['coapplicantIncome'])
```

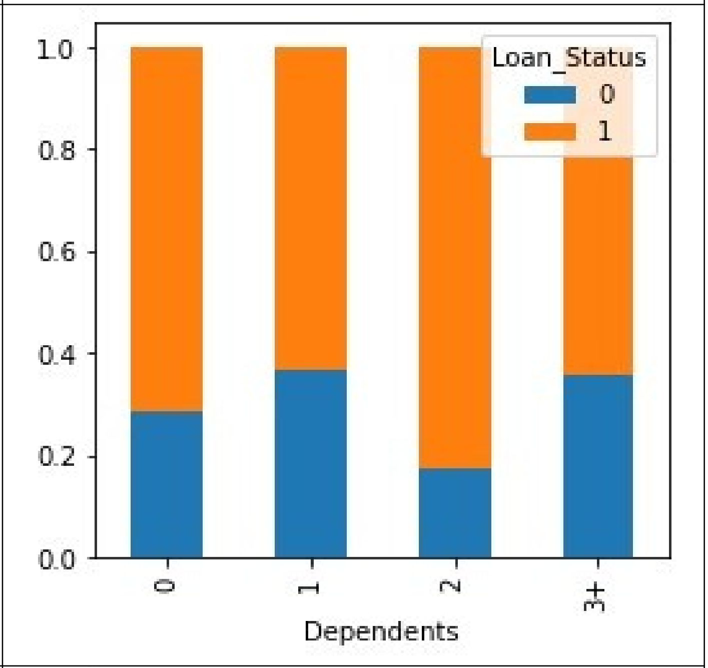
Fig.5.10 - Output Plots

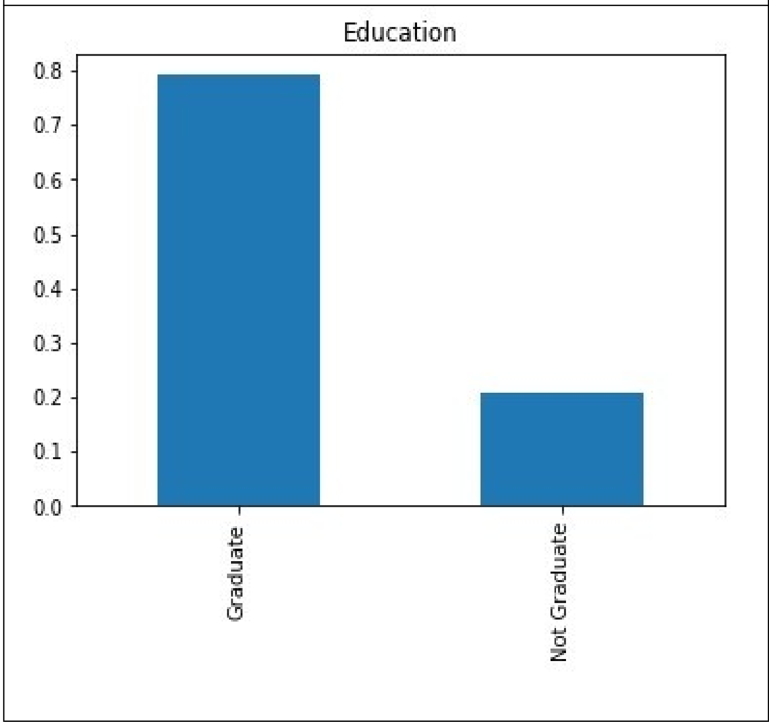
INITIAL	FINAL
	
	
	

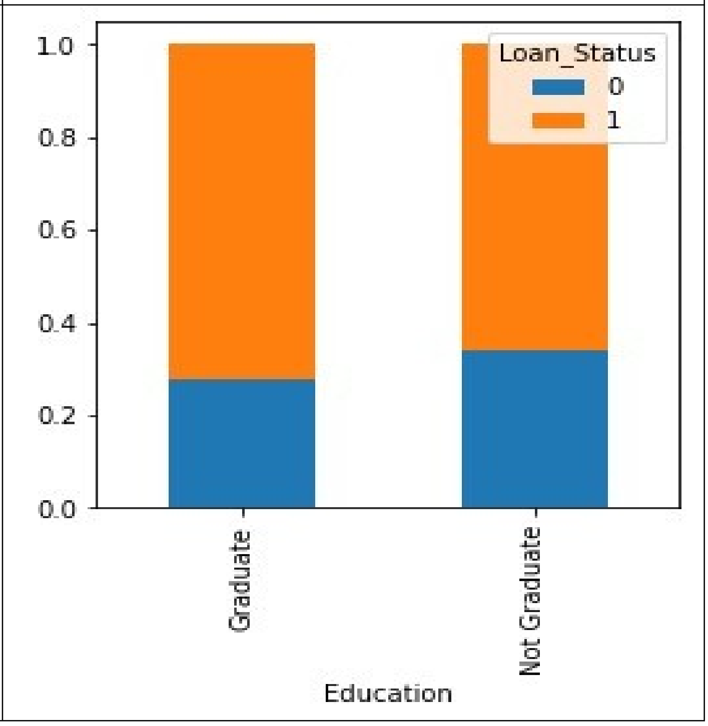

Credit_History
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1.0
0.0

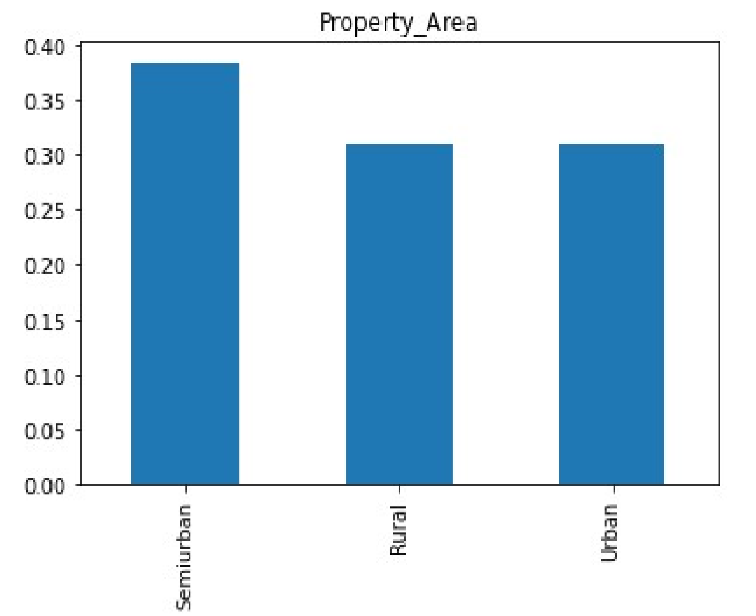

Loan_Status
0
1
0.0
0.2
0.4
0.6
0.8
1.0
0.0
1.0
Credit_History

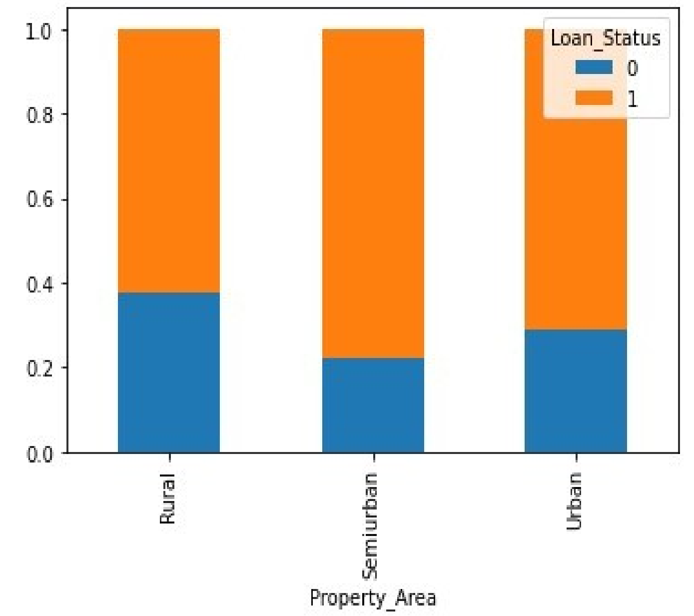

Dependents
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
2
3+

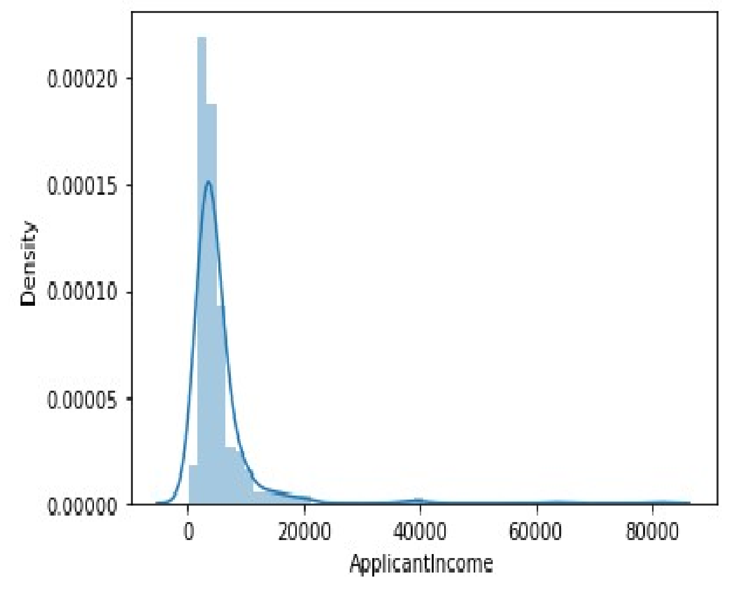

Loan_Status
0
1
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3+
Dependents

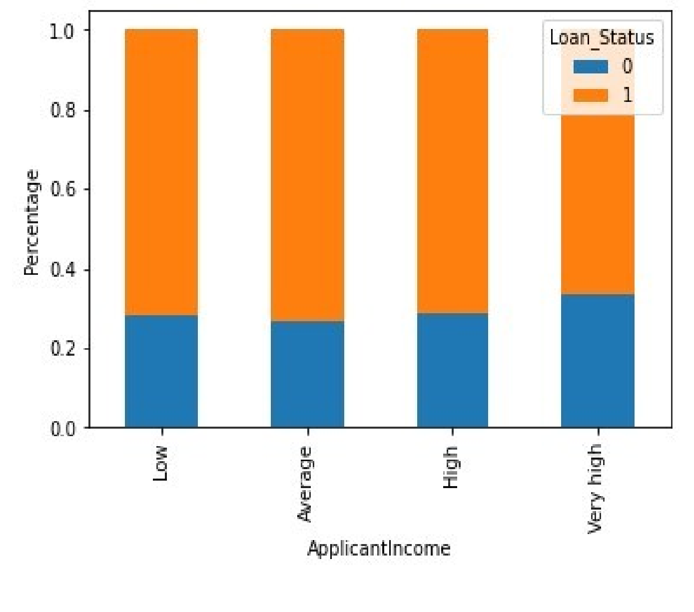

Education
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Graduate
Not Graduate

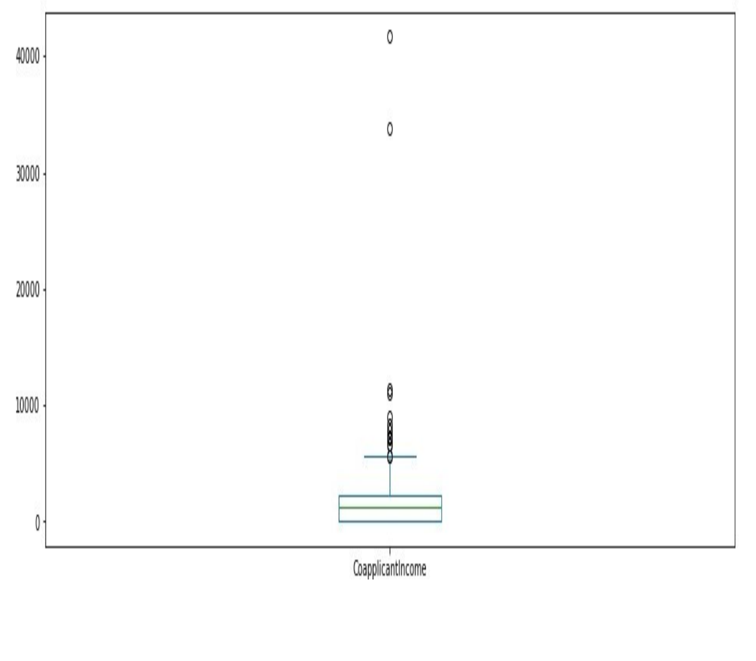

Loan_Status
0
1
0.0
0.2
0.4
0.6
0.8
1.0
Graduate
Not Graduate
Education

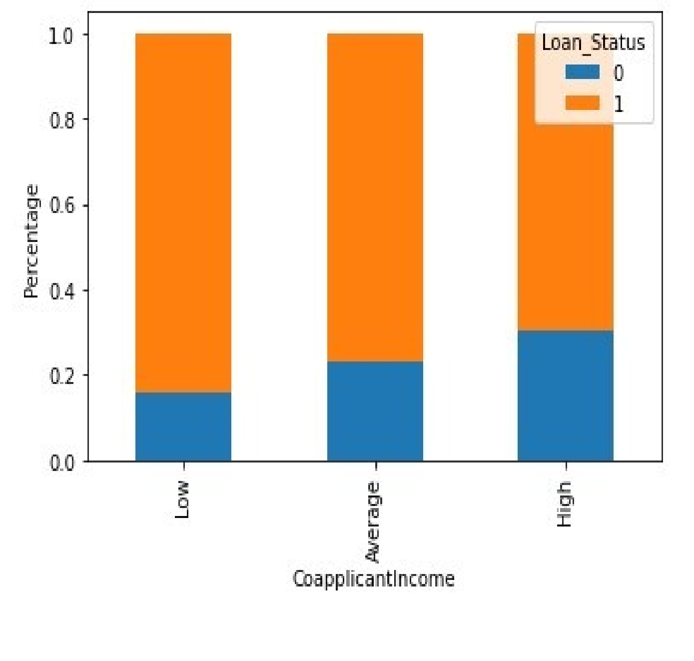

Property_Area
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Semiurban
Rural
Urban

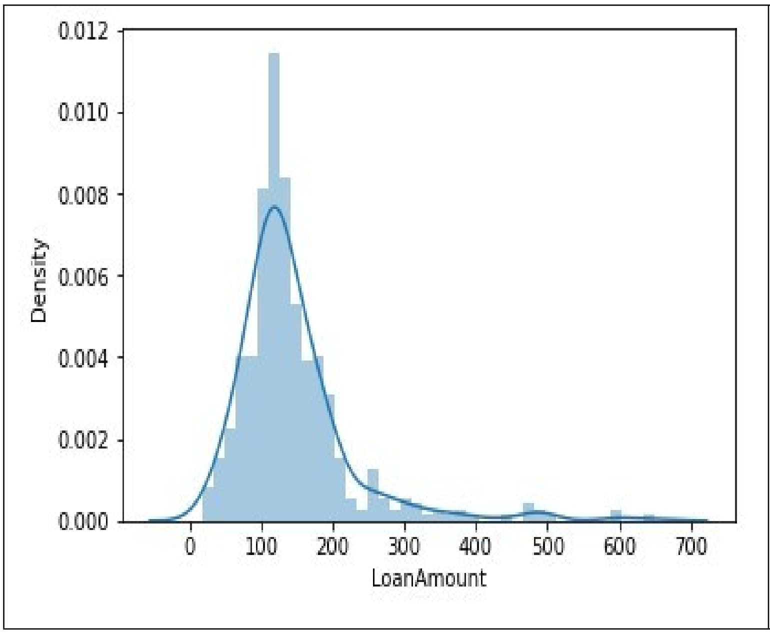

Loan_Status
0
1
1.0
0.8
0.6
0.4
0.2
0.0
Rural
Semiurban
Urban
Property_Area


Density
0.00020
0.00015
0.00010
0.00005
0.00000
0
20000
40000
60000
80000
ApplicantIncome


Loan_Status
0
1
Percentage
1.0
0.8
0.6
0.4
0.2
0.0
Low
Average
High
Very high
ApplicantIncome


40000
30000
20000
10000
0
CoapplicantIncome


Loan_Status
0
1
Percentage
1.0
0.8
0.6
0.4
0.2
0.0
Low
Average
High
CoapplicantIncome



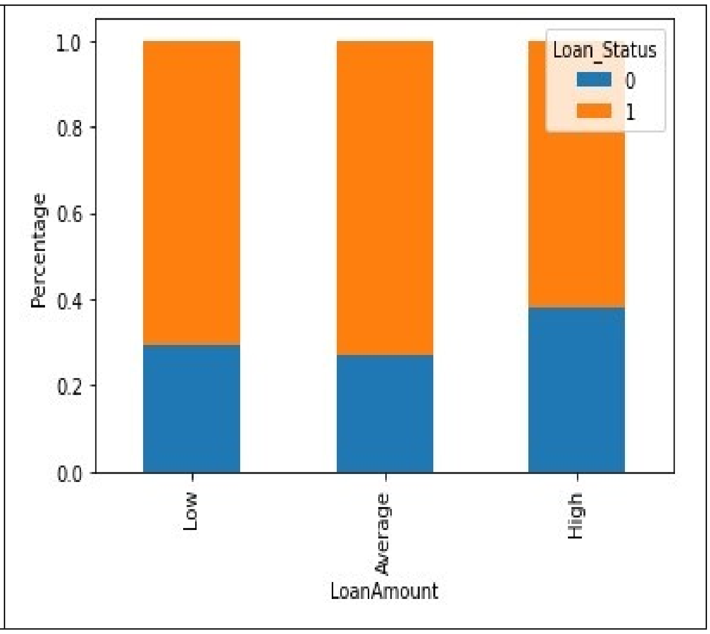


Fig.5.11 - Home Page

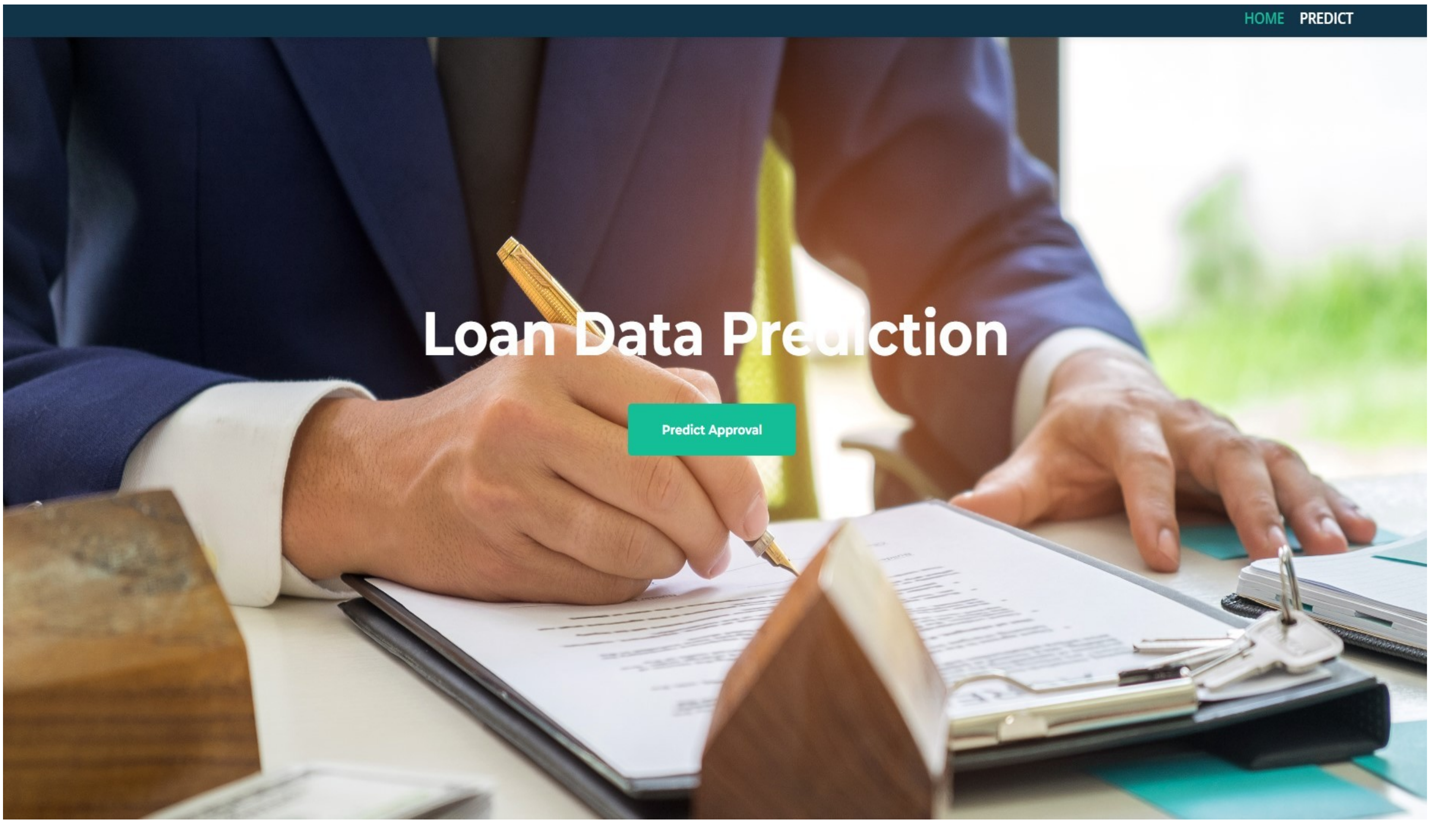


Fig.5.12 - Application Page

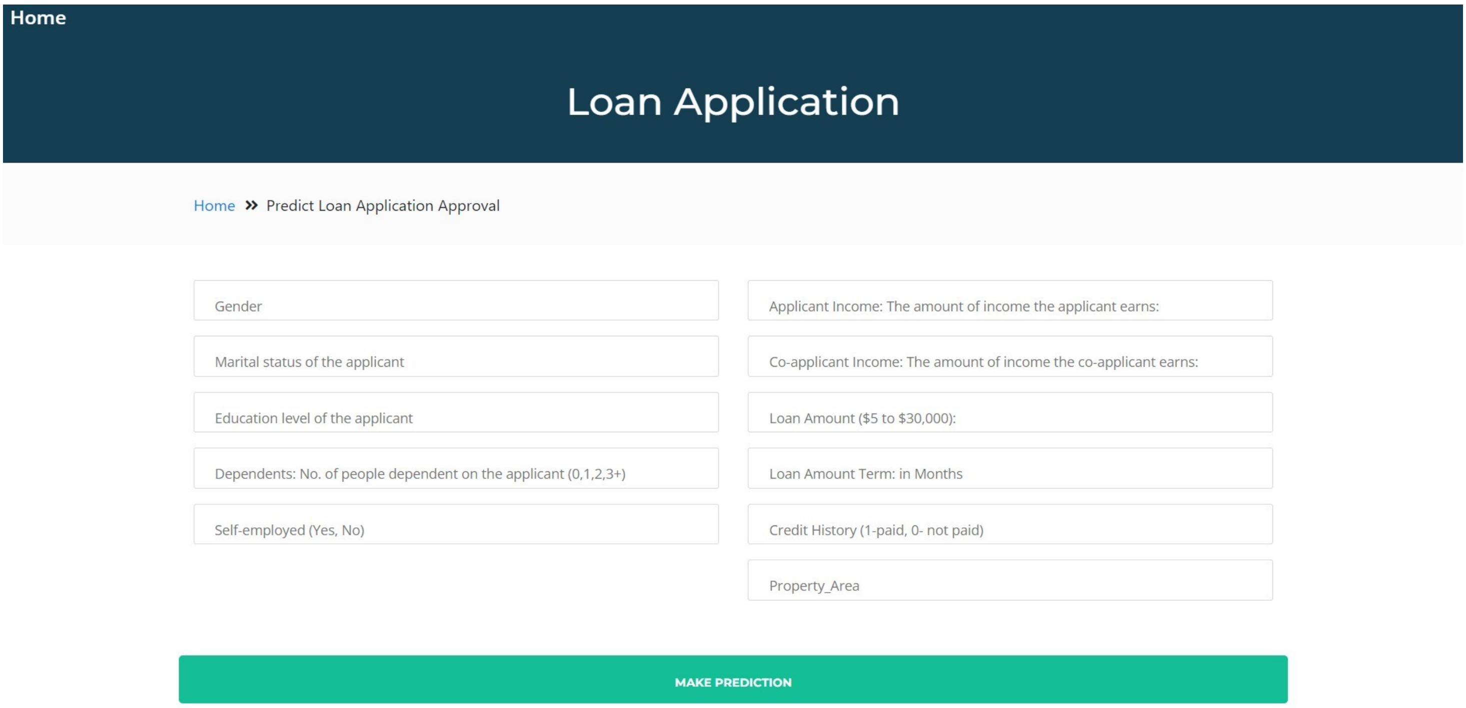
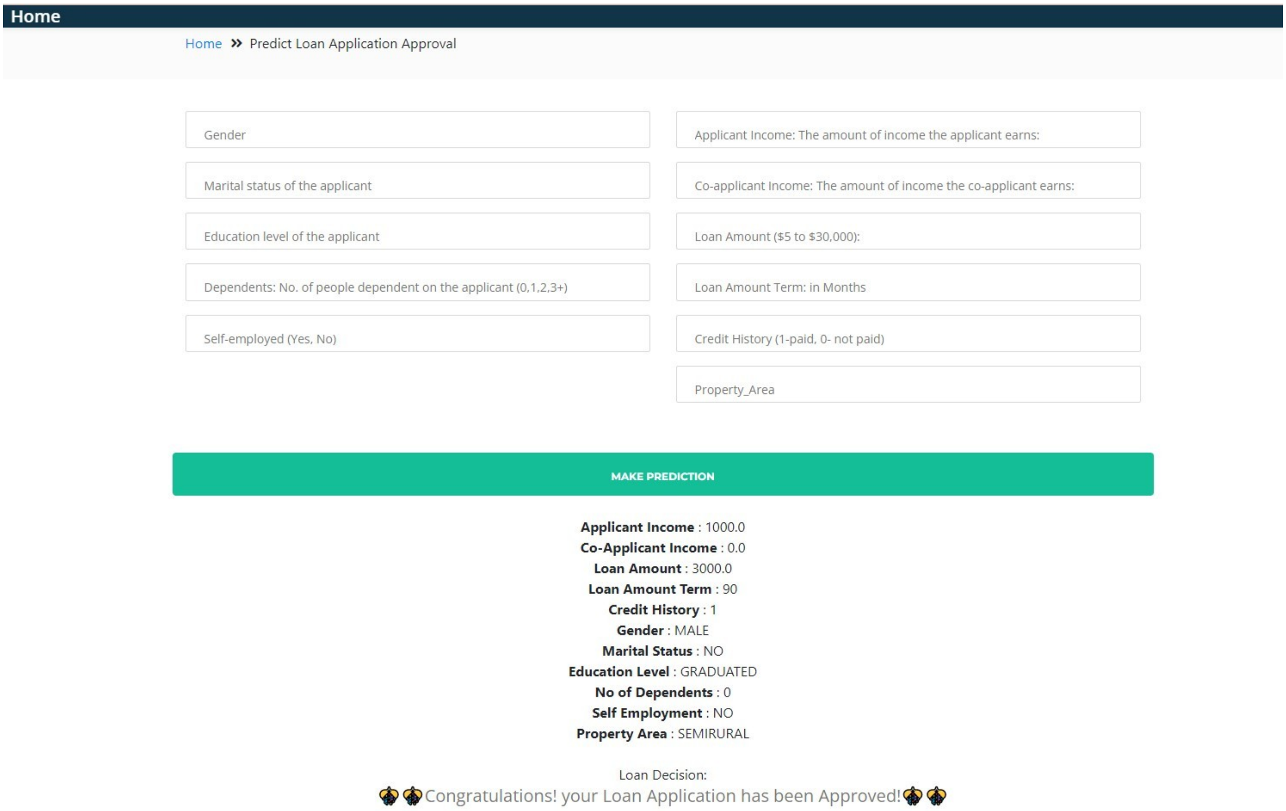
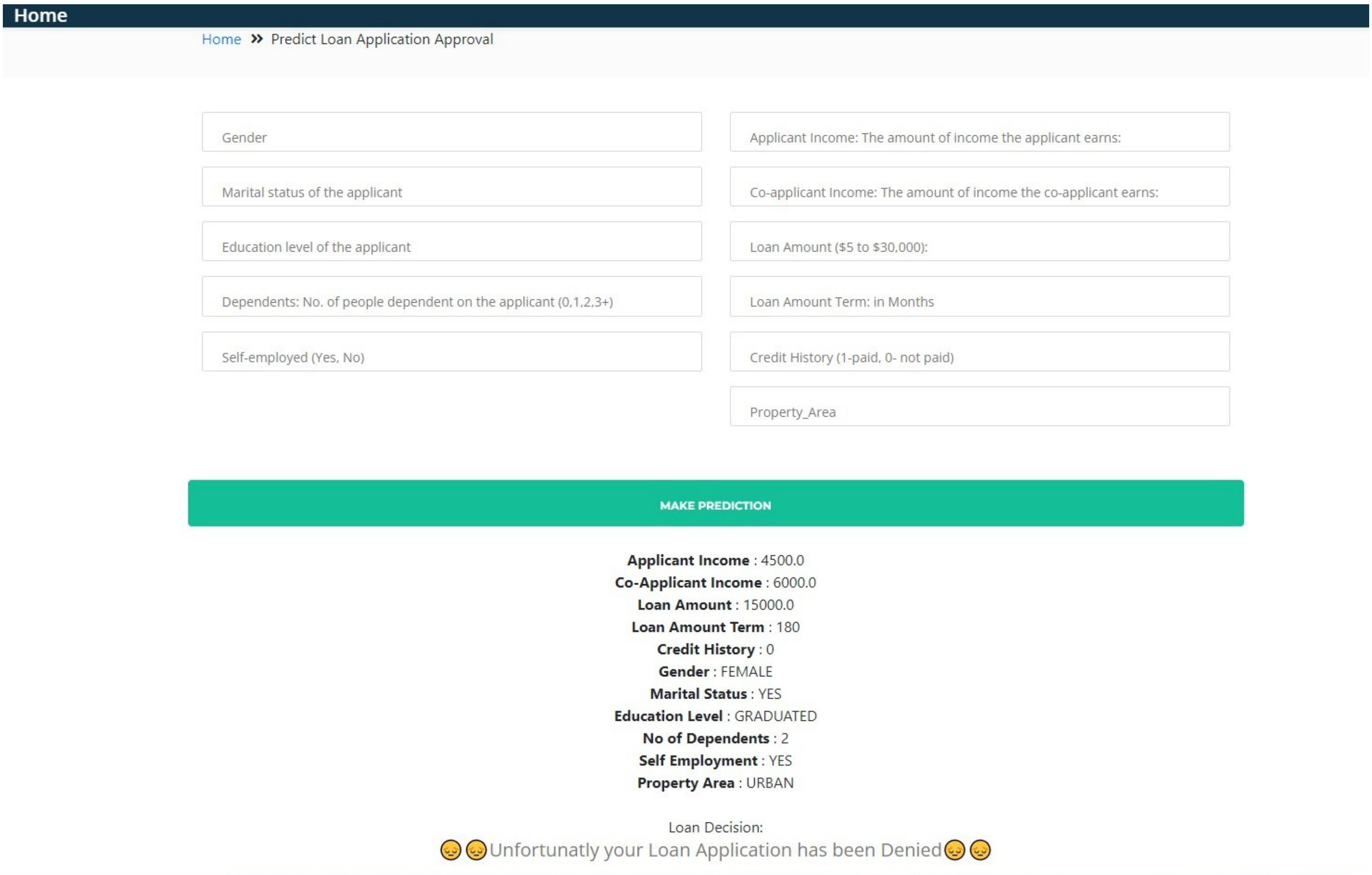
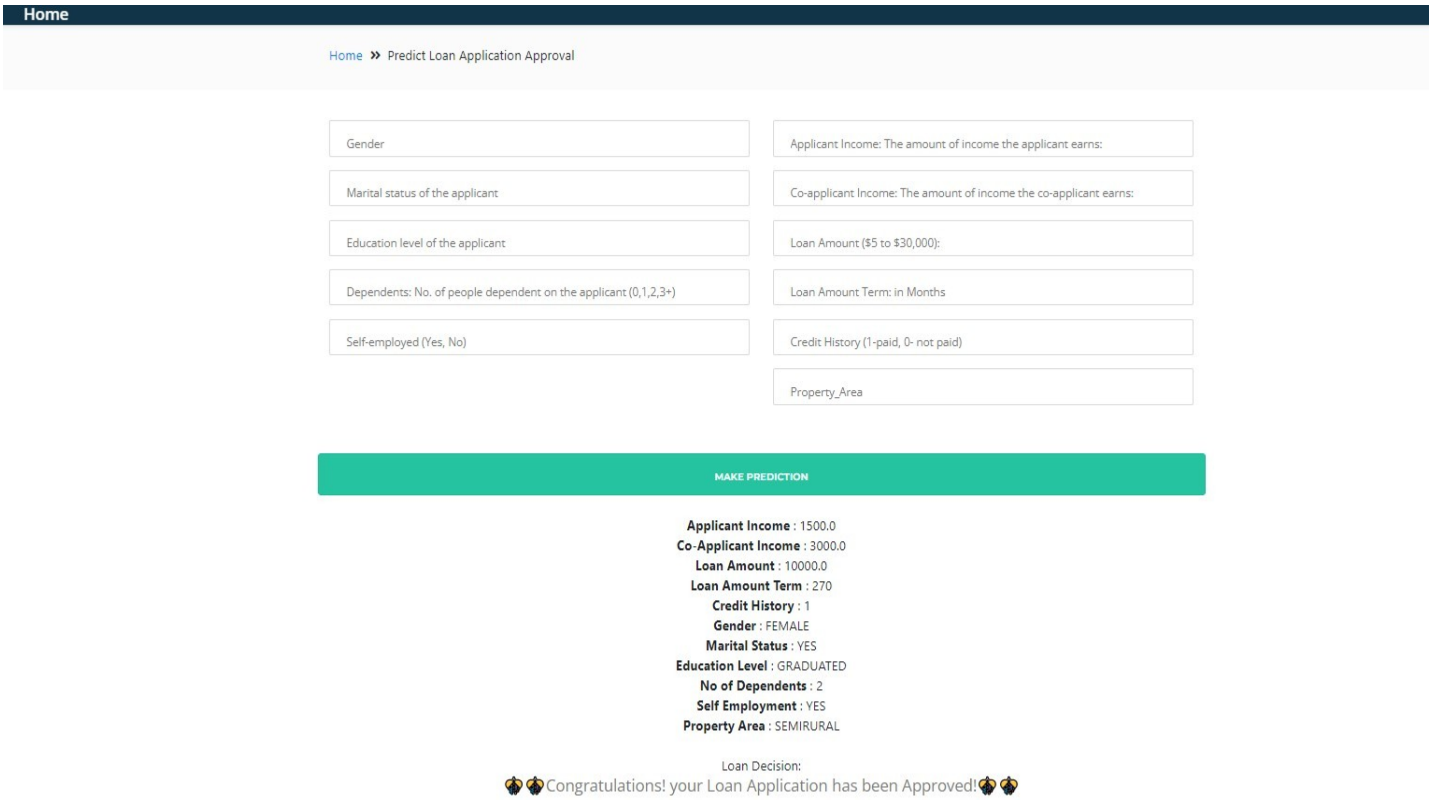


Fig.5.13 - Testing For Inputs




Home
Home » Predict Loan Application Approval
Gender
Marital status of the applicant
Education level of the applicant
Dependents: No. of people dependent on the applicant (0,1,2,3+)
Self-employed (Yes, No)
Applicant Income: The amount of income the applicant earns:
Co-applicant Income: The amount of income the co-applicant earns:
Loan Amount ($5 to $30,000):
Loan Amount Term: in Months
Credit History (1-paid, 0- not paid)
Property_Area
MAKE PREDICTION
Applicant Income : 4500.0
Co-Applicant Income : 6000.0
Loan Amount : 15000.0
Loan Amount Term : 180
Credit History : 0
Gender : FEMALE
Marital Status : YES
Education Level : GRADUATED
No of Dependents : 2
Self Employment : YES
Property Area : URBAN
Loan Decision:
Unfortunatly your Loan Application has been Denied


Home
Home » Predict Loan Application Approval
Gender
Marital status of the applicant
Education level of the applicant
Dependents: No. of people dependent on the applicant (0,1,2,3+)
Self-employed (Yes, No)
Applicant Income: The amount of income the applicant earns:
Co-applicant Income: The amount of income the co-applicant earns:
Loan Amount ($5 to $30,000):
Loan Amount Term: in Months
Credit History (1-paid, 0- not paid)
Property_Area
MAKE PREDICTION
Applicant Income : 1500.0
Co-Applicant Income : 3000.0
Loan Amount : 10000.0
Loan Amount Term : 270
Credit History : 1
Gender : FEMALE
Marital Status : YES
Education Level : GRADUATED
No of Dependents : 2
Self Employment : YES
Property Area : SEMIRURAL
Loan Decision:
Congratulations! your Loan Application has been Approved!

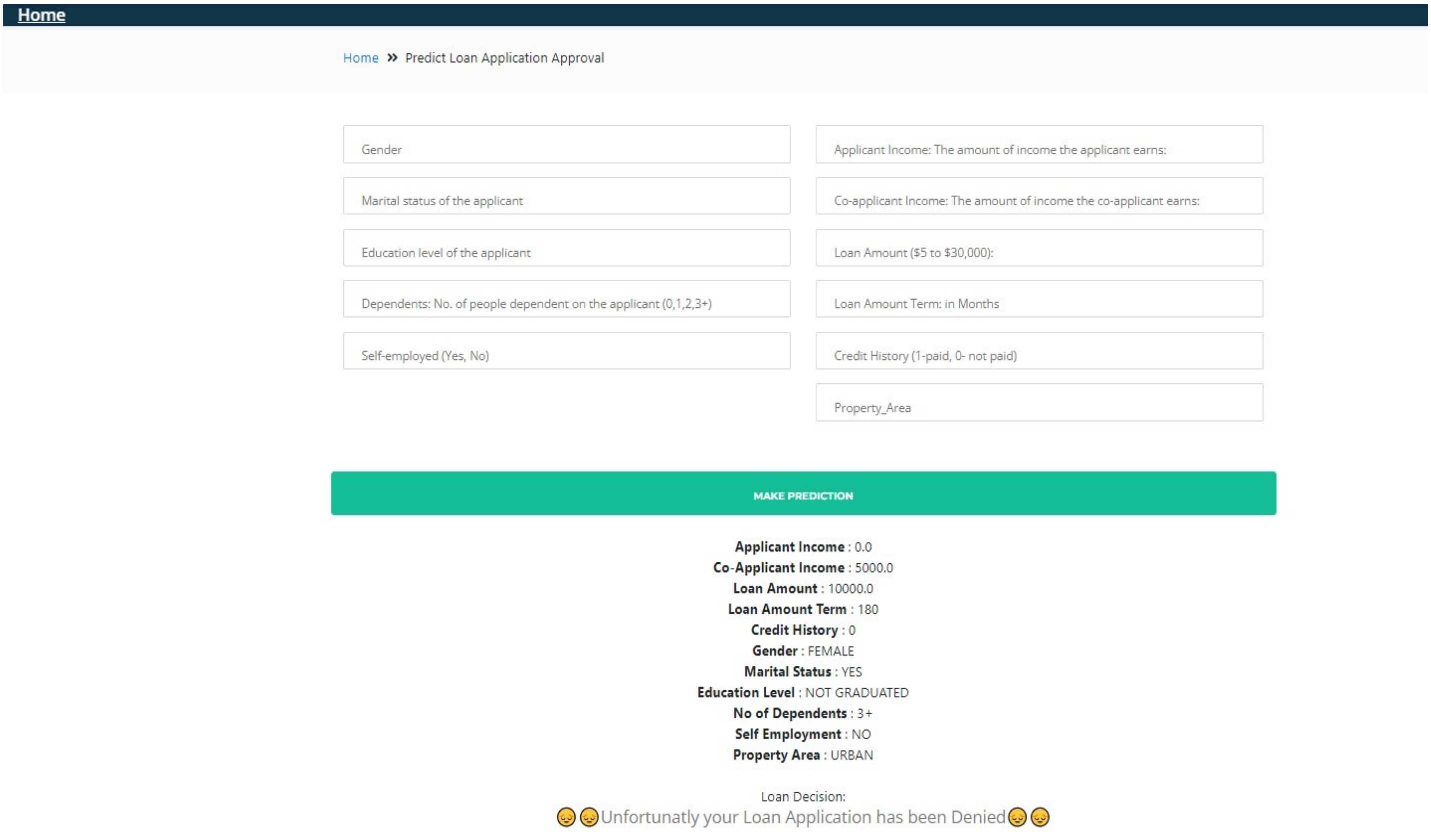
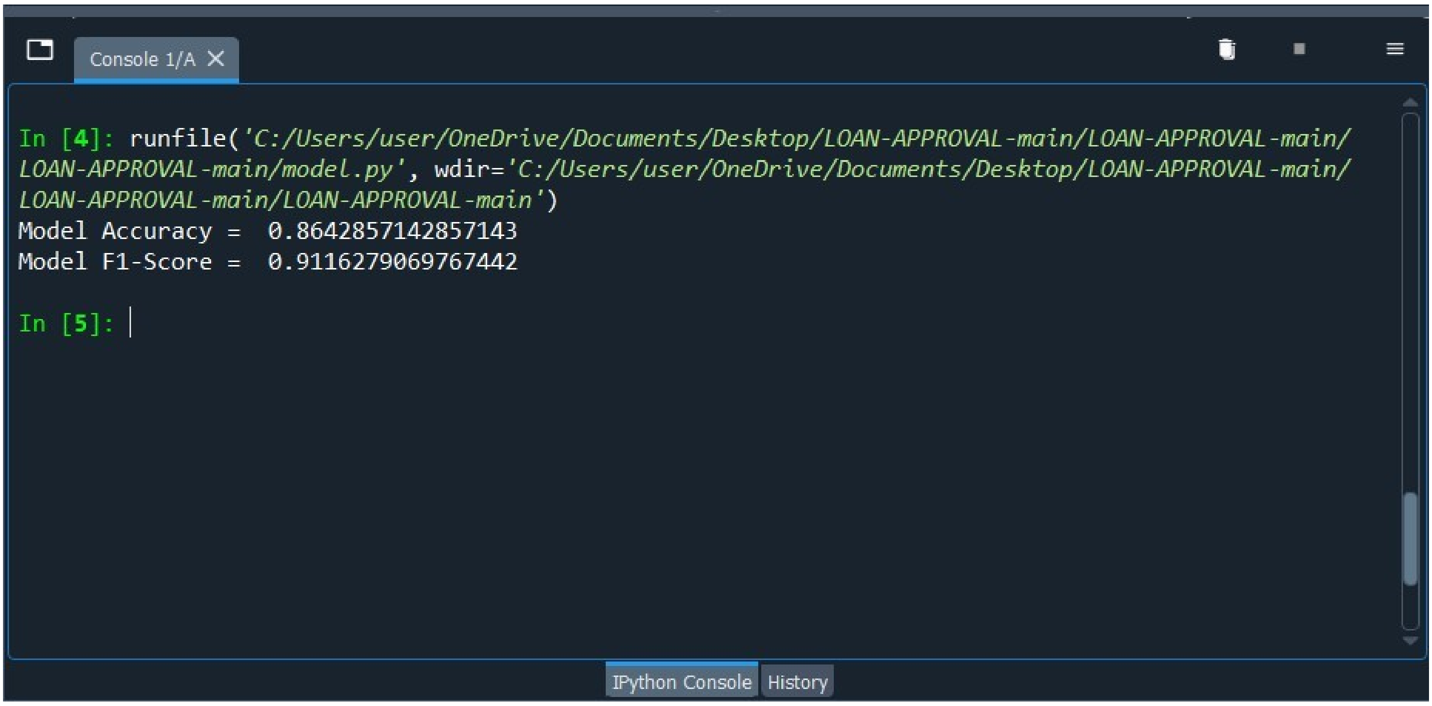


Fig.5.14 - Model Accuracy



CHAPTER 6

RESULTS AND DISCUSSIONS

6.1 Results And Discussions

Different tools were used to import machine learning models. Especially Sklearn tool is used to import the methodologies. The datasets are separated into training and testing sets in 70:30, and 80:20 ratio. The classifier is instructed and it is tested against the testing set after being trained on the training set. Classification performance was measured by computing scores for each classifier's accuracy, false-negative, and false-positive rates. Confusion matrix is also used to assess accuracy.

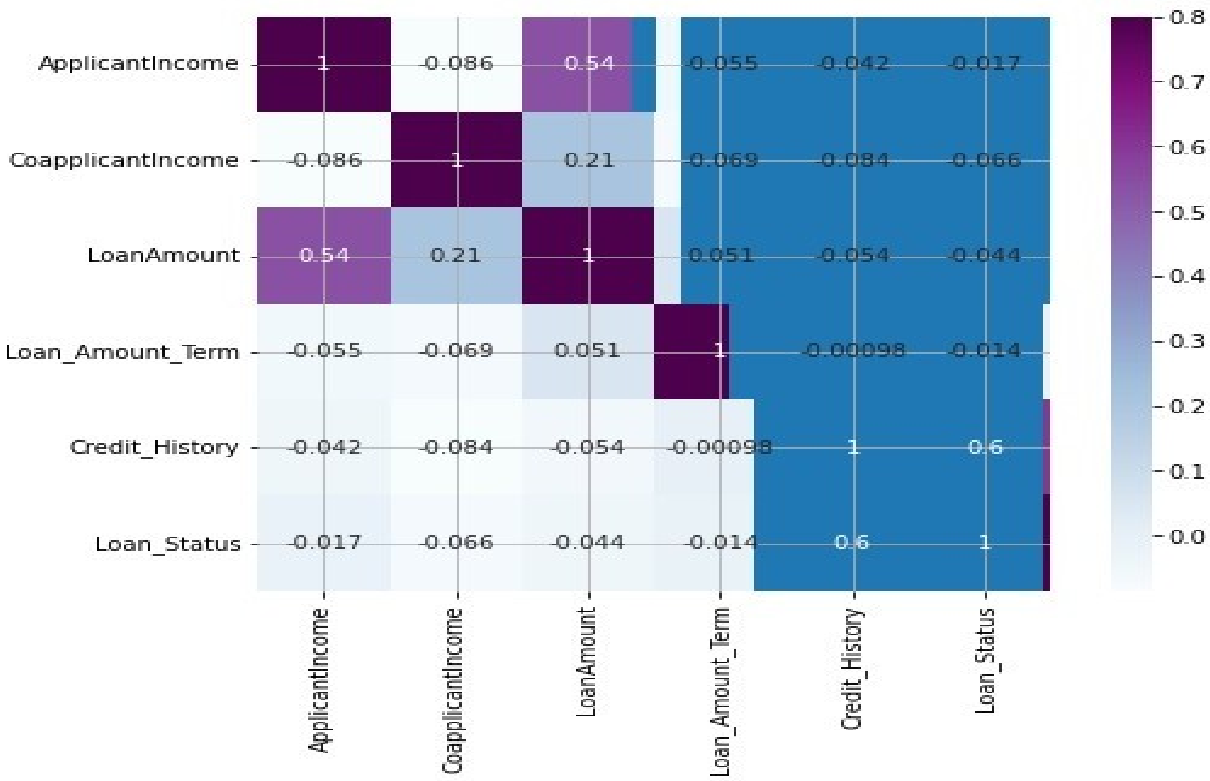


Fig - 6.1(Confusion Matrix)

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

This paper tries to make the detection techniques for loan prediction more effective. With logistic regression method, we had the lowest false positive rate and 82% detection accuracy. Additionally, the findings demonstrable classifiers function more effectively when a large amount of data is used as training data. Future loan prediction websites will be more successfully discovered by combining machine learning's logistic regression algorithm. After a thorough analysis of the product's benefits and drawbacks, it is safe to conclude that the component is a highly effective one. This application conforms with all Banker requirements and is operationally sound. It is straightforward to connect this component to a variety of other systems. Numerous software bugs, content issues, and—most significantly—the weight of features have successfully been corrected by automated prediction systems. Consequently, the alleged software may soon be upgraded to offer increased security, dependability, and dynamic weight adjustment. The automated processing system module can soon be integrated with this prediction module. Although After a defined period of time,the system can be updated in future software to incorporate fresh testing data in training data as it was trained on an outdated training dataset.

REFERENCES

[1] An Method for Predicting Loan Approval Using Machine Learning Algorithm, International Conference on Electronics and Sustainable Communication Systems(ICESC), M. Sheikh, A. Goel, and T. Kumar (2020).

[2] S. M. S. and R. Sunny T, "Loan Credibility Prediction System Based on Decision Tree Algorithm," International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 09, (2015).

[3] "Loan Approval Prediction Based on Machine Learning Method," IOSR Journal of Computer Engineering, by A. Kumar, I. Garg, and S. Kaur (2016).

[4] "Clustering Loan Applicants based on Risk Percentage using K-Means Clustering Approaches," by Dr. K. Kavitha Issue 2 of Volume 6 of IJARCSSE (2016). 65–80 in Library, vol. 38, no. 1, 2020.

[5] "Prediction of Loan Status in Commercial Bank Using Machine Learning Classifier," Proceedings of the International Conference on Intelligent Sustainable Systems, G. Arutjothi and Dr. C. Senthamarai (2017).

[6] "Loan Prediction by Using Machine Learning Models," International Journal of Engineering and Technology, P. Supriya, M. Pavani, N. Saisushma, N. Kumari, and K. Vikas (2019).

[7] "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thinking (IJCRT), R. Salvi, R. Ghule, T. Sanadi, and M. Bhajibhakare (2021).

[8] "Domain-specific modification of a partial least squares regression model for loan defaults prediction," 11th IEEE International Conference on Data Mining Workshops, B. Srinivasan, N. Gnanasambandam, S. Zhao, and R. Minhas (2011).

[9] "Neural Networks for Prediction of Loan Default Using Attribute Relevance Analysis," International Conference on Signal Acquisition and Processing, M. V. Reddy and Dr. B. Kavitha (2010).

[10] Business-Oriented Feature Selection for Hybrid Classification Model of Credit Scoring, IEEE Second International Conference on Data Stream Mining & Processing, August G. Chornous, I. Nikolskyi (2018).

[11] P.Dutta, "A STUDY ON MACHINE LEARNING ALGORITHM FOR ENHANCEMENT OF LOAN PREDICTION", International Research 3 ITM Web of Conferences 44, 03019 (2022) https://doi.org/10.1051/itmconf/20224403019 ICACC-2022 Journal of Modernization in Engineering Technology and Science, (2021).

[12] R. Salvi, R. Ghule, T. Sanadi, M. Bhajibhakare, "HOME LOAN DATA ANALYSIS AND VISUALIZATION," International Journal of Creative Research Thoughts (IJCRT), (2021).

[13] M. Sheikh, A. Goel, T. Kumar, "An Approach for Prediction of Loan Approval using Machine Learning Algorithm," International Conference on Electronics and Sustainable Communication Systems (ICESC), (2020).

[14] M.Aakanksha Saha, Tamara Denning, VivekSrikumar, Sneha Kumar Kasera. "Secrets inSource Code: Reducing False Positives usingMachine Learning", 2020 InternationalConference on Communication Systems &Networks (COMSNETS), 2020.

APPENDIX

A . CODING

CODE:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
#%matplotlib inline
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
#from google.colab import files
from sklearn.model_selection import StratifiedKFold
from sklearn import metrics
#from google.colab import files
from sklearn.tree import  DecisionTreeClassifier
from sklearn.metrics import accuracy_score, f1_score
import warnings
warnings.filterwarnings("ignore")

loan_data  = pd.read_csv("Data/loan_train.csv",index_col=False)
#loan_data = loan_data.drop(['Unnamed: 0'], axis=1)
loan_data.head()

test_data = pd.read_csv('Data/loan_test.csv')
test_data.head()

loan_data.shape

test_data.shape

loan_data['Loan_Status'].value_counts()

loan_data['Loan_Status'].value_counts(normalize=True)

loan_data['Loan_Status'].value_counts().plot.bar()

loan_data['Gender'].value_counts(normalize=True).plot.bar(title='Gender')
plt.show()
loan_data['Married'].value_counts(normalize=True).plot.bar(title='Married')
plt.show()
loan_data['Self_Employed'].value_counts(normalize=True).plot.bar(title='Self_Employed')
plt.show()
loan_data['Credit_History'].value_counts(normalize=True).plot.bar(title='Credit_History')
```

```
loan_data['Self_Employed'].value_counts(normalize=True).plot.bar(title='Self_Employed')
plt.show()
loan_data['Credit_History'].value_counts(normalize=True).plot.bar(title='Credit_History')
plt.show()

loan_data['Dependents'].value_counts(normalize=True).plot.bar( title='Dependents')
plt.show()
loan_data['Education'].value_counts(normalize=True).plot.bar(title='Education')
plt.show()
loan_data['Property_Area'].value_counts(normalize=True).plot.bar(title='Property_Area')
plt.show()

sns.distplot(loan_data['ApplicantIncome'])
plt.show()
loan_data['ApplicantIncome'].plot.box(figsize=(16,5))
plt.show()

loan_data.boxplot(column='ApplicantIncome', by = 'Education')
plt.suptitle("")

sns.distplot(loan_data['CoapplicantIncome'])
plt.show()
loan_data['CoapplicantIncome'].plot.box(figsize=(16,5))
plt.show()


loan_data.notna()
sns.distplot(loan_data['LoanAmount'])
plt.show()
loan_data['LoanAmount'].plot.box(figsize=(16,5))
plt.show()

Gender=pd.crosstab(loan_data['Gender'],loan_data['Loan_Status'])
Gender.div(Gender.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()

Married=pd.crosstab(loan_data['Married'],loan_data['Loan_Status'])
Dependents=pd.crosstab(loan_data['Dependents'],loan_data['Loan_Status'])
Education=pd.crosstab(loan_data['Education'],loan_data['Loan_Status'])
Self_Employed=pd.crosstab(loan_data['Self_Employed'],loan_data['Loan_Status'])
Married.div(Married.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
```

```
Education=pd.crosstab(loan_data['Education'],loan_data['Loan_Status'])
Self_Employed=pd.crosstab(loan_data['Self_Employed'],loan_data['Loan_Status'])
Married.div(Married.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()
Dependents.div(Dependents.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()
Education.div(Education.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()
Self_Employed.div(Self_Employed.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()

Credit_History=pd.crosstab(loan_data['Credit_History'],loan_data['Loan_Status'])
Property_Area=pd.crosstab(loan_data['Property_Area'],loan_data['Loan_Status'])
Credit_History.div(Credit_History.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True,figsize=(4,4))
plt.show()
Property_Area.div(Property_Area.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True)
plt.show()

loan_data.groupby('Loan_Status')['ApplicantIncome'].mean().plot.bar()

bins=[0,2500,4000,6000,81000]
group=['Low','Average','High','Very high']
loan_data['Income_bin']=pd.cut(loan_data['ApplicantIncome'],bins,labels=group)
Income_bin=pd.crosstab(loan_data['Income_bin'],loan_data['Loan_Status'])
Income_bin.div(Income_bin.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True)
plt.xlabel('ApplicantIncome')
P=plt.ylabel('Percentage')

bins=[0,1000,3000,42000]
group=['Low','Average','High']
loan_data['Coapplicant_Income_bin']=pd.cut(loan_data['CoapplicantIncome'],bins,labels=group)
Coapplicant_Income_bin=pd.crosstab(loan_data['Coapplicant_Income_bin'],loan_data['Loan_Status'])
Coapplicant_Income_bin.div(Coapplicant_Income_bin.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True)
plt.xlabel('CoapplicantIncome')
P=plt.ylabel('Percentage')

loan_data['Total_Income']=loan_data['ApplicantIncome']+loan_data['CoapplicantIncome']
bins=[0,2500,4000,6000,81000]
group=['Low','Average','High','Very high']
loan_data['Total_Income_bin']=pd.cut(loan_data['Total_Income'],bins,labels=group)
Total_Income_bin=pd.crosstab(loan_data['Total_Income_bin'],loan_data['Loan_Status'])
```

```
loan_data['Total_Income_bin']=pd.cut(loan_data['Total_Income'],bins,labels=group)
Total_Income_bin=pd.crosstab(loan_data['Total_Income_bin'],loan_data['Loan_Status'])
Total_Income_bin.div(Total_Income_bin.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True)
plt.xlabel('Total_Income')
P=plt.ylabel('Percentage')

bins=[0,100,200,700]
group=['Low','Average','High']
loan_data['LoanAmount_bin']=pd.cut(loan_data['LoanAmount'],bins,labels=group)
LoanAmount_bin=pd.crosstab(loan_data['LoanAmount_bin'],loan_data['Loan_Status'])
LoanAmount_bin.div(LoanAmount_bin.sum(1).astype(float), axis=0).plot(kind="bar",stacked=True)
plt.xlabel('LoanAmount')
P=plt.ylabel('Percentage')

loan_data=loan_data.drop(['Income_bin', 'Coapplicant_Income_bin', 'LoanAmount_bin', 'Total_Income_bin', 'Total_Income'], axis=1)
loan_data['Dependents'].replace('3+', 3,inplace=True)
test_data['Dependents'].replace('3+', 3,inplace=True)

matrix = loan_data.corr()
f, ax = plt.subplots(figsize=(9,6))
sns.heatmap(matrix,vmax=.8,square=True,cmap="BuPu", annot = True)

loan_data.isnull().sum()

loan_data['Gender'].fillna(loan_data['Gender'].mode()[0], inplace=True)
loan_data['Married'].fillna(loan_data['Married'].mode()[0], inplace=True)
loan_data['Dependents'].fillna(loan_data['Dependents'].mode()[0], inplace=True)
loan_data['Self_Employed'].fillna(loan_data['Self_Employed'].mode()[0], inplace=True)
loan_data['Credit_History'].fillna(loan_data['Credit_History'].mode()[0], inplace=True)

loan_data['Loan_Amount_Term'].value_counts()

loan_data['Loan_Amount_Term'].fillna(loan_data['Loan_Amount_Term'].mode()[0], inplace=True)

loan_data['LoanAmount'].fillna(loan_data['LoanAmount'].median(), inplace=True)

loan_data.isnull().sum()

test_data['Gender'].fillna(loan_data['Gender'].mode()[0], inplace=True)
test_data['Married'].fillna(loan_data['Married'].mode()[0], inplace=True)
test_data['Dependents'].fillna(loan_data['Dependents'].mode()[0], inplace=True)
```

```
loan_data['Loan_Amount_Term'].fillna(loan_data['Loan_Amount_Term'].mode()[0], inplace=True)

loan_data['LoanAmount'].fillna(loan_data['LoanAmount'].median(), inplace=True)

loan_data.isnull().sum()

test_data['Gender'].fillna(loan_data['Gender'].mode()[0], inplace=True)
test_data['Married'].fillna(loan_data['Married'].mode()[0], inplace=True)
test_data['Dependents'].fillna(loan_data['Dependents'].mode()[0], inplace=True)
test_data['Self_Employed'].fillna(loan_data['Self_Employed'].mode()[0], inplace=True)
test_data['Credit_History'].fillna(loan_data['Credit_History'].mode()[0], inplace=True)
test_data['Loan_Amount_Term'].fillna(loan_data['Loan_Amount_Term'].mode()[0], inplace=True)
test_data['LoanAmount'].fillna(loan_data['LoanAmount'].median(), inplace=True)

loan_data['LoanAmount_log']=np.log(loan_data['LoanAmount'])
loan_data['LoanAmount_log'].hist(bins=20)
test_data['LoanAmount_log']=np.log(test_data['LoanAmount'])

loan_data=loan_data.drop('Loan_ID',axis=1)
test_data=test_data.drop('Loan_ID',axis=1)

X = loan_data.drop('Loan_Status',axis=1)
y = loan_data.Loan_Status

train = loan_data.copy()
test = test_data.copy()

X = pd.get_dummies(X)
train=pd.get_dummies(train)
test=pd.get_dummies(test)

x_train, x_valid, y_train, y_valid = train_test_split(X,y, test_size=0.3)
model = LogisticRegression()
model.fit(x_train, y_train)

pred_cv = model.predict(x_valid)
print('Model Accuracy = ', accuracy_score(y_valid,pred_cv))
print('Model F1-Score = ', f1_score(y_valid,pred_cv))
```

```
#1. pip install markupsafe==2.0.1
#   pip install werkzeug==2.0.3
#2. pip install jinja2<3.1.0
#   from markupsafe import escape
#   pip install Flask==2.1.0


import flask
import pickle
import pandas as pd
import numpy as np
from markupsafe import escape


from sklearn.preprocessing import StandardScaler


#load models at top of app to load into memory only one time
with open('models/loan_application_model_lr.pickle', 'rb') as f:
    clf_lr = pickle.load(f)


# with open('models/knn_regression.pkl', 'rb') as f:
#     knn = pickle.load(f)
ss = StandardScaler()


genders_to_int = {'MALE':1,
                  'FEMALE':0}

married_to_int = {'YES':1,
                  'NO':0}

education_to_int = {'GRADUATED':1,
                  'NOT GRADUATED':0}

dependents_to_int = {'0':0,
                      '1':1,
                      '2':2,
                      '3+':3}
```

```python
self_employment_to_int = {'YES':1,
                          'NO':0}

property_area_to_int = {'RURAL':0,
                        'SEMIRURAL':1,
                        'URBAN':2}

loan_term = {'3':90,
             '6':180,
             '9':270,
             '12':365}


app = flask.Flask(__name__, template_folder='templates')
@app.route('/')
def main():
    return (flask.render_template('index.html'))

@app.route('/report')
def report():
    return (flask.render_template('report.html'))

@app.route('/jointreport')
def jointreport():
    return (flask.render_template('jointreport.html'))


@app.route("/Loan_Application", methods=['GET', 'POST'])
def Loan_Application():

    if flask.request.method == 'GET':
        return (flask.render_template('Loan_Application.html'))

    if flask.request.method =='POST':

        #get input
        #gender as string
        genders_type = flask.request.form['genders_type']
        #marriage status as boolean YES: 1 , NO: 0
        marital_status = flask.request.form['marital_status']
        #Dependents: No. of people dependent on the applicant (0,1,2,3+)
```

```
#gender as string
genders_type = flask.request.form['genders_type']
#marriage status as boolean YES: 1 , NO: 0
marital_status = flask.request.form['marital_status']
#Dependents: No. of people dependent on the applicant (0,1,2,3+)
dependents = flask.request.form['dependents']

#dependents = dependents_to_int[dependents.upper()]

#education status as boolean Graduated, Not graduated.
education_status = flask.request.form['education_status']
#Self_Employed: If the applicant is self-employed or not (Yes, No)
self_employment = flask.request.form['self_employment']
#Applicant Income
applicantIncome = float(flask.request.form['applicantIncome'])
#Co-Applicant Income
coapplicantIncome = float(flask.request.form['coapplicantIncome'])
#loan amount as integer
loan_amnt = float(flask.request.form['loan_amnt'])
#term as integer: from 10 to 365 days...
term_d = int(flask.request.form['loan_term'])
# credit_history
credit_history = int(flask.request.form['credit_history'])
# property are
property_area = flask.request.form['property_area']
#property_area = property_area_to_int[property_area.upper()]

#create original output dict
output_dict= dict()
output_dict['Applicant Income'] = applicantIncome
output_dict['Co-Applicant Income'] = coapplicantIncome
output_dict['Loan Amount'] = loan_amnt
output_dict['Loan Amount Term']=loan_term[str(term_d)]
output_dict['Credit History'] = credit_history
output_dict['Gender'] = genders_type
output_dict['Marital Status'] = marital_status
output_dict['Education Level'] = education_status
output_dict['No of Dependents'] = dependents
output_dict['Self Employment'] = self_employment
output_dict['Property Area'] = property_area
```

```python
x = np.zeros(21)

x[0] = applicantIncome
x[1] = coapplicantIncome
x[2] = loan_amnt
x[3] = term_d
x[4] = credit_history

print('------this is array data to predict-------')
print('X = '+str(x))
print('-----------------------------------------')

pred = clf_lr.predict([x])[0]

if pred==1:
    res = '🥳🥳Congratulations! your Loan Application has been Approved!🥳🥳'
else:
        res = '😔😔Unfortunatly your Loan Application has been Denied😔😔'



#render form again and add prediction
return flask.render_template('Loan_Application.html',
                             original_input=output_dict,
                             result=res,)




# temp = pd.DataFrame(index=[1])




# temp['genders_type'] = genders_to_int[genders_type.upper()]
# #marriage status as boolean YES: 1 , NO: 0
# temp['marital_status'] = married_to_int[marital_status.upper()]
# #Dependents: No. of people dependent on the applicant (0,1,2,3+)
# temp['dependents'] = dependents_to_int[dependents.upper()]
# #education status as boolean Graduated, Not graduated.
# temp['education_status'] = education_to_int[education_status.upper()]
```

```
        # #Applicant Income
        # temp['applicantIncome'] = applicantIncome
        # #Co-Applicant Income
        # temp['coapplicantIncome'] = coapplicantIncome
        # #loan amount as integer
        # temp['loan_amnt'] = loan_amnt
        # #term as integer: from 10 to 365 days...
        # temp['term_d'] =  term_d
        # # credit_history
        # temp['credit_history'] = credit_history
        # # property are
        # temp['property_area'] = property_area_to_int[property_area.upper()]

        # temp['loan_amnt_log']=np.log(temp['loan_amnt'])

        # Feature Engineering :
        #temp['Total_Income']= temp['applicantIncome']+temp['coapplicantIncome']
        #temp['Total_Income_log'] = np.log(temp['Total_Income'])
        #temp['EMI']= temp['loan_amnt']/temp['term_d']
        #temp['Balance Income'] = temp['Total_Income']-(temp['EMI']*1000)

        # Columns to drop and afterward Predict up on the feature engineered columns
        #temp = temp.drop(['applicantIncome', 'coapplicantIncome', 'loan_amnt', 'term_d'], axis=1)



        # Credit_History is the most important feature followed by Balance Income, Total Income, EMI.
        # So, feature engineering helped us in predicting our target variable.





        # #make prediction


if __name__ == '__main__':
    app.run()
```

```html
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
    <meta property="og:site_name" content="" />
    <meta property="og:site" content="" />
    <meta property="og:title" content="" />
    <meta property="og:description" content="" />
    <meta property="og:image" content="" />
    <meta property="og:url" content="" />
    <meta property="og:type" content="article" />

    <!-- Website Title -->
    <title>Loan Application</title>

    <!-- Styles -->
    <link href="https://fonts.googleapis.com/css?family=Montserrat:500,700&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="https://fonts.googleapis.com/css?family=Open+Sans:400,400i,600&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="{{url_for('static', filename='css/styles.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/fontawesome-all.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/swiper.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/magnific-popup.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/bootstrap.css')}}" rel="stylesheet">






    <!-- Favicon  -->
    <link rel="icon" href="{{url_for('static', filename='images/logo.svg')}}">
</head>

<body data-spy="scroll" data-target=".fixed-top">

    <!-- Preloader -->
    <div class="spinner-wrapper">
```

```
        </div>
    </div> -->



    <!-- Footer -->
    <div id=contact class="footer">
        <h4>Tools Used</h4>
        <a class="white" href="#">Python</a><br>
        <a class="white" href="#">Seaborn &amp; Matplotlib</a><br>
        <a class="white" href="#">Scikit-learn</a><br>
        <a class="white" href="#">Flask</a><br>
        <a class="white" href="#">SweetViz</a><br>
    </div> <!-- end of footer -->
    <!-- end of footer -->




    <!-- Scripts -->
    <script src="{{url_for('static', filename='js/jquery.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/popper.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/bootstrap.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/jquery.easing.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/jquery.magnific-popup.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/morphext.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/isotope.pkgd.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/validator.min.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
    <script src="{{url_for('static', filename='js/scripts.js')}}"></script>
    <!-- jQuery for Bootstrap's JavaScript plugins -->
</body>

</html>
```

```
function hideAllDetails()
{
    $(".container-feature-detail").hide();
    $(".container-df-associations").hide();
    $("span.bg-tab-summary-rollover").hide();
    $("#button-summary-associations-source, #button-summary-associations-compare").removeClass("button-assoc-selected");
    $("#button-summary-associations-source, #button-summary-associations-compare").addClass("button-assoc");
}


// GLOBAL EVENTS
// --------------------------------------------------------------------------------------------------------------------
// EVENT: [ANYWHERE] RIGHT-CLICK REMOVES SELECTION
// $(document).contextmenu(function() {
//     if (g_snapped != "")
//     {
//         g_snapped = "";
//         hideAllDetails();
//     }
//     if (g_lastHovered != "")
//     {
//         $(g_lastHovered).show();
//         //alert("#"+g_lastHovered);
//     }
//     return false;
// });

$("span.bg-tab-summary-rollover").hide();
hideAllDetails();

$(document).ready(function() {
// INITIALIZATION
// -------------------------------------------------------
hideAllDetails();
$("span.bg-tab-summary-rollover").hide();

// Make the detail column the same height, so the floating element has room
//$("#col2").height($("#col1").height());
$("#col1").height(g_height);
```

```
    <div class="text-med-bold" style="position: absolute; left:10px">TOP CATEGORIES</div>
    <div class="breakdown-hr"><hr></div>
    <div class="pair__header color-source" style="left: 66px"></div>
    <!-- Targets -->
        <div class="pair__header color-source-target"
             style="position: absolute; left: 162px; line-height: 11px;
                     text-align: right;">
                Loan_Status
        </div>
        <div class="pair__header color-compare" style="position: absolute; left: px">
            <!--None-->
        </div>
</div>

<!-- ROW CONTENT -->
    <div class="breakdown-row text-value " style="width:553px">
        <!-- Name -->
        <div class="text-label color-normal " style="position: absolute; left:10px;
        width: 66px; overflow: hidden; text-overflow: ellipsis; white-space: nowrap;">Male</div>

        <!-- Freq -->
        <div class="pair__col color-source" style="left: 66px">
            <div class="pair-pos__num dim">393</div>
            <div class="pair-pos__perc">82%</div>
        </div>
        <!-- Targets -->
                <div class="pair__col color-source-target" style="position: absolute; left: 162px">
                    <div class="pair-pos__num dim">281</div>
                    <div class="pair-pos__perc">72%</div>
            </div>
    </div>
    <div class="breakdown-row text-value row-colored" style="width:553px">
        <!-- Name -->
        <div class="text-label color-normal " style="position: absolute; left:10px;
        width: 66px; overflow: hidden; text-overflow: ellipsis; white-space: nowrap;">Female</div>

        <!-- Freq -->
        <div class="pair__col color-source" style="left: 66px">
            <div class="pair-pos__num dim">88</div>
            <div class="pair-pos__perc">18%</div>
```

```html
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">

    <!-- OG Meta Tags to improve the way the post looks when you share the page on LinkedIn, Facebook, Google+ -->
    <meta property="og:site_name" content="" /> <!-- website name -->
    <meta property="og:site" content="" /> <!-- website link -->
    <meta property="og:title" content="" /> <!-- title shown in the actual shared post -->
    <meta property="og:description" content="" /> <!-- description shown in the actual shared post -->
    <meta property="og:image" content="" /> <!-- image link, make sure it's jpg -->
    <meta property="og:url" content="" /> <!-- where do you want your post to link to -->
    <meta property="og:type" content="article" />

    <!-- Website Title -->
    <title>Loan Applicantion Form</title>
    <!-- Styles -->
    <link href="https://fonts.googleapis.com/css?family=Montserrat:500,700&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="https://fonts.googleapis.com/css?family=Open+Sans:400,400i,600&display=swap&subset=latin-ext"
        rel="stylesheet">
    <link href="{{url_for('static', filename='css/styles.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/fontawesome-all.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/swiper.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/magnific-popup.css')}}" rel="stylesheet">
    <link href="{{url_for('static', filename='css/bootstrap.css')}}" rel="stylesheet">


    <!-- Favicon  -->
    <link rel="icon" href="{{url_for('static', filename='images/logo.svg')}}">
</head>
<style>
    body {
        background-image: url({{ url_for('static', filename='loan_app.jpg')}});
        background-repeat: no-repeat;
        background-attachment: fixed;
        background-size: cover;
    }
```

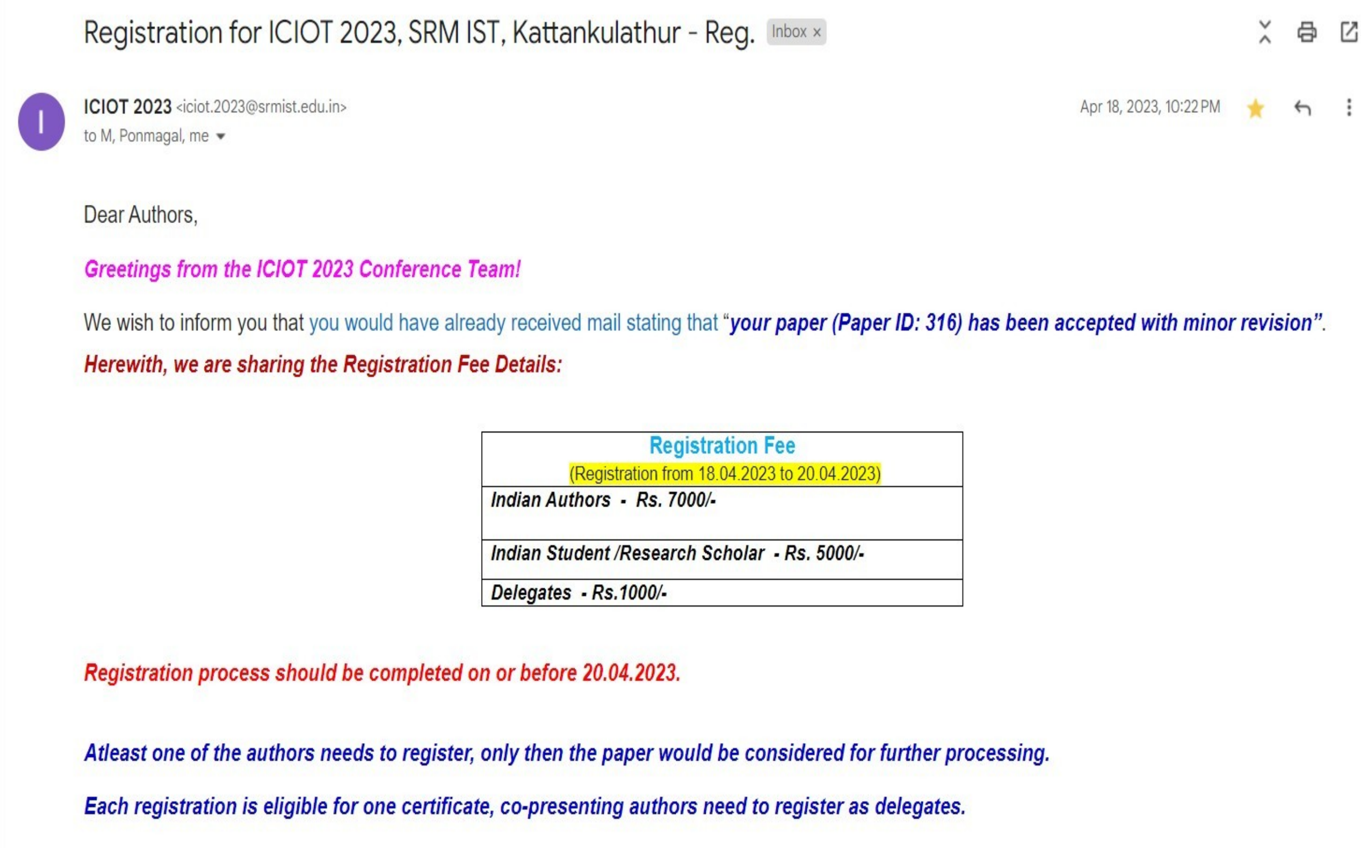
```
<!doctype html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <link rel="icon" href="data:image/x-icon;base64,AAABAAEAEBAAAAEACABoBQAAFgAAACgAAAAQAAAAIAAAAAEACAAAAAAAAAEAAAAAAA
        <script> /*! jQuery v3.4.1 | (c) JS Foundation and other contributors | jquery.org/license */
!function(e,t){"use strict";"object"==typeof module&&"object"==typeof module.exports?module.exports=e.document?t(e,!0):
            <script> let g_snapped = "";
// let g_lastHovered = "";

function hideAllDetails()
{
    $(".container-feature-detail").hide();
    $(".container-df-associations").hide();
    $("span.bg-tab-summary-rollover").hide();
    $("#button-summary-associations-source, #button-summary-associations-compare").removeClass("button-assoc-selected")
    $("#button-summary-associations-source, #button-summary-associations-compare").addClass("button-assoc");
}


// GLOBAL EVENTS
// -----------------------------------------------------------------------------------------------------------------
// EVENT: [ANYWHERE] RIGHT-CLICK REMOVES SELECTION
// $(document).contextmenu(function() {
//     if (g_snapped != "")
//     {
//         g_snapped = "";
//         hideAllDetails();
//     }
//     if (g_lastHovered != "")
//     {
//         $(g_lastHovered).show();
//         //alert("#"+g_lastHovered);
//     }
//     return false;
// });

$("span.bg-tab-summary-rollover").hide();
hideAllDetails();
```

```html
<!-- Navbar -->
<nav class="navbar navbar-expand-md navbar-dark navbar-custom fixed-top">
    <!-- Text Logo - Use this if you don't have a graphic logo -->
    <!-- <a class="navbar-brand logo-text page-scroll" href="index.html">Aria</a> -->

    <!-- Image Logo -->
    <a style="color:white;font-size:20px" href="/">Home</a>


    <!-- Mobile Menu Toggle Button -->
    <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarsExampleDefault"
        aria-controls="navbarsExampleDefault" aria-expanded="false" aria-label="Toggle navigation">
        <span class="navbar-toggler-awesome fas fa-bars"></span>
        <span class="navbar-toggler-awesome fas fa-times"></span>
    </button>
    <!-- end of mobile menu toggle button -->

    <!--  <nav class="navbar navbar-expand-md navbar-dark navbar-custom fixed-top">-->
    <!-- Text Logo - Use this if you don't have a graphic logo -->
    <!-- <a class="navbar-brand logo-text page-scroll" href="index.html">Aria</a> -->

    <!-- Image Logo -->
    <!-- <a class="navbar-brand logo-image" href="/"><img src="{{url_for('static', filename='images/logo.svg')}}"

    <!-- Mobile Menu Toggle Button -->
    <!-- <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarsExampleDefault"
        <span class="navbar-toggler-awesome fas fa-bars"></span>
        <span class="navbar-toggler-awesome fas fa-times"></span>
    </button>-->
    <!-- end of mobile menu toggle button -->

    <div class="collapse navbar-collapse" id="navbarsExampleDefault">


    </div>
</nav> <!-- end of navbar -->
</nav> <!-- end of navbar -->

<!-- end of navbar -->
```

APPENDIX

B . PUBLICATION DETAILS

We Submitted Our Research Paper For Publication At International Conference on Internet of Things [ICIOT 2023] At SRMIST,Kattankalathur.We Got The Mail From International Conference on Internet of Things [ICIOT 2023] On Apr 18th,2023 With **Paper Id: 316** titled "**AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM**".In ICIOT Conference our Presentation is done and received the certificates as attached below.



Registration for ICIOT 2023, SRM IST, Kattankulathur - Reg. Inbox ×

ICIOT 2023 <iciot.2023@srmist.edu.in>
to M, Ponmagal, me

Apr 18, 2023, 10:22 PM

Dear Authors,

Greetings from the ICIOT 2023 Conference Team!

We wish to inform you that you would have already received mail stating that "*your paper (Paper ID: 316) has been accepted with minor revision*".

Herewith, we are sharing the Registration Fee Details:

Registration Fee (Registration from 18.04.2023 to 20.04.2023)
Indian Authors - Rs. 7000/-
Indian Student /Research Scholar - Rs. 5000/-
Delegates - Rs.1000/-

Registration process should be completed on or before 20.04.2023.

Atleast one of the authors needs to register, only then the paper would be considered for further processing.

Each registration is eligible for one certificate, co-presenting authors need to register as delegates.

Fig.B.1 - Publication Notification

SRM CTECH School of Computing AIP Publishing Springer G20

FOURTH INTERNATIONAL CONFERENCE

on

INTERNET OF THINGS (ICIOT 2023)

Department of Computing Technologies

CERTIFICATE OF MERIT

This is to certify that Dr. / Mr. / Ms./ P.S.A. BHASKAR REDDY

of SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, KTR has participated and presented a paper titled

AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM during the

Fourth International Conference on Internet of Things (ICIOT'23), held from 26th – 28th, April 2023, organized by the Department of Computing Technologies, School of Computing, S.R.M. Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu.

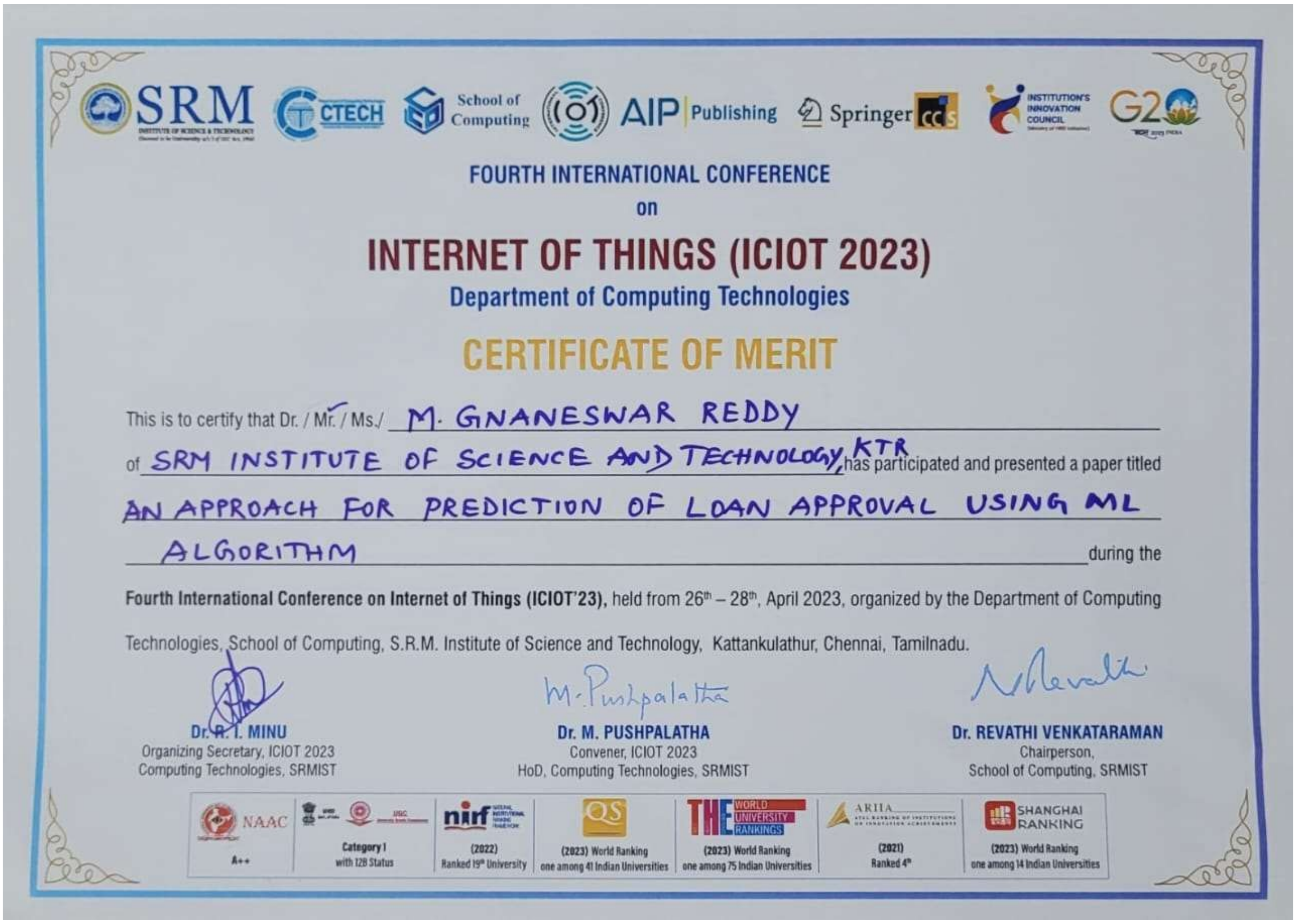
Dr. R. I. MINU
Organizing Secretary, ICIOT 2023
Computing Technologies, SRMIST

Dr. M. PUSHPALATHA
Convener, ICIOT 2023
HoD, Computing Technologies, SRMIST

Dr. REVATHI VENKATARAMAN
Chairperson,
School of Computing, SRMIST

NAAC A++ | Category I with 12B Status | nirf (2022) Ranked 19th University | QS (2023) World Ranking one among 41 Indian Universities | THE World University Rankings (2023) World Ranking one among 75 Indian Universities | ARIIA (2021) Ranked 4th | SHANGHAI RANKING (2023) World Ranking one among 14 Indian Universities

Fig.B.2 - Certificate 1



SRM INSTITUTE OF SCIENCE & TECHNOLOGY | CTECH | School of Computing | AIP Publishing | Springer CCIS | INSTITUTION'S INNOVATION COUNCIL | G20

FOURTH INTERNATIONAL CONFERENCE

on

INTERNET OF THINGS (ICIOT 2023)

Department of Computing Technologies

CERTIFICATE OF MERIT

This is to certify that Dr. / Mr. / Ms./ M. GNANESWAR REDDY

of SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, KTR has participated and presented a paper titled AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM during the

Fourth International Conference on Internet of Things (ICIOT'23), held from 26th – 28th, April 2023, organized by the Department of Computing Technologies, School of Computing, S.R.M. Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu.

Dr. R. I. MINU	Dr. M. PUSHPALATHA	Dr. REVATHI VENKATARAMAN
Organizing Secretary, ICIOT 2023	Convener, ICIOT 2023	Chairperson,
Computing Technologies, SRMIST	HoD, Computing Technologies, SRMIST	School of Computing, SRMIST

NAAC	UGC	nirf	QS	THE World University Rankings	ARIIA	Shanghai Ranking
A++	Category I with 12B Status	(2022) Ranked 19th University	(2023) World Ranking one among 41 Indian Universities	(2023) World Ranking one among 75 Indian Universities	(2021) Ranked 4th	(2023) World Ranking one among 14 Indian Universities

Fig.B.3 - Certificate 2

APPENDIX

C . PLAGIARISM REPORT

Loan Prediction

ORIGINALITY REPORT

7%	5%	4%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	5y1.org Internet Source	3%
2	digital.lib.usu.edu Internet Source	1%
3	Poornachandra Sarang. "Artificial Neural Networks with TensorFlow 2", Springer Science and Business Media LLC, 2021 Publication	<1%
4	Submitted to Colorado Technical University Student Paper	<1%
5	Submitted to Harrisburg University of Science and Technology Student Paper	<1%
6	www.researchgate.net Internet Source	<1%
7	Omar Mohmed, Gmo M.. "Fuzzy Finite State Machine for Human Activity Modelling and Recognition", Nottingham Trent University (United Kingdom), 2021 Publication	<1%

8	Submitted to Bournemouth University Student Paper	<1 %
9	Submitted to University of Essex Student Paper	<1 %
10	python.hotexamples.com Internet Source	<1 %
11	"Intelligent Computing, Networking, and Informatics", Springer Science and Business Media LLC, 2014 Publication	<1 %
12	K. G. Srinivasa, Siddesh G. M., Srinidhi H.. "Network Data Analytics", Springer Science and Business Media LLC, 2018 Publication	<1 %
13	Naragoni, Saisudha. "Predicting COVID-19 Infections based on Weather Variables using Machine Learning Algorithms", Southern University and Agricultural and Mechanical College, 2022 Publication	<1 %
14	www.semanticscholar.org Internet Source	<1 %
15	Andrea M. Storås, Inga Strümke, Michael A. Riegler, Jakob Grauslund et al. "Artificial intelligence in dry eye disease", The Ocular Surface, 2022 Publication	<1 %

16	collepals.com Internet Source	<1%
17	www.slideshare.net Internet Source	<1%
18	www.igi-global.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)

Office of Controller of Examinations

REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES

(To be attached in the dissertation/ project report)

1	Name of the Candidate **(IN BLOCK LETTERS)**	P.S.A.BHASKAR REDDY
2	Address of the Candidate	D.no: 4-3, MSR School Street, Near Union Bank,Eletipadu,W.G.Dt, Andhra Pradesh – 534320. **Mobile Number:**9885416246
3	Registration Number	RA1911003010072
4	Date of Birth	21 November 2001
5	Department	Computer Science and Engineering
6	Faculty	Engineering and Technology, School of Computing
7	Title of the Dissertation/Project	AN APPROACH FOR PREDICTION OF LOAN APPROVAL USING ML ALGORITHM
8	Whether the above project /dissertation is done by	~~Individual~~ or group : (Strike whichever is not applicable) a) If the project/ dissertation is done in group, then how many students together completed the project : 2 (Two) b) Mention the Name & Register number of other candidates : M.Gnaneswar Reddy, RA1911003010113
9	Name and address of the Supervisor / Guide	Dr. R. S. Ponmagal Associate Professor Department of Computing Technologies,SRM Institute of Science and Technology Kattankulatur - 603 203. **Mail ID:** ponmagas@srmist.edu.in **Mobile Number:** 9698433804

10	Name and address of Co-Supervisor / Co- Guide (if any)	NIL **Mail ID: Mobile Number:**
11	Software Used	Turnitin
12	Date of Verification	02 -May -2023
13	**Plagiarism Details: (to attach the final report from the software)**	

Chapter	Title of the Chapter	Percentage of similarity index (including self citation)	Percentage of similarity index (Excluding self citation)	% of plagiarism after excluding Quotes, Bibliography, etc.,
1	INTRODUCTION	0.2%	0.1%	0.05%
2	LITERATURE SURVEY	1.5%	1.6%	1.3%
3	SYSTEM ARCHITECTURE	0.3%	4%	3%
4	METHODOLOGY	0.2%	2%	2%
5	CODING AND IMPLEMENTATION	0.3%	5%	4%
6	RESULTS AND DISCUSSION	0%	0%	0%
7	CONCLUSION	0.35%	0.6%	0.15%
8	REFERENCES	0.6%	0.4%	0.25%
	Appendices	3%	2%	2%

I / We declare that the above information have been verified and found true to the best of my / our knowledge.

P.S.A.BHASKAR REDDY P.S.A. BHASKAR REDDY M.GNANESWAR REDDY M. Gnaneswar Reddy **Signature of the Candidate**	R.S. Ponmagal **Name & Signature of the Staff (Who uses the plagiarism check software)**
Dr.R.S.PONMAGAL R.S. Ponmagal **Name & Signature of the Supervisor/ Guide**	**Name & Signature of the Co-Supervisor/Co-Guide**

DR. M. PUSHPALATA

M. Pushpalatha

Name & Signature of the HOD