**A PROJECT REPORT**

**ON**

**LOAN PREDICTOR SYSYTEM**

**By**

**Anshul Tripathi**

**20021996**

**DEPARTMENT OF COMPUTER SCIENCE &** **ENGINEERING**

**GRAPHIC ERA DEEMED UNIVERSITY**

**DEHRADUN**

**2021-2022**

**DECLARATION**

I, Anshul Tripathi student of B-tech, Semester 4, Department of Computer Science and Engineering, Graphic Era Deemed University, Dehradun, declare that the technical project work entitled “**LOAN PREDICTOR SYSTEM**” has been carried out by me and submitted in partial fulfilment of the course requirements for the award of degree in B- tech of Graphic Era Deemed University during the academic year 2021-2022. The matter embodied in this synopsis has not been submitted to any other university or institution for the award of any other degree or diploma.

DATE : 14/07/2022

**ACKNOWLEDGEMENT**

Here by I am submitting the project report on **“Loan Predictor System”** as per the scheme of Graphic Era Deemed University, Dehradun.

I would like to express our sincere gratitude to **Dr. Devesh Pratap Singh,** Head of Dept. of Computer Science, for providing a congenial environment to work in and carry out our project.

I consider it mine cardinal duty to express the deepest sense of gratitude to **Prof. Kireet Joshi** , Department of Computer Science and Application for the invaluable guidance extended at every stage and in every possible way.

Finally I am very much thankful to all the faculty members of the Department of Computer Science and Technology, friends and our parents for their constant encouragement, support and help throughout the period of project conduction.

**Anshul Tripathi**

**Roll No - 20021996**

****

**CERTIFICATE**

This is to certify that the project report entitled “Loan Predictor System” is a bonafide project work carried out by Anshul Tripathi Roll no- 2017311 in partial fulfilment of award of degree of B.Tech of Graphic Era Deemed University, Dehradun during the academic year 2021-2022. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated. The project has been approved as it satisfies the academic requirements associated with the degree mentioned.

**Dr .Devesh Pratap Singh**

**HOD (COMPUTER SCIENCE)**

**PROBLEM STATEMENT AND MOTIVATION FOR PROJECT**

The two most pressing issues in the banking sector are: 1) How risky is the borrower? 2) Should we 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. We 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. This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status). These reports make a bank manager's job simpler and quicker.

LANGUAGE USED :

PYTHON

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn’t specialized for any specific problems. This versatility, along with its beginner-friendliness, has made it one of the most-used programming languages today. A survey conducted by industry analyst firm RedMonk found that it was the second-most popular programming language among developers in 2021 .

Python is commonly used for developing websites and software, task automation, data analysis, and data visualization. Since it’s relatively easy to learn, Python has been adopted by many non-programmers such as accountants and scientists, for a variety of everyday tasks, like organizing finances.

“Writing programs is a very creative and rewarding activity,” says University of Michigan and Coursera instructor Charles R Severance in his book *Python for Everybody.* “You can write programs for many reasons, ranging from making your living to solving a difficult data analysis problem to having fun to helping someone else solve a problem.”

Some features of python :

* Data analysis and machine learning
* Web development
* Automation or scripting
* Software testing and prototyping
* Everyday task

SKILL USED:

MACHINE LEARNING

Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values.

Recommendation engines are a common use case for machine learning. Other popular uses include fraud detection, spam filtering, malware threat detection, business process automation (BPA) and Predictive maintenance.

**Why it is important?**

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

### How to choose the right machine learning model

The process of choosing the right machine learning model to solve a problem can be time consuming if not approached strategically.

**Step 1:** Align the problem with potential data inputs that should be considered for the solution. This step requires help from data scientists and experts who have a deep understanding of the problem.

**Step 2:** Collect data, format it and label the data if necessary. This step is typically led by data scientists, with help from data wranglers.

**Step 3:** Chose which algorithm(s) to use and test to see how well they perform. This step is usually carried out by data scientists.

**Step 4:** Continue to fine tune outputs until they reach an acceptable level of accuracy. This step is usually carried out by data scientists with feedback from experts who have a deep understanding of the problem.

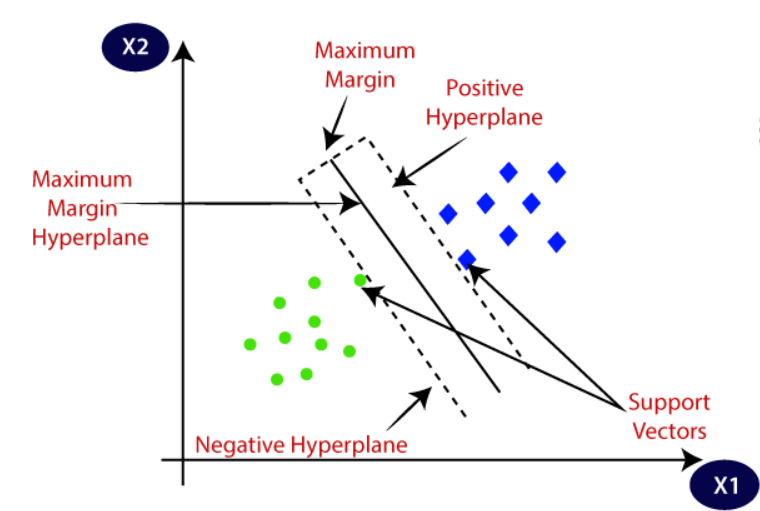
**METHODOLOGY FOLLOWED**

**ALGORUTHM USED : SVM(Support vector machine)**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:

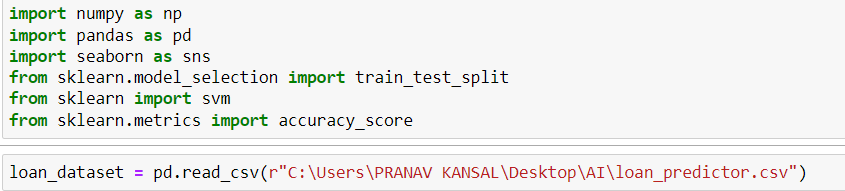


DATASET USED :

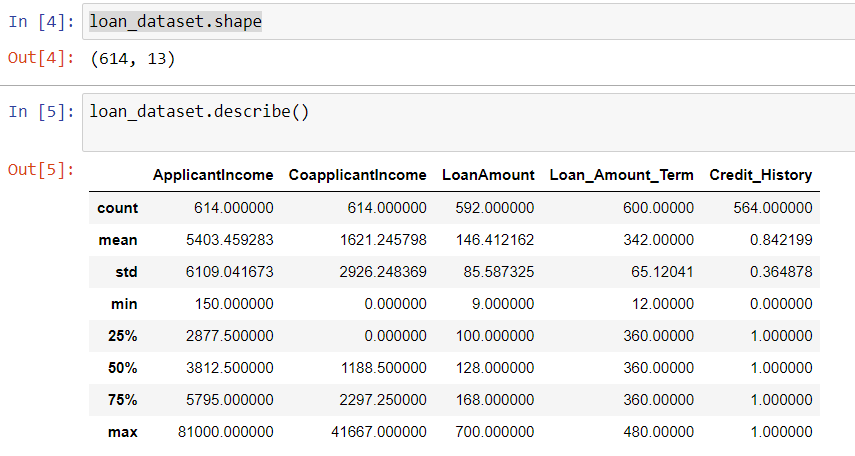
Dataset is collected from the URL:

https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset

**Data Reading:**

****

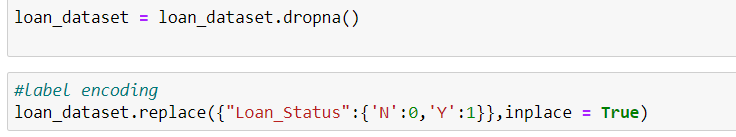
**Exploring Dataset:**

****

Dataset contains 614 rows and 13 columns like Loan Id, Gender, Married , Appplicant Income, Co- applicant Income, Dependants , Property Area etc.

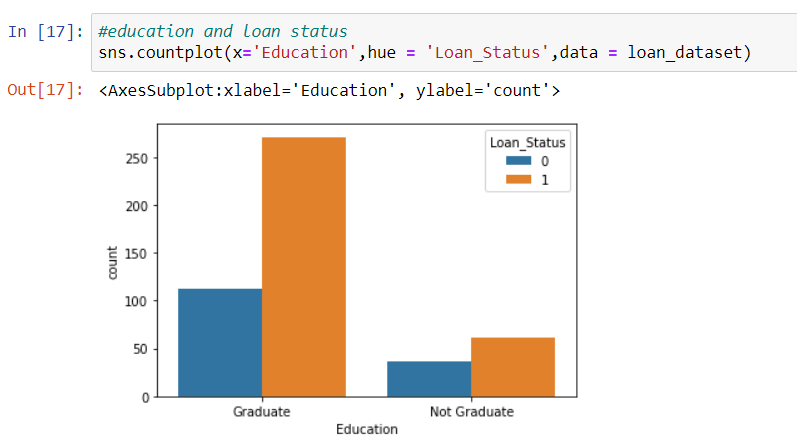
**Data Pre Processing** :

Before Proceeding Further We must pre-process our data such that when we need it after it must be sorted and processed. I have used dropna() and replace() functions to drop the null values and replace string values to processable integer value.



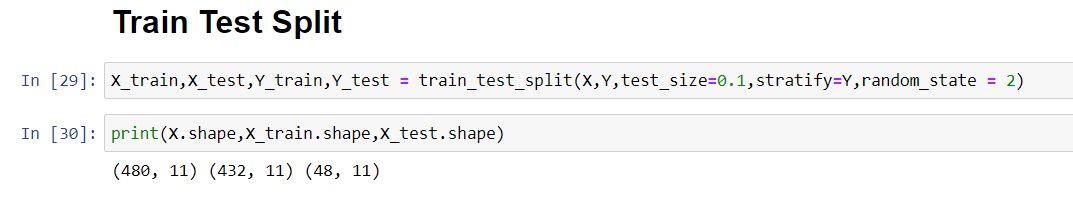
**Data Visualisation :**

Before moving further the dataset is visualised. And I have seen the relationship between different independent variables(Like Dependants,Gender,Married) and dependant variable(Like Loan Status).



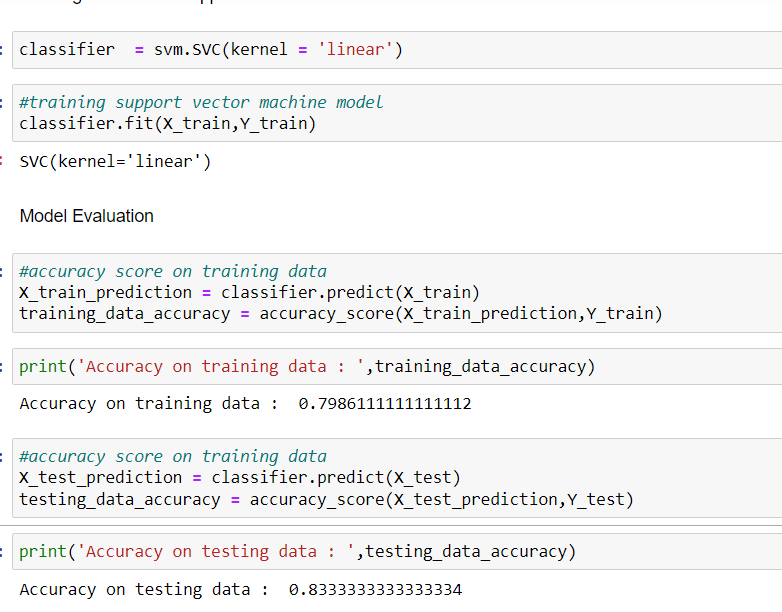
**Splitting data for training and testing :**

A function is created so that we can create the sequence for training and testing. In training and testing we separate each of them with the help of x\_train, y\_train and x\_test, y\_test.



**Implementation of SVM Model:**

In the next step I have trained and implemented the model SVM which have been imported from library sklearn. After implementation I have checked the accuracy of model on the dataset.



Now our Loan Predictor system is ready and can be converted to predicting system

**Conclusion:**

In this article, we explored SVM and built a Loan Predictor system that will be of great help in banking system.

**Refernce :**

1. <https://the-learning-machine.com/article/dl/long-short-term-memory>
2. <https://www.youtube.com/watch?v=XckM1pFgZmg>