# VISVESVARAYA TECHNOLOGICAL UNIVERSITY

**"JnanaSangama", Belgaum -590014, Karnataka.**

**PROJECT WORK-4 REPORT**
**on**

# (Customer Life Time Value) Prediction Using Machine Learning and Deep Learning

***Submitted by***


**SHASHANK KUMAR (1BM18CS098)**
**SHREYASH ADHIKARI(1BM18CS106)**
**SIDDHARTHA GHOSH (1BM18CS108)**
**MOHAMMAD ARAFAT KHAN (1BM18CS136)**


***Under the Guidance of***
**Prof. Vikrant B.M**
**Assistant Professor, BMSCE**

***in partial fulfillment for the award of the degree of***
**BACHELOR OF ENGINEERING**
***in***
**COMPUTER SCIENCE AND ENGINEERING**


**B. M. S. COLLEGE OF ENGINEERING**
**(Autonomous Institution under VTU)**
**BENGALURU-560019**

**Mar-2021 to Jun-2021**

**B. M. S. College of Engineering,**
**Bull Temple Road, Bangalore 560019**
(Affiliated To Visvesvaraya Technological University, Belgaum)

**Department of Computer Science and Engineering**

## CERTIFICATE

This is to certify that the project work entitled "**PROJECT TITLE**" carried out by **SHASHANK KUMAR (1BM18CS098), SHREYASH ADHIKARI(1BM18CS106), SIDDHARTHA GHOSH (1BM18CS108) AND MOHAMMAD ARAFAT KHAN (1BM18CS136)** who are bonafide students of **B. M. S. College of Engineering.** It is in partial fulfillment for the award of **Bachelor of Engineering in Computer Science and Engineering** of the Visveswaraiah Technological University, Belgaum during the year 2021. The project report has been approved as it satisfies the academic requirements in respect of **Project Work-4 (20CS6PWPW4)** work prescribed for the said degree.

Signature of the Guide
Prof. Vikranth B.M.
Assistant Professor
BMSCE, Bengaluru

Signature of the HOD
Dr. Umadevi V
Associate Prof. & Head, Dept. of CSE
BMSCE, Bengaluru

External Viva

Name of the Examiner — Signature with date

1._____ _____

2. _____ _____

# B. M. S. COLLEGE OF ENGINEERING

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING


***DECALARATION***

We, SHASHANK KUMAR (1BM18CS098), SHREYASH ADHIKARI (1BM18CS106), SIDDHARTHA GHOSH (1BM18CS108), MOHAMMAD ARAFAT KHAN (1BM18CS136) students of 5th Semester, B.E, Department of Computer Science and Engineering, B. M. S. College of Engineering, Bangalore, here by declare that, this Project Work-1entitled "Project Title" has been carried out by us under the guidance of Prof. Vikranth B.M., Assistant Professor, Department of CSE, B. M. S. College of Engineering, Bangalore during the academic semester Mar-2021-Jun-2021

We also declare that to the best of our knowledge and belief, the development reported here is not from part of any other report by any other students.

Signature

SHASHANK KUMAR (1BM18CS098)

SHREYASH ADHIKARI(1BM18CS106)

SIDDHARTHA GHOSH (1BM18CS108)

MOHAMMAD ARAFAT KHAN (1BM18CS136)

# 1. Introduction

The world of business has completely changed and revolves around its customers more than ever. The customer-centric approach is the new norm in today's market. The reason for that is the ample choices people have when choosing a product/service.

In this era of businesses fighting against each other to better serve and seize customers from their competitors, the need for them to grow and retain their existing customer base is very important.

But similar to the process of acquiring customers, there is a huge cost associated with the process of retaining existing customers too. (by giving discounts, targeted offers, etc.)

So, you might think, do they need to retain every single customer? Well, not really. In every business, some customers create more value for the business by being a loyal customer and some are just one-time buyers. Identifying such groups of customers and targeting only the high-value customers will help the business to at least sustain in this competitive market.

Now, the real challenge begins—**How to find the customer value?** Before answering this question, let's just define what does "customer value" means.

***What is Customer Value?***

Customer value or Customer Lifetime Value (CLV) is the total monetary value of transactions/purchases made by a customer with your business over his entire lifetime. Here the lifetime means the time period till your customer purchases with you before moving to your competitors.

- The customer lifetime value calculation for banking
- The key inputs into the customer lifetime value (CLV) banking calculation include:
- Average balances of loans and savings on a per customer basis
- Average interest rate margin (as a percentage)

- Average income/revenue per customer generated from non-interest income sources (e.g. fees, commissions, and other sales)
- Costs of providing customer services and access (which would include transaction costs, statement costs, and potentially a provision for infrastructure costs, and so on)

These inputs are used together to determine average annual profit on a per customer basis. This information is then combined with customer retention rates, other costs of retention and up selling, as well as initial customer acquisition costs – to determine the customer lifetime value (CLV) for the bank.

This is all calculated automatically for you in the Excel template for customer lifetime value for banks – which is available above for free download.

## Basic Formula

- *Gross margin * (Retention rate / [1+ Rate of discount – Retention rate]*

- *Gross Margin: Total revenue minus cost of acquisition and retention*
- *Retention Rate: Ratio of number of retained customers to number at risk*
- *Rate of discount: Interest rate used to calculate the present value of the future cash flow*

***Approach to the problem***

In general, there are two broad approaches to modelling the *CLV* problem*:*

***Historical Approach:***

*Aggregate Model* — calculating the CLV by using the average revenue per customer based on past transactions. This method gives us a single value for the CLV.

*Cohort Model* — grouping the customers into different cohorts based on the transaction date, etc., and calculate the average revenue per cohort. This method gives CLV value for each cohort.

***Predictive Approach:***

*Machine Learning Model* — using regression techniques to fit on past data to predict the CLV.

In this project, we will be using the Machine Learning Model to predict CLV using Regression Classification techniques.

With Predictive CLV, we will:

- Forecast the future value of existing customers with transaction history
- Predict the future value of first-time customers

# 2. Problem Definition and Algorithm

## 2.1 Task Definition

We will focus on the ML models that can be used to predict CLV values. Of customers new and present, by looking at their previous financial details.

Predicting the CLV is a multi-class classification problem and would require access to a labeled dataset. This would make the problem a supervised learning problem, in which each user entry in the training dataset has a label assigned to it.
We have decided to proceed with three different Algorithms, and select the algorithm, which gives the most satisfactory result.

**Machine Learning Models used**

1. Multiclass Classification using logistic regression
2. Multiclass Classification using Random Forest
3. Deep Learning - ANN

## 2.2 Algorithm Definition

Describe in reasonable detail the algorithm you are using to address this problem. A pseudocode description of the algorithm you are using is frequently useful. Trace through a concrete example, showing how your algorithm processes this example. The example should be complex enough to illustrate all of the important aspects of the problem but simple enough to be easily understood.

### *a) Multinomial Logistic Regression*

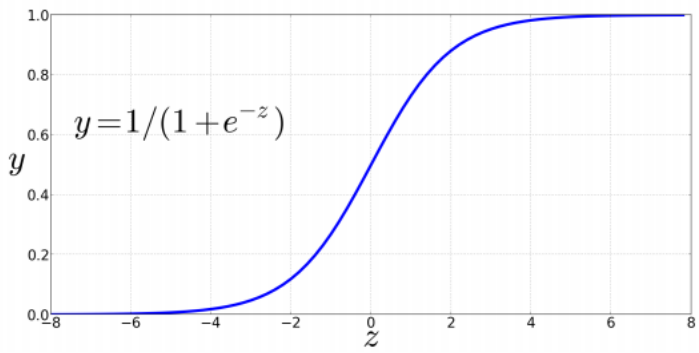
- Logistic regression is a classification algorithm.

- It is intended for datasets that have numerical input variables and a categorical target variable that has two values or classes. Problems of this type are referred to as binary classification problems.

- Logistic regression is designed for two-class problems, modeling the target using a binomial probability distribution function. The class labels are mapped to 1 for the positive class or outcome and 0 for the negative class or outcome. The fit model predicts the probability that an example belongs to class 1.

- By default, logistic regression cannot be used for classification tasks that have more than two class labels, so-called multi-class classification.

***Logistic Regression Model: -***




The mathematical function that essentially gives a binary output or 0 or 1. To do this, we can use the **logistic** or **sigmoid** function, which has the form:

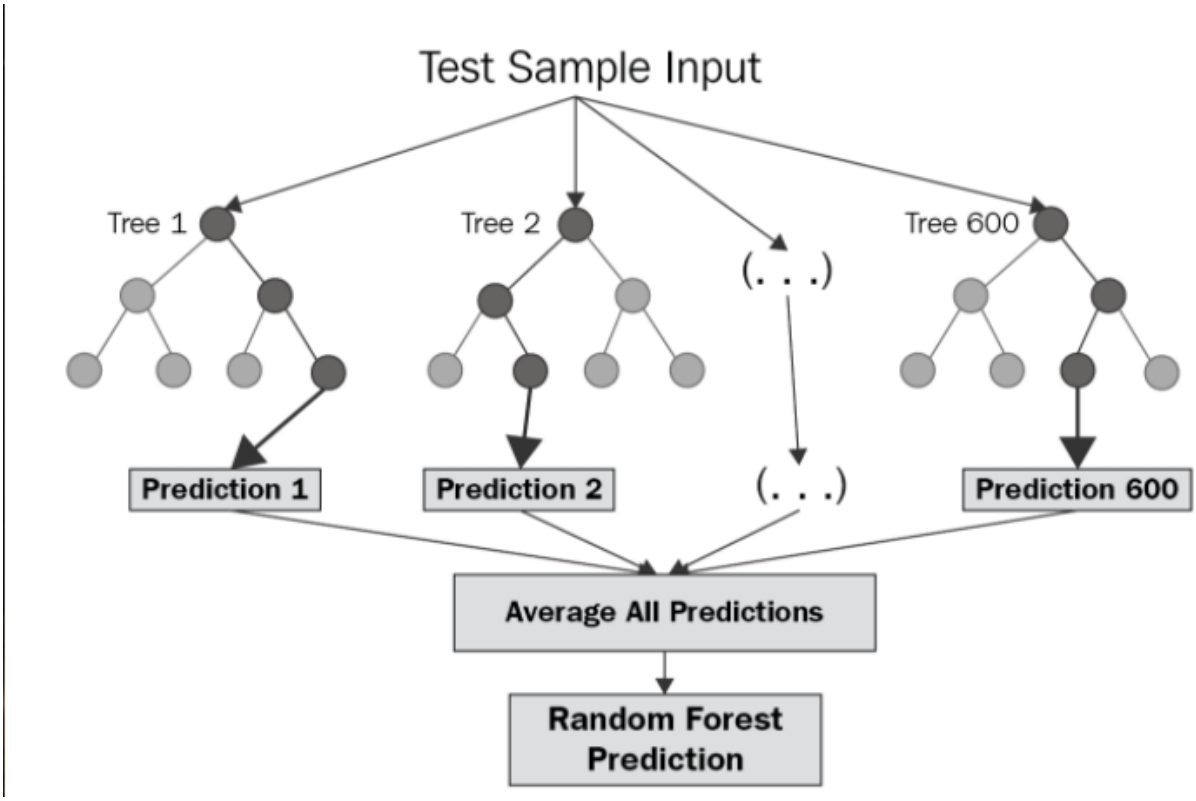
- Since we are having three categorical values Low, Medium and High that we need to predict, hence we have to modify the Logistic Regression to support multi-class classification problems.

- **Multinomial Logistic Regression** is the regression analysis to conduct when the dependent variable is nominal with more than two levels. Similar to multiple linear regression, the multinomial regression is a predictive analysis. Multinomial regression is used to explain the relationship between one nominal dependent variable and one or more independent variables.

- The **Multinomial Logistic Classifier** uses a generalization of the sigmoid, called softmax the softmax function, to compute the probability p(y = c|x). The softmax function takes a vector z = [z1,z2,...,zk] of k arbitrary values and maps them to a probability distribution, with each value in the range (0,1), and all the values summing to 1. Like the sigmoid, it is an exponential function.

## *b) Random Forest Regression*

- Random Forest Regression is a supervised learning algorithm that uses ensemble learning
- *Method for regression*. Ensemble learning method is a technique that combines predictions from
- Multiple machine learning algorithms to make a more accurate prediction than a single model.




A Random Forest operates by constructing several decision trees during training time and outputting the mean of the classes as the prediction of all the trees. To get a better understanding of the Random Forest algorithm, let's walk through the steps
Pick at random k data points from the training set.
Build a decision tree associated to these k data points.
Choose the number N of trees you want to build and repeat steps 1 and 2.
For a new data point, make each one of your N-tree trees predict the value of y for the data point in question and assign the new data point to the average across all of the predicted y values.

A Random Forest Regression model is powerful and accurate. It usually performs great on many problems, including features with non-linear relationships. Disadvantages, however, include the following: there is no interpretability, overfitting may easily occur, we must choose the number of trees to include in the model.

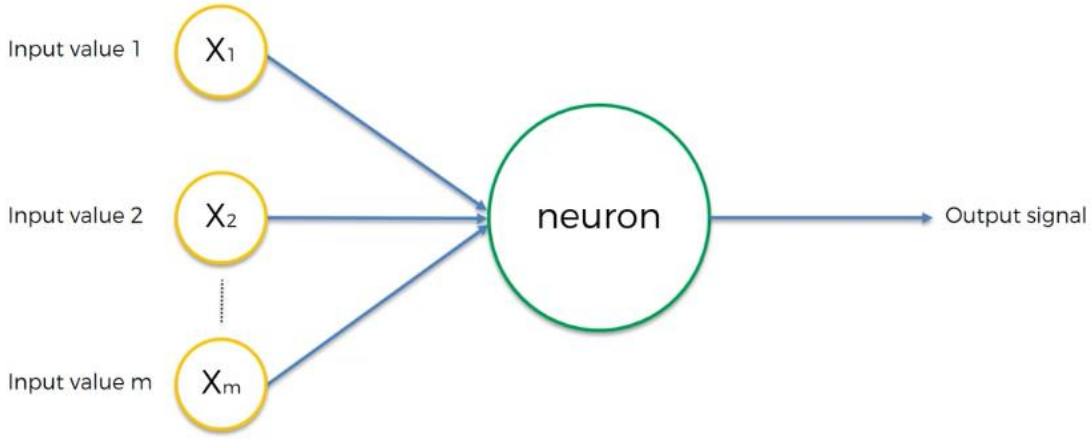
## *c) ANN Modelling using Deep Learning*

Deep learning is an artificial intelligence (AI) function that imitates the workings of the human brain in processing data and creating patterns for use in decision making.

Deep learning is a subset of machine learning in artificial intelligence that has networks capable of learning unsupervised from data that is unstructured or unlabeled. Also known as deep neural learning or deep neural network.

As a subset of machine learning, deep learning uses hierarchical neural networks to analyze data. Neuron codes are linked together within these hierarchical neural networks, similar to the human brain. Unlike other traditional linear programs in machines, the hierarchical structure of deep learning allows it to take a nonlinear approach, processing data across a series of layers which each will integrate subsequent tiers of additional information.

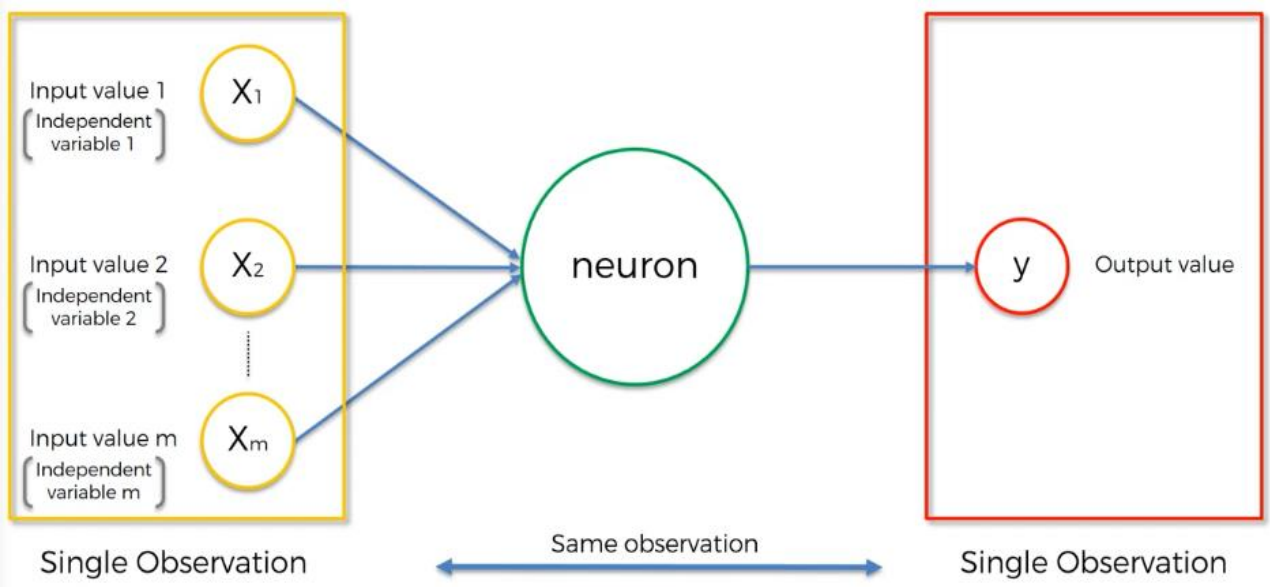
Neurons are building blocks of deep learning



The inputs here are independent variables, and all these are present for one single observation which we are training our model on. Also, these variables need to be either standardized (making sure they have a mean of zero and a variance one) or normalized making them to fit in about a range of values.

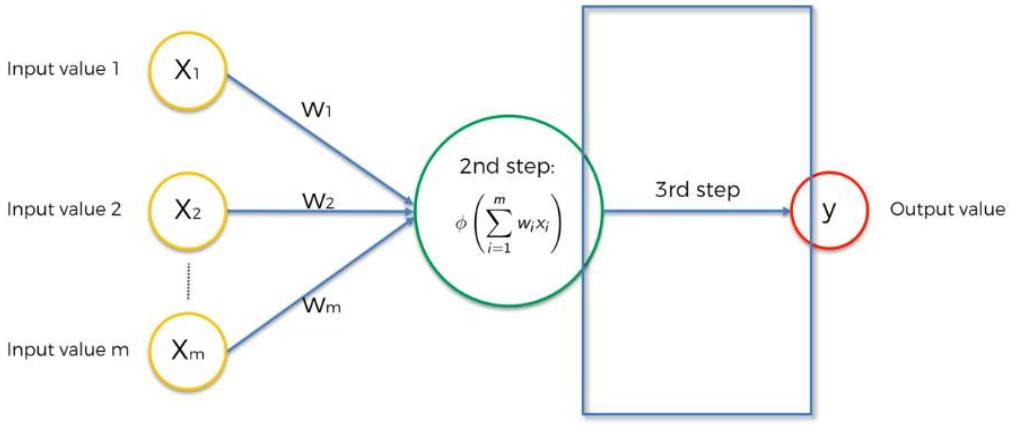
This is done as all these values are added up or multiplied in a neural network, hence it will be easier to process then if they're all about the same.

Here, the output values can be continuous, it can be binary or it can be categorical variable (several output values)



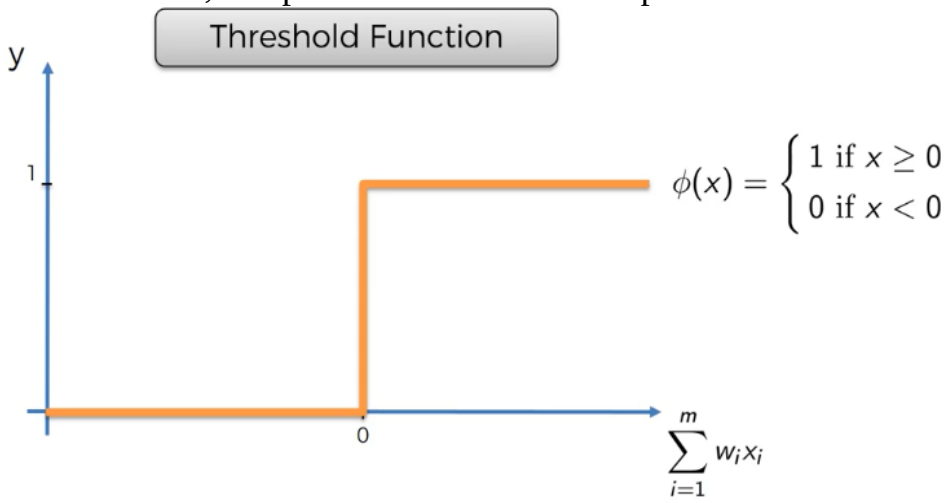

Weights are crucial in an artificial neural network and its functioning because weights are how neural networks learn - by adjusting the weights the NN decides in every single case what signal is poor and what is not important, what signal gets passed along and what signal doesn't get passed along. Hence, weights are the things that get adjusted through the process of learning, when you're training a NN it's basically adjusting the weights across the whole NN and that's where gradient descent and backpropagation comes into play.

The neuron takes all the input values and they are multiplied by the weight and added up, and then it applies an activation function after which it is decided if a signal is passed on or not.

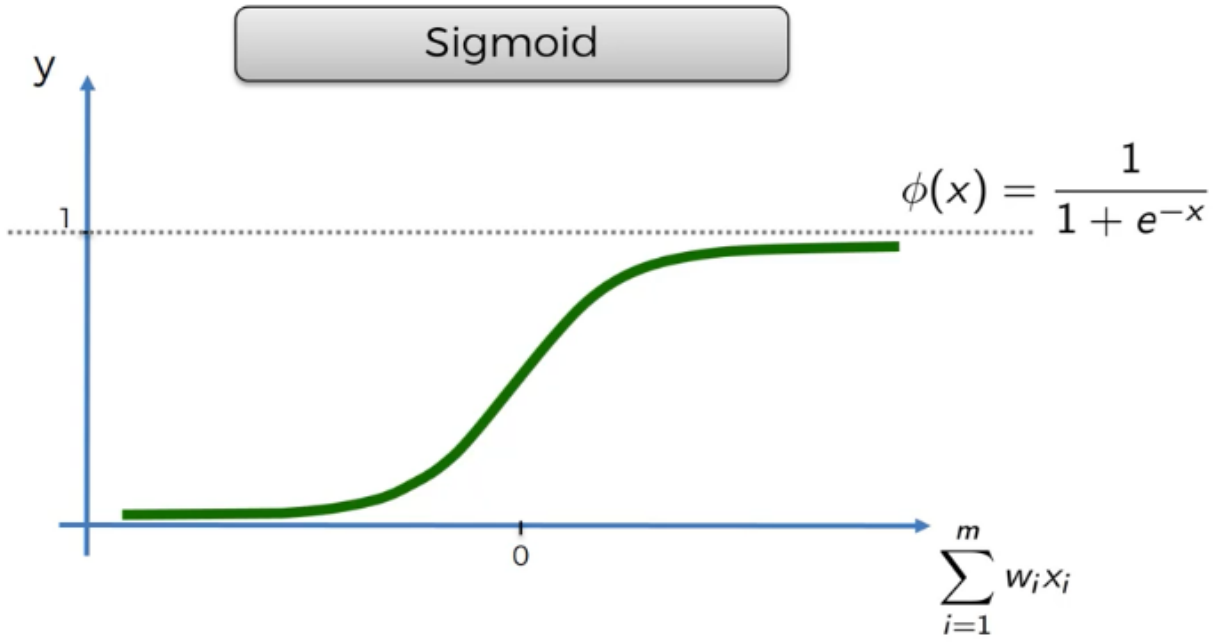



Types of activation function we can choose from:

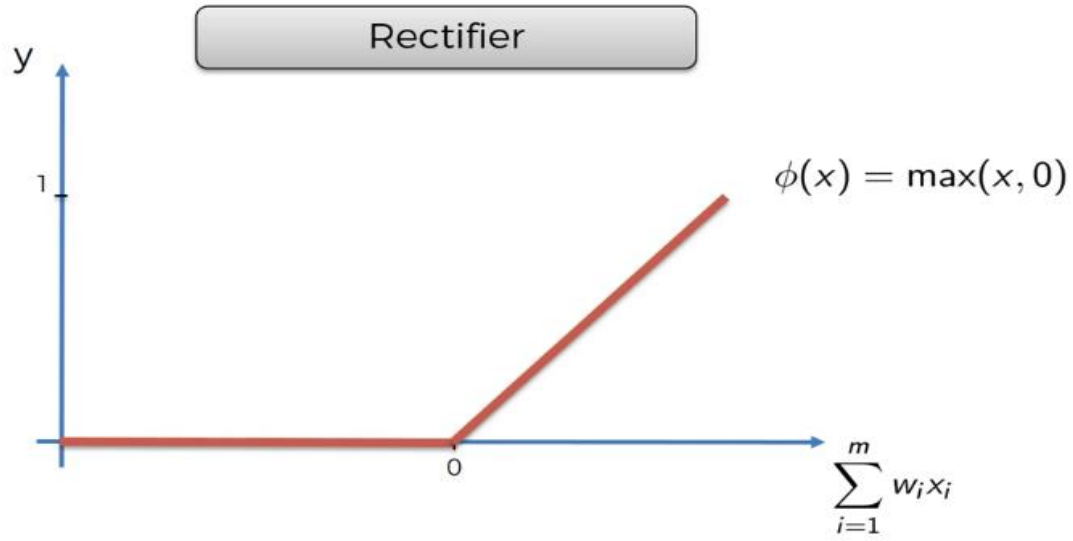
- ***Threshold function:***
  If value is less than zero, 0 is passed on otherwise 1 is passed




- ***Sigmoid Function:***
  Smooth and gradual progression, anything below zero drops off, above zero it approximates to one. It is very useful in the final layer (probability prediction)




- ***Rectifier:***
  As input value increases it gradually progresses. One of the most used function in ANN



# 3. Experimental Evaluation

Since we have already defined what our problem is, and what various algorithms we would be using to solve it, hence we need to define and acquire a dataset, which will have all the necessary attributes that we require.

***Acquiring the dataset***

As discussed the dataset would be the previous or current financial transcripts of each individual customer.

The key inputs into the customer lifetime value (CLV) banking calculation include:

- Average balances, loans and savings on a per customer basis
- Average interest rate margin (as a percentage)
- Average income/revenue per customer generated from non-interest income sources (e.g. fees, commissions, and other sales)
- Costs of providing customer services and access (which would include transaction costs, statement costs, and potentially a provision for infrastructure costs, and so on)

These are few of the inputs used together to determine average annual profit on a per customer basis.

This information is then combined with customer retention rates, other costs of retention and up selling, as well as initial customer acquisition costs – to determine the customer lifetime value (CLV) for the bank.

There are many online data sources where you can get free data sets to use in your project, hence we explored many websites to look for the perfect dataset, we would be needing some of these websites include:-

1. Kaggle
2. GitHub
3. DataCamp
4. Mockaroo*

However, unfortunately, we could not find the dataset of our choosing; hence, we figured out another way, instead of searching for an already existing dataset, we decided to create one. We used a website called Mockaroo to create our own dataset of 1000 customer records.

***Dataset Snapshot: -***

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Customer_no | Age | Bank_Balance | :age)Interest Rate Marg | Loan_Interest | Fees_Service | Total_Earning | tain_Amo | rvice_Spe | scount_Ra | Discount_Amount | Totl_Spend | Gross_Margin | ntention_ra | CLV | CLV_TYPE |
| 2 | 1 | 41 | 15,139.00 | 1.81 | $4,547.73 | $5,000.00 | $24,689.46 | $352.08 | $188.51 | 0.1 | $507.36 | $1,048.05 | $23,641.41 | 0.478 | 18,168.16 | L |
| 3 | 2 | 41 | 54,341.00 | 1.61 | $4,770.36 | $1,864.29 | $60,977.28 | $217.14 | $235.47 | 0.1 | $1,005.89 | $1,458.60 | $59,518.68 | 0.478 | 45,739.44 | M |
| 4 | 3 | 21 | 40,554.80 | 2.65 | $615.75 | $3,670.18 | $44,843.38 | $198.23 | $116.00 | 0.1 | $987.73 | $1,302.06 | $43,541.32 | 0.478 | 33,461.01 | M |
| 5 | 4 | 37 | 59,998.73 | 2.37 | $1,549.40 | $3,717.14 | $65,267.64 | $111.43 | $169.16 | 0.1 | $1,643.14 | $1,923.83 | $63,343.81 | 0.478 | 48,679.01 | M |
| 6 | 5 | 18 | 75,661.51 | 2.21 | $1,732.38 | $3,344.89 | $80,740.99 | $467.49 | $132.71 | 0.1 | $1,434.05 | $2,034.35 | $78,706.64 | 0.478 | 60,485.17 | M |
| 7 | 6 | 24 | 53,909.55 | 2.01 | $516.28 | $2,405.53 | $56,833.37 | $451.25 | $123.15 | 0.1 | $636.52 | $1,211.02 | $55,622.35 | 0.478 | 42,745.15 | M |
| 8 | 7 | 18 | 92,517.52 | 2.63 | $2,656.12 | $3,741.31 | $98,917.58 | $253.76 | $236.89 | 0.1 | $1,496.61 | $1,987.36 | $96,930.22 | 0.478 | 74,489.78 | M |
| 9 | 8 | 44 | 68,590.02 | 1.74 | $1,280.87 | $4,725.12 | $74,597.75 | $134.64 | $269.96 | 0.1 | $700.05 | $1,104.75 | $73,493.00 | 0.478 | 56,478.54 | M |
| 10 | 9 | 49 | 57,697.85 | 1.58 | $3,314.41 | $2,303.92 | $63,317.76 | $216.81 | $227.07 | 0.1 | $1,471.26 | $1,915.24 | $61,402.52 | 0.478 | 47,187.15 | M |
| 11 | 10 | 22 | 103,037.39 | 2.61 | $2,970.93 | $3,812.13 | $109,823.06 | $320.34 | $156.87 | 0.1 | $1,290.76 | $1,768.07 | $108,054.99 | 0.478 | 83,039.04 | H |
| 12 | 11 | 25 | 36,961.65 | 1.64 | $4,655.09 | $4,451.66 | $46,070.04 | $466.84 | $163.69 | 0.1 | $1,404.33 | $2,034.96 | $44,035.08 | 0.478 | 33,840.46 | M |
| 13 | 12 | 35 | 177,628.08 | 2.58 | $1,188.47 | $4,216.58 | $183,035.71 | $232.09 | $180.89 | 0.1 | $857.51 | $1,270.59 | $181,765.12 | 0.478 | 139,684.45 | H |
| 14 | 13 | 26 | 75,766.40 | 2.11 | $4,075.95 | $993.82 | $80,838.28 | $306.40 | $204.79 | 0.1 | $1,552.62 | $2,063.91 | $78,774.37 | 0.478 | 60,537.22 | M |
| 15 | 14 | 33 | 42,114.47 | 1.65 | $1,562.87 | $4,296.14 | $47,975.13 | $483.75 | $130.73 | 0.1 | $758.16 | $1,372.74 | $46,602.39 | 0.478 | 35,813.41 | M |
| 16 | 15 | 31 | 6,188.95 | 2.65 | $2,825.84 | $3,170.60 | $12,188.04 | $193.77 | $281.32 | 0.1 | $1,519.04 | $1,994.23 | $10,193.81 | 0.478 | 7,833.83 | L |
| 17 | 16 | 49 | 57,313.21 | 2.32 | $3,093.62 | $2,733.58 | $63,142.73 | $197.92 | $163.64 | 0.1 | $555.03 | $916.69 | $62,226.04 | 0.478 | 47,820.01 | M |

***Problem Type and Approach to solving it***

As discussed previously, it is a Multiclass Classification type.
We have records of 1000 customers, which includes all their financial data, from Bank Balance, Age, Interest Rate, Loans, Retention Rate etc:-
In addition to their existing Customer Lifetime Values.

Since it's a classification problem we need to classify new and existing customers into groups having a specific CLV score.

- All the customers whose, CLV values are in-group Low CLV are denoted by 0.
- All the customers whose, CLV values are in-group Medium CLV are denoted by 1.
- All the customers whose, CLV values are in-group High CLV are denoted by 2.

As a Banking firm, we need to target those customers that lie in the Group of Individual having **High CLV Values.**


***Defining the Dependent and the Independent Variables:***

Independent variables are variables that are manipulated or are changed by researchers and whose effects are measured and compared. The other name for independent variables is Predictor(s). The independent variables are called as such because independent variables predict or forecast the values of the dependent variable in the model.

The other variable(s) are also considered the dependent variable(s). The dependent variables refer to that type of variable that measures the affect of the independent variable(s) on the test units. We can also say that the dependent variables are the types of variables that are completely dependent on the independent variable(s). The other name for the dependent variable is the Predicted variable(s).

Usually when one is looking for a relationship between two things, one is trying to find out what makes the dependent variable change the way it does.
**The List of Dependent Variables are:**

1. ***Age*:** The Age of the customer
2. ***Bank Balance***: The sum total amount of all the account under the customer's name (Average)Interest Rate Margin (%) : Net interest margin (NIM) reveals the amount of money that a bank is earning in interest on loans
3. ***Loan Interest***: Loan amount to be paid as interest

4. ***Fees Service:*** All the charges that the customer pays to the bank, for the services offered to it by the bank.
5. ***Total Earning***: The sum total of the amount the customer contributes to the bank, which includes interests' payments, loan acquirements, and bank balance.
6. ***Retain Amount***: Banks can retain customer loyalty by working on the customer-brand relationship that goes beyond the minimum transactional services. This can be done in the form of better loan schemes, or gift vouchers, anything that keeps the customer happy and prevents them from defecting to another bank.
7. ***Service Spent***: The amount that the bank has to spend on each customer for the special services it provides to them.
8. ***Discount Rate***: The discount rate offered by the bank on the bank balance.
9. ***Discount Amount:*** The Discount amount the bank pays as interest to the customer.
10. ***Total Spend***: The Sum Total of all the amount the bank spends on the customer.
11. ***Gross Margin***: The total Profit the bank gets from each customer, after subtracting the Total Spend from Total Earning.
12. ***Retention rate***: Is often calculated on an annual basis, dividing the number of employees with one year or more of service by the number of staff in those positions one year ago. Positions added during the year would not be counted.

***Independent Variables***

The Independent variable is the **CLV_Type** value which is either **High**, **Medium** or **Low** that we will be predicting using the formula first then using machine learning models.

# 4. Conclusion

After running up our algorithm we, tested their accuracies and here are the results:

***Results (Model Accuracy) :-***

Logistical regression: 87%; Random forest: 99.3%; ANN Modelling: 97.3%

*Random forest Model* is Overfitting. Overfitting happens when a model learns the detail and noise in the training data to the extent that it negatively affects the performance of the model on new data. Hence we will not consider it.

*Logistic Regression* is neither Overfitting nor Underfitting hence gives us a decent accuracy, therefore we will consider it.

*ANN Modelling* also gives us a good prediction accuracy result therefore we will consider it as well.

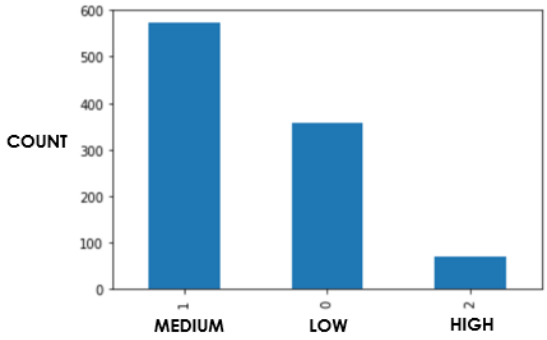
***1) Entering the values of a new customer***:

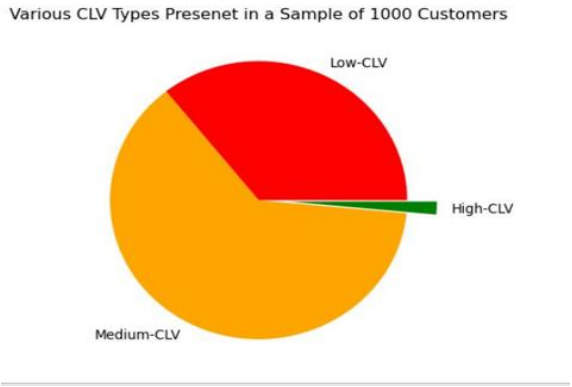
In [19]:
```
# Here we predict the value CLV from the inputs of a new Customer

a=float(input("Enter The Customer Bank Balance:"))
b=float(input("Enter The (Average)Interest Rate Margin(%) :"))
c=float(input("Loan_Interest :"))
d=float(input("Fees_Service :"))
e=a+b+c+d
f=float(input("Retain_Amount :"))
g=float(input("Service_Spent :"))
h=float(input("Discount_Rate :"))
i=float(input("Discount_Amount :"))
j=e-f-g-h-i
k=float(input("Retention_rate :"))
print("The Predicted Value of CLV is :",classifier.predict([[a,b,c,d,e,f,g,h,i,j,k]]))
f=m.predict([[a,b,c,d,e,f,g,h,i,j,k]])
if f==0:
  print("The CLV is Low")
elif f==1:
  print("The CLV is Medium")
else:
  print("The CLV is High")
```

```
Enter The Customer Bank Balance:98000
Enter The (Average)Interest Rate Margin(%) :1.81
Loan_Interest :4700
Fees_Service :24378
Retain_Amount :446
Service_Spent :119
Discount_Rate :0.1
Discount_Amount :555
Retention_rate :0.478
The Predicted Value of CLV is : [2]
The CLV is High
```

## *2) Count and Depiction of Various CLV in the Dataset*

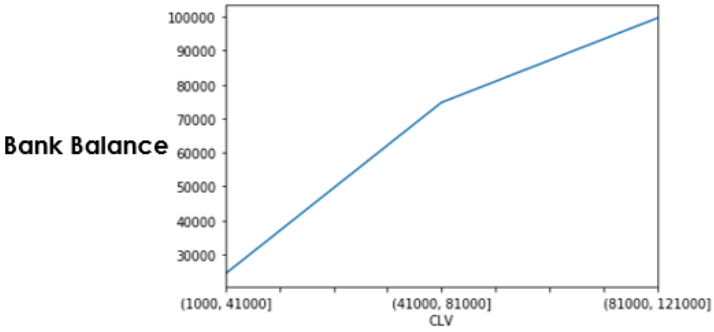
```
Out[55]: <matplotlib.axes._subplots.AxesSubplot at 0xeb1841ce4
```




## *3) Plot between Bank Balance and CLV*

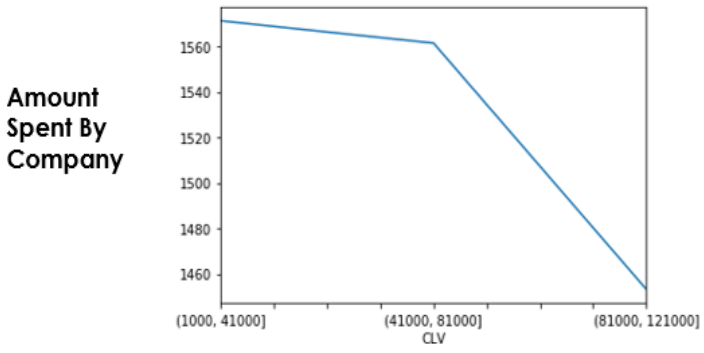
```
Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0xeb1f620088>
```




Customers with **Higher Bank Balance** tend to have a **Higher CLV** value.

## *4) Plot between CLV and Spent-Amount*

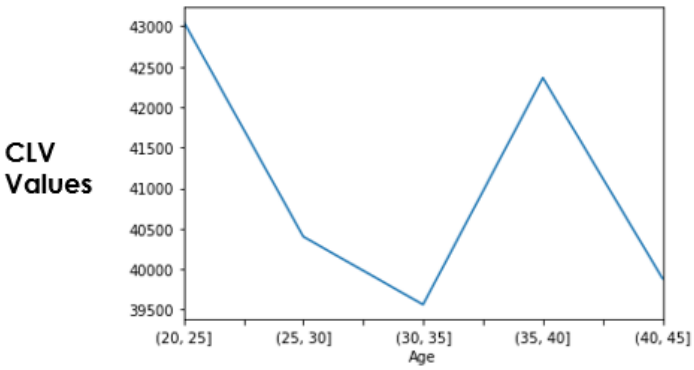
```
Out[76]: <matplotlib.axes._subplots.AxesSubplot at 0xeb20740608>
```




**High Valued Customers give more and take less**

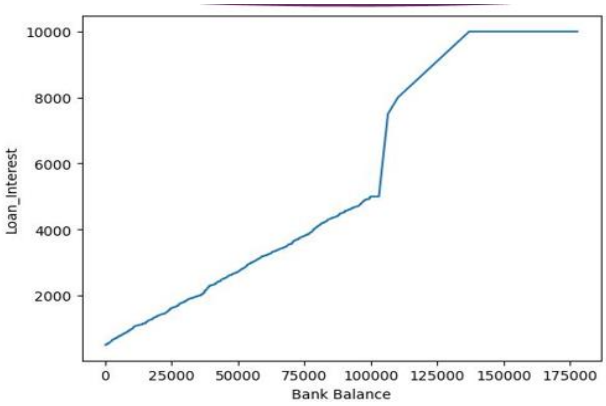
### *5) Plot between Age and CLV*

```
Out[79]: <matplotlib.axes._subplots.AxesSubplot at 0xeb20885d48>
```



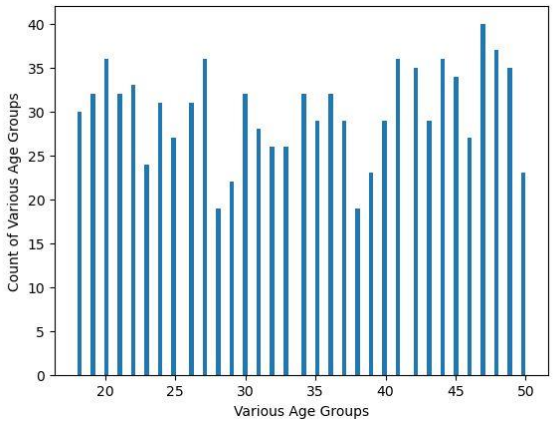

**People Within the age group of (30, 35) have a much lower mean salary then those Younger and Older**

### *6) Plot between Bank Balance and Loan Interest*




**People with Higher Bank Balance tend to pay more Interest on Loans**

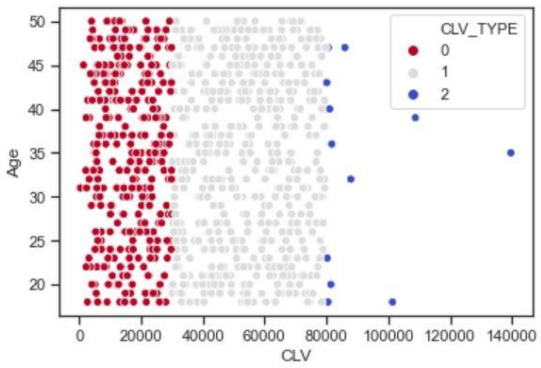
### *7) Histogram Plot of Count of Various Age Groups*




**Youngest Customer**: 18 year old
**Oldest Customer**: 50 year old
**Age Groups with the highest presence**: b/w 45-50
**Age Groups with the lowest presence**: b/w 18-20

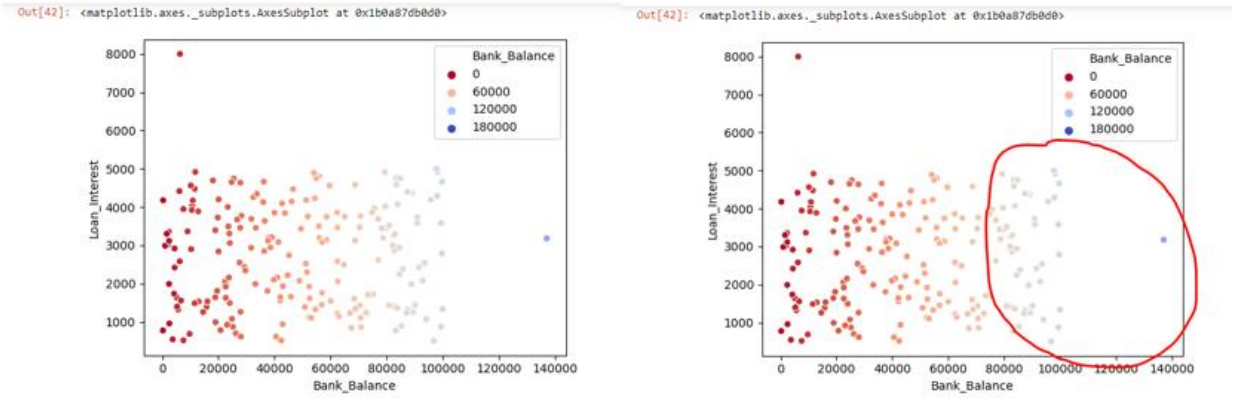
## *8) Scatter Plot of Various Age groups and their CLV's*

```
Out[75]: <matplotlib.axes._subplots.AxesSubplot at 0x205df150d00>
```




**We have analyzed the various age groups and their respective CLV, and have agreed that the Value of CLV is evenly distributed across various age groups**

## *9) Scatter Plot of Various Bank Balance and their Loan Interest Payment*




- We need to target the Customers that are profitable to the bank, and invest more into it i.e. in the form of a Higher Bank Balance and Paying more Loan Interest to the bank
- The points in the Scatter Plot surrounded by the Red Circle, are customers of HIGH Value to the Bank, and must be retained at all costs.

# 5. Related Work

Considerable research in CLV has concentrated mainly in the realm of direct marketing, primarily because direct marketing provided access to information about individual customers that enabled investigation of relevant issues. The information revolution is changing all this. As we move into the Internet era, presence of a company on the Web is no longer a luxury or competitive advantage; rather it is a must for survival. Better information on individual customers is available today because of this development. This new source of information has renewed interest in customer lifetime value among marketing researchers.

Reinartz and Kumar (2000) explained that the interest in customer lifetime value research has increased recently, mainly for three reasons. First, firms are now more and more interested in customer management processes for which an understanding of customer lifetime value concept is a prerequisite. Second, the Marketing Science Institute has elevated the topic to a capital research priority. Third, empirical evidence is particularly scarce in this CLV domain.

CLV models help quantify the relationship of the firm with its customers and subsequently allow the firm to make more informed decisions in a structured framework. CLV models also help a firm to know who its profitable customers are, and customer profitability provides a metric for the allocation of marketing resources to consumers and market segments. Marketing efforts are best directed at the most profitable consumers (Mulhern, 1999).

The era of mass marketing is being replaced by an era of targeted marketing. Knowledge of CLV enables firms to develop customer-specific marketing programs leading to an increase in efficiency and effectiveness of such programs. The Internet is undoubtedly a major instrument of such targeted marketing; the direct marketing concepts of CLV can be extended to be useful in interactive scenarios.

# 6. Future Work

1. ***Adding a User Interface:***

One significant challenge for data scientists, data analysts, and machine learning engineers is to showcase and demo their models to non-technical personnel.

Some libraries can handle this job for you and let you focus more on model development tasks. **Gradio** is a great example; it's a Python library that generates an easy-to-use UI for every machine learning model.

2. ***Implementing it on Business other than Banking***

Predicting customer lifetime value is very important for any form of business big or small, hence in this project we have only taken parameters that touch banking we can improve the model to accommodate predictions for other types of businesses.

3. ***Predict customer churn in a bank using Neural Designer***

Along with figuring out which customers to satisfy using Customer Lifetime Value, we also do some modifications in our existing dataset, and figure out what customers are likely to leave i.e. Estimate the Churn Rate of each customer, and try to figure out the policies that resulted in the customer leaving.

## Bilbiography


1. Machine Learning A to Z course on Udemy: https://www.udemy.com/courses/search/?src=ukw&q=Machine+learning+A+to+Z
2. Udemy Python Online Bootcamp: https://www.udemy.com/course/complete-python-bootcamp/learn/lecture/20568974?start=0
3. Ekinci, Yeliz & Uray, Nimet & Ulengin, Füsun. (2014). A customer lifetime value model for the banking industry: a guide to marketing actions. European Journal of Marketing.
4. Mental Health in Tech (2016): Exploratory Analysis using Data Visualization: https://www.kaggle.com/jchen2186/data-visualization-with-python-seaborn
5. Ambler, T. (2002). Comment: Customer lifetime values – credible, or utterly incredible? Journal of Marketing, 10, 201-202. http://doi.org/hb8Berger, P. D., & Nasr, N. I. (1998).
6. Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing, 12, 17-30. http://doi.org/hb9Blattberg, R. C., & Deighton, J. (1996).
7. A Definitive Guide for predicting Customer Lifetime Value (CLV): https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/
8. Manage marketing by the Customer Equity Test. Harvard Business Review, 74, 136-144.Blattberg, R. C., Getz, G., & Thomas, J. S. (2001).
9. Customer equity: Building and managing relationships as valuable assets. Boston: Harvard Business School Press.Cooper, R., & Kaplan, R. S. (1991).
10. Profit priorities from activity-based costing. Harvard Business Review, 69, 130-135.Doyle, P. (2000).
11. Value-based marketing. Journal of Strategic Marketing, 8, 299-311.http://doi.org/hcbDwyer, R. F. (1997).
12. Customer lifetime valuation to support marketing decision making. Journal of Direct Marketing, 11, 6-13.Fornell, C. (2000).
13. Customer asset management, capital efficiency, and shareholder value. Keynote Speech, July 20, 2000, Performance measurement, past, present and future. Cambridge University, England.Gupta, S., & Lehmann, D. R. (2003).
14. Customer as assets. Journal of Interactive Marketing, 17, 9-24. http://doi.org/hccGupta, S., Lehmann, D. R., & Stuart, J. A. (2004).
15. Valuing customers. Journal of Marketing Research, 41, 7-18. http://doi.org/frkxt9Guracaronu, C., & Ranchhod, A. (2002).
16. Measuring customer satisfaction: A platform for calculating, predicting and increasing customer profitability. Journal of Targeting, Measurement & Analysis for Marketing, 10, 203-219.
17. Logistic Regression: https://www.google.com/search?q=logistic+regression&rlz=1C1GCEA_enKW787KW787&ei=VszJYLXuBNmE4-EPu9uJ0AE&oq=logistic+regression+&gs_lcp=Cgdnd3Mtd2l6EAMYADIHCAAQsQMQQzIHCAAQsQMQQzIHCAAQsQMQQzIFCAAQsQMyBwgAELEDEEMyBAgAEE

MyBAgAEEMyBAgAEEMyBAgAEEMyAggAOgcIABBHELADOgUIABCRAjoFCC4QkQI6CggAELEDEIMBEEM6BQguELEDOggILhCxAxCDAToLCC4QxwEQrwEQkQI6DQguELEDEMcBEKMCEENQ9awCWMq8AmD2vQJoA3ACeAGAAYsDiAHECpIBBzAuOC4wLjGYAQCgAQGqAQdnd3Mtd2l6yAEIwAEB&sclient=gws-wiz