**An**

**Internship Report**

**ON**

**“Machine Learning”**

From

“Tech-gyan Technologies”

Including

“Loan Status Prediction”

**SUBMITTED BY:**

Nilay Pandey, 2001330100176

**Under Supervision of**

**Ibrar Ahmed**  
Assistant Professor

Department of Computer Science and Engineering



November 2022

DEPARTMENT OF INFORMATION TECHNOLOGY

**NOIDA INSTITUE OF ENGINEERING AND TECHNOLOGY,  
GREATER NOIDA**

**Content Table**

|  |  |  |
| --- | --- | --- |
| **For Internship S. No.** | **TITLE** | **Pg No.** |
| 1. | Internship Certificate | 3 |
| 2. | Abstract | 4 |
| 3. | Technology Background | 5-8 |
| 4. | Project Problem Background | 9 |
| 5. | Project Modules | 10-12 |
| 6. | Snapshots of project | 13-14 |
| 7. | Applications | 15 |
| 8. | Reference | 16 |

**Internship Certificate**



**Abstract**

When any financial institution lends the money to the person, it is always been a high risk. Today data is increasing with the rapid pace in the banks, therefore the bankers need to evaluate the person’s data before giving the loan. It can be a big headache to evaluate the data. This problem is solved by analyzing and training the data by using one of the Machine Learning algorithms. For this, we have generated a model for the prediction that the person will get the loan or not. The primary objective of this paper is to check whether the person can get the loan or not by evaluating the data with the help of decision tree classifiers which can gives the accurate result for the prediction. Keywords—Loan, Machine Learning, Decision tree, Data training.

**Introduction**

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans: The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.



The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that “it gives computers the ability to learn without being explicitly programmed”. And in 1997, Tom Mitchell gave a “well-posed” mathematical and relational definition that “A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E. Machine learning involves a computer to be trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train a computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to the computer whether a picture is cat or not. Then if we show the computer a new image, then from the above training, the computer should be able to tell whether this new image is a cat or not.

**Technology Background**

• Model A model is a specific representation learned from data by applying some machine learning algorithm. A model is also called hypothesis.

• Feature A feature is an individual measurable property of our data. A set of numeric features can be conveniently described by a feature vector. Feature vectors are fed as input to the model. For example, in order to predict a fruit, there may be features like color, smell, taste, etc. Note: Choosing informative, discriminating and independent features is a crucial step for effective algorithms. We generally employ a feature extractor to extract the relevant features from the raw data.

• Target (Label) A target variable or label is the value to be predicted by our model. For the fruit example discussed in the features section, the label with each set of input would be the name of the fruit like apple, orange, banana, etc.

• Training The idea is to give a set of inputs(features) and it’s expected outputs(labels), so after training, we will have a model (hypothesis) that will then map new data to one of the categories trained on.

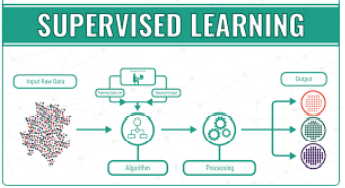
• Prediction Once our model is ready, it can be fed a set of inputs to which it will provide a predicted output(label).

**Supervised Learning**

• I like to think of supervised learning with the concept of function approximation, where basically we train an algorithm and in the end of the process we pick the function that best describes the input data, the one that for a given X makes the best estimation of y (X -> y). Most of the time we are not able to figure out the true function that always make the correct predictions and other reason is that the algorithm rely upon an assumption made by humans about how the computer should learn and this assumptions introduce a bias.

• Here the human experts act as the teacher where we feed the computer with training data containing the input/predictors and we show it the correct answers (output) and from the data the computer should be able to learn the patterns.

• Supervised learning algorithms try to model relationships and dependencies between the target prediction output and the input features such that we can predict the output values for new data based on those relationships which it learned from the previous data sets.



List of Common Algorithms

• Nearest Neighbour

• Naive Bayes

• Decision Trees

• Linear Regression

• Support Vector Machines (SVM)

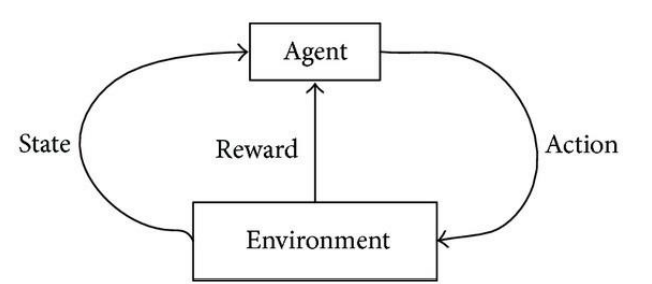
• Neural Networks.

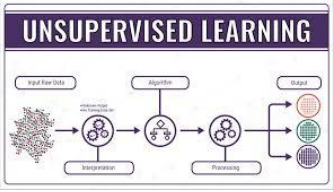
**Unsupervised Learning**

• The computer is trained with unlabelled data.

• Here there’s no teacher at all, actually the computer might be able to teach you new things after it learns patterns in data, these algorithms a particularly useful in cases where the human expert doesn’t know what to look for in the data.

• are the family of machine learning algorithms which are mainly used in pattern detection and descriptive modelling. However, there are no output categories or labels here based on which the algorithm can try to model relationships. These algorithms try to use techniques on the input data to mine for rules, detect patterns, and summarize and group the data points deriving.





**Draft**

• Descriptive Model

• The main types of unsupervised learning algorithms include Clustering algorithms and Association rule learning algorithms.

**List of Common Algorithms**

• k-means clustering, Association Rules.

**Project Problem Background**

**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 algorithms**](https://www.projectpro.io/article/common-machine-learning-algorithms-for-beginners/202)**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**.

The immense increase in capitalism, the fast-paced development and instantaneous changes in the lifestyle has us in awe. Emi, loans at nominal rate, housing loans, vehicle loans, these are some of the few words which have skyrocketed from the past few years. The needs, wants and demands have never been increased this before. People gets loan from banks; however, it may be baffling for the bankers to judge who can pay back the loan nevertheless the bank shouldn’t be in loss. Banks earn most of their profits through the loan sanctioning. Generally, banks pass loan after completing the numerous verification processes despite all these, it is still not confirmed that the borrower will pay back the loan or not. To get over the dilemma, I have built up a prediction model which says if the loan has been assigned in the safe hands or not. Government agencies like keep under surveillance why one person got a loan and the other person could not. In Machine Learning techniques which include classification and prediction can be applied to conquer this to a brilliant extent. Machine learning has eased today’s world by developing these prediction models. Here we will be using the fine techniques of machine learning – Decision tree algorithm to build this prediction model for loan assessment. It is as so because decision tree gives accuracy in the prediction and is often used in the industry for these models.

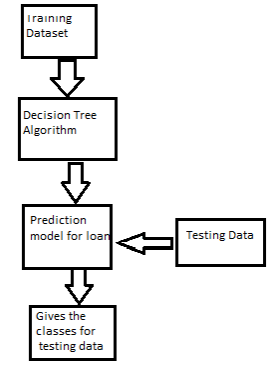
**Project Modules**

**EXISTING SYSTEM**

Banks need to analyze for the person who applies for the loan will repay the loan or not. Sometimes it happens that customer has provided partial data to the bank, in this case person may get the loan without proper verification and bank may end up with loss. Bankers cannot analyze the huge amounts of data manually; it may become a big headache to check whether a person will repay its loan or not. It is very much necessary to know the person getting loan is going in safe hand or not. So, it is pretty much important to have a automated model which should predict the customer getting the loan will repay the loan or not.

**PROPOSED SYSTEM**

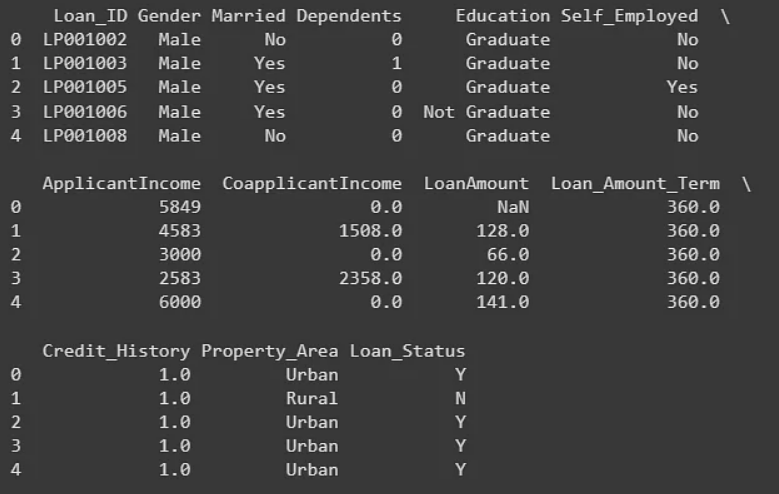
I have developed a prediction model for Loan sanctioning which will predict whether the person applying for loan will get loan or not. The major objective of this project is to derive patterns from the datasets which are used for the loan sanctioning process and create a model based on the patterns derived in the previous step. This model is developed by using the one of the machine learning algorithms.



**DATA SET**

Datasets are gathered from Kaggle. Data set is now provided to Machine learning models on the basis of this facts this version is trained. Data sets are divided into Existing and New Customers. Every new applicant info act as a fact test set. After the operation of testing, model expect whether the brand-new applicant is in case for approval of the loan or now not primarily based upon the inference it concludes on the idea of training information sets.

**Dataset for Loan status prediction-**



**Processes for Loan Prediction:**

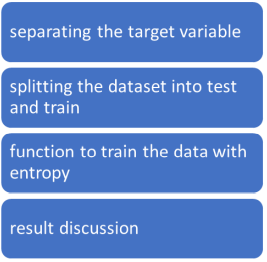
Separating the target variable

X= balance\_data. values[:, 1:5]

Y= balance\_data. values[:, 0]

Splitting the dataset into test and train

X\_train, X\_test, Y\_train, y\_test= train\_test\_split (X, Y, test\_size = 0.3, random\_state= 100)



**Five major steps of this project:**

Data Collection

Data Preparation

Data Visualization

Model Selection

Creating Model

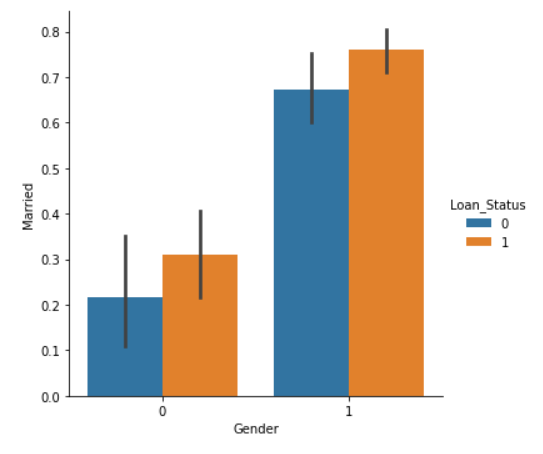
**System Requirement:**

**1.** Windows 7/8/10/11

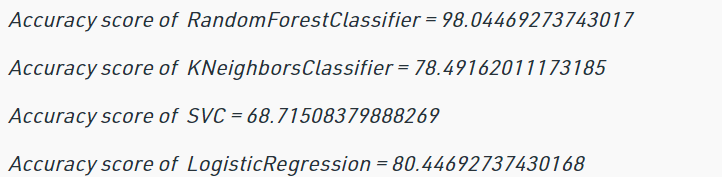
**2.** Python 3 with numpy, pandas, sklearn, matplotlib libraries installed.

**3.**pycharm IDE

**Snapshot of Project**



**Fig (1)**



**Fig (2)**

Chart, treemap chart

Description automatically generated **Fig (3)**

**Application**

Dream Housing Finance company deals in all kinds of home loans. They have a presence across all urban, semi-urban and rural areas. The customer first applies for a home loan and after that, the company validates the customer eligibility for the loan.

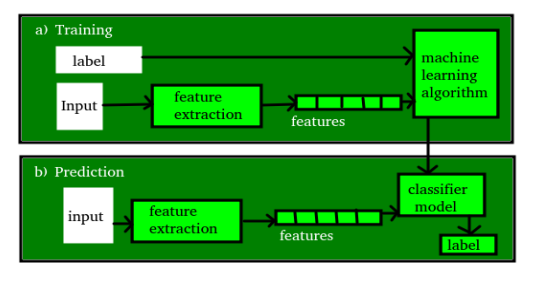
The company wants to automate the loan eligibility process (real-time) based on customer detail provided while filling out online application forms. These details are Gender, Marital Status, Education, number of Dependents, Income, Loan Amount, Credit History, and others.

To automate this process, they have provided a dataset to identify the customer segments that are eligible for loan amounts so that they can specifically target these customers.

Loan status can have two values: Yes or NO.

Yes: if the loan is approved

NO: if the loan is not approved



**References**

**.** [**www.kaggle.com**](http://www.kaggle.com)

**.** [**www.geeksforgeeks.org**](http://www.geeksforgeeks.org)

**.** [**www.wikipedia.org**](http://www.wikipedia.org)

**.** [**www.researchgate.net**](http://www.researchgate.net)

**And many more…**