Documentation

Introduction

After taking loans from the bank, some people are unable to repay their borrowed amount, resulting in a loss to the bank.

Data at hand:

We have been provided a Dataset of 804 loan takers and some information about them, out of which ***only 3 had missing values(0.37% of total data) and hence were removed***

The information provided was as follows:

- Customer ID
- Amount of money in checking account
- Duration of loan borrowed
- Credit score to ascertain the risk profile of a borrower
- Sex of customer
- Whether the customer is married or not
- Whether the customer has taken car loan or not
- Whether the customer has taken personal loan or not
- Whether the customer has taken home loan or not
- Whether the customer has taken education loan or not
- Whether the customer is employed or not
- Value of loan amount borrowed
- Balance of the savings account
- Total duration of employment
- Age of the customer
- Number of credit accounts a customer has
- Whether the customer defaulted or not

What can be done?

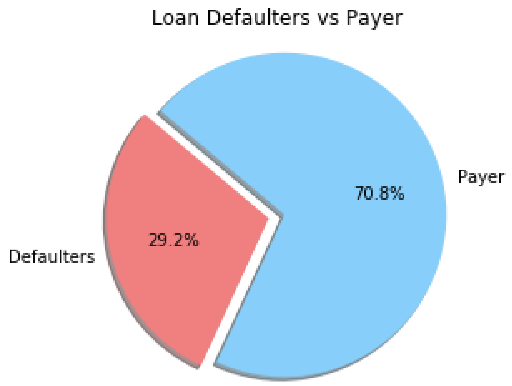
The aim of this project is to try to predict if someone will be a *'defaulter'* i.e. would be unable to repay the loan.

This will be done by analysing a given dataset consisting of various information about the previous borrowers (viz. Age, Employment status,credit score, etc.) and whether they were able to pay back the loan or not.

We will then be using a Neural network to fabricate a model which would predict if someone would be a defaulter, thus, helping the bank in deciding whether loan should be given or not.

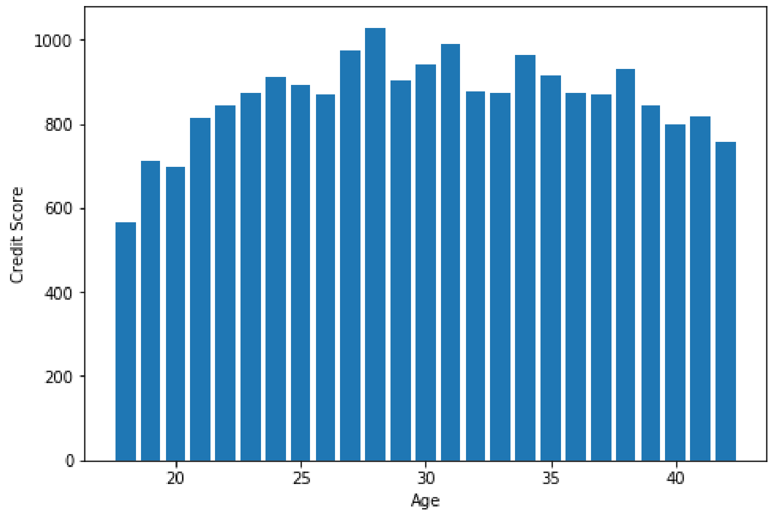
Data analysis

1. Percentage of loan defaulters vs Payers:



The above pie chart gives us an idea about the percentage of people who have not paid their loan in the past.

2. Credit score vs. Age:

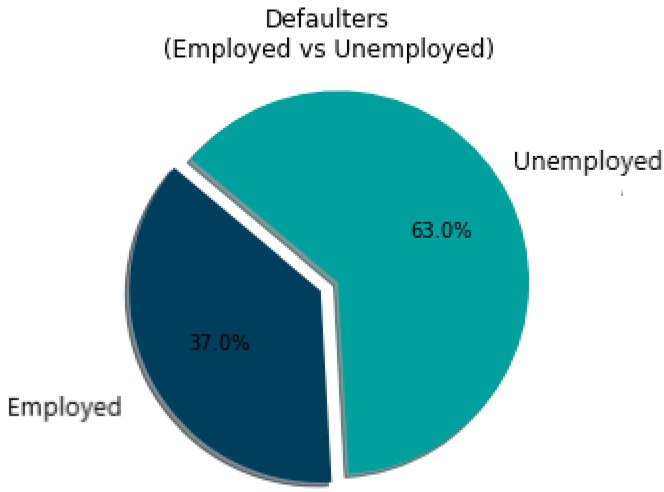


e

This graph shows us that the people between the ages 27 and 35 tend to have a slightly greater credit score when compared to other groups.

Note: The further analysis will be focused on the previous defaulters, thus giving us a better idea about the trend/patterns common in defaulters.

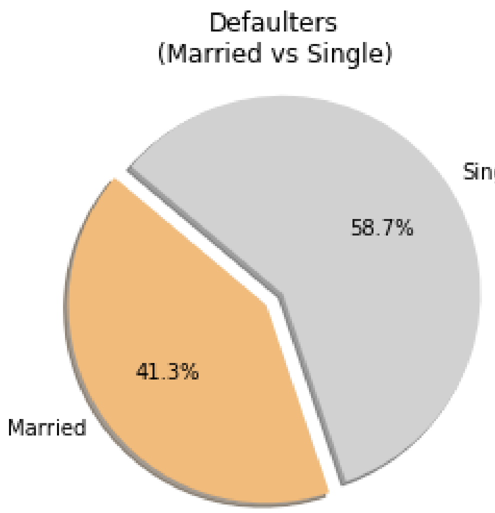
3. Defaulters (Employed vs unemployed)



This pie chart shows that the number of Unemployed defaulters is greater than that of Employed ones.

This difference does not seem extremely significant, one reason for that maybe that the Unemployed population is influenced by students who might have taken education loan.

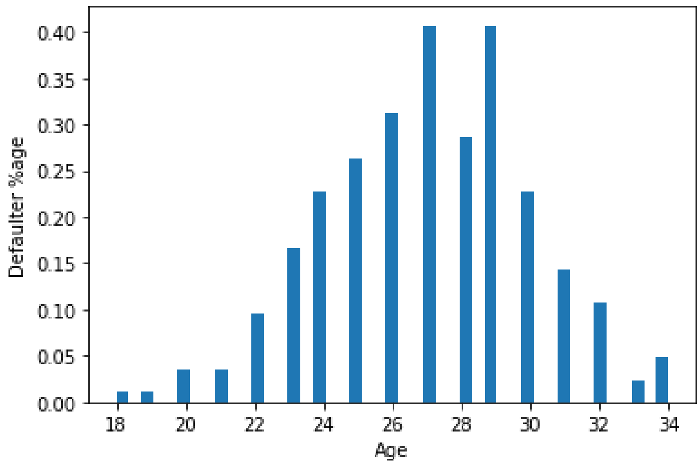
4. Defaulters (Married vs. Single)



We see that the number of Defaulters is greater with the single population.

This may be because most of the married population tend to make their decisions more cautiously and are more responsible in general.

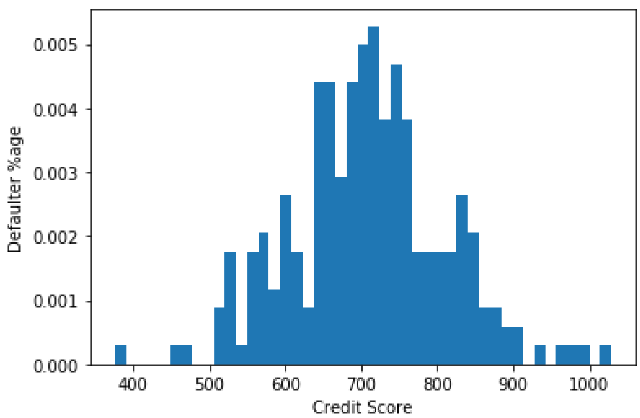
5. Defaulter %age vs Age:



The above graph shows that the population at the age between 25 to 29 is a major contributor to the defaulters.

The average age is 31.27, thus this metric is not affected by the sheer number of samples. The people in the age group 25-29 really tend to be defaulters.

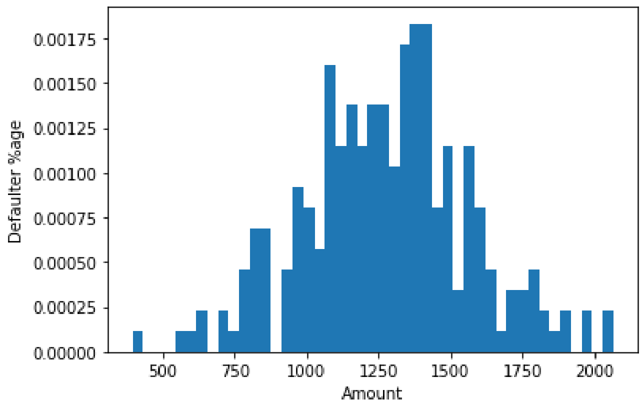
6. Defaulter %age vs Credit score:



We can infer from the above graph that people with credit score above 750 have a lot less contribution to the defaulters.

The lower values of 400-650 credit score can be explained by the fact that banks do not tend to give loans to people having score less than 500, thus they are not present in the dataset.

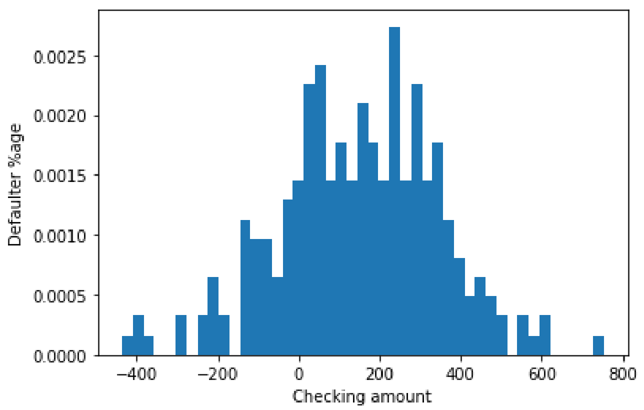
7. Defaulter %age vs Amount



This shows us that the people taking loan for the sum between 1000 & 1500 seem to be the major contributor to the defaulter population.

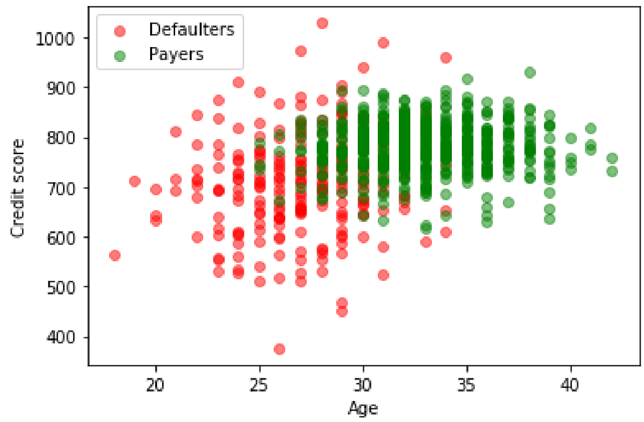
This graph may also be influenced by the average amount borrowed which is: *1217.63*

8. Defaulter %age vs checking amount:



Here we see that people with checking amount of 0-300 tend to be defaulters.

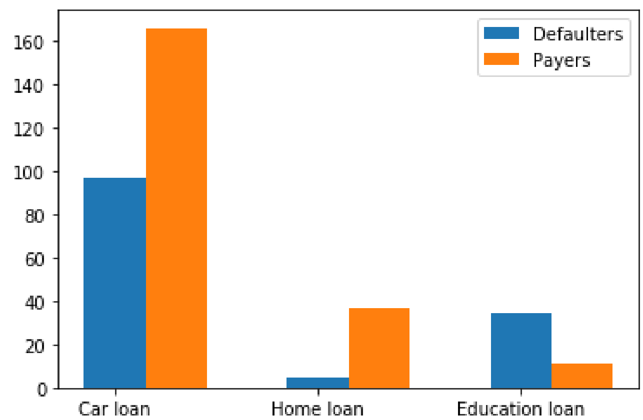
9. Credit score vs Age



The above scatter plot shows a clear distinction that:

- People with credit score of 600-700 and in the age of 20-25 seem to be defaulters,
- Those above the age of 35 do not indulge in Defaults at all.

10. Number of defaulters and payers w.r.t types of loans:



Here we can clearly see:

- Maximum number of loan taken is car loan.
- Defaults tend to be less in case of car and home loans
- Education loan has a higher number of defaulters than payers.

Conclusion

By the above analysis, we can conclude that:

- Employment status does not have a significant weightage in defining defaulters
- Marital status does not have a significant weightage in defining defaulters
- The amount borrowed may be a determining factor but the fact that average amount borrowed is 1217.63, reduces the credibility of this factor.
- Age is a determining factor for defaulters.
- When observing the scatter plot of age vs credit score, some clear boundaries are visible:
  - People with credit score of 600-700 and in the age of 20-25 seem to be defaulters,
  - Those above the age of 35 do not indulge in Defaults at all. Thus this is also a determining factor

- Also, we see that:
  - Maximum number of loans taken is car loan.
  - Defaults tend to be less in case of car and home loans
  - Education loan has a higher number of defaulters than payers.

Feature Selection

After the above data analysis, we inferred that certain features about the bank customers had a lower weightage and were not a determining factor for the defaults.

To further help the training process of the neural network we applied the SelectKPercentile

Feature selection model which reduced our features to only 4 namely: **Checking amount, Credit score, Age and Savings amount,** which is similar to our data analysis conclusion.

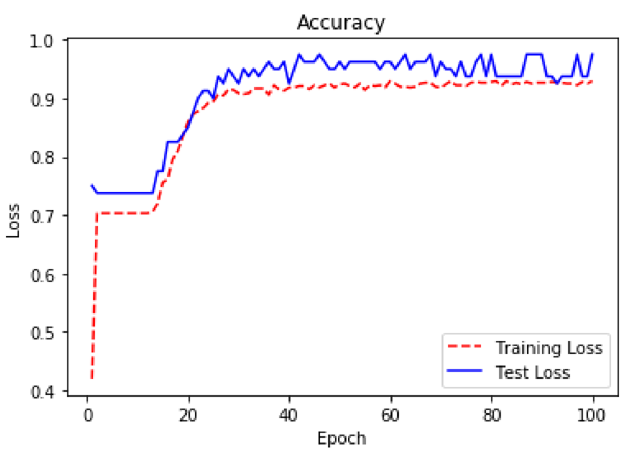
Neural Network modelling

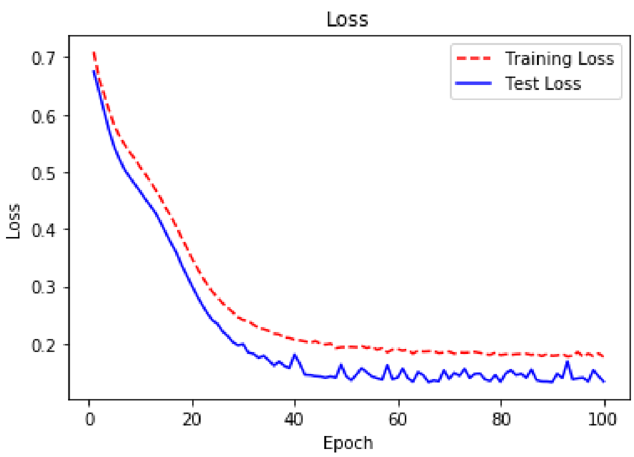
We chose neural network for making the model, as, this kind of problem requires a higher accuracy, this can be a deciding factor for the annual profits of the bank

A simple Deep neural network was created with 5 dense layers.

The Network had an accuracy score of 97.5%

The accuracy and Loss curves are given below:






```
# f1: 2 tp / (2 tp + fp + fn)
f1 = f1_score(y_test, yhat_classes)
print('F1 score: %f' % f1)

Accuracy: 0.975000
Precision: 1.000000
Recall: 0.904762
F1 score: 0.950000
```

Submission by:

Amanul Haque
B.I.T ,Durg, (c.g)
8839807138
amanul.h.2030@gmail.com

Bhumi Agrawal
JDMC,D.U.
6264856770
bhumiagrawal20209@gmail.com