PROJECT REPORT

On

Loan Prediction

Submitted in partial fulfilment of the

Requirements for the award of the Degree of

Bachelor of Technology

In

Computer science and Engineering

Under the esteemed guidance of

Srinivas Marimganti

By

K.Siddharth – 180330202
S.Vardhan – 180330225
KrishnaVamsi – 180330207
K.NITIN RAJ - 180330203

(DST-FIST Sponsored Department)

K L EDUCATION FOUNDATION

R.V.S Nagar, Moinabad-Chilkur Rd, near AP Police

Academy, Aziznagar, Telangana

2020-2021

K L EDUCATION FOUNDATION

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(DST-FIST Sponsored Department)

CERTIFICATE

This is to certify that this project based lab report entitled **"LOAN PREDEICTION** " is a bonafide work done by **180330202,180330225,180330207,180330203** in the course **18CS3166S MACHINE LEARNING** in partial fulfilment of the requirements for the award of Degree in Bachelor of Technology in **COMPUTER SCIENCE AND ENGNEERING** during the Even Semester of Academic year 2020-2021.

Faculty in Charge **Head of the Department**

K L EDUCATION FOUNDATION

DEPT OF COMPUTER SCINCE AND ENGINEERING

(DST-FIST Sponsored Department)

DECLARATION

We hereby declare that this project based lab report entitled **"LOAN PREDEICTION"** has been prepared by us in the course **18CS3166S MACHINE LEARNING** in partial fulfilment of the requirement for the award of degree bachelor of technology in **COMPUTER SCIENCE AND ENGINEERING** during the Even Semester of the academic year 2020- 2021. We also declare that this project-based lab report is of our own effort and it has not been submitted to any other university for the award of any degree.

Date: 13-11-2020

Place: Hyderabad

ACKNOWLEDGEMENT

Our sincere thanks to **Dr.KG Suma** in the Lab for her outstanding support throughout the project for the successful completion of the work.

We express our gratitude to **Dr M.Srinivas** Course Co-ordinator for **18CS3166S MACHINE LEARNING** course in the Computer Science and Engineering Department for providing us with adequate planning and support and means by which we can complete this project-based Lab.

We express our gratitude to **Mrs Anu Priya,** Head of the Department for computer science and Engineering for providing us with adequate facilities, ways and means by which we can complete this project-based Lab.

We would like to place on record the deep sense of gratitude to the honourable Vice Chancellor, K L University for providing the necessary facilities to carry the project-based Lab.

Last but not the least, we thank all Teaching and Non-Teaching Staff of our department and especially our classmates and our friends for their support in the completion of our project-based Lab.

K.Siddharth

Vardhan

Krishna

K.NITIN RAJ

TABLE OF CONTENTS

ABSTRACT

CHAPTER 1: INTRODUCTION

CHAPTER 2: LITERATURE SURVEY

CHAPTER 3: NOVELITY OF WORK

CHAPTER 4: IMPLEMENTATION WITH CODE

CHAPTER 5: RESULT AND ANALYSIS

CHAPTER 6: CONCLUSION OF WORK

ABSTRACT

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only, so finding out to whom the loan can be granted which will be a safer option for the bank is a typical process. So in this project we try to reduce this risk factor behind selecting the safe person so as to save lots of bank efforts and assets. This is done by mining the Big Data of the previous records of the people to whom the loan was granted before and on the basis of these records/experiences the machine was trained using the machine learning model which give the most accurate result. The main objective of this project is to predict whether assigning the loan to particular person will be safe or not. This paper is divided into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing. In this paper we are predict the loan data by using some machine learning algorithms they are classification, logistic regression, Decision Tree and gradient boosting.

Keywords: Machine learning, Decision Tree, prediction, Python.

INTRODUCTION:

Distribution of the loans is the core business part of almost every banks. The main portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking environment is to invest their assets in safe hands where it is. Today many banks/financial companies approves loan after a regress process of verification and validation but still there is no surety whether the chosen applicant is the deserving right applicant out of all applicants. Through this system we can predict whether that particular applicant is safe or not and the whole process of validation of features is automated by machine learning technique. The disadvantage of this model is that it emphasize different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system.

Loan Prediction is very helpful for employee of banks as well as for the applicant also. The aim of this Project is to provide quick, immediate and easy way to choose the deserving applicants. It can provide special advantages to the bank. The Loan Prediction System can can automatically calculate the weight of each features taking part in loan processing and on new test data same features are processed with respect to their associated weight. A time limit can be set for the applicant to check whether his/her loan can be sanctioned or not. Loan Prediction System allows jumping to specific application so that it can be check on priority basis. This Project is exclusively for the managing authority of Bank/finance company, whole process of prediction is done privately no stakeholders would be able to alter the processing. Result against particular Loan Id can be send to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities.

LITERATURE SURVEY

A. Vaidya proposed a method for approval of loan prediction using logistic regression [1]. Logical regression is a machine learning technique which is very useful in prediction system. The approval of loan is a very important processin banking system. A. Vaidya solves the problem by applying machine learning in a sample data set for loan approval applications. It also opens other areas on which machine learning is applicable. A. Li and Q. Sun[2] find a method to calculate risk involved in loan approvals for SMEs. A concept of loan consuming radius was introduced which was based upon supply chain in consumer market. F. M. Isiket al. develop a loan approval system using Business Process Execution Language BPEL[3]. The concept of BPELis very useful in business firms. A reasoning engine was developed which removes some services from the BPEL process which are not necessary to complete a process. The system was applied on oan approval which involve many processes. [4]V. C. T. Chanet al. proposed a credit approval system using web services. The system approved credit for the customers. With credit application the customer submits some other useful information's. This information's are processed by Credit Approval System which finally give credit score to the applicant. The paper developed a web services based solution of this problem. J. Lohokareet al. [5] proposed a system which automatically collect data for an applicant and decides the credit score. The system work on the social media to collect information about the user. R. Yanget al. [6] analyzed that whether the credit default behavior of a SME depends upon credit features of its owner or not. The author concluded that features of the owner behaves as valuable parameters to calculate risk of a loan for SMEs. [7]M. Bayraktaret al. [7] proposed a method for credit risk analysis using machine learning. Boltzman machine was used to make the analysis for risk calculation of loan. H. A. P. Pérezet al. [8] introduced fuzzy model for calculation of credit score of the customer. The information collected by the system for calculation of the credit score was converted into gradual values using fuzzy sets. The fuzzy based method performs better for calculation of the credit score of the applicants. S. Yadav and S. Thakur[9] applied Big Data approach for loan analysis. The techniques of big data analysis was applied on customer data to calculate bank loan analysis. Hadoop based method was used in theloan analysis. Y. Lin[10] analysis of the effect of the political approaches effect the loans of state banks. The paper investigated that in state owned banks, the political relationship plays a considerable role. [11]Ruifen Zhaoworked on approval of college loans. Education loans are very common among students because of rise in the cost of education. The paper investigated the issues in loan approval of college students. M. Houshmand and M. D. Kakhki[12] proposed an expert system which evaluates the loan approvals.

The system used rule base approaches for loan approval decisions. L. Hui-ling[13] analyze the relation between characteristics of the banks, firms and loans approval. The paper investigated that there is a strong relationship between approval of loans and characteristics of business firm who apply the loan and characteristics of the bank. C. Yin[14] apply fuzzy logic to calculate the bank loan risks. A new pattern recognition system using fuzzy logic was developed which evaluate the risks involves in the approval of bank loans for applicants. J. Ma and Y. Cheng[15] proposed Markov Chain based model for risk management of bank loans. A. V. Gutierrez[16] proposed a model for housing loan. The model was worked for green housing loans. J. Chen and W. Guo[17] worked on loan limit of the loan applicants. The model worked on supply chain for financing decision making. G. Arutjothi and C. Senthamarai, [18] used machine learning classifier for prediction of loan approval status in banks. The machine learning based prediction system was applied on commercial banks. The paper conclude that the machine learning approach is very useful in loan status prediction. Y.

Shi and P. Song[19] proposed a method for evaluating project loans using risk analysis. The method evaluate the risk involved in loans of commercial banks. R. ZhangandD. Li [20] used machine learning approached in prediction systems. The machine learning approach was used for assessment of water quality. The paper concluded that machine learning is a very unimportant tool in prediction systems. C. Franket al. [21] used machine learning in prediction of smoking status. Different machine learning approaches were applied and investigated for finding the smoking status. From the results its was ensured that logistic regression performs better. R. Lopeset al. applied machine learning approach for the prediction of credit recovery [22]. Credit recovery is very important issue for banking system. The prediction of credit recovery is a challenging tasks. Different machine learning approach was applied to predict the credit recovery and gradient expansion algorithms (GBM)outperformed the other machine learning approaches. After going through this literature it is found that loan approval prediction problem is very important for banking system. Machine learning algorithm are very useful in predicting outcomes even when data is very big in size. This paper investigated some machine learning algorithms and applied ML on test data set of loan approvals. Next section discussed the three machine learning approaches.

LINKS

http://sersc.org/journals/index.php/IJAST/article/view/460/423

https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.ijitee.org/wp-content/uploads/papers/v9i1/A4881119119.pdf&ved=2ahUKEwiB0_Dun8HsAhXO6XMBHfp0BFAQFjANegQIExAB&usg=AOvVaw2gLK4mc0hylHlp9uZYUJHO

https://www.ijarcce.com/upload/2016/march-16/IJARCCE%20128.pdf

https://www.kaggle.com/yaheaal/loan-status-with-different-models

https://ieeexplore.ieee.org/document/8845444

http://www.sciencedirect.com/science/article/pii/S037872060100091X

NOVELITY OF WORK

Unlike other projects we have implemented 16 algorithms vs 10.

we have done both methods where we apply the algorithms taking all the features as well as taking the important ones only.

not only filling in the empty values but also downscaling and upscaling where necessary.

we have improved upon the random forest algorithm to match with the accuracy of logistic regression.

IMPLEMENTATION

- A Correlation Heat Map is been plotted to find the required features for applying machine learning algorithms with them.

- Data Standardization and Data Normalization has been done in this project.

- A total of 20 algorithms been applied using these features obtained and processes such as Up Sampling and Down Sampling of data is been done on 10 algorithms.

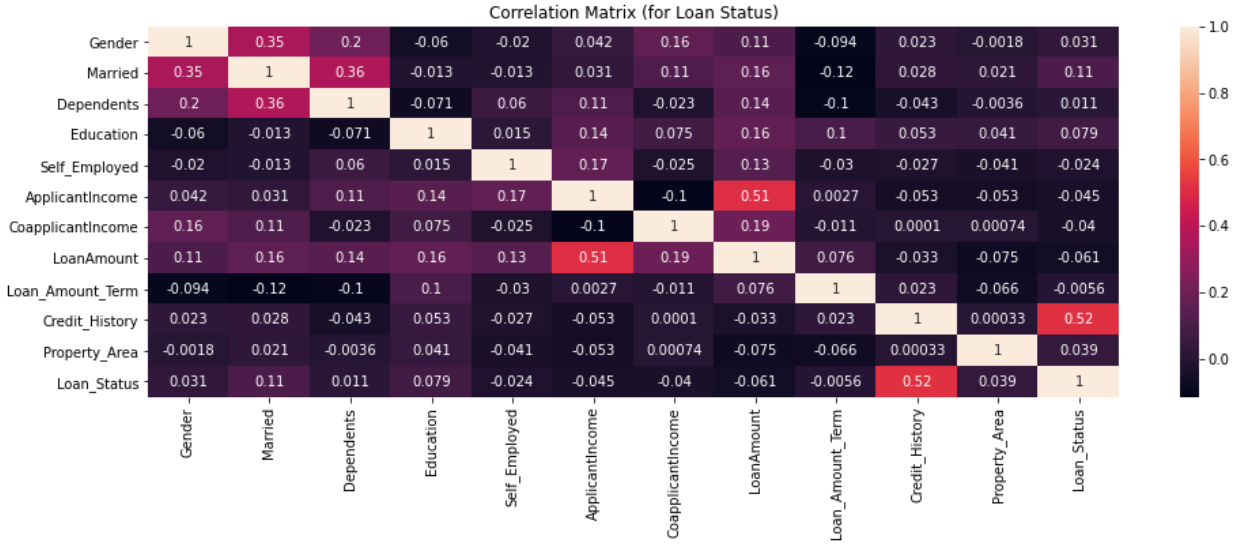
- List of Algorithms Applied for this project:
  - Logistic Regression
  - Decision Tree Classifier
  - Random Forest Classifier
  - Extra Tree Classifier
  - K-NN Clustering
  - K-Means Clustering
  - Ada Boosting Algorithm
  - Linear Discriminant
  - Gradient Boosting
  - Voting Classifier
  - Ensemble Learning (Logistic Regresion)
  - Ensemble Learning (Decision Tree)
  - Ensemble Learning (Voting Classifier)
  - Single-Layer Perceptron
  - Multi-Layer Perceptron
  - Multi-Layer Perceptron Classifier
  - Hierarchical Clustering
  - Naive Bayes
  - Support Vector Machine (SVM)
  - XGBoost Classifier

- Single-Layer perceptron and Multi-Layer perceptron had been implemented where 1-input,1-hidden, and 1-output layers for Single-Layer using TensorFlow and 1-input, 2-Hidden and 1-Output layers for Multi-Layer perceptron with both TensorFlow and keras has been Implemented.

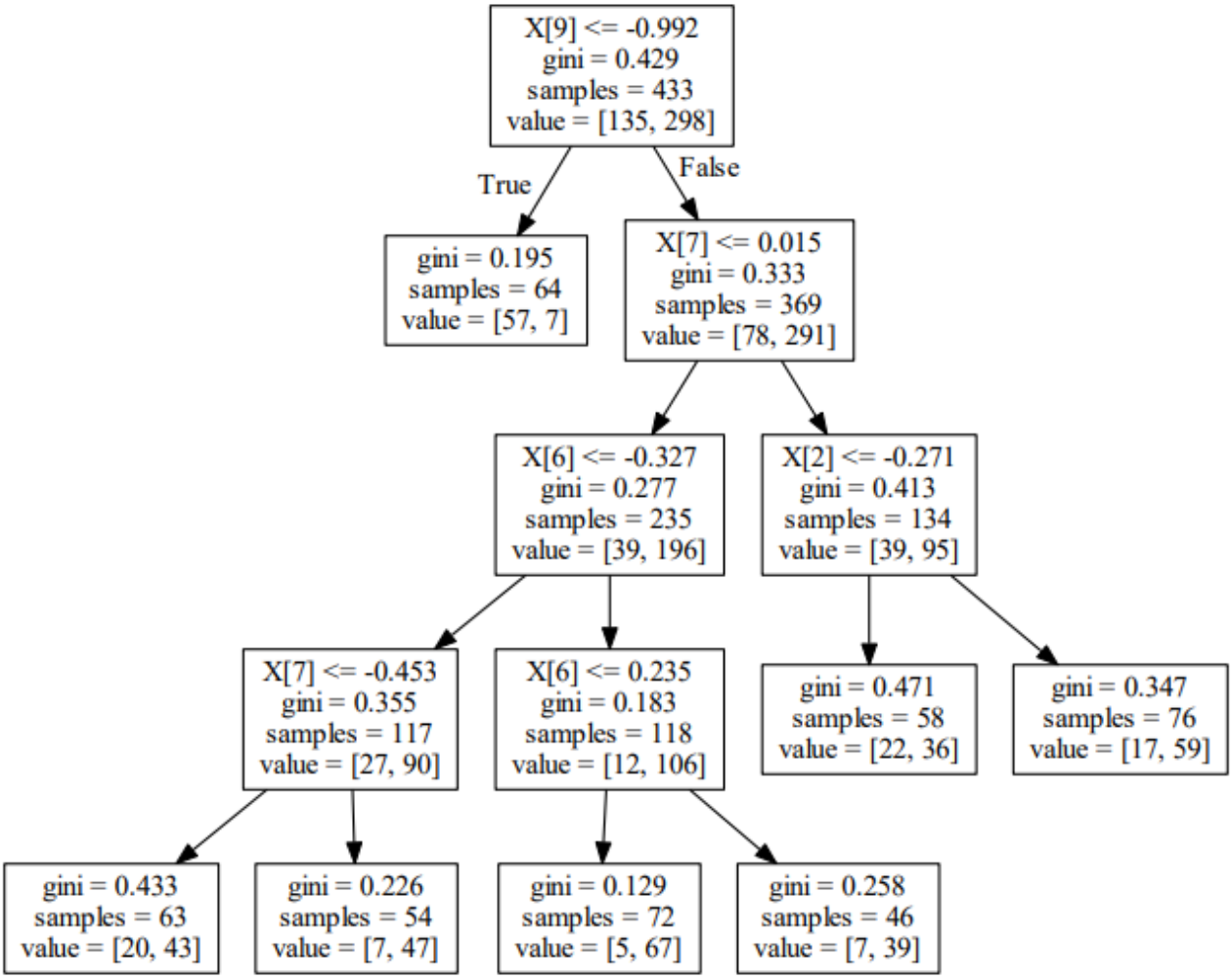
- Many other algorithms like Ensemble Learnings, Naïve Bayes, Extra Tree etc. are been implemented.

- For some of the algorithms we have done feature selection process to know about how features effect the loan approval prediction.

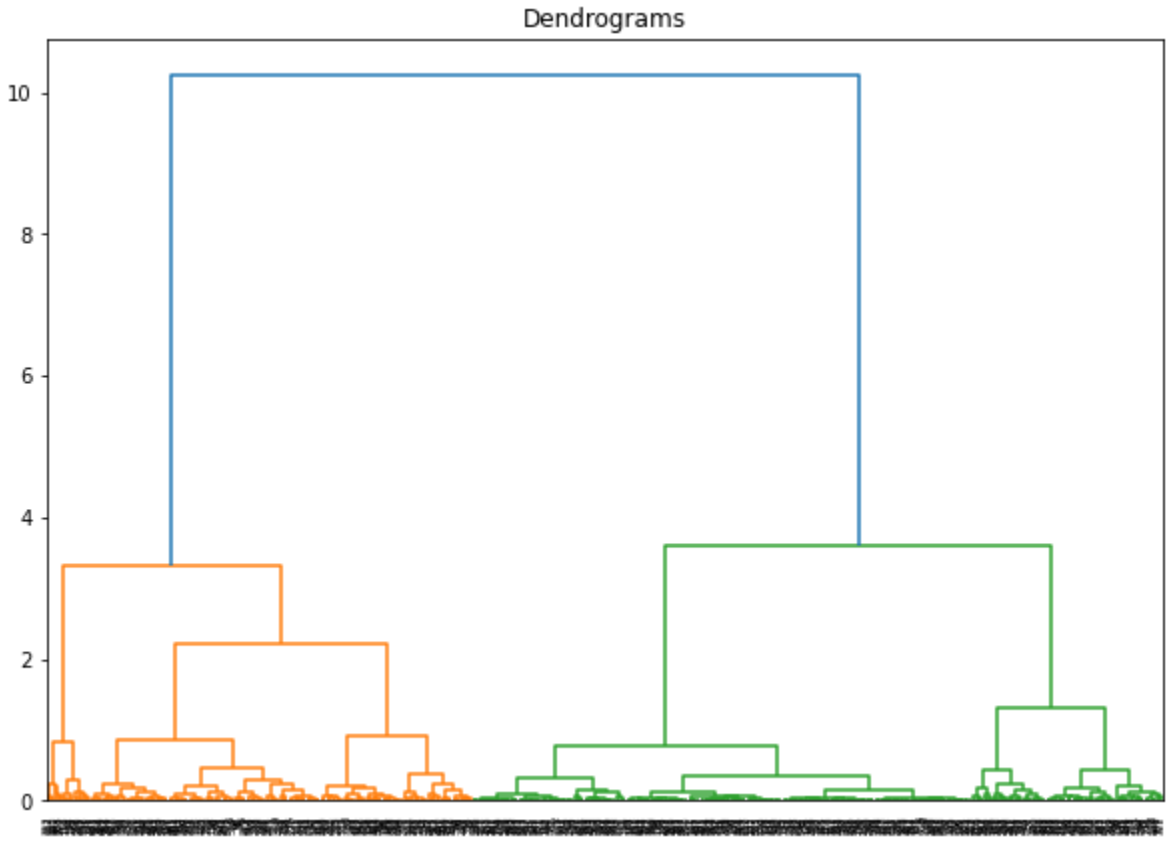
Correlation Matrix (for Loan Status):

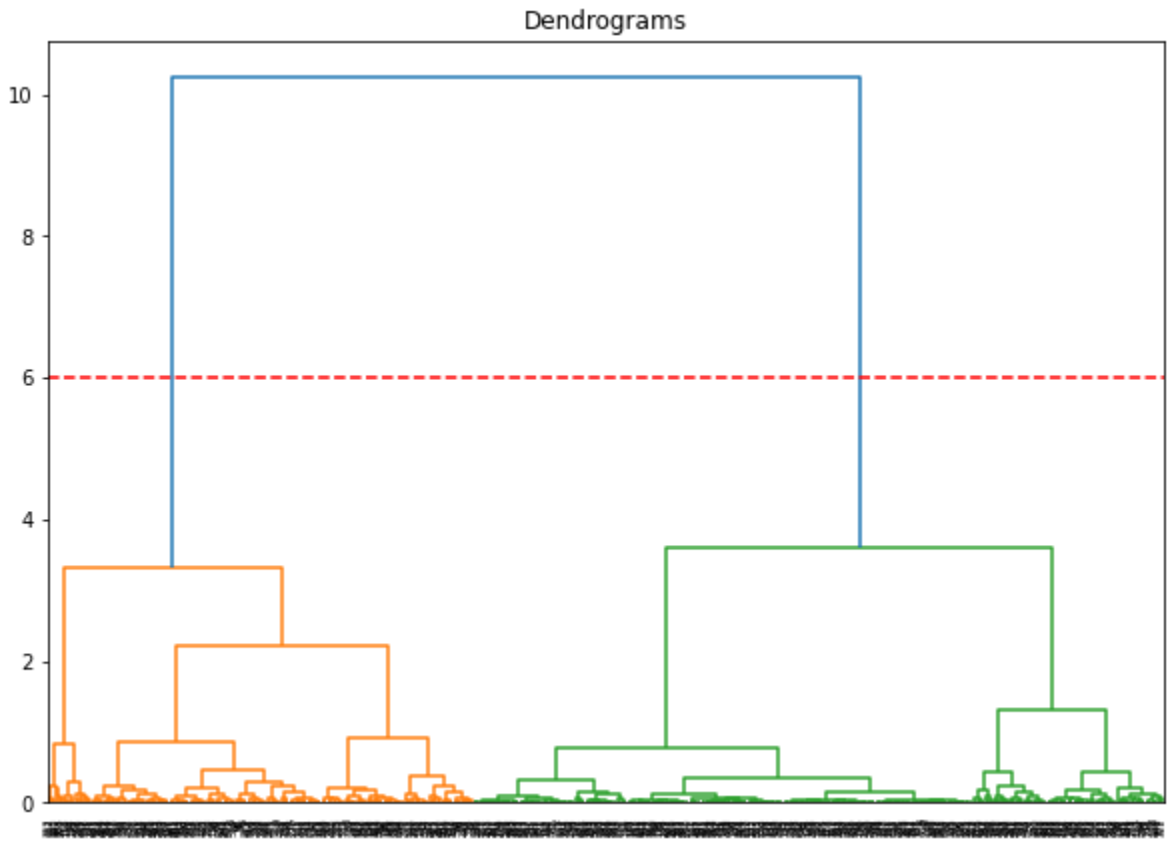


Decision Tree:

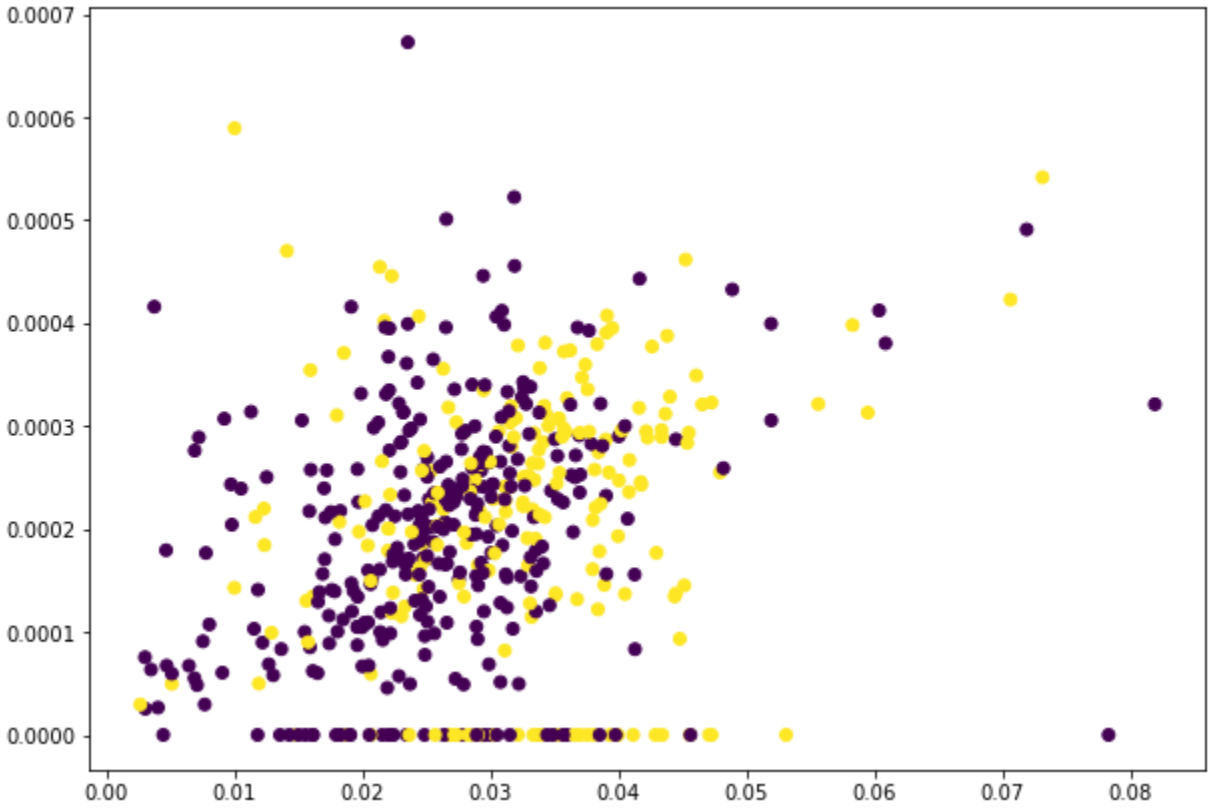


Hierarchical Clustering:

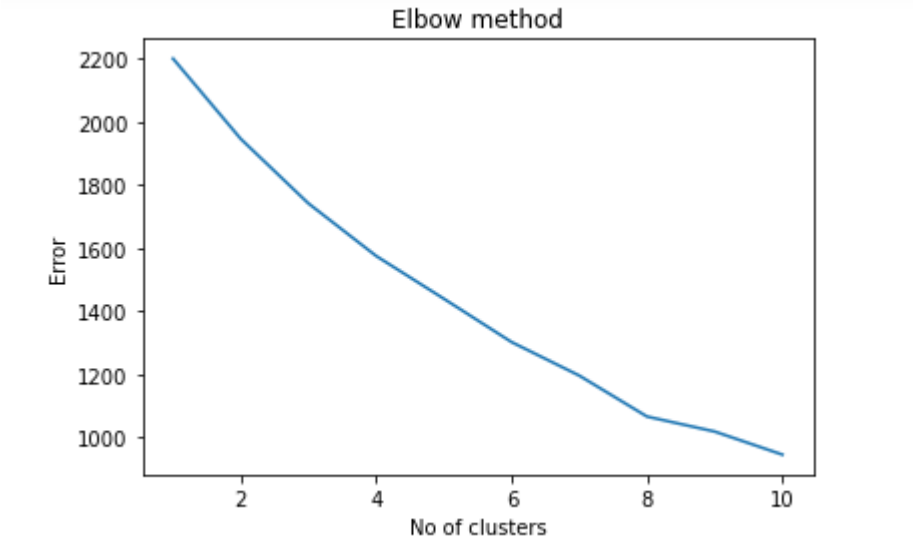

Dendrograms
10
8
6
4
2
0


Dendrograms
10
8
6
4
2
0

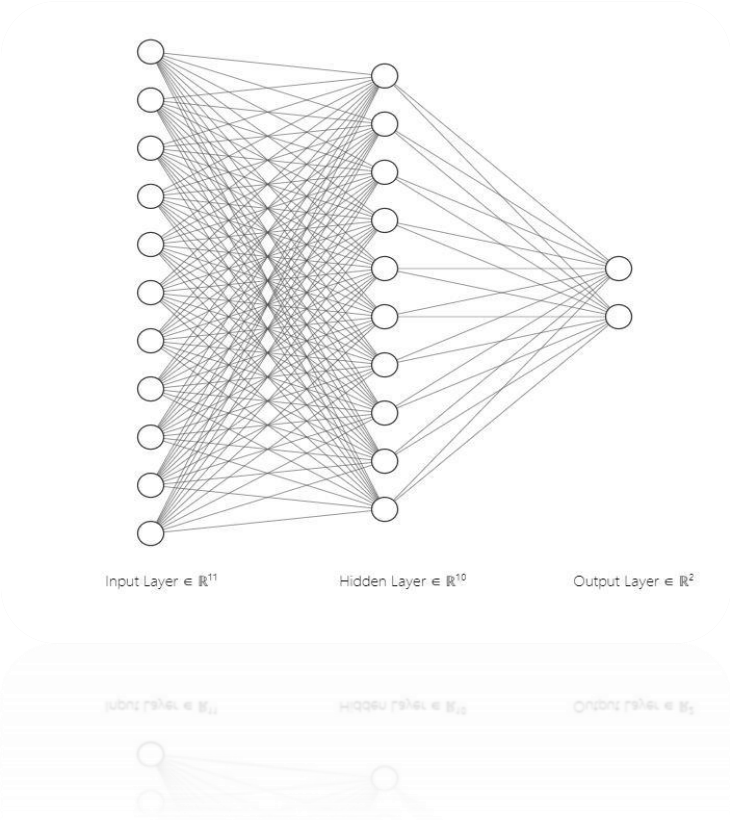
Scatter Plot:



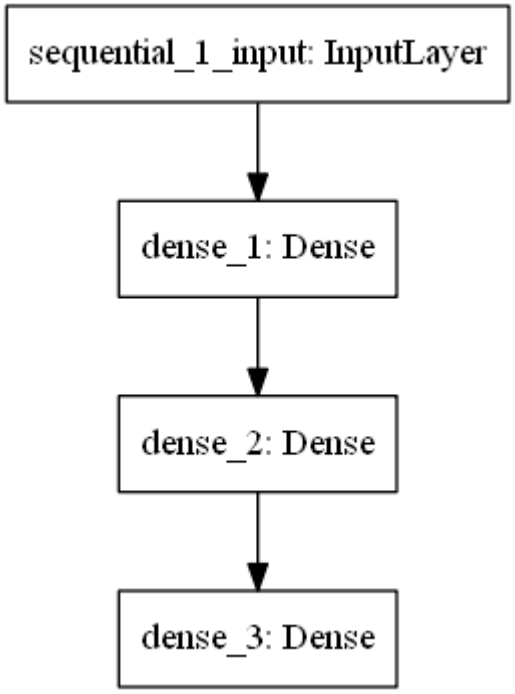
K-Means Elbow Method:

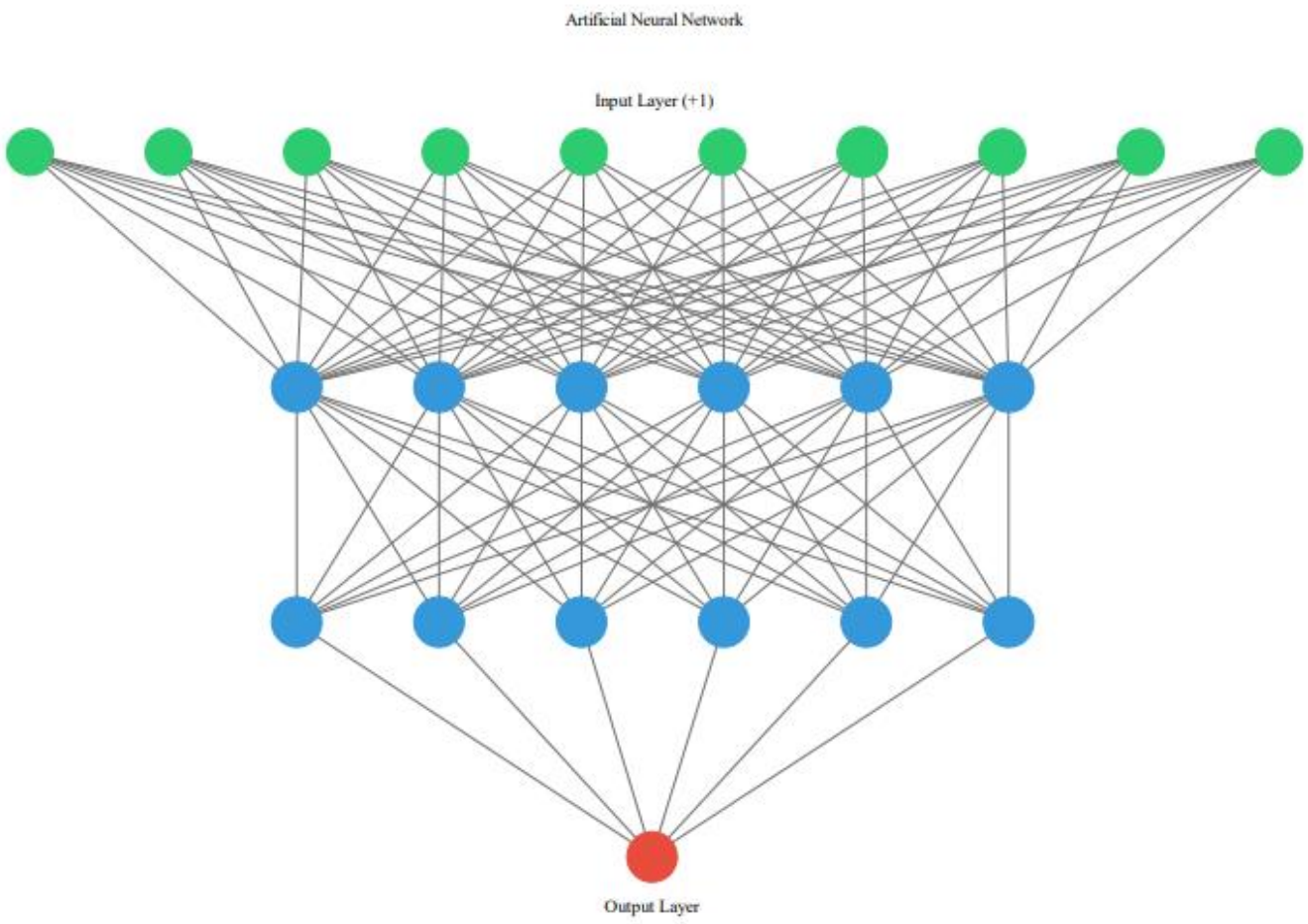


Single-Layer Perceptron:



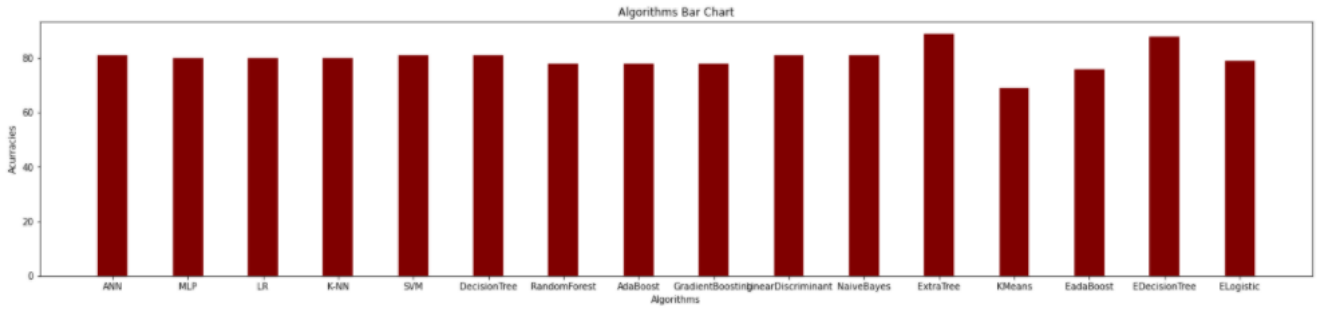
Multi-Layer Perceptron:



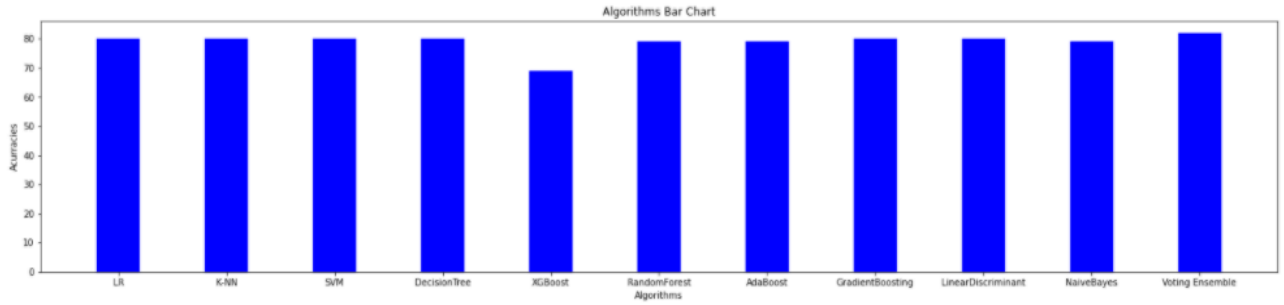

Artificial Neural Network
Input Layer (+1)
Output Layer

Results and Analysis

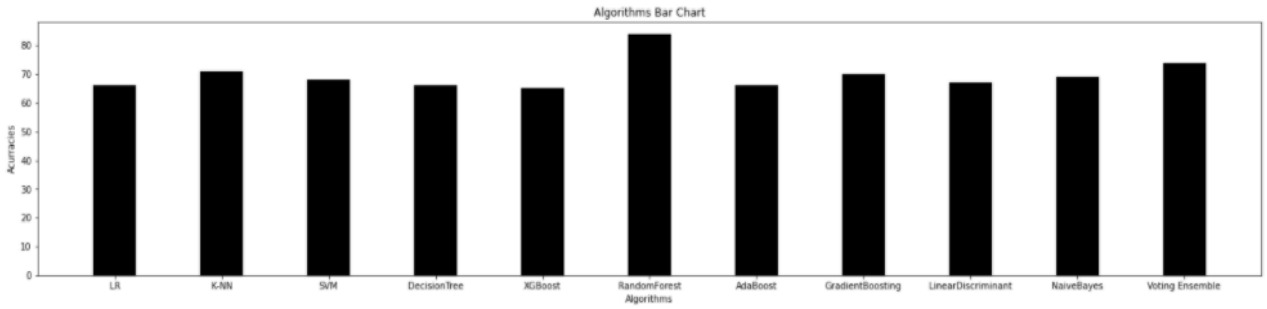
All Features Bar Chart (Algorithms):



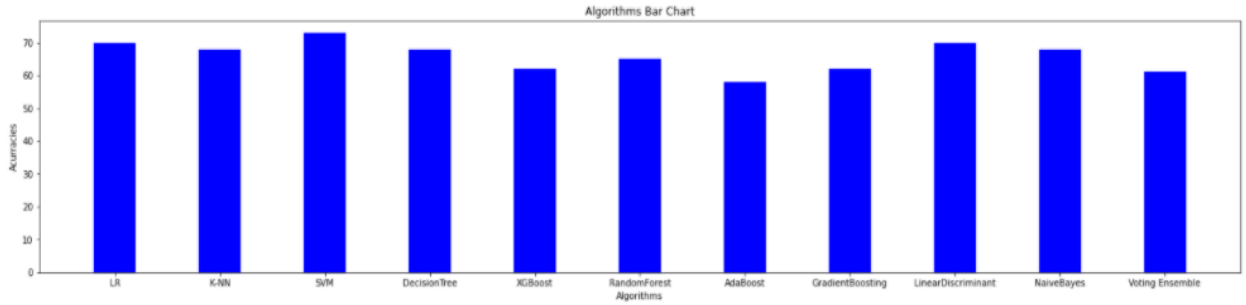
Features Selection Bar Chart (Algorithms):



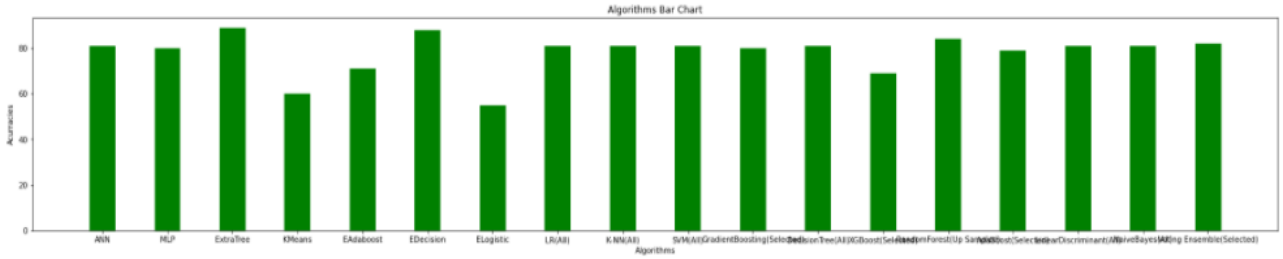
Up Sampled Data Bar Chart (Algorithms):



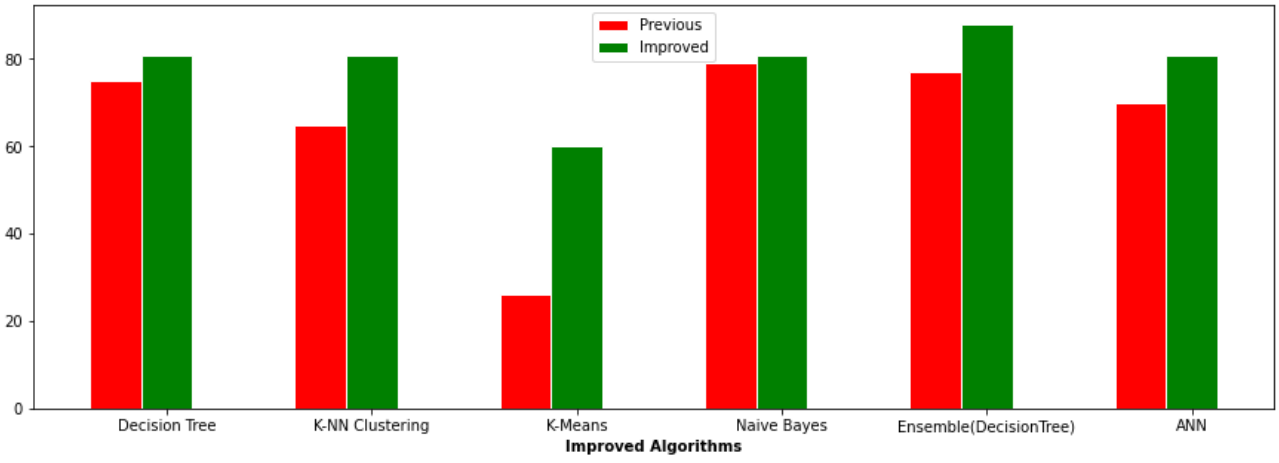
Down Sampled Data Bar Chart (Algorithms):

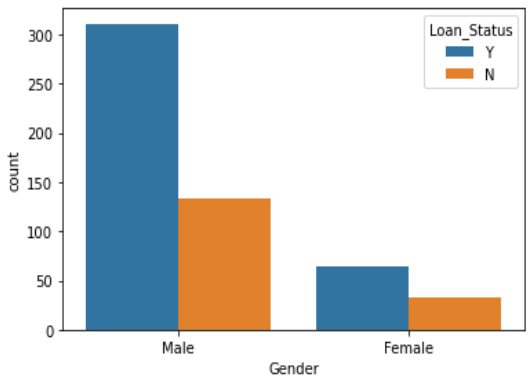


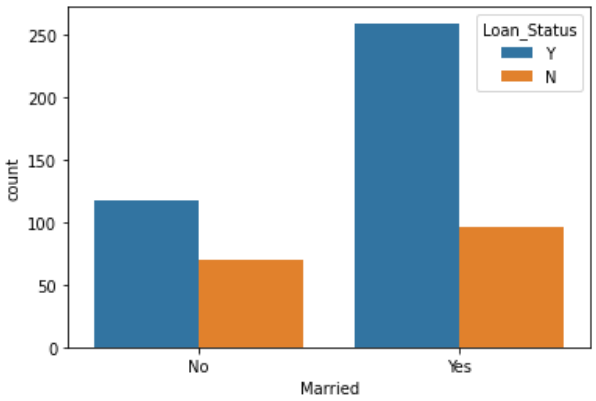
All Algorithms Bar chart (Algorithms):

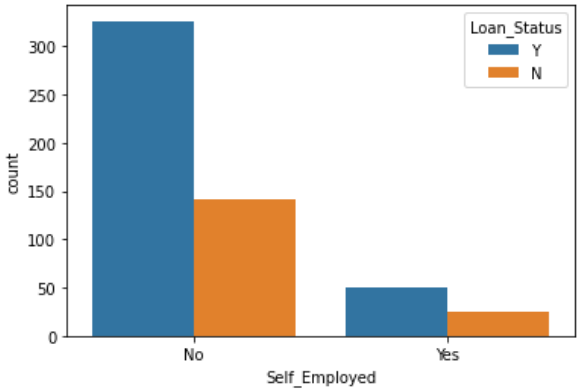


Improved Algorithms Bar Chart:

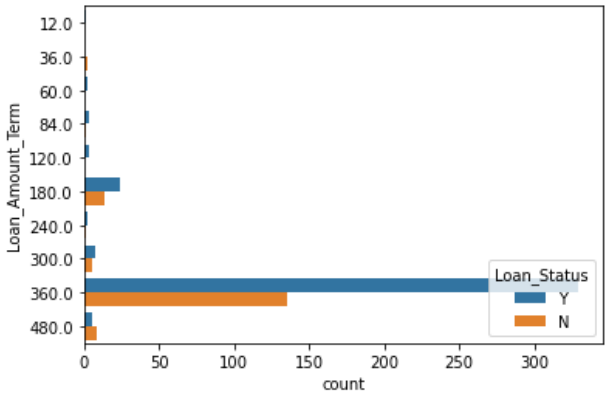



Loan_Status
Y
N
count
300
250
200
150
100
50
0
Male
Female
Gender


Loan_Status
Y
N
count
250
200
150
100
50
0
No
Yes
Married


Loan_Status
Y
N
count
300
250
200
150
100
50
0
No
Yes
Self_Employed

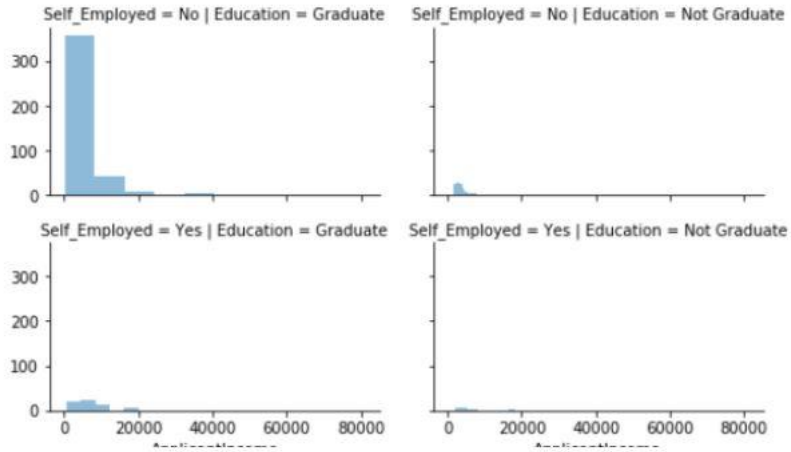

Loan_Status
Y
N
count
160
140
120
100
80
60
40
20
0
Urban
Rural
Semiurban
Property_Area




```
grid = sns.FacetGrid(df_train, row='Self_Employed', col='Education', size=2.2, aspect=1.6)
grid.map(plt.hist, 'ApplicantIncome', alpha=.5, bins=10)
grid.add_legend()
```

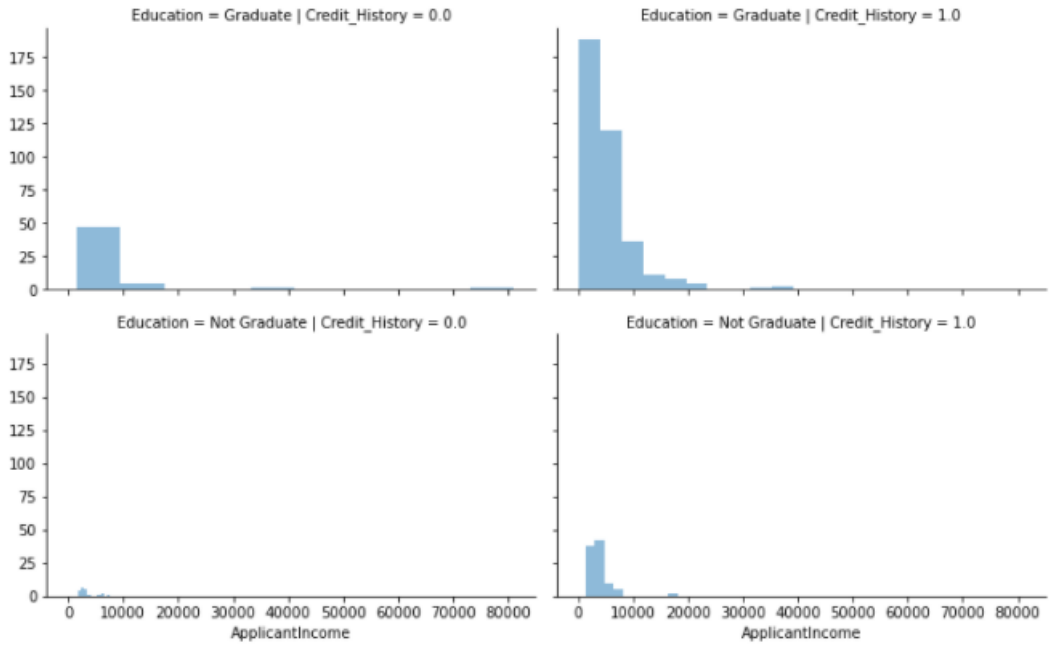
```
C:\Users\user\Anaconda3\lib\site-packages\seaborn\axisgrid.py:230: UserWarning: The `size` para
`; please update your code.
  warnings.warn(msg, UserWarning)
```

```
<seaborn.axisgrid.FacetGrid at 0x465265fd30>
```

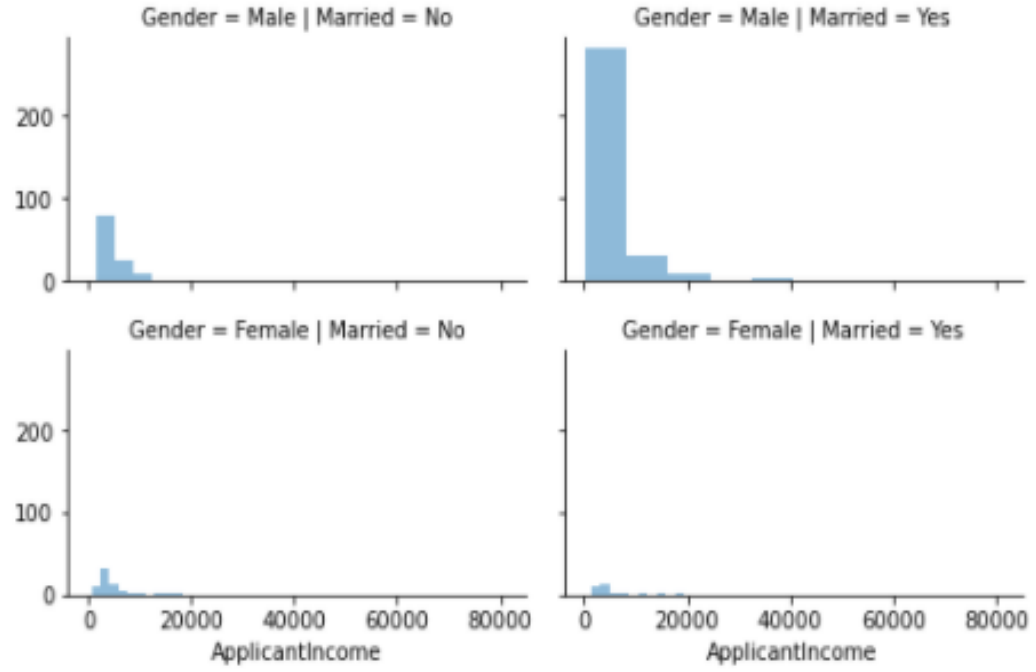


A graduate but not self-employed has more income

```
<seaborn.axisgrid.FacetGrid at 0x1d172e199c8>
```




```
<seaborn.axisgrid.FacetGrid at 0x1d16e7be148>
```



- After Collecting and cleaning the data, through various visualization tools available in python we can come up with assertive statements for market analysis and customer knowledge such as owning urban and semi urban properties is much more likely grant you a loan as compared to owing a rural property.
- We can also perform analysis on the workers such as if we compare the total number of applicant's male vs female and approved male vs female, we can see that male ratio is higher than female ratio.
- Logistic regression, Extra tree and Random forest have better accuracies.

CONCLUSION OF WORK
WITH IMPROVEMENTS

We have improved upon the random forest algorithm to match with the accuracy of logistic regression.

Implemented both supervised and unsupervised algorithms (and matched certain accuracies)

Improved upon the accuracies of all algorithms

Gave feature importance for loan prediction

THANK

YOU