# VISVESVARAYA TECHNOLOGICAL UNIVERSITY



JNANA SANGAMA, BELAGAVI – 590 018

### An Internship Project Report on

***Loan Prediction***

Submitted in partial fulfillment of the requirements for the VII Semester of degree of **Bachelor of Engineering in Information Science and Engineering** of Visvesvaraya Technological University, Belagavi

### Submitted By

**Akash Anand**

## 1RN18IS010

### Under the Guidance of

**Mrs. Hema N Assistant Professor Department of ISE**

**Department of Information Science and Engineering**

**RNS Institute of Technology**

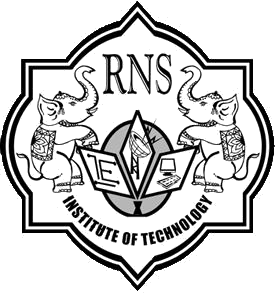
**Dr. Vishnuvaradhan Road, Rajarajeshwari Nagar post, Channasandra, Bengaluru-560098**

**2021-2022**

RNS INSTITUTE OF TECHNOLOGY

**Dr. Vishnuvaradhan Road, Rajarajeshwari Nagar post, Channasandra, Bengaluru - 560098**

**DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING**

****

# CERTIFICATE

Certified that the Internship work entitled ***Loan Prediction*** has been successfully completed by **Akash Anand (1RN18IS010)** a Bonafide student of **RNS Institute of Technology, Bengaluru** in partial fulfillment of the requirements of 8th semester for the award of degree in **Bachelor of Engineering in Information Science and Engineering** of **Visvesvaraya Technological University, Belagavi** during academic year **2021-2022**. The internship report has been approved as it satisfies the academic requirements in respect of internship work for the said degree.

|  |  |  |
| --- | --- | --- |
| **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** | **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** |
| **Mrs. Hema N** | **Dr. Suresh L** | **Dr. M K Venkatesha** |
| Internship Guide | Professor and HoD | Principal |
| Assistant Professor | Department of ISE | RNSIT |
| Department of ISE | RNSIT |  |
|  | **External Viva** |  |
| **Name of the Examiners** |  | **Signature with Date** |
| **1.** | | **1. \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ \_\_\_ \_\_ \_\_** |
| **2.** | | **2. \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_ \_\_\_ \_\_ \_\_** |

I, **Akash Anand [USN: 1RN18IS010]** student of VII Semester BE, in Information Science and Engineering, RNS Institute of Technology hereby declare that the Internship work entitled ***Loan Prediction*** has been carried out by us and submitted in partial fulfillment of the requirements for the *VII Semester degree of* ***Bachelor of Engineering in Information Science and Engineering*** *of Visvesvaraya Technological University, Belagavi* during academic year 2021-2022.

Place: Bengaluru Date:

### Akash Anand (1RN18IS010)

Loans are the core business of banks. The main profit comes directly from the loan’s interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don’t have assurance if the applicant is able to repay the loan with no difficulties.

The two most pressing issues in the banking sector are: 1) How risky is the borrower? 2) Should i lend to the borrower given the risk? The response to the first question dictates the borrower's interest rate. Interest rate, among other things (such as time value of money), tests the riskiness of the borrower, i.e. the higher the interest rate, the riskier the borrower. I will then decide whether the applicant is suitable for the loan based on the interest rate. Lenders (investors) make loans to creditors in return for the guarantee of interest-bearing repayment. That is, the lender only makes a return (interest) if the borrower repays the loan. However, whether he or she does not repay the loan, the lender loses money. Banks make loans to customers in exchange for the guarantee of repayment.

Some would default on their debts, unable to repay them for a number of reasons. The bank retains insurance to minimize the possibility of failure in the case of a default. The insured sum can cover the whole loan amount or just a portion of it. Banking processes use manual procedures to determine whether or not a borrower is suitable for a loan based on results. Manual procedures were mostly effective, but they were insufficient when there were a large number of loan applications. At that time, making a decision would take a long time. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies.

At the very onset i would like to place our gratefulness to all those people who helped me in making the Internship a successful one.

Coming up, this internship to be a success was not easy. Apart from the sheer effort, the enlightenment of the very experienced teachers also plays a paramount role because it is, they who guided us in the right direction.

First of all, i would like to thank the **Management of RNS Institute of Technology** for providing such a healthy environment for the successful completion of internship work.

In this regard, i express sincere gratitude to our beloved Principal **Dr. M K Venkatesha,** for providing us all the facilities.

I am extremely grateful to our own and beloved Professor and Head of Department of Information science and Engineering, **Dr. Suresh L**, for having accepted to patronize us in the right direction with all her wisdom.

I place our heartfelt thanks to **Ms. Hema N** AssistantProfessor, Department of Information Science and Engineering for having guided internship and all the staff members of the department of Information Science and Engineering for helping at all times.

I thank **Mr. Aman Upadhayaya, Instructor, NASTECH**, for providing the opportunity to be a part of the Internship program and having guided me to complete the same successfully.

I also thank our internship coordinator **Dr. R Rajkumar,** Associate Professor, Department of Information Science and Engineering. We would thank our friends for having supported us with all their strength and might. Last but not the least, we thank our parents for supporting and encouraging me throughout. I have made an honest effort in this assignment.

[CERTIFICATE](#_bookmark0) [1](#_bookmark0)

[DECLARATION](#_bookmark1) [2](#_bookmark1)

[ABSTRACT](#_bookmark2) [I](#_bookmark2)

[ACKNOWLEDGMENT](#_bookmark3) [II](#_bookmark3)

[TABLE OF CONTENTS](#_bookmark4) [III](#_bookmark4)

[LIST OF FIGURES](#_bookmark5) [IV](#_bookmark5)

[ABBREVIATIONS](#_bookmark7) [V](#_bookmark7)

1. [INTRODUCTION](#_bookmark8) [1](#_bookmark8)
   1. [BACKGROUND](#_bookmark9) [2](#_bookmark9)
   2. [REQUIREMENTS](#_bookmark10) [3](#_bookmark10)
2. [LITERATURE](#_bookmark11) SURVEY [3](#_bookmark11)

3.1[CURRENT DESIGN](#_bookmark12) [4](#_bookmark12)

3.2[PROPOSED DESIGN](#_bookmark13) [4](#_bookmark13)

1. [SYSTEM DESIGN](#_bookmark14) [5](#_bookmark14)
2. [IMPLEMENTATION](#_bookmark17) 6

4.2 [CODE SEGMENT](#_bookmark21) 7

1. [RESULTS](#_bookmark23) 11
2. [CONCLUSION AND FUTURE ENHANCEMENT](#_bookmark24) 12

[REFERNCES](#_bookmark25) 13

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Description** | **Page No.** |
| 3.3 | System Design | 5 |
| 4.1.1 | Import the data | 7 |
| 4.1.2 | Plotting the histogram of credit policy | 7 |
| 4.1.3 | Plotting the histogram of purpose of loan | 8 |
| 4.1.4 | Graph of fico Vs Interest rate | 8 |
| 4.1.5 | Graph of fico Vs Interest rate w.r.t fully\_paid | 9 |
| 4.1.6 | Train test splitting | 10 |
| 4.1.7 | Training for decision tree and random forest | 10 |

# ABBREVIATIONS

# 

# INTRODUCTION

Python is developed by Guido van Rossum. Guido van Rossum started implementing Python in 1989. Python is a very simple programming language so even if you are new to programming, you can learn python without facing any issues.

Interesting fact: Python is named after the comedy television show Monty Python’s Flying Circus. It is not named after the Python snake.

Merits of Python:

1. Readable: Python is a very readable language.

2. Easy to Learn: Learning python is easy as this is a expressive and high level programming language, which means it is easy to understand the language and thus easy to learn.

3. Cross platform: Python is available and can run on various operating systems such as Mac, Windows, Linux, Unix etc. This makes it a cross platform and portable language.

4. Open Source: Python is a open source programming language.

5. Large standard library: Python comes with a large standard library that has some handy codes and functions which we can use while writing code in Python.

6. Free: Python is free to download and use. This means you can download it for free and use it in your application. See: Open Source Python License. Python is an example of a FLOSS (Free/Libre Open Source Software), which means you can freely distribute copies of this software, read its source code and modify it.

7. Supports exception handling: If you are new, you may wonder what is an exception? An exception is an event that can occur during program exception and can disrupt the normal flow of program. Python supports exception handling which means we can write less error prone code and can test various scenarios that can cause an exception later on.

8. Advanced features: Supports generators and list comprehensions. We will cover these features later.

9. Automatic memory management: Python supports automatic memory management which means the memory is cleared and freed automatically. You do not have to bother clearing the memory.



1

## BACKGROUND

There are so many different types of Machine Learning systems that it is useful to classify them in broad categories, based on the following criteria:

• Whether or not they are trained with human supervision (supervised, unsupervised, semi supervised, and Reinforcement Learning)

• Whether or not they can learn incrementally on the fly (online versus batch learning)

• Whether they work by simply comparing new data points to known data points, or instead by detecting patterns in the training data and building a predictive model, much like scientists do (instance-based versus model-based learning)

In finance, a loan is the lending of money by one or more individuals, organizations, or other entities to other individuals, organizations, etc. The recipient (i.e. the borrower) incurs a debt, and is usually liable to pay interest on that debt until it is repaid, and also to repay the principal amount borrowed. To read more check out Wikipedia. The whole process of ascertaining if a burrower would pay back loans might be tedious hence the need to automate the procedure.

**2.** **LITERATURE SURVEY**

1. “Loan Prediction using Decision Tree and Random Forest “Author- Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar Year-2020.

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

2. “Loan Prediction System Using Decision Tree and Random Forest Algorithms”

Authors- Shubham Chaudhary, Vishal Baliyan, Yatharth Katheria

the idea of this undertaking is to gather loan data from the Lending Club website and use machine learning techniques on this data to extract important information and predict if a customer would be able to get a loan from a bank or not. Most banks are renewing their business models and switching to Machine Learning methodology. In this paper, I have discussed classifiers based on Machine and deep learning models on real data in predicting loan default

probability. The most important features from various models are selected and then used in the modelling process to test the stability of Random Forest classifiers and Decision Tree Classifier by comparing their performance on data.

3. A. Goyal and R. Kaur, “Accuracy Prediction for Loan Risk Using Machine Learning Models”.

Estimating the probability that an individual would default on their loan, is useful for banks to make a decision whether to approve a loan to the individual or not. In this paper, I find the accuracy of several models in R language and evaluate it to establish the finest model to forecast the finance status for an organization. I did the experiment five times on the same

data set and find the experimental results that show the Tree Model for Genetic Algorithm is the best model for forecasting the finance for costumers.

4. A. Goyal and R. Kaur, “A survey on Ensemble Model for Loan Prediction”, International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.

In this paper I found the ensemble model that is combination of two or more algorithms and give better results as compared to stand alone models. The performance is also enhanced through the ensemble model.

## REQUIREMENTS

Software Requirements

|  |  |
| --- | --- |
| **Name of Components** | **Specification** |
| OperatingSystem | Windows 10, Ubuntu 20.0. |
| python | conda |
| Browser | Chrome, IE |
| Integrated Development Environment | Jupytr notebooks |

Hardware Requirements

|  |  |
| --- | --- |
| **Name of Components** | **Specification** |
| Processor | 10TH Gen CORE i3 Processor |
| RAM | 8GB |
| Hard Disk | 1. GB SSD |

# SYSTEM DESIGN

## CURRENT DESIGN

Banking processes use manual procedures to determine whether or not a borrower is suitable for a loan based on results. Manual procedures were mostly effective, but they were insufficient when there were a large number of loan applications. At that time, making a decision would take a long time.

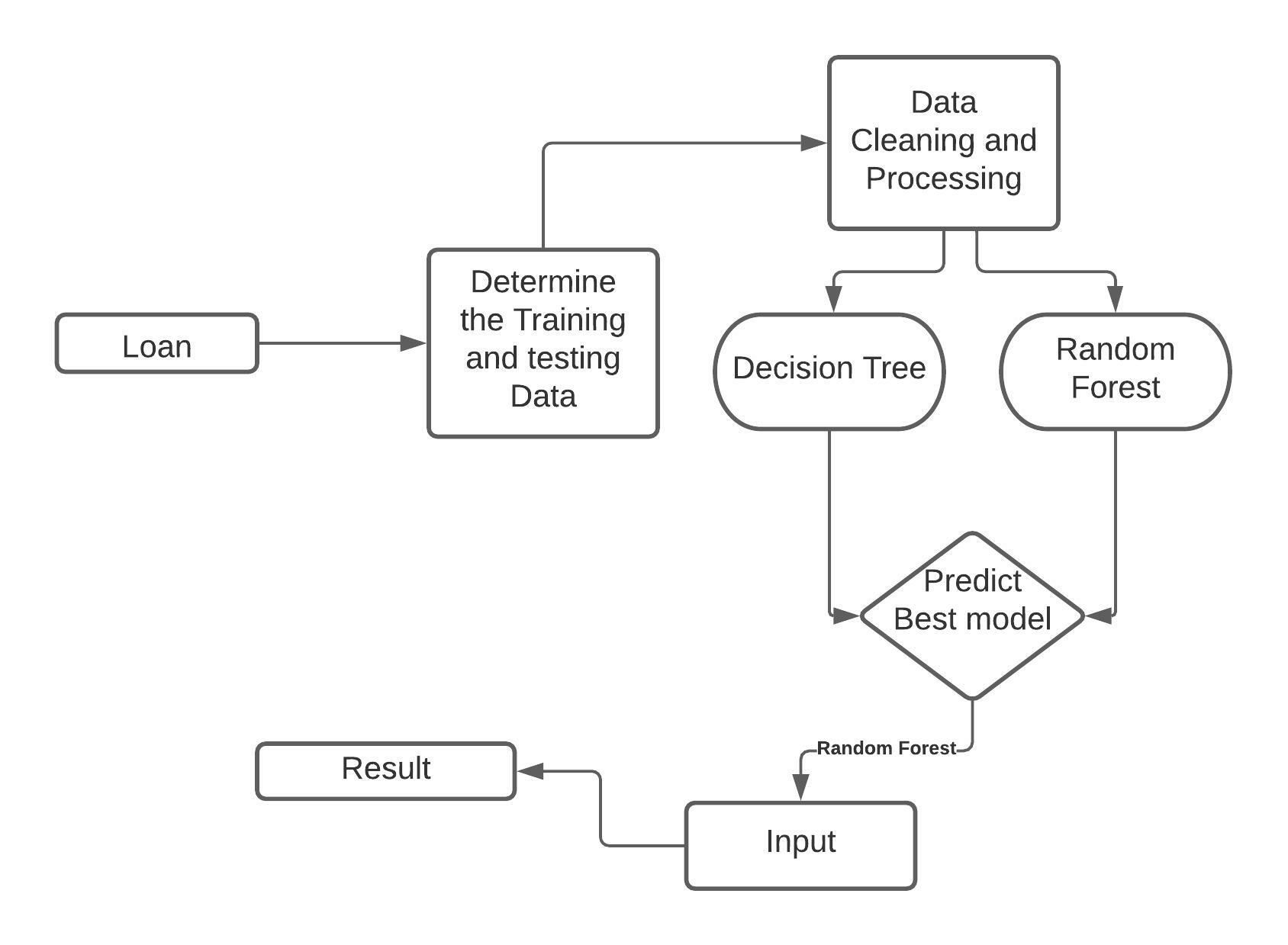
## 3.2 PROPOSED DESIGN

Our aim from the project is to make use of pandas, matplotlib, & seaborn libraries from python to extract insights from the data and scikit-learn libraries for machine learning.

Secondly, to learn how to hypertune the parameters using grid search cross validation for the Random forest machine learning model.

And in the end, to predict whether the loan applicant can replay the loan or not using voting ensembling techniques of combining the predictions from multiple machine learning algorithms.

# 3.3 SYSTEM DESIGN



3.3 System design/architecture

Phase 1 – Collection of data:

Through a Kaggle competition.

Phase 2 – Data preprocessing:

I have to perform various tasks to observe the patterns in images. discriminate some images which are not useful. Images with more number of detection, images with law brightness, images without bounding box.

Phase 3 – Comparing different models and training

Phase 4 – Result/Deploying the model

# 4.IMPLEMENTATION

**4.1. EXPLORATORY DATA ANALYSIS**

* Purpose : The purpose of the loan (takes values "credit\_card", "debt\_consolidation", "educational", "major\_purchase", "small\_business", and "all\_other").
* int.rate : The interest rate of the loan, as a proportion (a rate of 11% would be stored as 0.11). Borrowers judged by LendingClub.com to be more risky are assigned higher interest rates.
* Installment : The monthly installments owed by the borrower if the loan is funded.
* log.annual.inc : The natural log of the self-reported annual income of the borrower.
* Dti : The debt-to-income ratio of the borrower (amount of debt divided by annual income).
* Fico : The FICO credit score of the borrower.
* days.with.cr.line : The number of days the borrower has had a credit line.
* revol.bal : The borrower's revolving balance (amount unpaid at the end of the credit card billing cycle).
* revol.util : The borrower's revolving line utilization rate (the amount of the credit line used relative to total credit available).
* inq.last.6mths : The borrower's number of inquiries by creditors in the last 6 months.
* delinq.2yrs : The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
* pub.rec : The borrower's number of derogatory public records (bankruptcy filings, tax liens, or judgments).

**4.2. MACHINE LEARNING METHODS**

Decision tree: A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a coin toss comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes).

Random forest: Random forest or random decision forests are an ensemble learning method used for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.

* 1. **CODE SEGMENT**

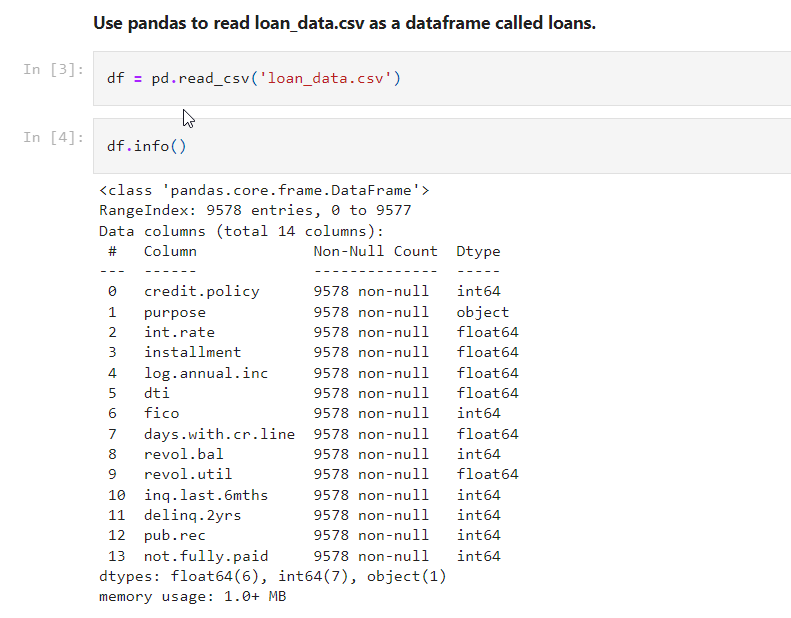
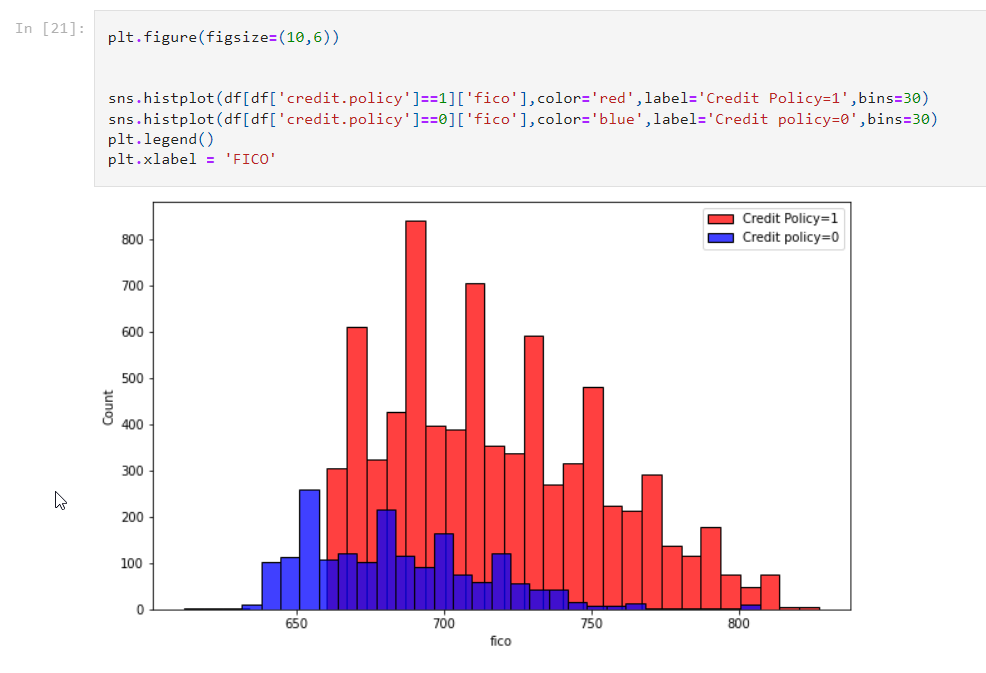
1. **Get the data**. 

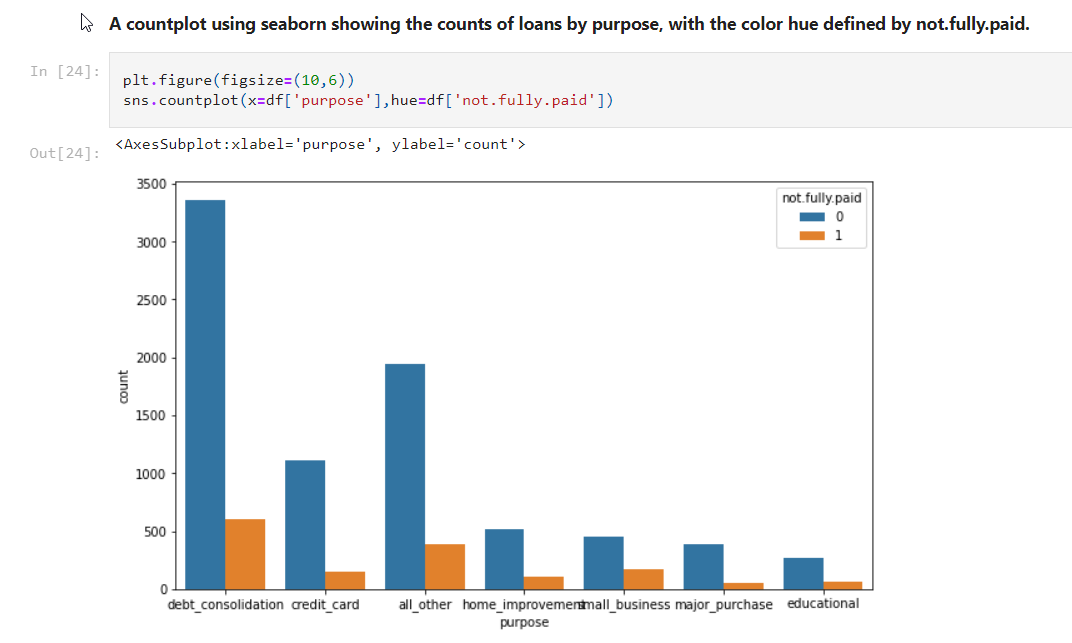
Fig 4.1.1 importing the data.

1. **Exploratory data analysis.**

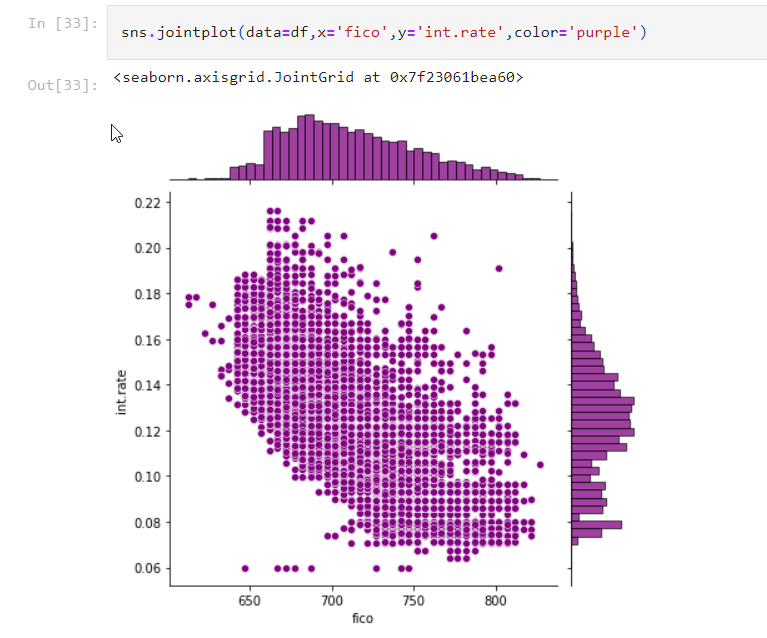


4.1.2 plotting histogram for credit policy

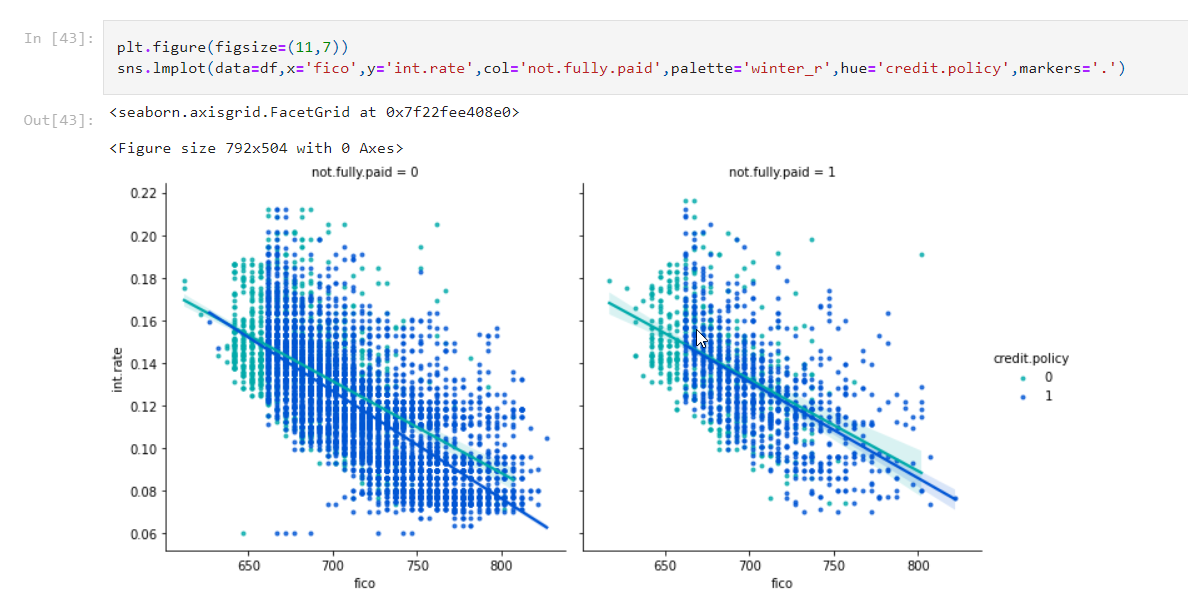
**-**



4.1.3 plotting histogram for purpose of loan

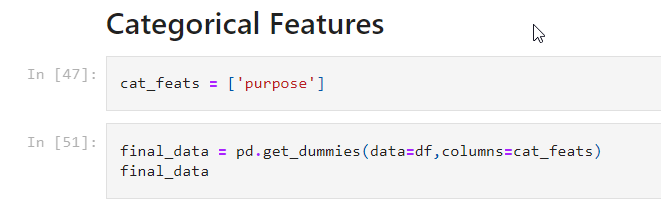


4.1.4 graph for fico Vs interest rate



4.1.5 graph on fico Vs Int\_rate w.r.t fully paid

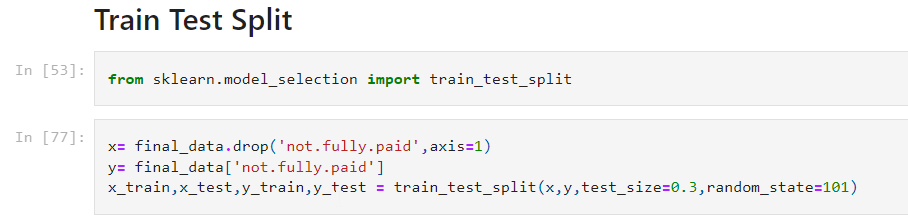
1. **Categorical Features**



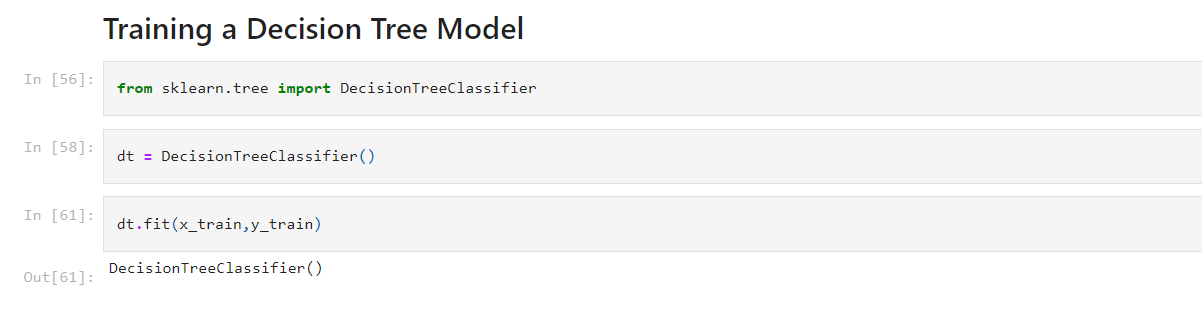
credit.policy, int.rate, installment, log.annual.inc, dti, fico, days.with.cr.line, revol.bal,

revol.util, inq.last.6mths, delinq.2yrs, pub.rec, not.fully.paid, purpose\_all\_other, purpose\_credit\_card, purpose\_debt\_consolidation, purpose\_educational, purpose\_home\_improvement , purpose\_major\_purchase, purpose\_small\_business

1. **Training**

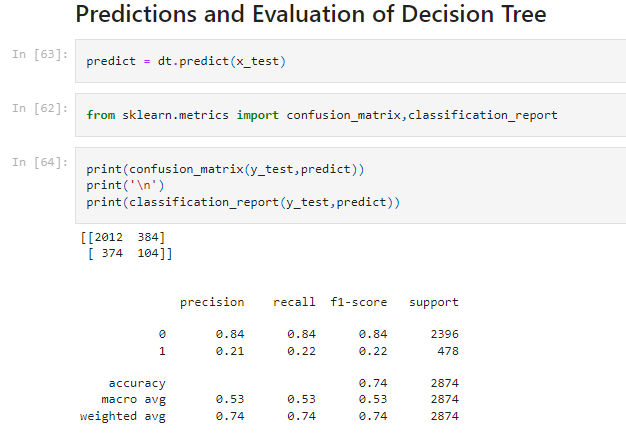
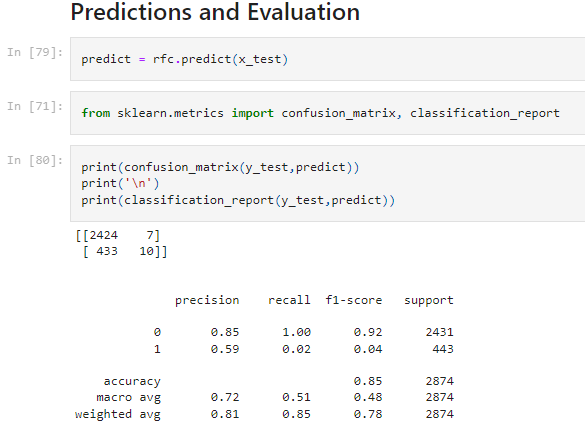


4.1.6 train test splitting

4.1.7 Training for decision and training random forest

# RESULTS

Of Random Forest

# CONCLUSION AND FUTURE ENHANCEMENT

The main purpose of this work is to classify and analyze the nature of the loan applicants. From a proper analysis of available data and constraints of the banking sector, it can be concluded that by keeping safety in mind that this modelling is much effective or highly efficient. This application is operating efficiently and fulfilling all the major requirements of Banker. Although the application is flexible with various systems and can be plugged effectively.

The module can be enhanced to cover below business scenarios

This paper work can be extended to higher level in future so the software could have some better changes to make it more reliable, secure, and accurate.

Thus, the system is trained with a present data sets which may be older in future

so, it can also take part in new testing to be made such as to

pass new test cases.

# REFERNCES

* Loan Prediction using Decision Tree and Random Forest Kshitiz Gautam, Arun Pratap Singh, Keshav Tyagi, Mr. Suresh Kumar, IRJET
* Loan Prediction System Using Decision Tree and Random Forest Algorithms ,Shubham Chaudhary, Vishal Baliyan, Yatharth Katheria, IJETIE
* A. Goyal and R. Kaur, “Accuracy Prediction for Loan Risk Using Machine Learning Models”.
* A. Goyal and R. Kaur, “A survey on Ensemble Model for Loan Prediction”, International Journal of Engineering Trends and Applications (IJETA), vol. 3(1), pp. 32-37, 2016.
* <https://en.wikipedia.org/wiki/Exploratory_data_analysis>
* https://www.experian.com/blogs/ask-experian/credit-education/score- basics/what-is-a-good-credit-score/