# Leveraging the Efficiency of the Customer Retention Process: A Deep Learning Approach

## Saurabh Khanolkar[a],Mayuresh Gaitonde[a],Vishal Dabgotra[a]

*[a]Computer Department , Vidyalankar Institute of Technology, Mumbai 400037, Maharashtra, India*

-------------------------------------------------------------------***-------------------------------------------------------------------

**Abstract -** *With an increase in competition, the risk of losing existing customers to competitors i.e. Customer churn has become a major challenge to Subscription-based businesses. It is more economical for subscription-based companies to maintain existing customers than to acquire new customers. Even a slight increase or decrease in the churn rate can make a huge impact on the exchequer of a company. This makes Customer Relationship Management (CRM), a very crucial component of a subscription-based business. Use of mass customer retention strategies (i.e. focusing the retention campaign on the entire customer base) will result in a prodigal spending of resources. Hence the Retention Strategies must be focused on only those who are more likely to churn. In our research, we propose a system that uses demographic segmentation to segment the entire customer base into two segments. One segment will contain the customers that will churn and the other will contain the customers that will not churn. This system will be implemented using a four layered Feed Forward Neural Network which will decide if a customer will churn or not. For the trained model we were able to achieve an accuracy of 86.25%. Also, by using retention strategies only for people who are going to churn instead of the use of mass retention strategies, we came to the conclusion that the proposed solution had a potential to save 79.75% of the company's resources.*

*Key Words*: churn prediction,customer retention,deep learning,demographic segmentation, Artificial Neural Networks.

## 1. Introduction

A business model defines how a business operates, makes money and achieves its targets. The technical progress and the increasing number of operators raised the level of competition [1].

Few commonly used businesses models are Brick and mortar, e-commerce, bricks and clicks, Nickels and dime, Subscription based, drop shipping, Agency based, Aggregator, etc. In this paper we'll only focus on Subscription based business models. Subscription based businesses use the idea of selling a product or service on a monthly or a yearly basis to acquire recurring payments from a single customer. Such Subscription based products or services focus on making revenue through multiple payments from a single customer. Many companies are moving from the classic business revenue model of one- time payment for a product to a subscription-based model for consistent and recurring revenue. But this comes with its own set of challenges. There are many challenges that can be faced by a subscription-based business. For instance, the credit cards used for payment might have exceeded the limit or even expired, there is constant need for quality content, also subscription fatigue may take place. Failure in any one of these might lead to churning of subscribers, which is the biggest threat faced by a subscription-based business.

Churn is defined as the defection of customers to a competitor in the same market segment. The rate of this defection is termed as churn rate. A number of factors are responsible for affecting the churn rate. Even a small month-on-month increase in churn can be hazardous to growth of a business, Employee churn is the overall turnover, which refers to people leaving their jobs in an organization. Employee churn also can be called attrition [3]. So, understanding what churn is and how to analyze it is crucial. For any subscription- based business, the acquisition of new customers is more expensive than retaining existing ones. In this age of extreme competition and similarities in products from various competitors, the importance of Customer Relationship Management (CRM) is increasing. Even a small increase in the churn rate can have a negative effect on the company's ability to grow. Hence the analysis of churn and understanding reasons for the churn is fundamental in addressing and reducing the churn rate. Traditional organizations use regressive methods for preventing churn and retaining customers. Regressive methods inflict huge costs on the organization. Switching to modern methods of customer retention will save a huge amount of resources for the organization. The aim of this research is to study the savings of an organization using modern customer retention strategies as opposed to regressive methodologies.

## 2. Proposed Work

### 2.1 Traditional approach

Traditionally, Customer churn has been handled using regressive techniques like Surveys, Emails, Customer feedback and focus groups.

### 2.2 Problem with traditional approach

Traditional approach relies heavily on mass retention strategies. This means that an organization will spend its resources on the entire consumer base of the organization instead of only the customers most likely to churn out. This is a clear waste of the company's time, efforts and most importantly it's capital! Thus, problem with regressive retention strategies is that the organization has to spend huge amount of resources on retention campaigns. This is not a viable approach since spending resources on a customer who is not going to leave in the organization in the first place is nothing more than a waste of the company's valuable resources. Therefore, the need of the hour is to curb this unnecessary spending of company's resources which will indirectly create more value for the organization.

### 2.3 Traditional approach

Our aim is to optimize the complexity of the process of customer retention applied using conventional techniques. A modern solution to the problem discussed above will be to classify the customers. Thus, there arises a need for a novel approach which focuses only on the customers that have a high probability of leaving. Instead of spending resources on the entire customer base, the low hanging fruits i.e. the customers that are most likely to defect need to be identified. This therefore turns out to be a classification problem. Hence the customer base needs to be classified into two segments. Segment 1 containing the customers that are most likely to churn and Segment 2 consisting of customers that are not likely to churn. This process is termed as demographical segmentation. This result can be achieved using an Artificial Neural Network (ANN)i.e. deep learning-based model. More specifically we will be making use of a four layered feed forward neural network which will operate as a classifier that will decide if a customer will churn or not. A Neural Network has been chosen because of the hypothesis that it gives more accuracy as compared to a traditional machine learning model.

## 3. Implementation and Analysis

### 3.1. Inputs and Materials required

Dataset: The dataset is a set of 10,000 customer records of a European bank. This dataset consists of 14 columns. The columns include: "RowNumber"," CustomerId"," Surname"," CreditScore"," Geography"," Gender"," Age", "tenure"," Balance"," NumOfProducts"," HasCrCard"," IsActiveMember"," EstimatedSalary"," Exited". Out of these, "Exited" is the dependent variable. The dataset is divided into 2 parts. The training set and the test set. 80% of the entire dataset is the training set and the remaining 20% is the test set.

### 3.2. Model Description

Today there are a lot of Deep learning models that are in use. Solving non-linear problems is the most powerful feature of artificial neural network technology in which other classical techniques do not resolve with [2]. The typical models include Autoencoder (AE), Deep Belief Network (DBN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). The type that encompasses them all is called an Artificial Neural Network. Let us view what a neural network system is: A common neural system has something from some dozen to a whole lot, thousands, or maybe an oversized variety of neurons referred to as units organized in an exceedingly progression of layers, all of that interfaces with the layers on either facet.

The input units receive varied data from the surroundings that has to be learned by the model, in order to recognize or to be processed. The output unit is on the opposite side of the network and displays how the network responds to certain inputs. In between the input units and output units are one or more layers of hidden units, which, together, form the majority of the artificial processing unit. The flow of information in a neural network

takes place mainly in the two possible ways. When the model is being trained, the input of the network is being loaded with patterns gathered, which are used for triggering the hidden units' layers, and then arrive at the output.

### 3.2.1. Working of a Neural Network

### 3.2.1.1. Neuron

A Neuron is a fundamental unit of an Artificial neural network. Neurons contain 3 major parts to them. They have a number of inputs, a processor and outputs. A neuron accumulates all its incoming inputs, processes them (using the activation function) and if it goes over a certain threshold it fires a signal through the output. The thing to note about neurons is that neurons learn. They have multiple inputs, for each input there's a weight (the weight of that particular connection). Inputs are multiplied by its corresponding connection weight, added up and then the result is passed on to the processor



$$s_j = \sum_i w_{ij} \cdot y_i$$
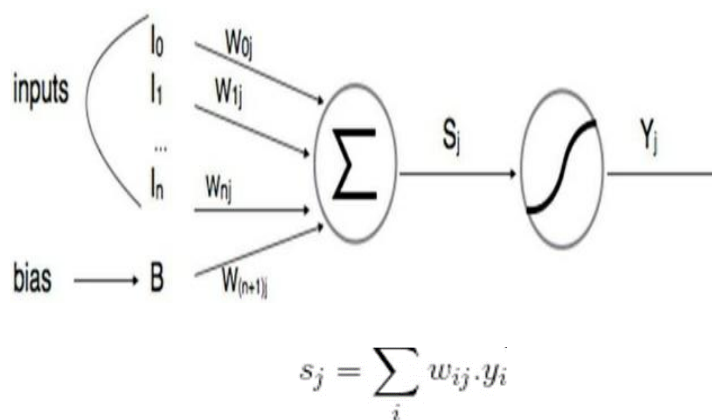
Here yi is all the inputs

After computing the resulting sum, the neuron passes it through its activation function

### 3.2.1.2. Activation Function

The activation function decides if the neuron will fire up or not. If the input value for the processor is above a threshold, the neuron fires up and produces the desired output. There are 4 types of activation functions:
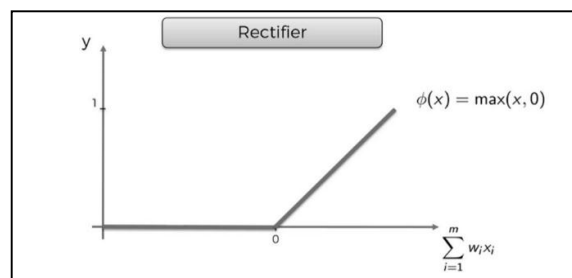
- Threshold Function



$$\phi(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$
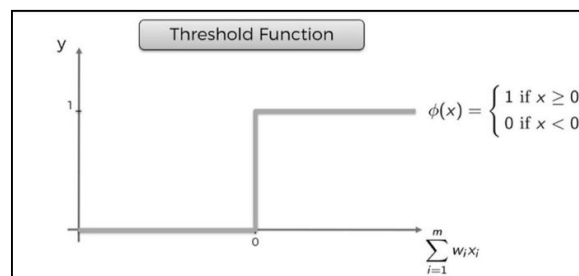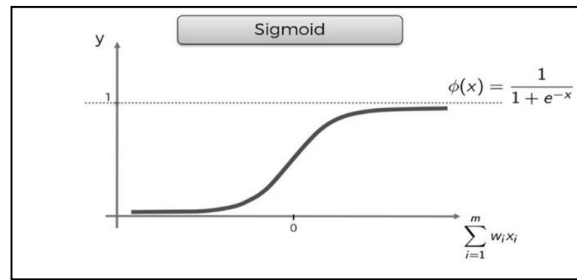
- Sigmoid Function



- Rectifier Function
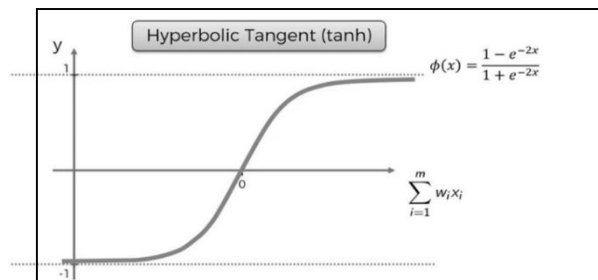


- Hyperbolic Function



Fig. 2. (a. Threshold Function), (b. Sigmoid Function), (c. Rectifier Function), (d. Hyperbolic Function)

*3.2.1.3. Feed Forward Neural Network*

   This is considered to be the simplest neural network. Neurons are organized in layers. Neurons in a particular layer are connected to neurons in the next one. Hence, the output of each layer becomes the input for the next layer.

Fig. 3. Layers of Artificial Neural Network
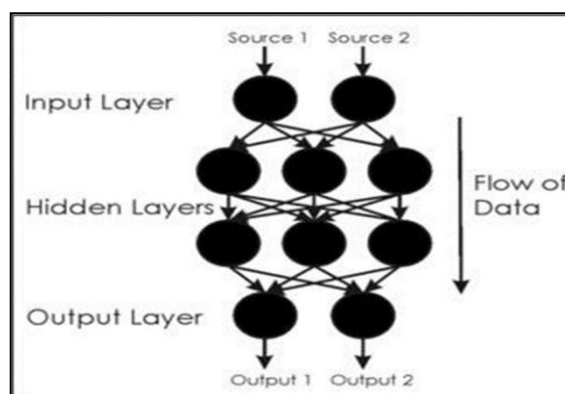
### 3.2.1.4. Learning process of a neural network

A neural network is trained to learn. The algorithm used by neural network to learn is called backpropagation. When the neural network is given an input, it produces an output. Now the network is taught what should have been the appropriate output for the input that was given. The network takes this ideal output and then it adjusts the weights to produce a more appropriate output next time. This process starts from the output layer and goes backwards until the input layer is reached. Next time if the same input is shown to the network, it will give an output closer to that ideal one. This process will be repeated for many iterations or epochs until the difference between the actual output (ideal output) and the predicted output becomes small enough. The difference between the actual output (ideal output) and the predicted output is called as "cost function". In other words: This process will be repeated for many iterations or epochs until the cost function is minimized.

### 3.2.1.5. The process of Back Propagation

In this algorithm, gradient descent calculation is used to calculate the weights. Gradient or instant slope of the actual value of the weight, is created by this algorithm and it moves it in the direction that will lead to a lower error (red dot in the image). This process will be repeated for every weight in the network. A graphic of weight vs error in a networks output is shown below:
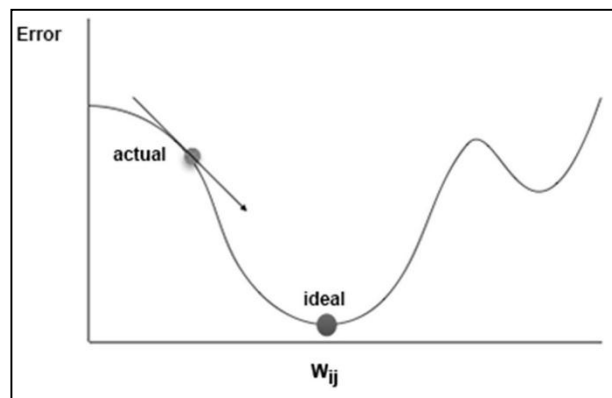


Fig. 4. A graph of weight vs error in a networks output

### 3.2.2. Implementing of Neural Network

### 3.2.2.1. Data Pre-processing

- The dataset has 14 columns. Columns starting from "credit score" to "estimated salary" are taken as x - variables (independent variables) depending on their correlation with the dependent variable/y-variable i.e. "Exited"
- Sklearn library has been used for data pre-processing.
- Categorical data is encoded using "OneHotEncoder" and "LabelEncoder" functions of the sklearn library.
- The dataset has been split into training and test set with an 80-20 split. Therefore, the dataset is divided into 4 parts: X_train, Y_train, X_test and y_test.
- The dependent variables have been standardized using the "Standard Scalar" function.
- Standardization is the process of getting the values of all the variables in a form where mean is 0 and standard deviation is 1.
- It is a necessary step in order to convert the data into a uniform format.
- Classification algorithms require a dataset that has been trained and organized into some classes [4].

### 3.2.2.2. Construction of ANN model

- The model is initialized using a sequential classifier.
- The "Dense" module from the keras library helps to initialize the weights of the ANN and is also used to add layers of neurons to our neural network.
- The neural network has 4 layers of neurons:
    - i. The input layer: 11 neurons (due to 11 inputs/independent variables)
    - ii. Hidden layer1: 6 neurons
    - iii. Hidden layer2: 6 neurons

iv. The output layer: 1 neuron (binary prediction)
- The activation function used for the hidden layer is "Relu" (Rectifier function)
- The activation function used for output function is sigmoid function (so that we get our prediction in terms of probability which can later be converted to 0s and 1s)
- The classifier uses stochastic gradient descent algorithm (randomized gradient descent) for back propagation.
- The neural network classifier has then been compiled, fit to the training set and trained.
- The trained classifier is then tested on the test set to predict the dependent variables i.e. "Exited" for each observation/ tuple of the dataset.

*3.2.2.3. Result Visualization*

The accuracy of the predictions is visualized using the confusion matrix, analysis of which has been done in the "Result Analysis" section.

*3.2.3. Deep Learning Frameworks used*

TensorFlow, Keras, Pytorch, Theano, Caffe are the packages generally used for deep learning. TensorFlow works on images along with sequence-based data. Pytorch is much more effective as opposed to TensorFlow. Caffe works very well if the data is in the form of images. Caffe lags behind the other frameworks when it comes to recurrent neural networks and language models. Theano and Tensor Flow are the two amazing mathematical platforms which give the premise to deep learning research and development. However, these two powerful libraries can get some what complicated to get started with. In our paper, we have made use of an open source library, which is mainly used for deep learning and written using python. It is the Keras library. It runs on top of TensorFlow. Keras python library gives a perfect and advantageous approach to make a scope of deep learning models over Theano and Tensor Flow. Hence keras is the library chosen for our research.

## 4. Result Analysis

The output layer of the neural network uses sigmoid activation function as discussed above. The sigmoid activation function gives a vector of probabilities of the customers leaving the bank. A high value in one of the vector cells suggest that there is a high probability of that particular customer churning. As opposed to this, a lower value in the matrix suggests that there is a lower value of the customer churning.
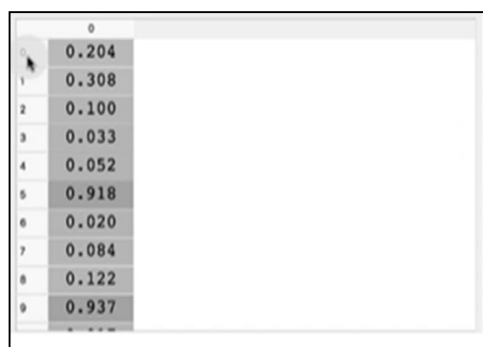
The result vector :



Fig. 5. Result Vector

An interesting thing that an organization can do is it can rank the customers according to their probabilities of leaving. Once such a ranking is created, the organization can bin the customers according to their ranks. This process forms the very basis of demographical segmentation. The organization can apply different customer retention strategies on the different bins/segments. For example; An organization can segment the top 10% of their customers who are most likely to churn and then apply data mining techniques to identify the reason for the same.

Our approach here has been binary. We have taken the threshold of 0.5. Any probability above 0.5 will be considered as "1". This means that anyone getting a probability of above 0.5 will be considered as a customer who is likely to churn. Anyone getting a probability less than 0.5 will be considered as "0" i.e. less likely to churn. Thus, the

entire customer base has been divided into 2 segments; one which is likely to churn and another which isn't. We can calculate the accuracy of our model by looking at the confusion matrix which states the number of correct predictions v/s the number of incorrect predictions.

The confusion matrix :



Fig. 6. Confusion Matrix

### 4.1. Analysis of Confusion Matrix

- The cell cm [0][0] i.e. the cell (0,0) states that there are 1550 of the 2000 observations that are not likely to leave the bank and the model predicted them correctly.
- The cell cm [0][1] i.e. the cell (0,1) states that there are 45 of the 2000 observations that are likely to leave the bank but the model predicted them incorrectly.
- The cell cm [1][0] i.e. the cell (1,0) states that there are 230 of the 2000 observations that are not likely to leave the bank but the model predicted them incorrectly.
- The cell cm [1][1] i.e. the cell (1,1) states that there are 175 of the 2000 observations that are likely to leave the bank and the model predicted them correctly.

### 4.2. Calculating the accuracy of the model

- There is a total of 1550 + 175 i.e. 1725 correct predictions.
- The total number of observations in our test set are 2000.
- Calculating the accuracy of our model: (1725/2000) *100 = 86.25%
- Therefore, the model is 86% accurate in predicting whether a customer will churn or not.

### 4.3. Analyzing the resource saved for the organization

- Consider that the resources spent on the customer retention campaign for 2000 customers using the traditional approach is "X".
- According to the proposed solution, the model has predicted that, 199+206; i.e. 405 customers are likely to churn out. (view confusion matrix)
- Therefore, the new amount of resources spent on the customer retention campaign will be (405/2000) *X
- Thus, the percent of resources saved is 100-{[(405*X)/2000] *100}/X
- This comes out to be 79.75%
- Hence our proposed method has the potential to save a total of 79.75% of resources for the organization.

## 5. Conclusion

Customer churn is a major issue that organizations face. In this modern world, there are many organizations that use regressive techniques for retaining their customers. Businesses spend unnecessary resources on the entire customer base for retaining their customers[5]. This leads to a prodigal spending of resources and in turn a

financial burden for the organization. The solution to this issue is a "focused customer retention strategy". This is a classification problem. The objective of classification can be achieved very well using neural networks with a high accuracy. Thus, neural networks have brought an unprecedented accuracy in a field which previously functioned on uncertainty.

## 6. Competing Interests

The Authors declare that they have no Competing Interests

## Acknowledgements

## References

[1] Gerpott TJ, Rams W, Schindler A, "Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market.", Telecommunications Policy, vol. 25, issue 4, pp. 249–269, May 2001.

[2] Yildiz B, Yezegel A, "Fundamental analysis with artificial neural network", *The International Journal of Business and Financial Research,* vol. 4, pp. 149-158, 2010.

[3] Martins PS, "Working Churning and Firms' Wage Policies," International Journal of Manpower, vol. 29, no. 1, pp. 48-63, 2008.

[4] Arusada MDN, Putri NAS and Alamsyah A, "Training Data Optimization Strategy for Multiclass Text Classification," International Conference on Information and Communication (ICoICT), vol. 5, 2017.

[5] "What is churn prediction?" https://www.appier.com/blog/churn-prediction/

[6] Longanecker C, Why You Should Use a Subscription Business Model. Entrepreneur Media, Inc, 2015.

[7] Rozum JA, Defining and Understanding Software Measurement Data [J]. Pittsburgh: Software Engineering Institute Carnegie Mellon University, pp. 2-3, 2002.

[8] Martin G, "*The Importance of Marketing Segmentation*", American Journal of Business Education vol. 4, no. 6, June 2011.

[9] Gigli M, "*What is Demographic Segmentation in Marketing?",* lesson 7, chapter 3.

https://study.com/academy/lesson/what-is-demographic-segmentation-in-marketing-definition-advantages-disadvantages.html

[10] Mialki S, "*What is Demographic Segmentation & why is it necessary for Marketing Campaigns?", January 28,2020.*

https://instapage.com/blog/demographic-segmentation

[11] Fripp G, "*Advantages and limitations of the main market segmentation bases"*, Marketing Study Guide.

https://www.marketingstudyguide.com/advantages-and-limitations-of-the-main-market-segmentation-bases/