***Mini Project Report on***

**Credit Risk Analysis using Artificial Neutral Networks**

By

# Y Ningthoibi Devi (Reg. No. 201700218)

# Bidisha Borgohain (Reg. No.-201700349)

# Pushkar Srivastava (Reg. No.-201700331)

# Group Id - 22

*In partial fulfillment of requirements for the award of degree in Bachelor of Technology in Computer Science and Engineering (2020)*

Under the Project Guidance of

# Chitrapriya N

## Assistant Professor

**Sikkim Manipal Institute of Technology, Majhitar DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY**

(A constituent college of Sikkim Manipal University) MAJITAR, RANGPO, EAST SIKKIM – 737136

# PROJECT COMPLETION CERTIFICATE

This is to certify that the below mentioned students of Sikkim Manipal Institute of Technology has worked under my supervision and guidance from **8th January 2020 to 15th May 2020** and has successfully completed the project entitled **“Credit Risk Analysis using Artificial Neural Networks”** in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

|  |  |  |
| --- | --- | --- |
| University Registration No | Name of Student | Course |
| **201700218**  **201700349**  **201700331** | **Y Ningthoibi Devi**  **Bidisha Borgohain**  **Pushkar Srivastava** | **B.Tech (CSE)**  **B.Tech (CSE)**  **B.Tech (CSE)** |

t

## Chitrapriya N

Assistant Professor

Department of Computer Science and Engineering Sikkim Manipal institute of Technology

Majhitar, Sikkim – 737136

I

# PROJECT REVIEW CERTIFICATE

This is to certify that the work recorded in this project report entitled **“Credit Risk Analysis using Artificial Neural Neutrals”** has been jointly carried out by **Y Ningthoibi Devi (Reg. 201700218), Bidisha Borgohain (Reg. 201700349) and Pushkar Srivastava (Reg. 201700331)** of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering. This report has been duly reviewed by the undersigned and recommended for final submission for Mini Project Viva Examination.

## Chitrapriya N

Assistant Professor

Department of Computer Science and Engineering Sikkim Manipal Institute of Technology

Majhitar, Sikkim – 737136

II

# CERTIFICATE OF ACCEPTANCE

This is to certify that the below mentioned students of Computer Science & Engineering Department of Sikkim Manipal Institute of Technology (SMIT) have worked under the supervision of **Chitrapriya N** of Assistant Professor, Department of Computer Science and Engineering from **8th January 2020 to 15th May 2020** on the project entitled **“Credit Risk Analysis using Artificial Neural Networks”.**

The project is hereby accepted by the Department of Computer Science & Engineering, SMIT in partial fulfillment of the requirements for the award of Bachelor of Technology in Computer Science and Engineering.

|  |  |  |
| --- | --- | --- |
| University Registration No | Name of Student | Project Venue |
| **201700218** | **Y Ningthoibi Devi** | **SMIT** |
| **201700349** | **Bidisha Borgohain** | **SMIT** |
| **201700331** | **Pushkar Srivastava** | **SMIT** |

## Prof. (Dr.) Kalpana Sharma

Professor & Head of Department

Computer Science & Engineering Department Sikkim Manipal Institute of Technology Majhitar, Sikkim – 737136

III

# DECLARATION

We, the undersigned, hereby declare that the work recorded in this project report entitled “**Credit Risk Analysis using Artificial Neural Networks**” in partial fulfillment for the requirements of award of B.Tech (CSE) from Sikkim Manipal Institute of Technology (A constituent college of Sikkim Manipal University) is a faithful and bonafide project work carried out at “**SIKKIM MANIPAL INSTITUTE OF TECHNOLOGY**” under the supervision and guidance of **Chitrapriya N,** Assistant Professor, Department of Computer Science and Engineering.

The results of this investigation reported in this project have so far not been reported for any other Degree or any other Technical forum.

The assistance and help received during the course of the investigation have been duly acknowledged.

# Y Ningthoibi Devi

# (Reg. No.-201700218)

# Bidisha Borgohain

# (Reg. No.-201700349)

# Pushkar Srivastava

# (Reg. No.-201700331)

IV

**ACKNOWLEDGMENT**

We take this opportunity to acknowledge indebtedness and deep sense of

gratitude to our guide **Chitrapriya N** whose valuable guidance and kind supervision gave us throughout

the course which shaped the present work as it shows.

We pay our deep sense of gratitude to **Prof. (Dr.) Kalpana Sharma, H.O.D, Computer Science &**

**Engineering Department, Sikkim Manipal Institute of Technology** for giving us the opportunity to

work on this project and provided all support required.

We obliged to our project coordinators **Dr. Sandeep Gurung, Mr. Santanu Kr. Misra, Mr. Biraj**

**Upadhyaya and Ms. Nitisha Pradhan** for elevating, inspiration and kind supervision in completion of

our project.

We would also like to thank any other staff of **Computer Science & Engineering**

**Department, Sikkim Manipal Institute of Technology** for giving us continuous

support and guidance that has helped us in completion of our project.

**Y Ningthoibi Devi (Reg. No.-201700218)**

**Bidisha Borgohain (Reg. No.-201700349)**

**Pushkar Srivastava (Reg. No.-201700331)**

V

**DOCUMENT CONTROL SHEET**

|  |  |  |
| --- | --- | --- |
| 1 | Report No | CSE/Mini Project/Internal/B.Tech/B/Group ID:22/2020 |
| 2 | Title of the Report | Credit Risk Analysis using  Artificial Neutral Networks |
| 3 | Type of Report | Technical |
| 4 | Author | Y Ningthoibi Devi  Bidisha Borgohain  Pushkar Srivastava |
| 5 | Organizing Unit | Sikkim Manipal Institute of Technology |
| 6 | Language of the Document | English |
| 7 | Abstract | “Credit Risk Analysis Using Artificial Neural Networks” aims to analyse the risk associated with granting loans to customers in a financial institutions using Artificial Neural Networks and predict whether a customer is risky or non risky depending on how much he/she is capable of paying back the loan. |
| 8 | Security Classification | General |
| 9 | Distribution Statement | General |

**TABLE OF CONTENTS**

**TABLE OF CONTEN**

VI

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapter** | **Title** | **Page No.** |
|  | **Abstract** |  |
| **1** | **Introduction** |  |
|  | 1.1 General Overview of the Problem | 1 |
|  | 1.2 Literature Survey | 2,3 |
|  | 1.3 Problem Definition | 4 |
|  | 1.4 Proposed Solution Strategy | 5 |
| **2** | Implementation | 6 |
| **3** | Experimental Results | 11 |
| **4** | Testing and validation | 12 |
| **5** | Conclusion | 13 |
| **6** | Limitations and Future Scope | 14 |
| **7** | Gantt Chart | 15 |
| **8** | References | 16 |

VII

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Name** | **Page No.** |
| 1 | Workflow flowchart | 5 |
| 2,3,4 | Dataset Information | 6 |
| 5,6 | Dependent Feature | 6 |
| 7,8 | Missing values | 7 |
| 9 | Categorical Data | 7 |
| 10 | Encoding of the categorical value | 8 |
| 11  12 | Handling of missing values  Feature Scaling | 8  8 |
| 13 | Splitting of train and test set | 9 |
| 14,15 | Feature Selection | 9 |
| 16 | Hyper-parameter tuning | 10 |
| 17  18  19 | Neural network model summary  VIII  Training phase  Testing &validation | 10  10  12 |
|  | IX |  |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Title Name** | **Page no.** |
| 1 | Training and validation data accuracy | 11 |
| 2 | Testing and the validation of the predictions | 12 |

**X**

**ABSTRACT**

1. Credit risk or credit default indicates the probability of non-repayment of bank financial services that have been given to the customers.​
2. Credit risk plays a crucial role for banks and financial institutions, especially for commercial banks and it is always difficult to interpret and manage the credit risk.​
3. Neural network algorithms has the ability to tackle the problem of analysing credit default and predicting the credit worthiness of the loan application over a time period.​

**1.INTRODUCTION**

**1.1 General overview**

Our project entitled “Credit Risk Analysis Using Artificial Neural Networks” aims to analyse the risk associated with granting loans to customers in a financial institutions using Artificial Neural Networks and predict whether a customer is risky or non risky depending on how much he/she is capable of paying back the loan.

Artificial Neural Network connection between two neurons indicates the pathway for the flow of information. Each connection has a weight, an integer number that controls the signal between the two neurons.

If the network generates a “good or desired” output, there is no need to adjust the weights. However, if the network generates a “poor or undesired” output or an error, then the system alters the weights in order to improve subsequent results.

In this project, we aim to use Supervised Learning technique.

Supervised learning is a learning in which we teach or train the machine using data which is well labelled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data (set of training examples) and produces a correct outcome from labelled data.

**LITERATURE SURVEY**

|  |  |  |  |
| --- | --- | --- | --- |
| **SI No.​** | **Paper and Author Details​** | **Findings​** | **Relevance to the Project​** |
| 1.​ | Name:Credit Risk Prediction Using Artificial Neural Networks  Author: Shruti Goyal  [I]  ​  ​  ​  ​ | 1.This paper has studied the Artificial Neural Networks and Linear Regression models for credit default.  2.Both the models had equal effect on the dataset and gave efficient results. Artificial Neural Networks gave an accuracy of 97.67% and linear regression of 97.69%.  3.While fitting a model using the neural network process, user needed to take extra care of the attributes and the normalization of the data to improve the performance.  ​  ​ | 1.Efficiency of the Artificial Neural Networks to improve the performance of the model.  2.The importance of the attributes and data normalization in determining the efficiency of the neural network. |
| 2. | Name:Measuring Credit Risk of Bank Customers Using Artificial Neural Network  Authors:Mohsen Nazari, Mojtaba Alidadi  [II] | 1.This paper determines the most important discriminants for rating the customers in an Iranian Bank.  2.Individual loan frequency and the loan amount had the most importance and status of the customer’s bank account ,previous relation with the bank had the least importance in the determination of classification of bad and good customers.  3. The model has been build using one hidden layer and hyper tangent activation function. | 1.The architecture of the ANN model.  2.The importance of attributes like loan frequency and loan amount in the classification of good and bad customers. |
| 3. | Name:Evaluating Credit Risk Using Artificial Neural Networks  Author: Qeethara K. Al-Shayea and Ghaleb A.El-Refae  [III] | 1.Two well known and available datasets have been used (German and Australian dataset) to test the neural network using feedforward back propagation neural network.  2.The accuracy on the German dataset is 77% and on the Australian Dataset is 86%.  3.The results of applying the proposed neural networks to distinguish between the customers based upon selected parameters showed very good abilities of the network to learn patterns. | 1.Use of the neural network model with learning algorithm: the feed forward back propagation neural network with supervised learning in terms of their ability to classify bad or good customers. |
| 4. | Name:Credit Risk Prediction: A comparative study between Discriminant Analysis and the neural network approach.  Author: ihem Khemakhem ,and Younés Boujelbènea  [IV] | 1.The neural network approach differs from the conventional method of credit scoring in its nature of the black box and its ability to handle the non- linear relations between variables.  2.The back-propagation algorithm was applied on the entire learning dataset.  3.Neural network outperforms the discrimant analysis in the credit risk prediction. | 1.use of back-propagation algorithm to decrease the error |

**PROBLEM DEFINITION**

* Credit risk develops from the probability that the borrowers may be unwilling or unable to fulfill their contractual obligations.​
* There is a crucial need For any credit-granting institution, such as commercial banks and certain retailers, the ability to discriminate good customers from bad ones.

**SOLUTION STRATEGY**

* An analysis of the data of the bank customers are done.
* A feedforward artificial neural network is used for developing the credit classification model. Here the information flow is unidirectional. There are no feedback loops. The inputs and outputs are fixed.
* A training or learning algorithm , that is , the back-propagation algorithm is used to develop the model. This algorithm learns by example. We submit to the algorithm what the network wants to do, that is, to differentiate between the risky and non-risky customers. It changes the networks weights so that it can produce the desired output for a particular input.

Training Dataset

Hyper-parameter tuning

Build the Neural Network

Data Pre-Processing

Collection of data of bank Customers

Training Phase

Non Accurate

Attributes

Testing Dataset

Accurate

Risky

Classification

Testing Phase

Non-Risky

Figure 1

IMPLEMENTATION

1. A brief look at the dataset

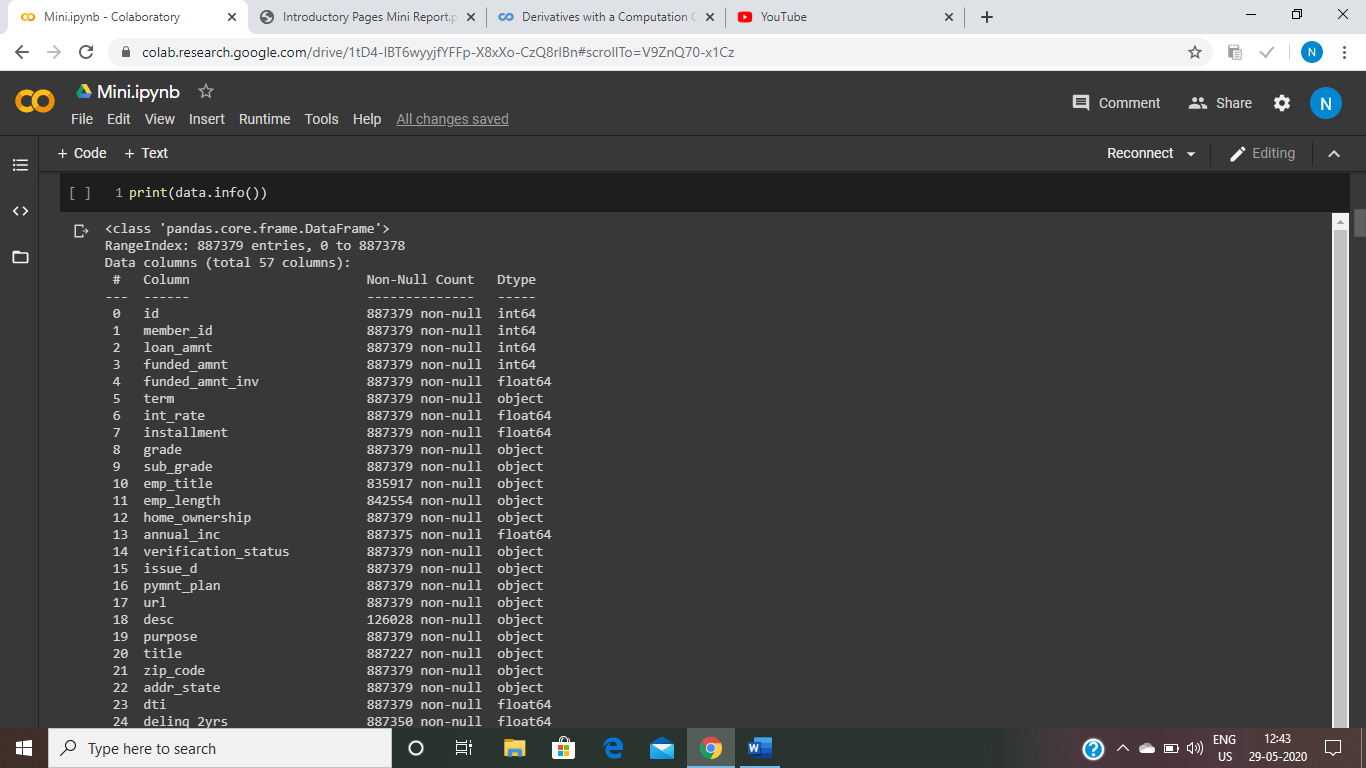
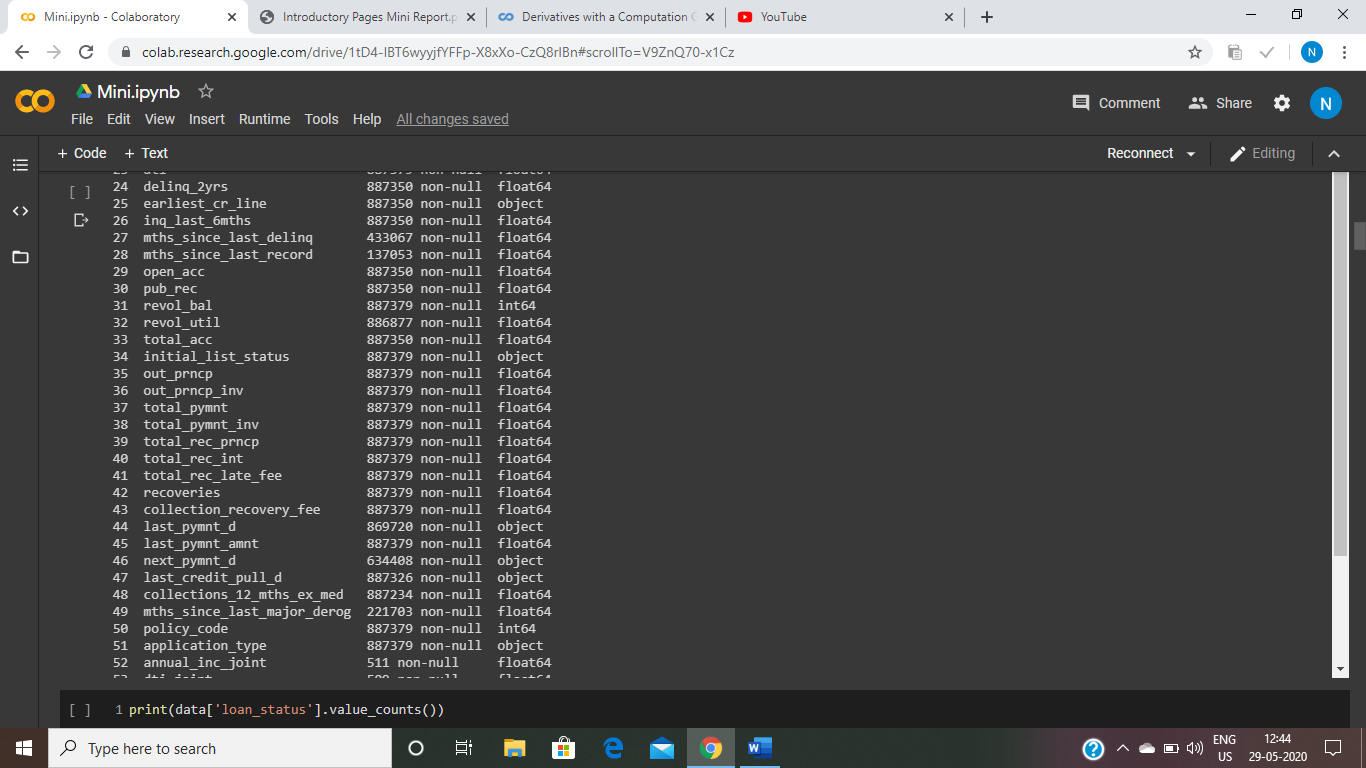
 

Figure 2 Figure 3

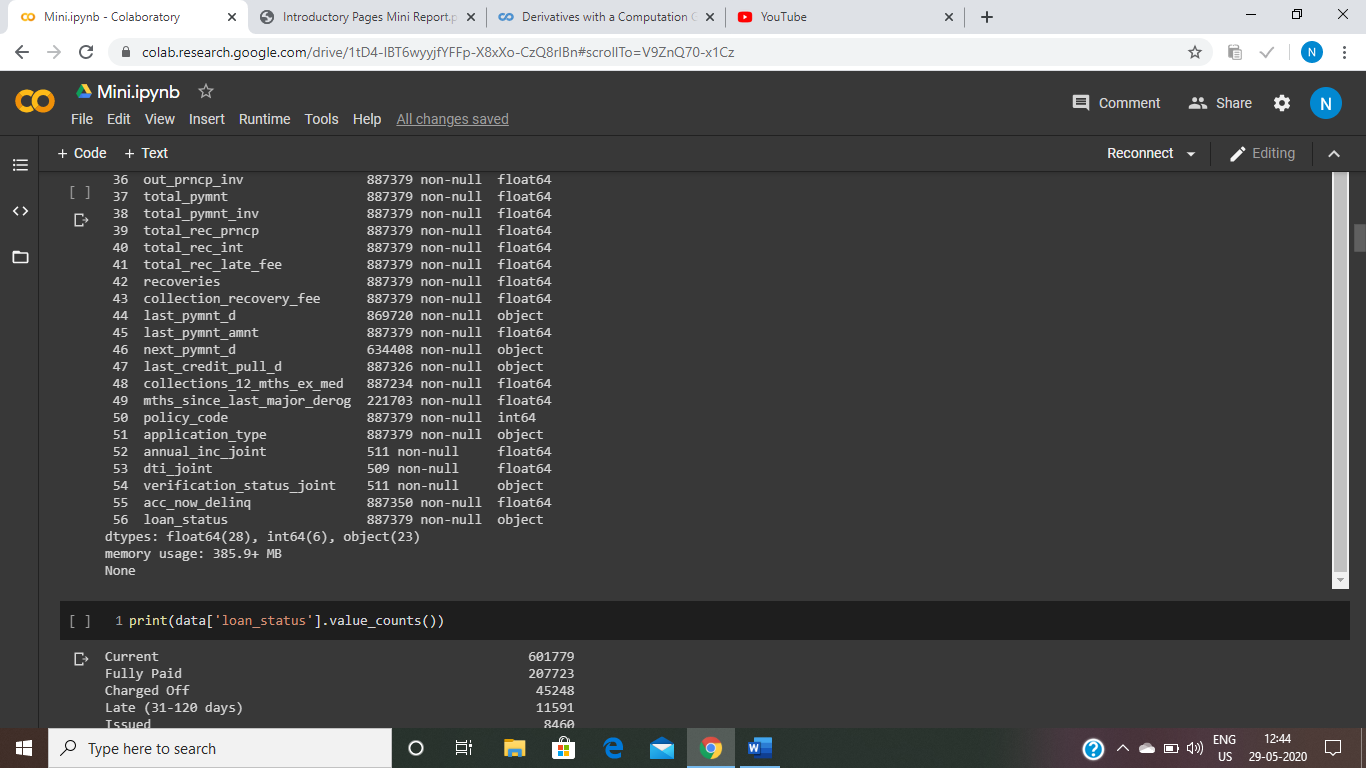


Figure 4

Total no of features: 57 features(56 independent feature, 1 dependent feature)

Total no of rows(total no of customer data): 887379

1. A brief look at the dependent feature

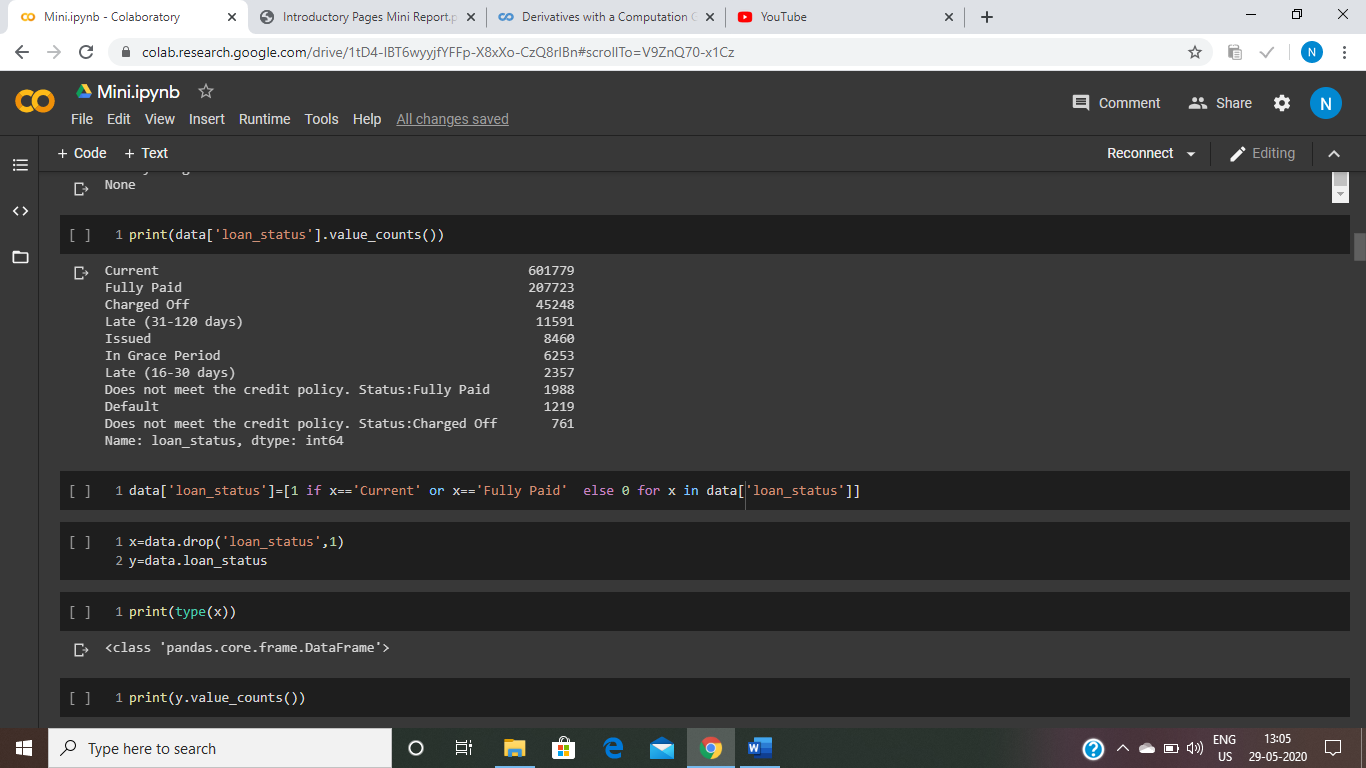


Figure 5

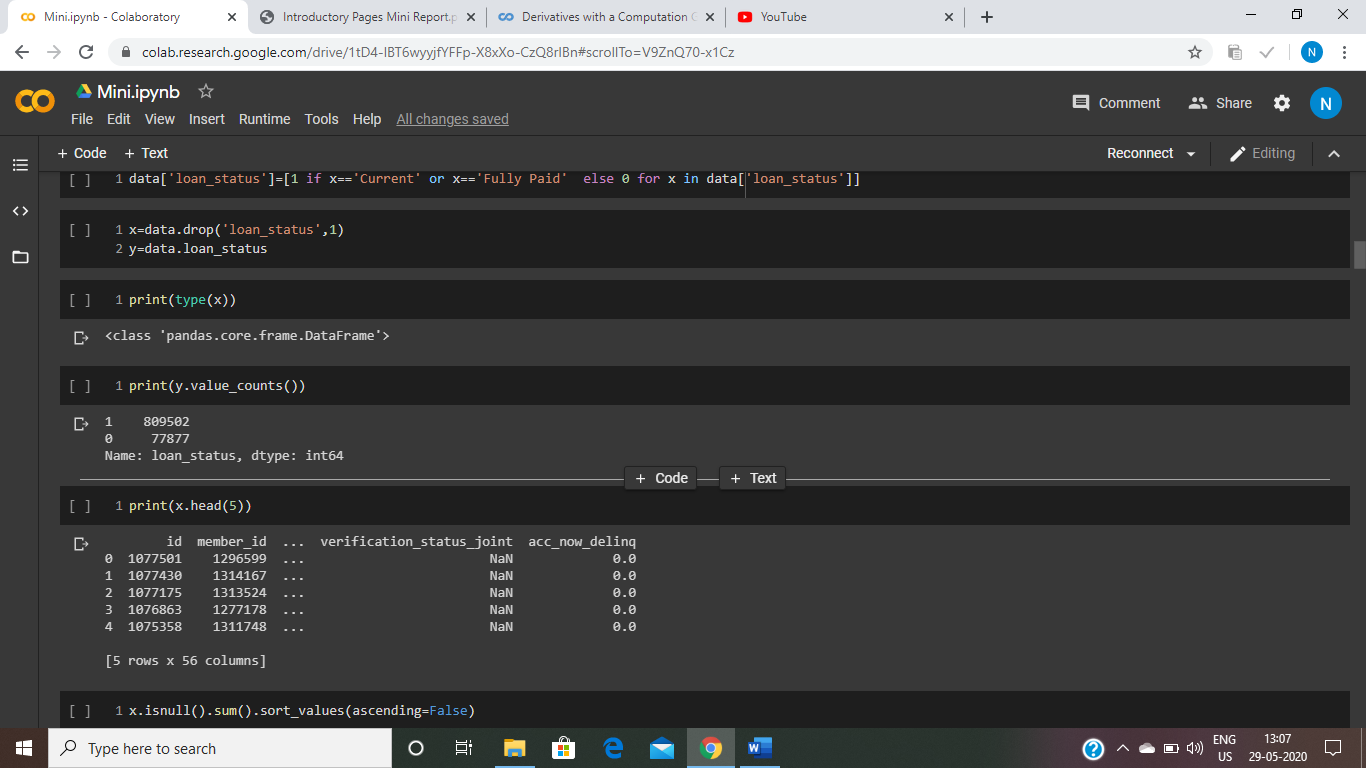


Figure 6

Loan status is the dependent feature,i.e, the target outputs.

The label ‘1’ denotes the non risky customers; label ‘0’ denotes the risky customers.

Total no of risky customers:77877

Total no of non risky customers:809502

1. A brief look at the missing values

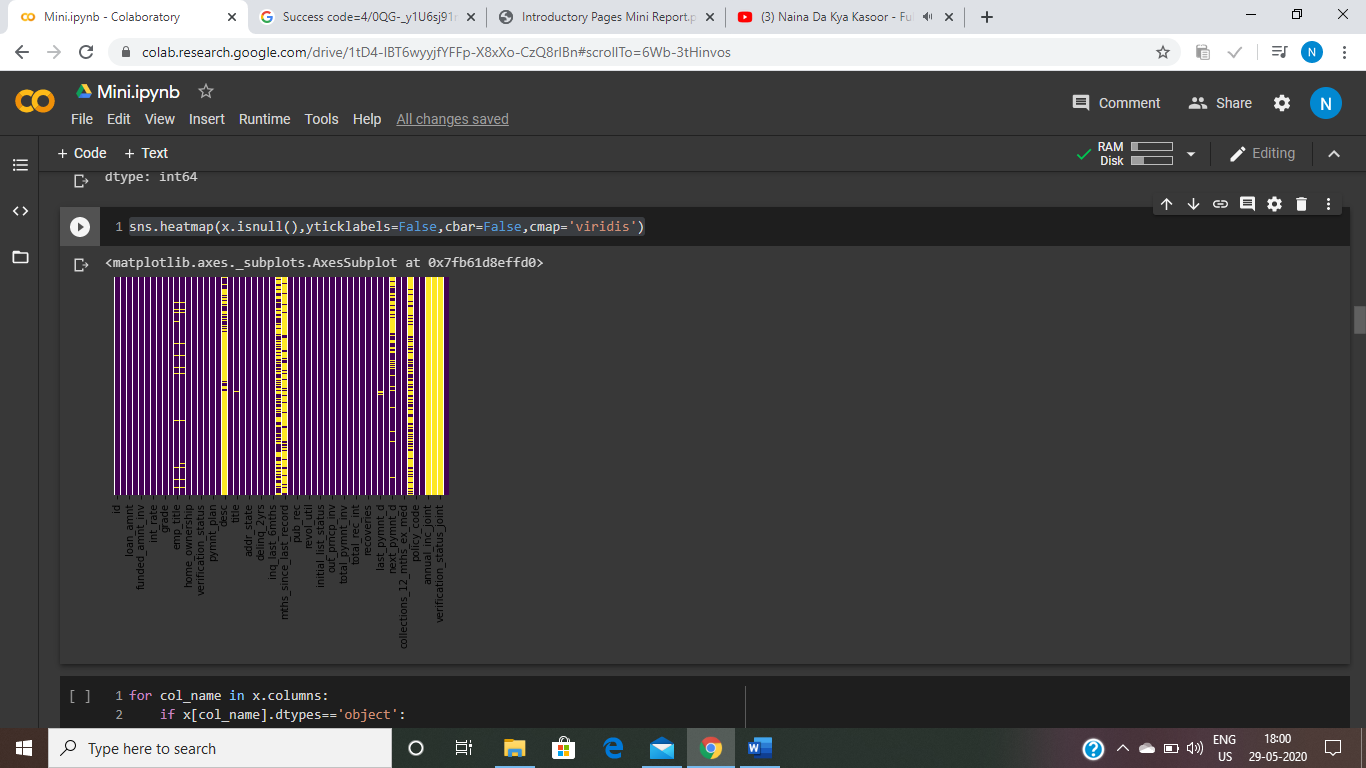
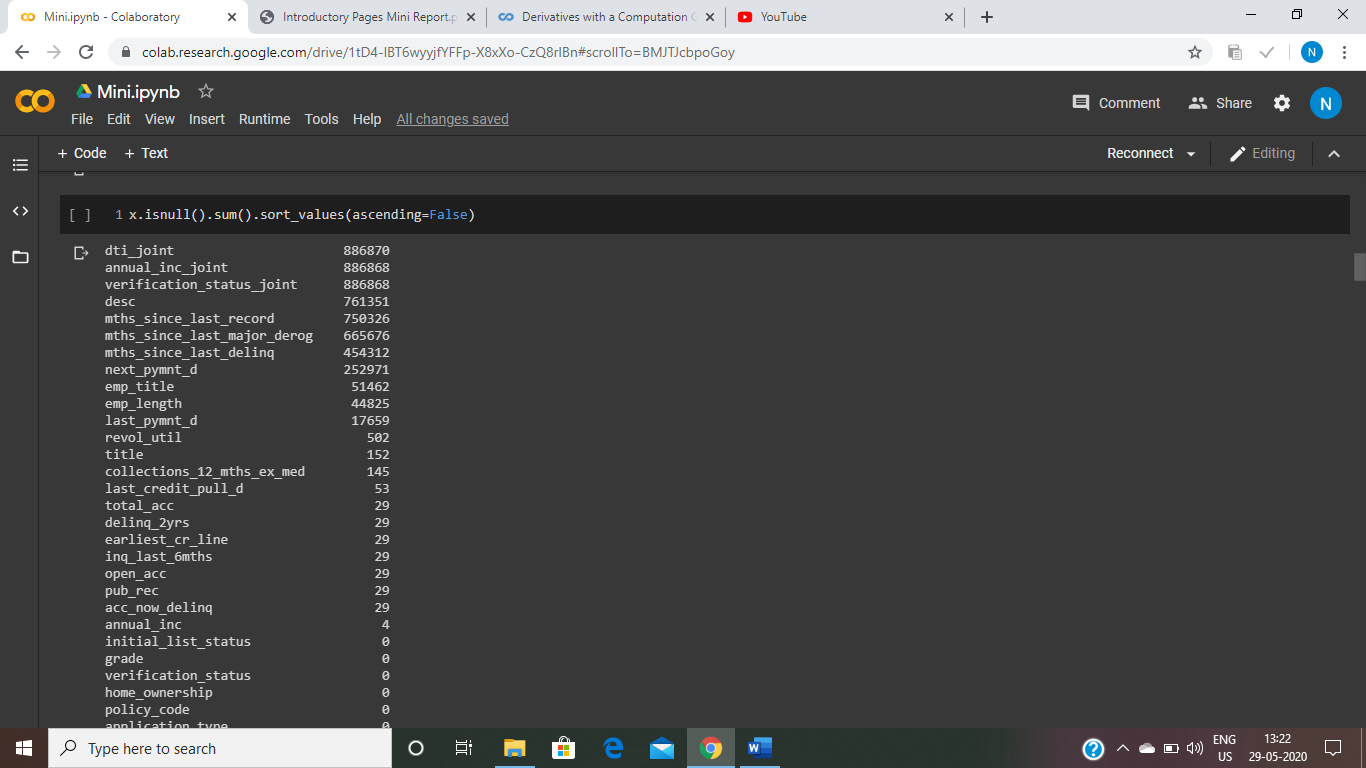
 

Figure 7 Figure 8

1. A brief look at the categorical data

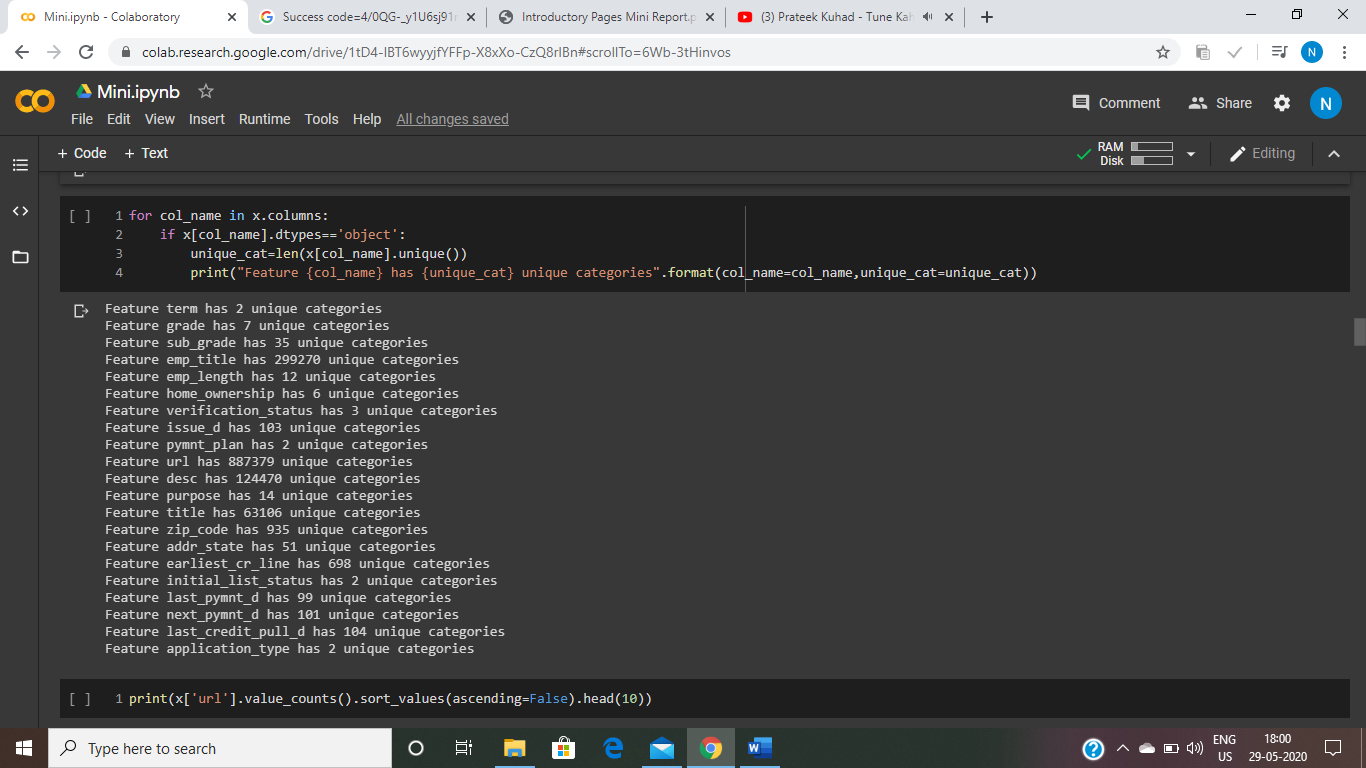


Figure 9

1. Data pre-processing:
   1. Handling the categorical data

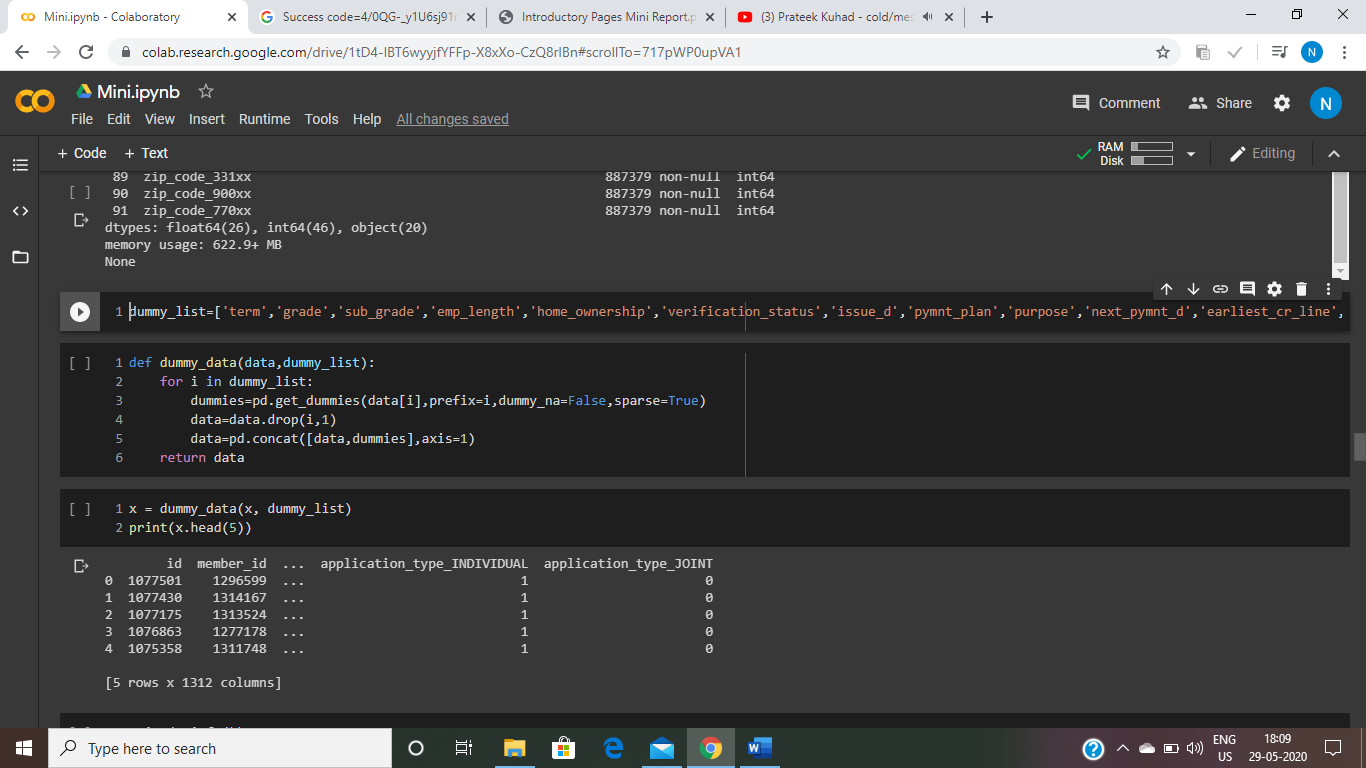


Figure 10

* 1. Handling the missing values:

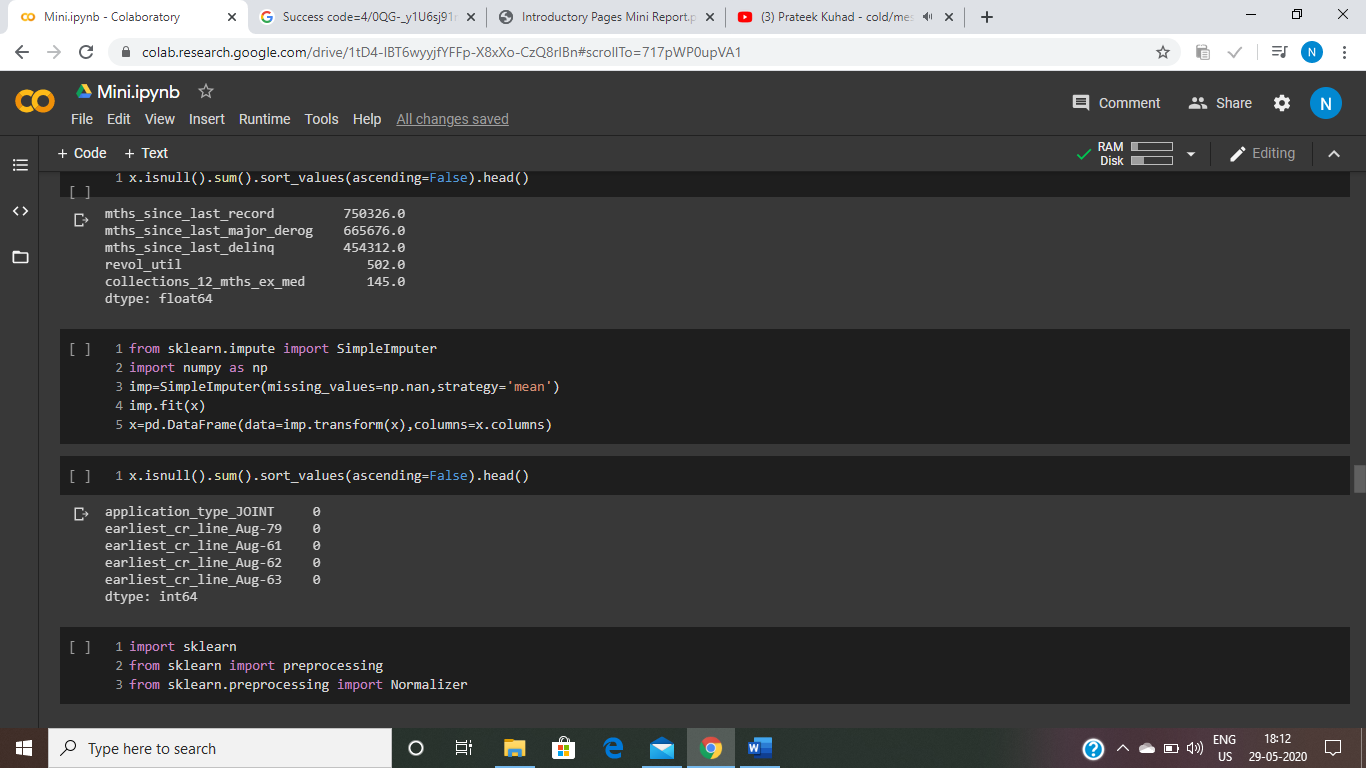


Figure 11

* 1. Scaling down the features

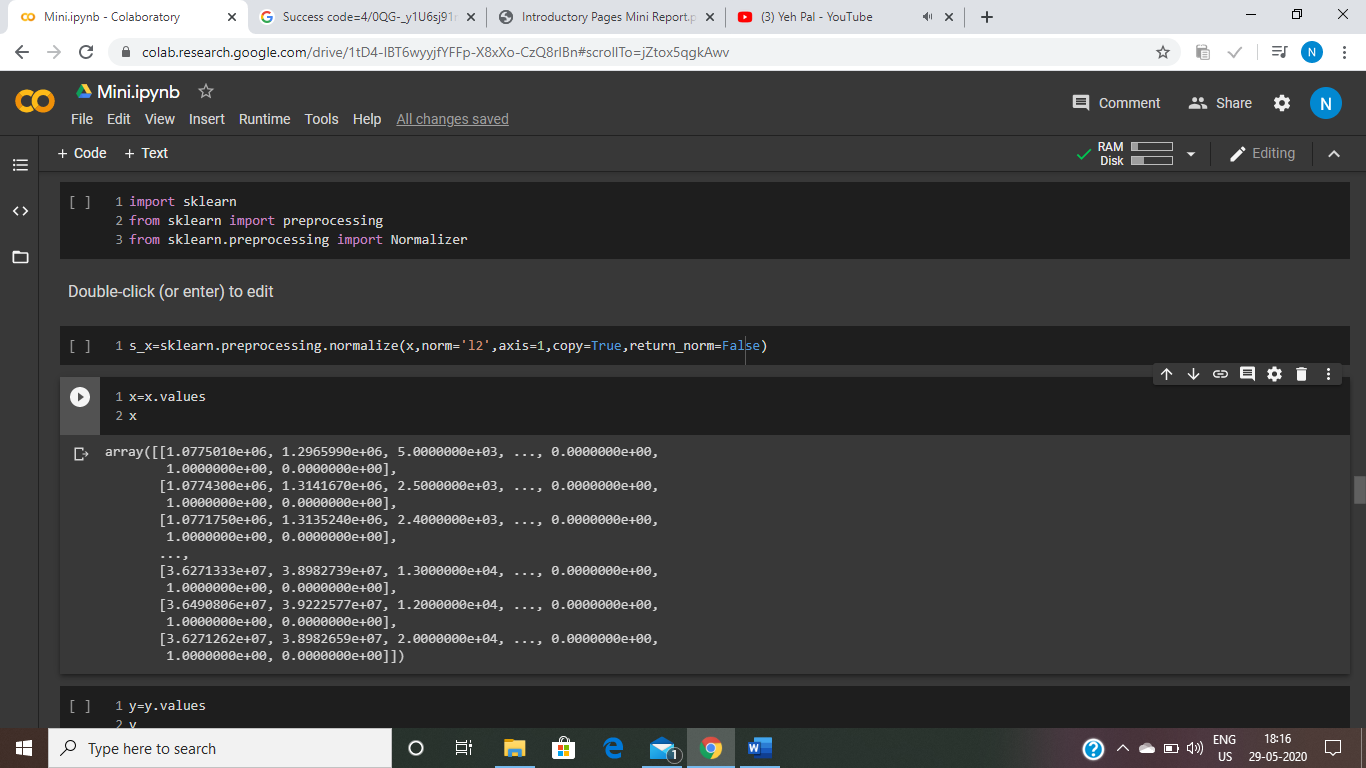


Figure 12

* 1. Splitting of the training and testing set

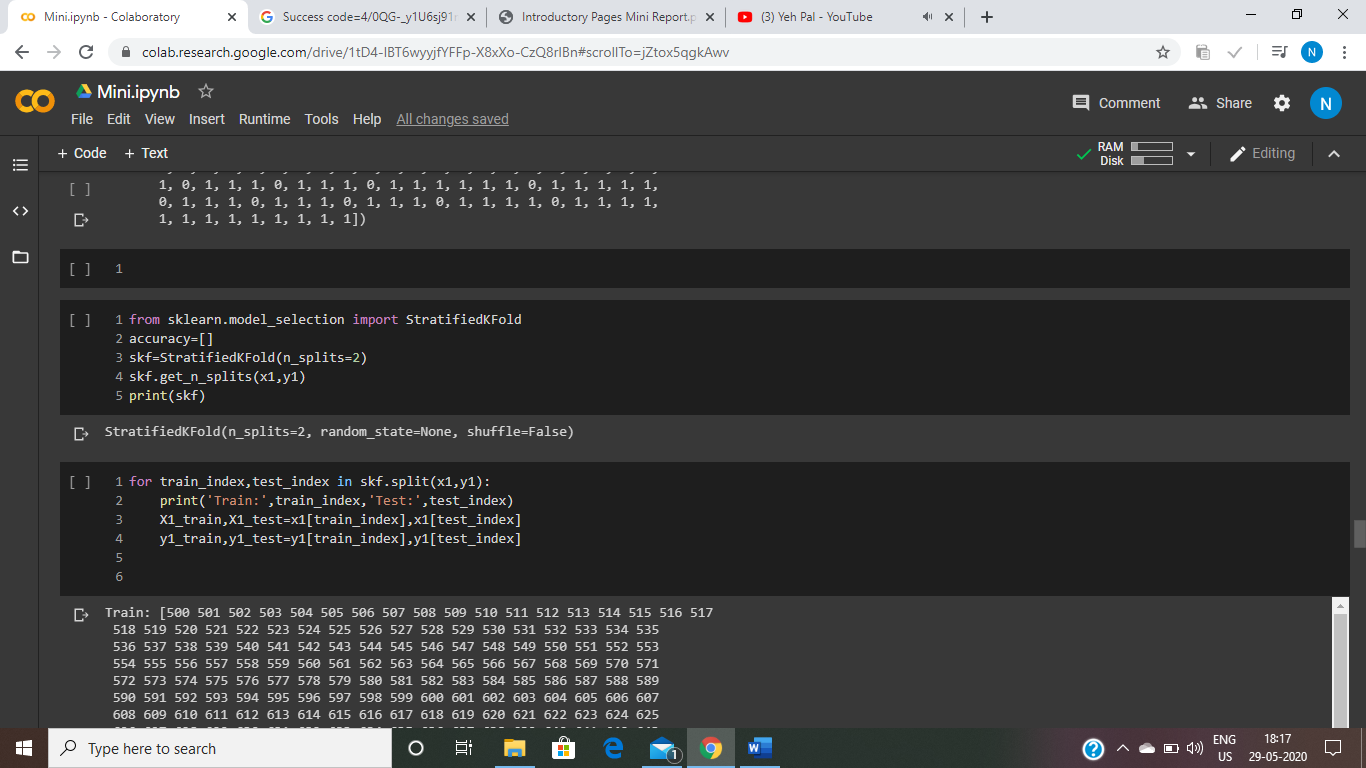


Figure 13

1. Feature Selection

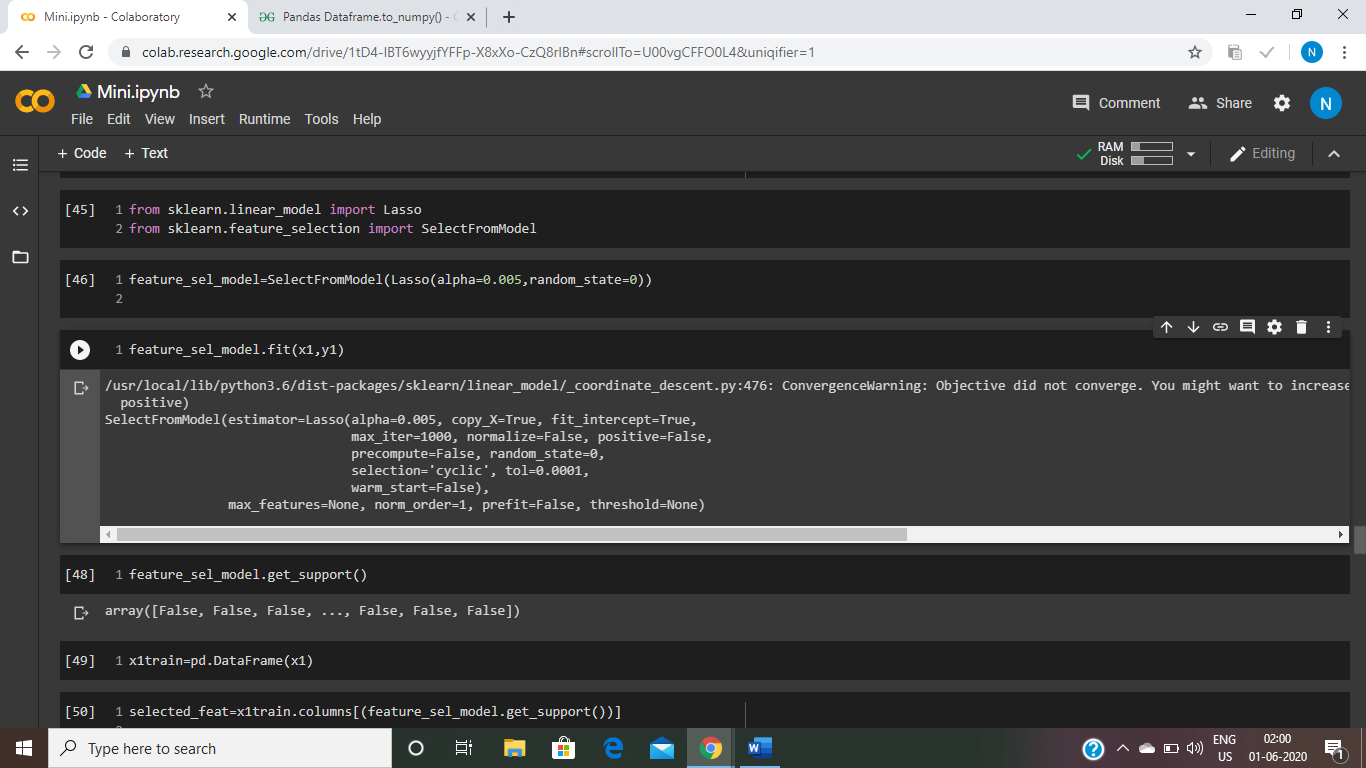


Figure 14

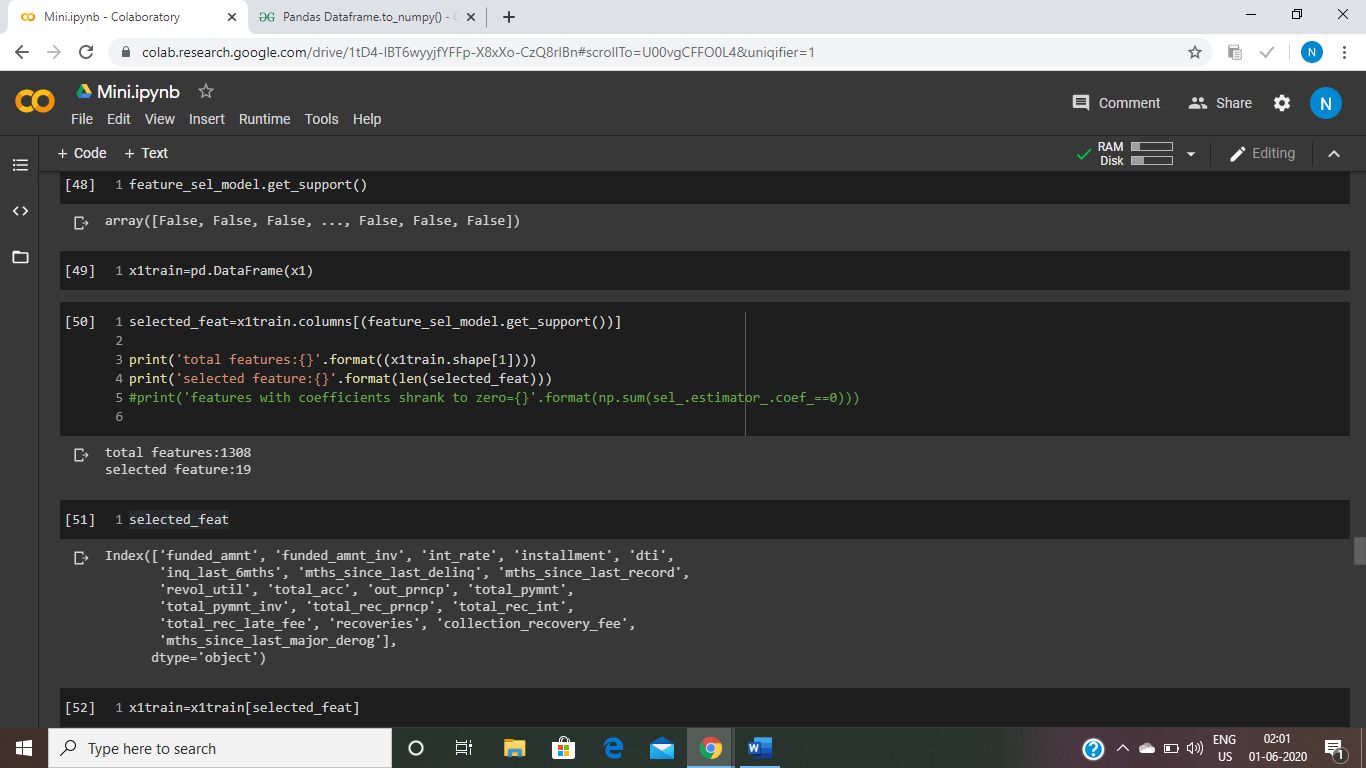


Figure 15

Total no of features that contributes to the prediction =19 features

1. Hyperparamter Tuning

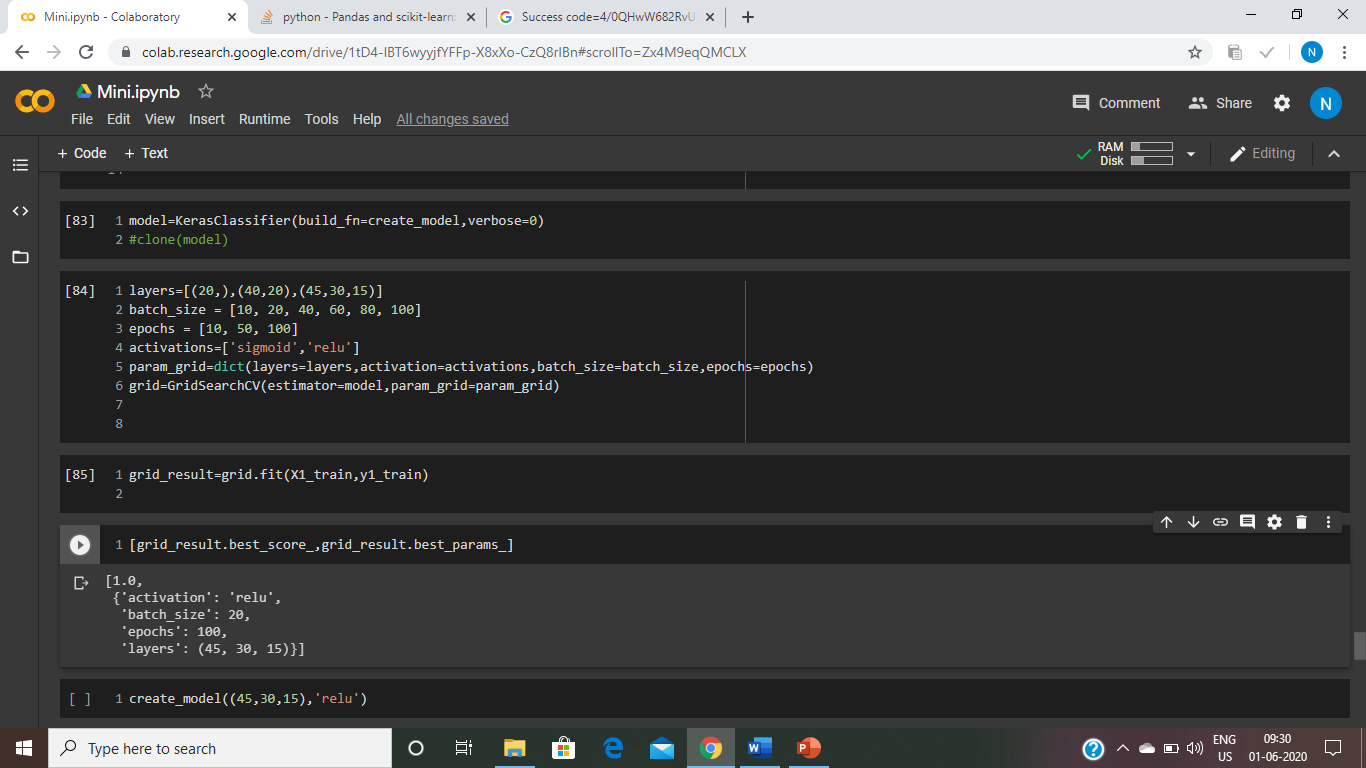


Figure 16

Total no of hidden layers:3

Batch size=20

Epochs=100

Activation function to be used at the hidden layer:relu

1. Building the neural network

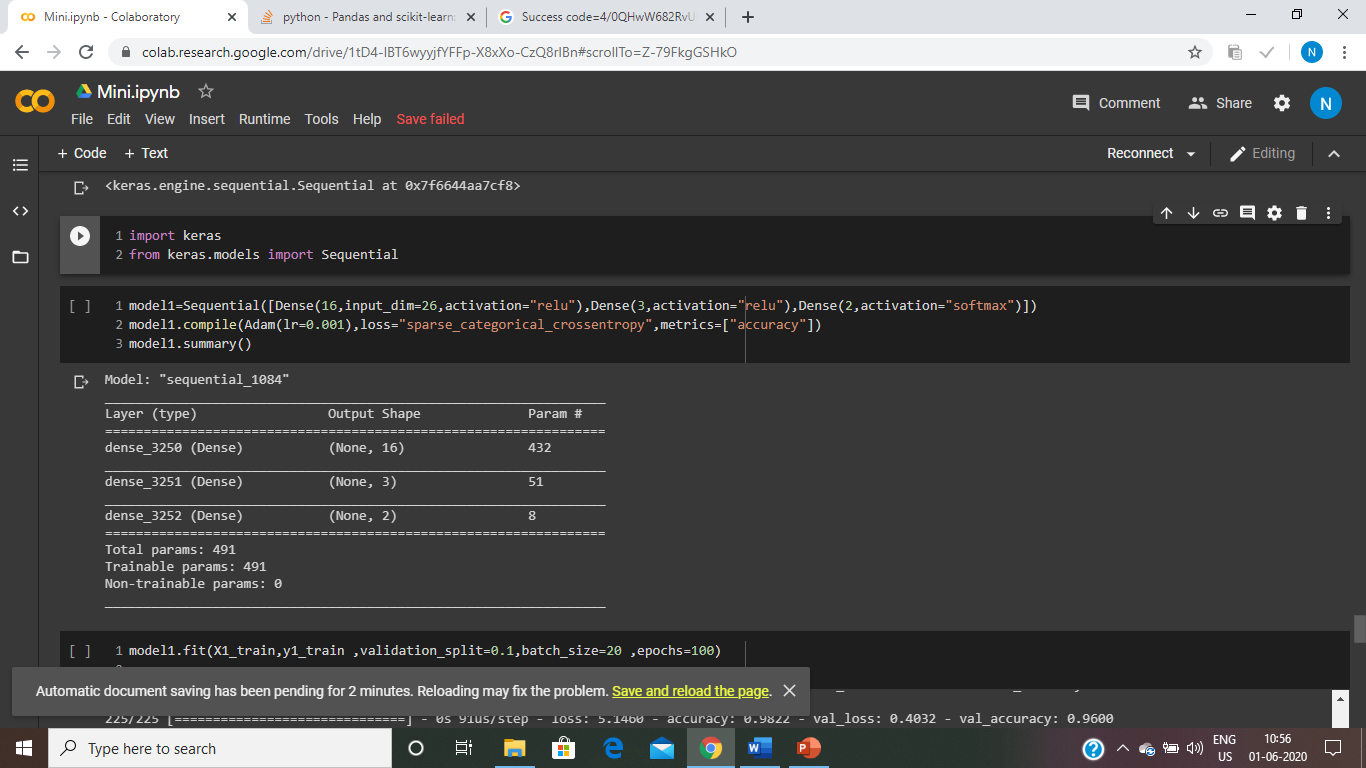


Figure 17

Learning rate taken=0.001

Optimiser used=adam

Loss=sparse categorical crossentropy

1. Training the neural network

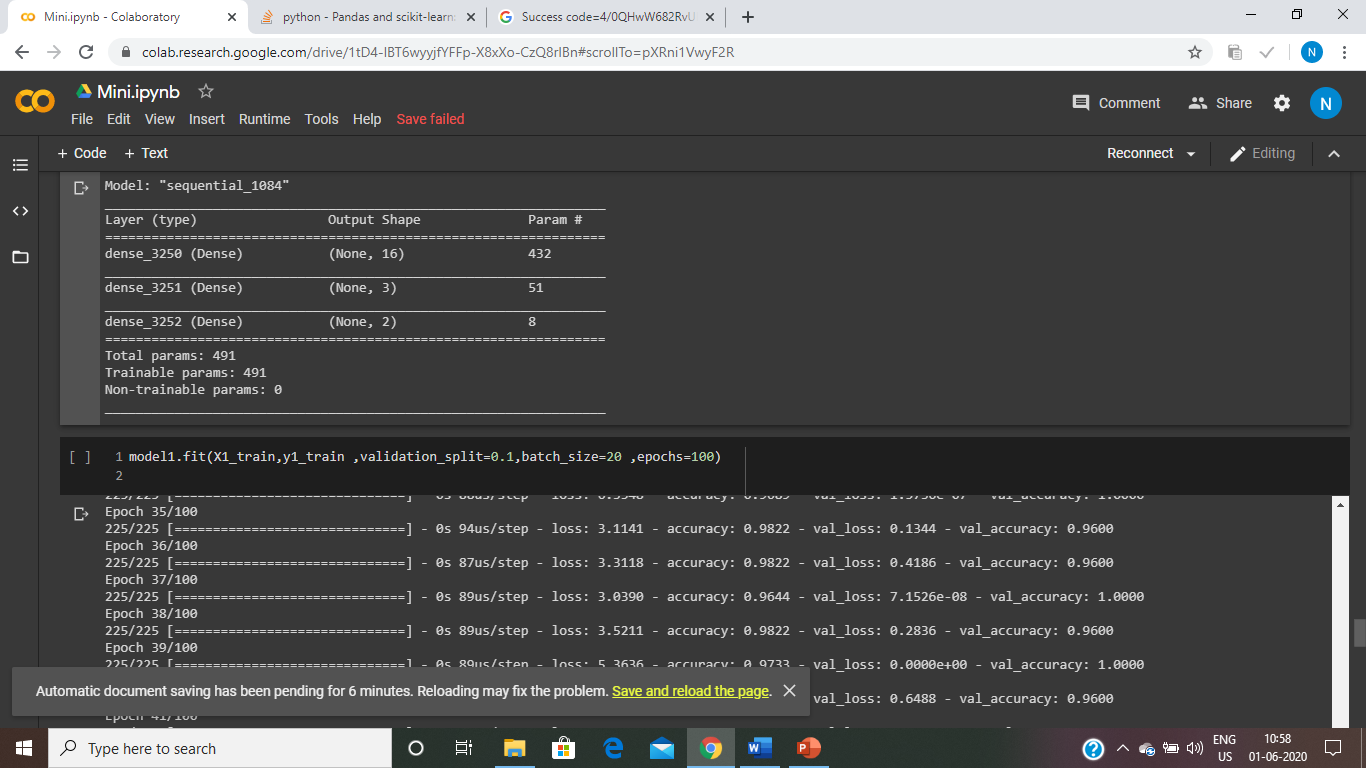


Figure 18

90% of the training set is used for training and 10% for validation

EXPERIMENTAL RESULTS

|  |  |  |  |
| --- | --- | --- | --- |
| No of training sample | No of validation  sample | Training accuracy | Validation accuracy |
| 450 | 50 | 0.9689 | 0.9600 |
| 900 | 100 | 0.9667 | 0.9500 |
| 2250 | 250 | 0.8293 | 0.8800 |
| 4500 | 500 | 0.8338 | 0.9460 |
| 22500 | 2500 | 0.8497 | 0.8232 |
| 45000 | 5000 | 0.8392 | 0.8700 |

Table 1

TESTING AND VALIDATION

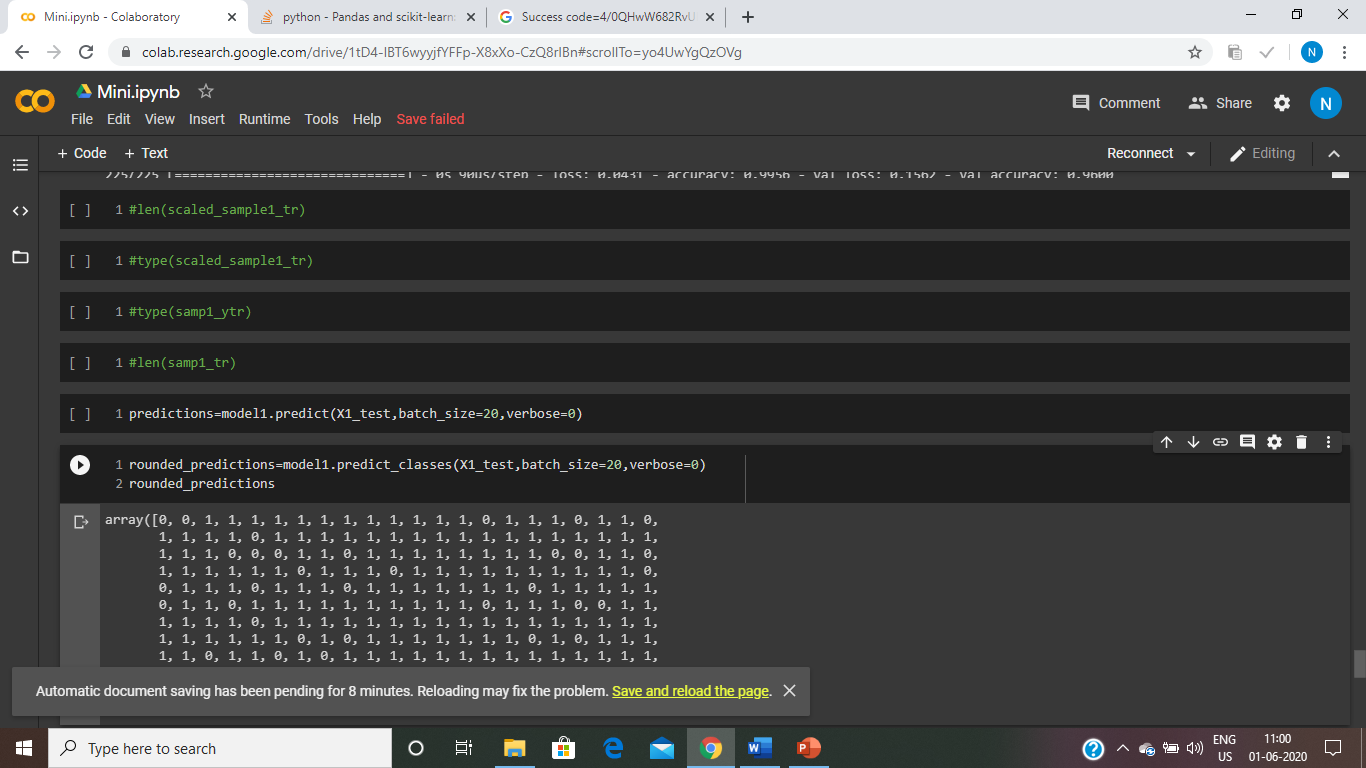


Figure 19

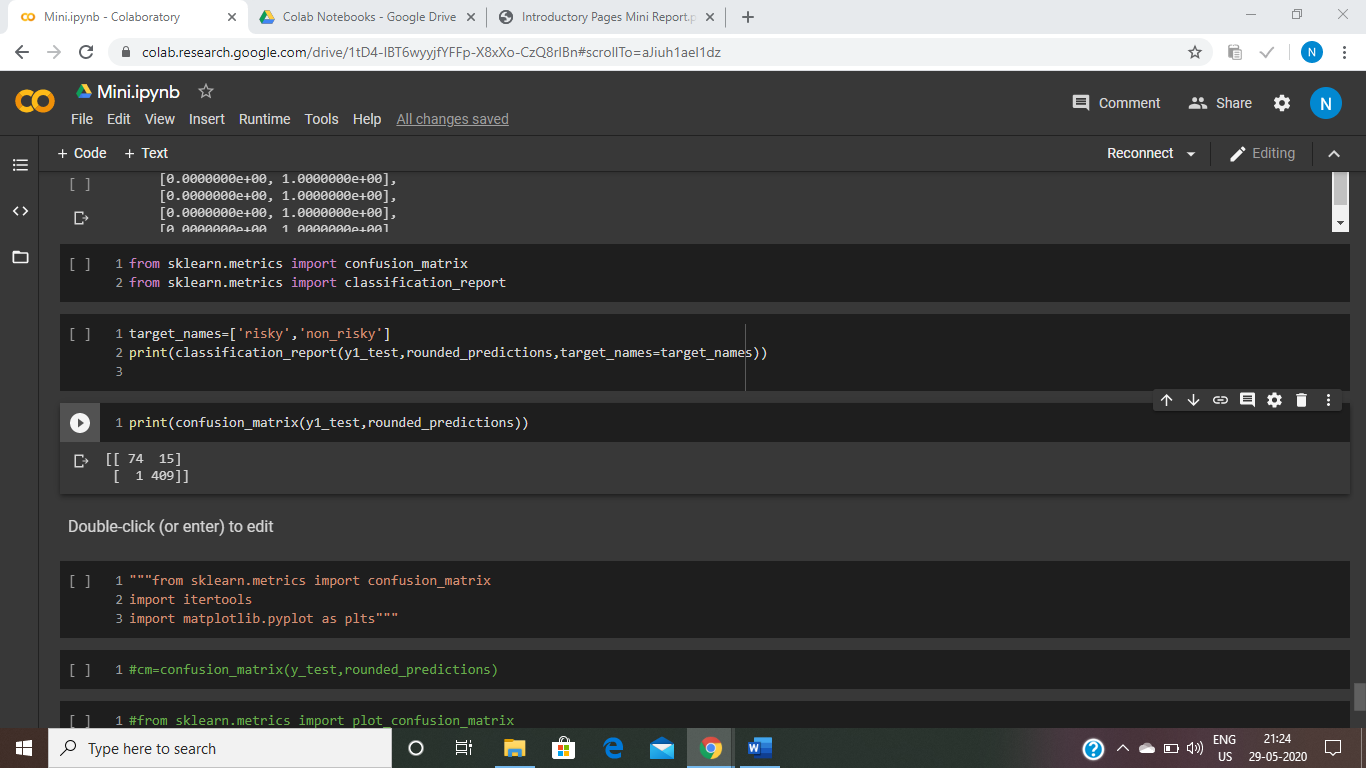


Figure 20

To check the accuracy of the prediction, confusion matrix is used

Testing and validation results

|  |  |  |
| --- | --- | --- |
| No of training samples | Precision | Recall |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Risky | Non Risky | Risky | Non risky |
| 450 | 0.96 | 0.97 | 0.92 | 1.00 |
| 400 | 1.00 | 0.95 | 0.76 | 1.00 |
| 2250 | 1.00 | 0.83 | 0.01 | 1,00 |
| 4500 | 0.88 | 0.85 | 0.01 | 1.00 |
| 22500 | 0.09 | 0.81 | 0.00 | 1.00 |
| 45000 | 0.00 | 0.85 | 1.00 | 0.92 |

Table 2

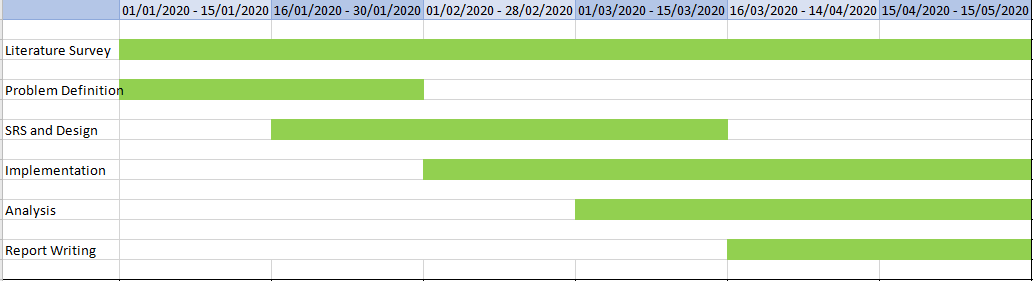
CONCLUSION

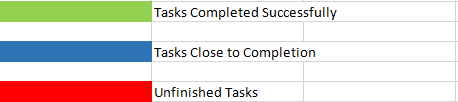
* Credit risk analysis system is an critical financial decision making system.
* In our study of credit risk, we take different no of data samples and feed it to our artificial neural network and train the network and observe its prediction.
* As, we feed more data sample in the neural network, the model is able to generalize better.
* From the testing and validation result, we observe that the bank dataset we used, gives better prediction for non-risky customers.(refer table 2)

LIMITATIONS AND FUTURE SCOPE

* We can integrate an user interface to make the credit risk analysis system more user friendly for the bank
* The data and features of a bank customer that determines their debt repayment capacity may change with time. Hence, the dataset used might go outdated.

GANTT CHART





REFERENCES

1. Credit Risk Prediction Using Artificial Neural Network Algorithm.

Author Name: Shruti Goyal

Posted at : datasciencecentral.com

Posted on: March 14,2018 at 11:30 am

Accessed at:17.01.20, 10:45 pm

<https://www.datasciencecentral.com/profiles/blogs/credit-risk-prediction-using-artificial-neural-network-algorithm>

2. Measuring Credit Risk of Bank Customers Using Artificial Neural Network

Authors:Mohsen Nazari, Mojtaba Alidadi

Published at: Journal of Management Research ISSN 1941-899X

Published on: April 1, 2013

Accessed at: 17.0120, 7:32 pm

<http://www.macrothink.org/journal/index.php/jmr/article/view/2899/2621>

3. Evaluating Credit Risk Using Artificial Neural Networks

Author: Qeethara K. Al-Shayea and Ghaleb A.El-Refae

Published at:www.getview.org

Published on:September 2011

Accessed at:18.01.2020, 6:30pm

<https://www.researchgate.net/publication/298563719_Evaluating_Credit_Risk_Using_Artificial_Neural_Networks>

4. :Credit Risk Prediction: A comparative study between Discriminant Analysis and the neural network approach.

Author: Ihem Khemakhem ,and Younés Boujelbènea

Published at:Accounting and Management Information Systems Vol. 14, No. 1, pp. 60-78

Published on:2015

Accessed at:18.01.2020, 9:30 pm

<https://ideas.repec.org/a/ami/journl/v14y2015i1p60-78.html>