# Predictive Model for determining whether a person should be granted a Loan or not

**NAME : CHETHAN P**

**USN : 1BY22MC011**

**COLLEGE: BMS Institute of Technology and Management, Bengaluru**

**PROJECT CODE : P4**

# 1.INTRODUCTION

Generally, loan prediction involves the lender looking at various background information about the applicant and deciding whether the bank should grant the loan. Parameters like credit score, loan amount, lifestyle, career, and assets are the deciding factors in getting the loan approved. If, in the past, people with parameters similar to yours have paid their dues timely, it is more likely that your loan would be granted as well.

Machine learning algorithm can exploit this dependency on past experiences and comparisons with other applicants and formulate a data science problem to predict the loan status of a new applicant using similar rules.

Several collections of data from past loan applicants use different features to decide the loan status. A machine learning model can look at this data, which could be static or time-series, and give a probability estimate of whether this loan will be approved. Let's look at some of these datasets.

# 2. Methodology

**Data Collection:**

To develop the churn prediction model, historical data including customer attributes and churn status is required. Data can be collected from various sources, such as customer databases, CRM systems, or transactional records. The dataset should include both churned and active customers, along with relevant features

**Data Preprocessing:**

Before applying the KNN algorithm, the dataset needs to be preprocessed. This step involves handling missing values, encoding categorical variables, and normalizing numerical features. Additionally, data should be divided into training and testing sets to evaluate the model's performance accurately.

**K-Nearest Neighbour Algorithm:**

The KNN algorithm is a non-parametric classification technique that assigns labels to new data points based on their proximity to the nearest neighbors in the training set. The key steps involved in the KNN algorithm are as follows:

a) Compute the distance between the new data point and each instance in the training set.

b) Select the K nearest neighbors based on the calculated distance.

c) Assign the most common class label among the K neighbors as the predicted class label for the new data point.

# 3.TOOLS EXPOSED

## 3.1.JUPYTER NOTEBOOK:

The jupyter notebook app is a server-client application that allows editing and running notebook documents via a web browser. The jupyter notebook app can be executed on a local desktop requiring no internet access or can be installed on a remote server and accessed through the internet. In addition to displaying/editing/running notebook documents, the jupyter notebook app has a dashboard, a control panel showing local files and allowing to open notebook documents or shutting down their kernels. A notebook kernel is a computational engine that executes the code contained in a notebook document. The jupyter kernel referenced in this guide executes python code. Kernels for many other language exist. When you open a notebook document the associated kernel is automatically launched. When the notebook is executed the kernel performs the computation and produces the results. Depending on the type of computations the kernel may consume significant CPU and RAM. Note that the RAM is not released until the kernel is shut down. The notebook dashboard is the component which is shown first when you launch jupyter notebook app. The notebook dashboard is mainly used to open notebook documents and manage the running kernels. The jupyter notebook extends the console based approach to interactive computing in a qualitatively new direction, providing a web based application suitable for capturing the whole computation process: developing, computing and executing code as well as communicating the results. The jupyter notebook combines two components a web application and notebook documents. A web application: A web browser based tool for interactive authoring of documents which combine explanatory text, mathematics, computations and their rich media output. Notebook documents: A representation of all content visible in the web application, including inputs and outputs of the computations, explanatory text, mathematics, images and rich media representation of objects.

## 3.2.GOOGLE COLAB:

Collaborator or Colab for short, is a product from Google research. Colab allows anybody to write and execute arbitrary python code through the browser and is especially well suited to machine learning, data analysis and education. More technically Colab is a hosted Jupiter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs. Colab resources are not guaranteed and not unlimited, and the usage limits sometimes fluctuate. This is necessary for Colab to be able to provide resources free charge. Resources in Colab are prioritized for interactive use cases.

We prohibit actions associated with bulk compute, actions that negatively impact others as well as actions associated with bypassing the policies. Jupyter is the open source project on which the Colab is based. Colab allows you to use and share Jupyter notebooks with others without having to download, install or run anything. You can search Colab notes using google drive. Clicking on the Colab logo at the top left of the notebook view will show all notebooks in drive. You can also search for notebooks that you have opened recently by clicking on file and then open notebook. Google drive operations can time out when the number of folders or subfolders in a folder grows too large. If thousands of items are directly contained in the top level "My drive" folder then mounting the drive will likely time out. Repeated attempts may eventually succeed as failed attempts cache partial state locally before timing out. Colab is able to provide resources free of cost in part by having dynamic usage limits that sometimes fluctuate this means that overall usage limits as well as idle timeout periods, maximum VM lifetime, GPU types available and other factors vary over time. Colab does not publish these limits in parts because they can vary quickly. This is necessary for Colab to be able to provide access these resources free of charge. Colab works with most of the major browsers and is most thoroughly tested with the latest versions of Chrome, Firefox and Safari.

## 4.TASK PERFORMED

### 4.1.GENERAL STEPS:

• Extracting the data form data set.

• Analysis of the data.

• Performing the basic operations.

• Developing the predictive model.

### 4.2.IMPORTANCE OF DATA ANALYSIS:

While analyzing data sets, it is important to define the objectives so that further steps become clearer. Analysis lets us pose questions about data. For questioning data, it is important to have data collection on which further operations will be carried out. After the above steps, "Data Analysis" comes into picture. Data analysis o is the process of raw data cleaning and conversion so that further operations become easier to carry on and then the conclusions can be drawn from the results. For Today, data has become the backbone of all research in almost every field. Research and analysis is no more limited to just the area of sciences, but has grown to be a part of businesses – startups and established organizations, government works and more.

### 4.3.DATA SET:

A data set is a collection of similar and related data or information. It is organised for better accessibility of an entity. Data sets are used for data analytics as they provide related information in a united form. It can be structured or unstructured.

## 5.ALGORITHMS USED FOR LOAN PREDICTION

### 1.SUPPORT VECTOR MACHINE FOR LOAN PREDICTION:

Support Vector Machine (SVM) is a supervised machine learning algorithm that generates a hyperplane (a decision boundary) to separate classes even in a high-dimensional vector space. It can capture different non-linear relationships between the features and the target variable. It decides a class for a sample based on the sign of w[T]+b. In the equation, w (weights) represents the negative and positive hyperplane margin, and b is the bias. SVM is particularly useful in loan prediction because this task usually has several features that need to be considered for the final decision

### 2. XG BOOST FOR LOAN PREDICTION:

"Boosting" is a method that combines individual models in an ensemble manner to gain higher performance gain. AdaBoost and Stochastic Gradient Boosting are the most popular boosting algorithms where higher weights are assigned to wrong classified instances during training. At the same time, SGB adds randomness as an integral part of training. Extreme Gradient Boost (XGBoost) is an improvement over Gradient Boost and is very popular in binaryclassificationalgorithms. The decision trees are built in parallel in XGBoost than in series, giving it an edge over normal Decision Trees and Boosting algorithms.

### 3..RANDOM FOREST FOR LOAN PREDICTION:

The random forest algorithm improves the flexibility and decisionmaking capacity of individual trees. It is another machine learning algorithm incorporating the ensemble learning theorem as its foundation, combining results from various decision trees to optimize training. In some use cases of loan and credit risk prediction, some features are more important than the rest or, more specifically, some features whose removal would improve the overall performance

## 6.PYTHON LIBRARIES USED FOR LOANPREDICTION

### 1.PANDAS

Pandas is the most straightforward and powerful package for beginners for data loading, cleaning, and processing. Modules in Pandas will help us treat null values, handle categorical variables, get an overview of the dataset, and perform exploratory data analysis if needed.

## 2..SCIKIT-LEARN:

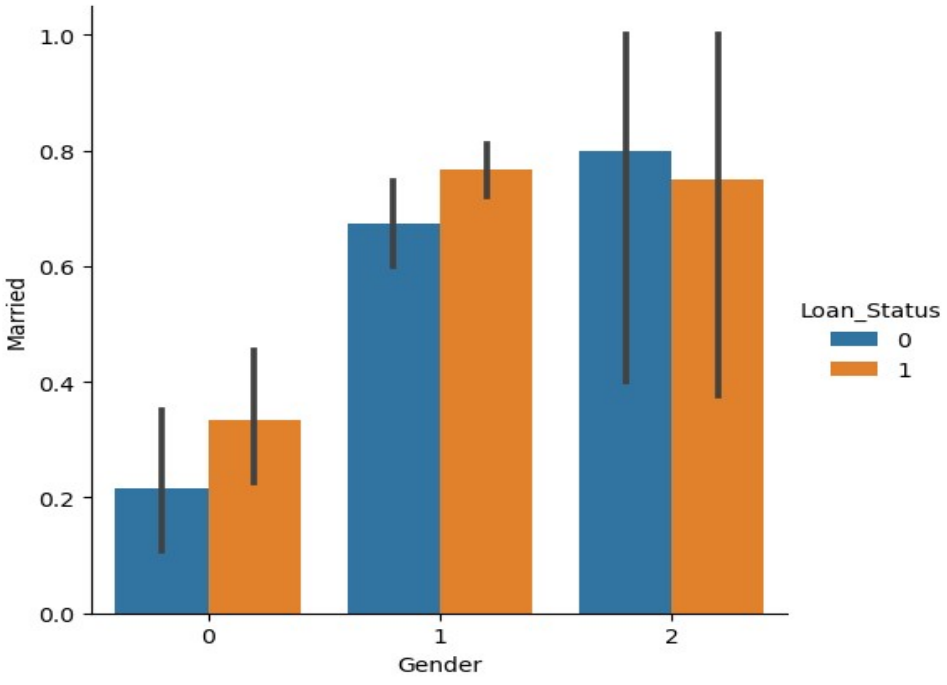
Perhaps the most accessible library Python for machine learning beginners, scikit-learn has ready-to-use modules for most machine learning-related tasks, from data preparation to model building, optimized training, and evaluation. To build our machine learning model, we use the existing modules available in sklearn. We use them through a module called RandomizedSearchCV, which computes cross-validation accuracy to find the best set of hyperparameters for every model.

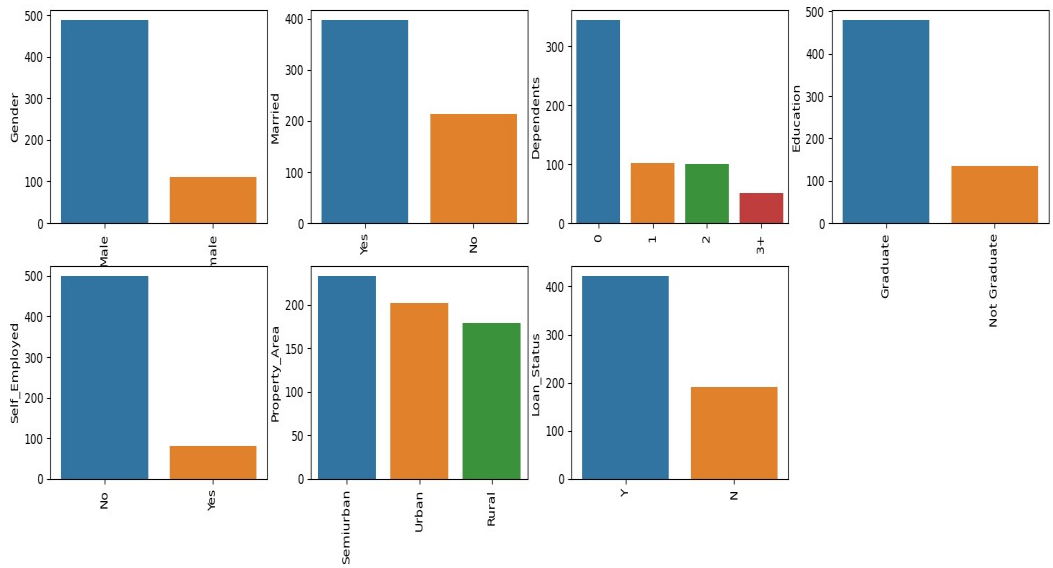
## 3.XGBOOST:

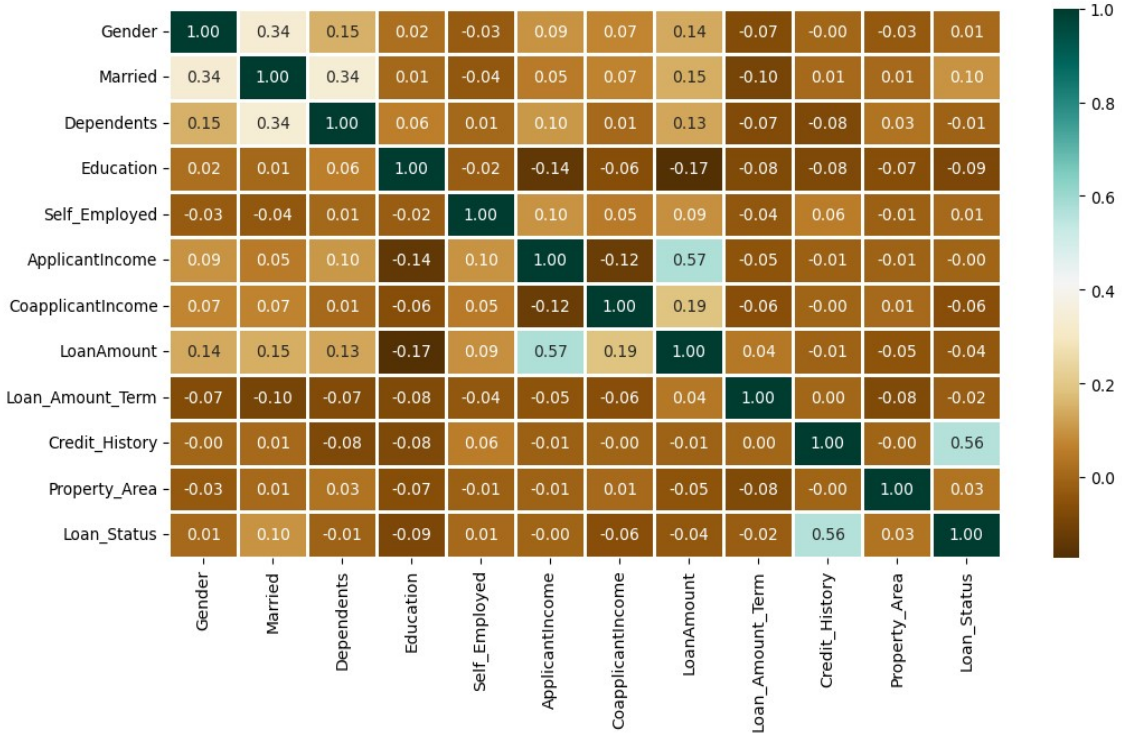
The XGBoost package available outside of sklearn has a faster and more accessible implementation of the boosting algorithm. We install it separately and use the XGBClassifier module from it.

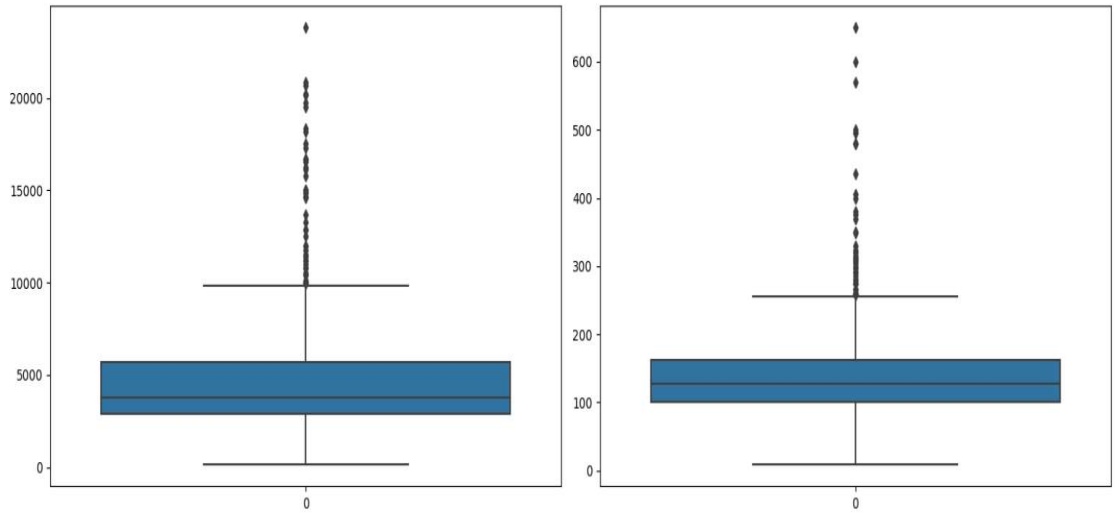
**NumPy** and **Matplotlib** are used for standard data processing and visualization tasks, respectively.

## 7.PLOTS FOR DIFFERENT DATA




Gender
Male
Married
Yes
No
Dependents
0
1
2
3+
Education
Graduate
Not Graduate
Self_Employed
No
Yes
Property_Area
Semiurban
Urban
Rural
Loan_Status
Y
N


Gender 1.00 0.34 0.15 0.02 -0.03 0.09 0.07 0.14 -0.07 -0.00 -0.03 0.01
Married 0.34 1.00 0.34 0.01 -0.04 0.05 0.07 0.15 -0.10 0.01 0.01 0.10
Dependents 0.15 0.34 1.00 0.06 0.01 0.10 0.01 0.13 -0.07 -0.08 0.03 -0.01
Education 0.02 0.01 0.06 1.00 -0.02 -0.14 -0.06 -0.17 -0.08 -0.08 -0.07 -0.09
Self_Employed -0.03 -0.04 0.01 -0.02 1.00 0.10 0.05 0.09 -0.04 0.06 -0.01 0.01
ApplicantIncome 0.09 0.05 0.10 -0.14 0.10 1.00 -0.12 0.57 -0.05 -0.01 -0.01 -0.00
CoapplicantIncome 0.07 0.07 0.01 -0.06 0.05 -0.12 1.00 0.19 -0.06 -0.00 0.01 -0.06
LoanAmount 0.14 0.15 0.13 -0.17 0.09 0.57 0.19 1.00 0.04 -0.01 -0.05 -0.04
Loan_Amount_Term -0.07 -0.10 -0.07 -0.08 -0.04 -0.05 -0.06 0.04 1.00 0.00 -0.08 -0.02
Credit_History -0.00 0.01 -0.08 -0.08 0.06 -0.01 -0.00 -0.01 0.00 1.00 -0.00 0.56
Property_Area -0.03 0.01 0.03 -0.07 -0.01 -0.01 0.01 -0.05 -0.08 -0.00 1.00 0.03
Loan_Status 0.01 0.10 -0.01 -0.09 0.01 -0.00 -0.06 -0.04 -0.02 0.56 0.03 1.00
Gender
Married
Dependents
Education
Self_Employed
ApplicantIncome
CoapplicantIncome
LoanAmount
Loan_Amount_Term
Credit_History
Property_Area
Loan_Status
1.0
0.8
0.6
0.4
0.2
0.0




## 8. SKILLS ACQUIRED

1. Understand, evaluate, design and implement artificial intelligence models.

2. Implement contemporary artificial intelligence techniques, from knowledge representation, to deep learning, developing in demand skills and leadership qualities for an exciting career in AI.

3. Apply the legal, ethical, social and philosophical context for practical AI projects.

4. Extend knowledge in artificial intelligence through research, experimentation and analysis.

5. Practical or hands on experience in training an ML model.

Gain expertise in technical drawing to visualize concepts.

## 1.TECHNICAL OUTCOMES:

- Machine learning involves computations on large data sets, hence we learnt strong basic fundamental knowledge such as computer architecture, algorithms and data structure complexity. Getting in depth into the python language and exploring new commands.
- Synthesize visual perception skills along with drawing skills to visually communicate ideas. Deconstruction of designs for its motives and inspirations. To learn to synthesize data and make connections within the data points using the available frameworks.
- To frame an appropriate actionable problem statement with reference to user needs and contextual alignments.

- Data analysis of different data sets and to understand the concepts on a real world basis to implement and make use of AI/ML in our upcoming career
- To train different models and to make sure the requirement of the respective clients and make to implement a model according to their requirements.

## 9. TIME MANAGEMENT:

Time management helps you allocate time for the most important tasks. When we follow a schedule we don't have to spend time and energy on what to do. Instead we can focus on what matters and do well. The quality of the work will suffer if we are constantly worrying about meeting the deadlines. Time management helps to prioritize the tasks, so we can have enough time to focus on each project to put in the effort and produce high quality outcomes. Many software companies have to work against tight timelines. Proper time management will allow us to allocate enough time to meet each deadline. Planning ahead also keeps us calm and think freely to work more in an efficient way. Personality development is referred to as a process of developing and enhancing one's personality. It helps one to gain confidence and high self esteem. It is essential to think positive and don't get upset over minor things, to be a little flexible and always look at the broader perspectives of life. Do not think of harming others and share whatever you know. Always help others. Be a patient listener and never interrupt when others are speaking. Try to imbibe good qualities of others. Confidence is the key to a positive personality. Exude confidence and positive aura wherever you go. Personality development teaches you to be calm and composed even at stressful situations. Never over react. Avoid finding faults in others. Learn to be a little broad minded and flexible.

## 10.CONCLUSION

Manual processing of loan applications is a long, cumbersome, error-prone, and often biased process. It might lead to financial disaster for banks and obstruct genuine applicants from getting the needed loans. Loan Prediction using machine learning tools and techniques can help financial institutions quickly process applications by rejecting high-risk customers entirely, accepting worthy customers, or assigning them to a manual review. Such processes with loan prediction using machine learning intact can reduce loan processing times by nearly 40%.

Loan prediction analysis uses specific parameters about a loan application to determine whether or not the loan should get approved. Approved loans usually have a good credit history, decent applicant income, and reliability in other factors. Banks use statistical and manual methods to verify these factors and decide about the applicant's loan status.

We saw some existing approaches and datasets used to approach loan eligibility prediction and how AI might help smoothen this process. Finally, we built an endto-end loan prediction machine learning project using a publicly available dataset from scratch. At the end of this project, one would know how different features influence the model prediction and how specific attributes affect the decision more than the other features. Only building machine learning projects from scratch, even as beginners, will naturally bring such insights to light and give a comprehensive view of a machine learning problem.