**TRAINING REPORT**

**On**

**CHURN MODELLING ON BANK CUSTOMER**

Submitted to MAHARAJA RANJIT SINGH PUNJAB TECHNICAL

UNIVERSITY in partial fulfillment of the requirement for the award of the degree of

**B.TECH**

**in**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted By**

**CHANDRA BHUSHAN**

**Roll No.: 15110151**

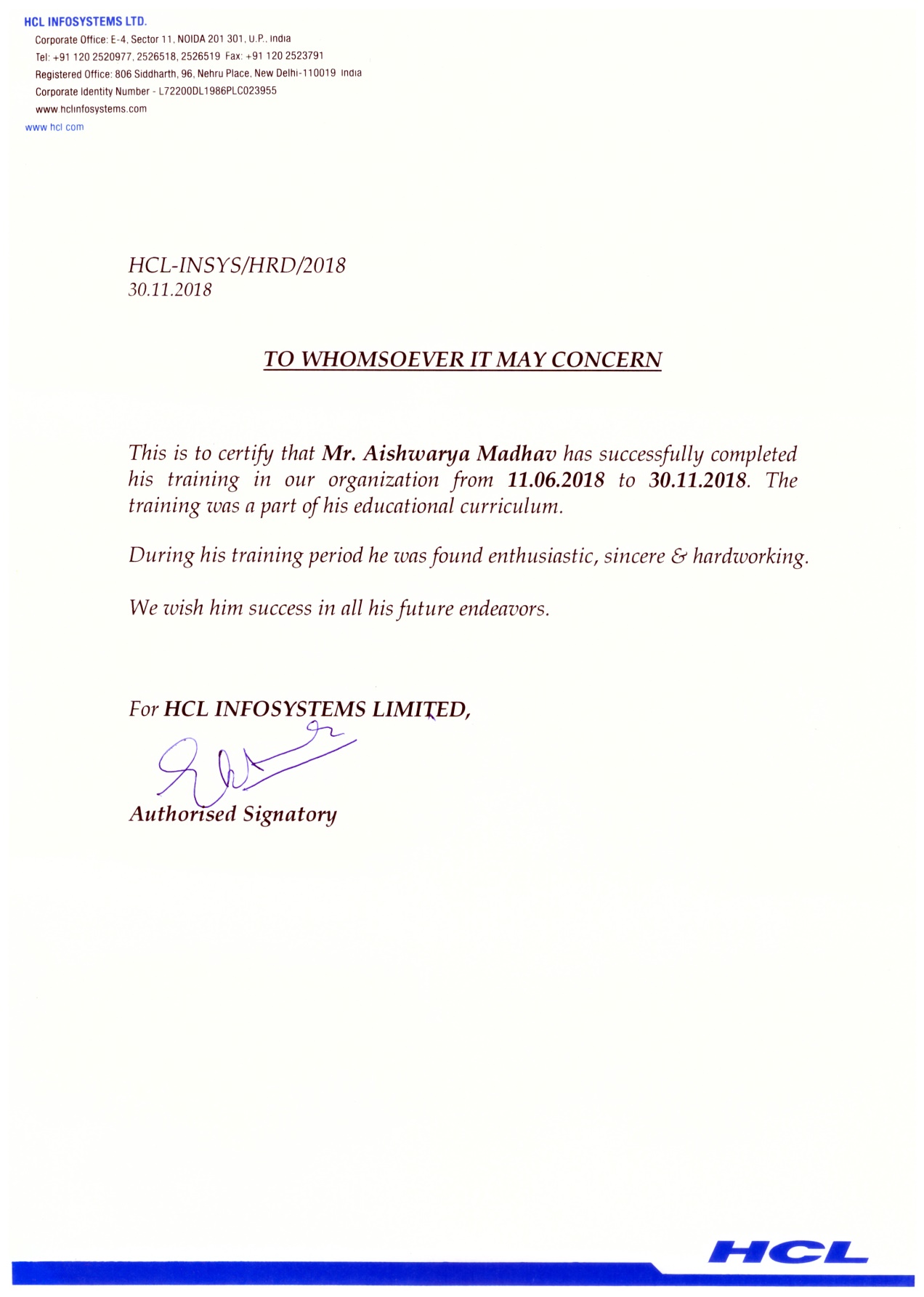
****

**DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING**

**GIANI ZAIL SINGH CAMPUS COLLEGE OF ENGINEERING & TECHNOLOGY, BATHINDA-151001**

**DEC 2018**

**CERTIFICATE BY COMPANY**



**PREFACE**

Training is an integral part of B.Tech each and every student has to undergo the training for 6 months in a company.

This record is concerned about our practical training during the 7th semester of our B.Tech. We have taken our practical training in (Data Science- RStudio, Python3.5, Anaconda, Spyder, Hive, Tableau, ).

During this training, we got to learn many new things about the industry and the current requirements of companies. This training proved to be a milestone in our knowledge of present industry. Every moment was an experience in itself, an experience which theoretical study can’t provide.

**ACKNOWLEDGEMENT**

It is my pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced my thinking, behaviour and acts during the course of study.

I express my sincere gratitude to ***Dr.Naresh Garg*** worthy HOD and ***Er.Dinesh Kumar*** and ***Er. Manpreet Kaur***. Training & Placement In-charge for providing me an opportunity to undergo summer training at Naresh I Technology.

I am thankful to ***Naveen Sir*** for support, cooperation, and motivation provided to me during the training for constant inspiration, presence and blessings.

I also extend my sincere appreciation to ***Naveen Sir*** who provided his valuable

suggestions and precious time in accomplishing my training report.

Lastly, I would like to thank the almighty and my parents for their moral support and my friends with whom I shared my day-to day experience and received lots of suggestions that my quality of work.

**CHANDRA BHUSHAN**

**CANDIDATE’S DECLARATION**

I, Chandra Bhushan, Roll No. 15110151, B.Tech(Semester-VII) of the **Giani Zail Singh Campus College** **of Engineering & Technology, Bathinda** hereby declare that Training Report entitled “**CHURN MODELLING ON BANK CUSTOMER”** is an original work and data provided in study is authentic to the best of my knowledge.This report has not been submitted to any other Institute for the award of any other degree.

**Chandra Bhushan**

(RollNo. 15110151)

**Place:** **BATHINDA**

**Date: 12-11-2018**

**INDEX**

|  |  |
| --- | --- |
| **Chapters** | **Page No.** |
| 1. **Introduction** |  |
| * 1. CRM   2. Introduction to the technique of churn modelling   3. CART   4. Logistic Regression   5. Artificial Neural Network      1. Problem Statement and Motivation      2. Thesis Structure Flow Chart  1. **Literature Survey**    1. Customer Chrn Management    2. Customer Satisfaction    3. Churn Management Framework    4. Feature Selection    5. Search Phase   The Evaluation Phase  Development of Predictive Model  Traditional Methods  Decision Trees  Regression Analysis  Neural Network Development  Validation Methods   1. **System Requirement Specification**    1. Non-functional requirements    2. Specific Requirements    3. Software requirements    4. Hardware requirements      1. **System Design**    1. Dataflow Diagram 2. **Implementation**    1. The Proposed Neural Network (NN) Based Approach    2. Artificial Neural Network    3. Dataset    4. Training and Testing of Neural Network 3. **Performance Analysis** 4. **Result & Discussion** 5. **Screenshots** 6. **Code** 7. **Future Enhanceent** 8. **Conclusion**   **References** | 11  12  12  13  13  13  14  15  15  17  17  18  18  18  18  19  19  20  20  21  23  23  24  24  25  26  27  28  14  16  28  30  30  33  34  35  37  41  42 |

**LIST OF FIGURES AND TABLES**

|  |  |  |
| --- | --- | --- |
| 1.5.1 | Thesis Structure | 14 |
| 2.1 | Main churn contribution to the Telecommunication Industry | 16 |
| 2.3 | The stage of Churn Management Framework | 17 |
| 2.5 | Predictive modeling and Segmentation technique | 21 |
| 4.1 | Dataflow diagram of proposed system | 26 |
| 4.2 | The implementation of Neural Network using clemantine-12.0 | 27 |
| 4.3 | Alyuda Network topology | 27 |
| 5.1.1 | Neural Network with one hidden layer | 28 |
| 5.1.3 | The implementation of Neural Network using Clementine-12.0 | 32 |
| 6.1 | Accuracy | 33 |
| 8.1 | Dataset | 35 |
| 8.2 | Processed data | 35 |
| 8.3 | Result | 36 |

**ABSTRACT**

Marketing literature states that it is more costly to engage a new customer than to retain an existing loyal customer. Churn prediction models are developed by academics and practitioners to effectively manage and control customer churn in order to retain existing customers. As churn management is an important activity for companies to retain loyal customers, the ability to correctly predict customer churn is necessary. As the cellular network services market becoming more competitive, customer churn management has become a crucial task for mobile communication operators. This paper proposes a neural network (NN) based approach to predict customer churn in subscription of cellular wireless services. The results of experiments indicate that neural network based approach can predict customer churn with accuracy more than 92%. Further, it was observed that medium sized NNs perform best for the customer churn prediction when differ ent neural network’s top ologies were experimented.

The aim of this article is to present a case study of usage of one of the data mining methods, neural network, in knowledge discovery from databases in the banking industry. Data mining is automated process of analysing, organization or grouping a large set of data from different perspectives and summarizing it into useful information using special algorithms. Data mining can help to resolve banking problems by finding some regularity, causality and correlation to business information which are not visible at first sight because they are hidden in large amounts of data. In this paper, we used one of the data mining methods, neural network, within the software package Alyuda NeuroInteligence to predict customer churn in bank. The focus on customer churn is to determinate the customers who are at risk of leaving and analysing whether those customers are worth retaining. Neural network is statistical learning model inspired by biological neural and it is used to estimate or approximate functions that can depend on a large number of inputs which are generally unknown. Although the method itself is complicated, there are tools that enable the use of neural networks without much prior knowledge of how they operate. The results show that clients who use more bank services (products) are more loyal, so bank should focus on those clients who use less than three products, and offer them products according to their needs.

**Chapter-1**

**INTRODUCTION**

In today’s business world companies are recognising that customer value and increased revenue is more likely to come from their existing customer base than from new customer acquisition. The reasoning behind this is that companies know their existing customers, already have a relationship with them, and amplitude of data on them. In common recognition of this problem, industry has seen an emergence of customer relationship management (CRM) products. Software companies have realized that CRM has become a receptive topic within industry and therefore an area of potential wealth and opportunity for the development and promotion of products that promise to not only boost customer retention, but also for increasing selling opportunities. There are three core components that are typically covered by CRM: cross-selling, up-selling and customer retention. Each of these areas of CRM is described below. The focus of this research is on customer retention.

**1.1. CRM**

Cross-selling is attempting to promote products to the customer that the customer does not usually purchase. For example a company might recognise that a customer has aation. They are also not very sensitive to assumptions about error terms and they can tolerate noise and chaotic components. Notably, SVMs are increasingly used in materials science, the design of engineering systems and financial risk prediction.

Also, most methods that are in use are only applicable to a small portion of stock markets and usually such models do not generalize well to all stocks. Additionally, existing libraries are highly efficient in obtaining the optimal hyperparameters to be used in LSSVM and other algorithms.history of purchasing books from a specific author. Cross-selling could take the genre that the author typically covers in his/her writing and use that as a basis of recommending books to the customer from alternative authors who write in the same genre, attempting to persuade the customer to try something different. Up-selling attempts to promote repeat purchases of the same products that the customer has purchased in the past. Up-selling also involves the promotion of more expensive versions of the same service as a way of increasing revenue. The focus of this research is on customer retention, which is the process of retaining existing customers. There are several reasons why retaining existing customers is important. The first reason is that markets have become saturated to a point where new custom to the industry is scarce. The second reason is cost. New customer acquisition can be costly to a business for numerous reasons.

**1.2. Introduction to the Techniques for Modelling Customer Churn**

In order to manage customer churn within a company it is important to build an effective and accurate customer churn model. There are several modelling techniques available that can aid in the prediction of customer churn. The most common techniques have been identified from literature as:

1. Classification and regression trees (CART)
2. Logistic regression models (LRM)
3. Artificial neural networks (NN)

**1.2.1. CART**

CART (also known as recursive partitioning regression) is a popular classification technique used for predicting events. CART works by recursively dividing the response variables into increasingly homogenous subsets based on significant thresholds of the predictor variables (Lozano et al., 2008). CART development usually consists of two phases: tree building and tree pruning. The tree building phase consists of recursively partitioning the training set according to the values of the attributes. The partitioning process continues until all or most of the records in each of the partitions contain identical values. Certain branches may need to be removed because it is possible that they may contain noisy data. The trees are allowed to grow as large as possible to avoid early stopping where important rules could be missed. The trees are then pruned backwards to avoid over fitting (Gray and Fan, 2008). The pruning phase involves selecting and removing the branches that contain the largest estimated error rate.

**1.2.2. Logistic Regression**

Logistic regression model (LRM) is an extension of multiple regressions. It provides an output that is in the form of a probability between the values 0 and 1 (Nefeslioglu et al., 2008). A description of a LRM is provided by Nannings et al. (2008) who state “One reason for the popularity of the LRM is the interpretation that is given to a covariate coefficient i in terms of an *odds ratio*. For an event with probability *p* its odds are *p*/(*1*−*p*). Regression is the study of dependence and is the central part of many research projects. It can answer questions about the dependence of a response variable on one or more predictor variables, including predictions of future values (Weisberg, 2005).

**1.2.3. Artificial Neural Networks**

Artificial neural networks consist of basic elements known as ‘neurons’. These neurons consist of three main components: (i) weight, (ii) bias and (iii) activation function. Each neuron receives an input on which it applies a weight value. This weight holds the key to the neural network’s overall performance because it provides the strength of the connection to the specific input. Each input is assigned a unique weighted value per neuron connection and all inputs are connected to all neurons in the first hidden layer. Each input is multiplied by its corresponding weight. These values are summed and a bias value that is a constant non-zero value is added. This summation is transformed using scalar-scalar functions known as the activation or transfer functions. These functions help to establish non-linearity into the NN, contributing to the immense power of this technique. A myriad of NN architectures have emerged over time, each identifiable by their topology (Cevik and Guzelbey, 2008).

**1.3. Problem Statement and Motivation**

From the issues imposed through market saturation and cost implications as described inthe first section of this chapter, there has been an identification of a need for a computerbased churn prediction methodology that is capable of accurately identifying a loss of customer in advance, so that proactive retention strategies can be deployed in a bid to retain the customer. The churn prediction has to be accurate because retention strategies can be costly. A limitation of current research is that other studies have focussed almost exclusively on churn capture, neglecting the issue of misclassification of non-churn aschurn. Retention campaigns commonly include making service based offers to customers in a bid to retain them. These offers can be costly, so offering them to customers who do not intend to churn can have a considerable impact on the total cost of a retention strategy. A further limitation of current research is that it is usually based on a single output in the form of 0 for non-churn and 1 for churn. This has been recognised as a limitation because it restricts analysis possibilities.

**1.5.1. Thesis Structure Flow Chart**



**Figure 1.5.1: Thesis Structure**

**Chapter 2**

**LITERATURE SURVEY**

As markets have become increasingly saturated, companies have acknowledged that their business strategies need to focus on identifying those customers who are most likely to churn. In order to maximise business profits a company should focus on minimising customer churn. Typical churn for a service provider is usually around 4% each month. For the wireless telecommunications industry it has been suggested that customer churn cost the service sector around four billion dollars in 2006 (Chu et al.,2007). This statement is reinforced by Sweeney and Swait (2008) who state, “Customer churn is an ever-growing issue in the relational services sector (e.g., retail banking, telecommunications), where business models ultimately depend on long-term relationships with customers as the basis for profitability”. This is supported by Eriksson and Vaghult (2000) who state, “customer retention leads to reduced sales and marketing costs compared to selling to new customers”. This in turn explains the statement by Hidalgo et al. (2008) who claim, “a 1% improvement in the customer retention rate improves firm value by 5%”.

**2.1. Customer Churn Management**

It is becoming common knowledge in business, that retaining existing customers is the best core marketing strategy to survive in industry (Kim et al., 2004, Lariviere and Van Den poel, 2004). Retained customers generate more financial returns than new customers, which is why businesses should make every effort to retain their existing customer base, rather than investing valuable revenue in attempting to capture new subscribers (Buckinx and Van Den Poel, 2004). Seo et al. (2008) expand on this by claiming “in addition to the reduced costs, there is potential and opportunity value of customers which is gained over a long period of time. Because wireless telecommunications is not a onetime sale like commodity products, service providers can offer additional services over the length of a customer’s tenure to generate more revenue”. This is supported by Eriksson and Vaghult (2000) who state, “customer retention leads to increased sales and reduced marketing costs compared to selling to new customers”. Ahn et al. (2006) support this claim by stating “with an increase in customer retention rates of just 5%, the average net present value of a customer increases by 35% for software companies and 95% for advertising agencies. Therefore, in order to be successful in the maturing market, the strategic focus of a company ought to shift from acquiring customers to retaining customers by reducing customer churn”. Further to the increase of sales and profits that are generated by loyal customers it has also become apparent that when a customer churns from his/her current service provider cost are imposed on that service provider that are in most cases unrecoverable (Gans, 2000).

A number of products exist for customer relationship management (CRM) which aims at analysing a company’s customer base. CRM is not a new concept, beginning in the mid 1990’s with IT based systems being developed to track multiple customer activities (Minami and Dawson, 2008). Today there are many commercial products available for the purpose of CRM with the amount of money being invested in this field exploding over recent years (Ang and Buttle, 2002). With so many products emerging organisations should take great care when deciding to purchase one of these solutions off-the-shelf’. Chen and Popovich (2003) state that “CRM vendors might entice organisations with promises of all powerful applications. To date there is no 100% solution”. The author agrees with the quote from Chen and Popovich (2003); however, it is believed that the main reason no 100% solution is available is due to the uncertainty and complexity involved in churn prediction, and the fact that churn can be a result of many varying factors. The main contributors to churn within the telecommunications service sector have been reported by Chu et al. (2007). Figure 2 illustrates the significance of each of these contributors:



**Figure 2: Main Churn Contributors For The Telecommunications Industry (Chu et al., 2007).**

**2.2.Customer Satisfaction**

Referring again to the pie chart presented in Figure 2 it is shown that the cause of 35% of churn within the telecommunications industry is related to customer service, quality and coverage issues. These three categories can be linked to customer satisfaction, as customer satisfaction is a reflection of the customer’s perceived service in relation to the service that he/she expects to receive. Hahm et al. (1997) state, “satisfaction is really a gap measure between performance and expectation”. This statement is backed up by Kim et al. (2007) who state “a completely satisfied customer perceives their service to meet or exceed expectation”.

Finding a suitable measurement of customer satisfaction is a major problem for organisations and has been a focus of research for quite some time (Yi, 1989). This statement is reinforced by Siskos et al. (1998) who state, “measuring customer satisfaction is a major problem for every firm or organisation, especially within the frame of marketing management practices. Satisfaction of customer needs is the main objective according to the principles of modern marketing science”.

**2.3.Churn Management Framework**

A five stage model for developing a customer churn management framework has been identified (Datta et al., 2001). These stages are illustrated in Figure 3:

****

**Figure 3: The Stages a Churn Management Framework (Datta et al., 2001).**

**2.3.3. Feature Selection**

Feature selection is the process of identifying the fields which are most suitable for predicting an event, described by Sun et al. (2004) as a critical process. It is an important stage because it helps with both data cleansing and data reduction by including the important features of a database and excluding the redundant, noisy and less informative ones (Yan et al., 2004). There are two main stages to feature selection. The first stage is a search strategy for the identification of the feature subsets and the second is an evaluation method for testing their integrity, based on some criteria.

**2.3.3.1. Search Phase**

The search phase of feature selection can be split into three categories, optimal, heuristic and randomised searches. The optimal search method is straight forward; however it is an exhaustive search method. With this process the number of possible subsets grows rapidly, making it unusable for even moderately sized feature sets. There are optimal search methods that avoid the exhaustive approach, using for example the branch and bound algorithm (Sun et al., 2004).

**2.3.3.2. The Evaluation Phase**

Evaluation strategies can be divided into two categories, the first is called filter and the second is called wrapper. A wrapper evaluation method is where the feature subset evaluation is performed using a learning algorithm that is incorporated in the classification design, while the filter approach uses feature subset evaluation external to the classification design. Filter approaches are more efficient than wrapper approaches because they evaluate the fitness of features using criteria that can be tested quickly. However these approaches could lead to non-optimal features, especially when the features are dependent on the classifier, leading to poor classifier performance (Sun et al., 2004, Kavzoglu and Mather, 2001).

**2.4.Development of predictive model**

A predictive model is defined as one that takes patterns that have been discovered in the database, to predict a future event (Rygielski et al., 2002). According to Crespo and Weber (2004) the most important predictive modelling techniques include decision trees and neural networks. The popularity of these technologies are reinforced by Baesens et al. (2004) who claim, neural networks and decision trees are typically traditional classification technologies. The following subsections provide an overview of both traditional and soft computing techniques used for predictive modelling.

**2.4.1. Traditional Methods**

This section covers the most common techniques that have been found in literature which have commonly been used for predictive analysis and data mining.

**2.4.1.1. Decision Trees**

The most popular type of predictive model is the decision tree. Decision trees have become an important knowledge structure, used for the classification of future events (Muata and Bryson, 2004a). Decision tree development usually consists of two phases, tree building and tree pruning. The tree-building phase consists of recursively partitioning the training sets according to the values of the attributes. The partitioning process continues until all, or most of the records in each of the partitions contain identical values. Certain branches may need to be removed because it is possible that they could consist of noisy data. The pruning phase involves selecting and removing the branches that contain the largest estimated error rate. Tree pruning is known to enhance the predictive accuracy of the decision tree while reducing complexity (Au et al., 2003). Pruning should be regarded as a process of experimentation because it is possible that pruning the tree could decrease the accuracy of the output rather than enhance it.

A classification and regression tree (CART) is constructed by recursively splitting theinstance space into smaller sub-groups until a specified criterion has been met. The decrease in impurity of the parent node against the child nodes defines the goodness of the split. The tree is only allowed to grow until the decrease in impurity falls below a user-defined threshold. At this time the node becomes terminal, or leaf node (Bloemer et al., 2002).

**2.4.1.2. Regression Analysis**

Regression analysis is a popular technique used by the researchers dealing with predicting customer satisfaction. It provides a first step in model development. Mihelis et al. (2001) developed a method to determine customer satisfaction using an ordinal regression based approach. Another model for assessing the value of customer satisfaction was developed by Rust and Zahorik (1993). They used logistic regression to link satisfaction with attributes of customer retention. They claim that the logistic function can be interpreted as providing the retention probability. Kim and Yoon (2004) use a binomial logit model to determine subscriber churn in the telecommunications industry, based on discrete choice theory. Discrete choice theory is the study of behaviour in situations where decision makers must select from a finite set of alternatives. According to Au et al. (2003) regression analysis is fine for determininga probability for prediction; however it is unable to explicitly express the hidden patterns in a symbolic and easily understandable form.

**2.5. Neural Network Development**

Artificial neural networks are loosely mathematically based on what is known about physical, biological cognition and learning. Although much is understood regarding biological systems, biological brains are morphologically and chemically extremely complex, with invasive studies ethically prohibited. Because of this, huge portion of information are missing from the mystery of the human mind. Therefore even the most complex neural mathematical models should not be regarded as accurate, however can be viewed as providing the smallest number of conflicts with respect to what is currently known (Wythoff, 1993).

Artificial neural networks have been defined as by Cevik and Guzelbey (2008) as “a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. The main component of this

model is the structure of its information processing unit”.

Artificial neural networks have been successfully used to estimate intricate non-linear functions. A neural network is an analogous data processing structure that possesses the ability to learn. The concept is loosely based on a biological brain and has successfully been applied to many types of problems, such as classification, control, and prediction (Behara et al., 2002).



**Figure 7: Predictive Modelling and Segmentation techniques (Verhoef, 2001)**

**2.6. Validation Methods**

Validation is extremely important for data mining models as it is the only real way to ensure that the model is not simply remembering each data instance that was used in training (Bellazzi and Zupan, 2008).

There are several methods documented for validating a customer churn model. Some popular methods are discussed below:

* + Cross-fold validation: Hwang et al. (2004) performed validation by creating a 70/30 divide of the data. The 70% divide created the training set, and the 30% divide created the validation set. Cross-fold validation is based on the principle of using the available data for both training and validation. Cross-fold validation is most suitable in those cases in which there is a scarcity of data.

**Chapter 3**

**SYSTEM REQUIREMENT SPECIFICATION**

To be used efficiently, all computer software needs certain hardware components or other software resources to be present on a computer. These prerequisites are known as (computer) system requirements and are often used as a guideline as opposed to an absolute rule. Most software defines two sets of system requirements: minimum and recommended. With increasing demand for higher processing power and resources in newer versions of software, system requirements tend to increase over time. Industry analysts suggest that this trend plays a bigger part in driving upgrades to existing computer systems than technological advancements.

Software requirements specification establishes the basis for an agreement between customers and contractors or suppliers on how the software product should function (in a market-driven project, these roles may be played by the marketing and development divisions). The software requirements specification lays out functional and non-functional requirements, and it may include a set of use cases that describe user interactions that the software must provide.

**3.1 Non-functional requirements**

Non-functional requirements are the functions offered by the system. It includes time constraints and constraints on the development process and standards. The non-functional requirements are as follows:

* **Speed:** The system should process the given input into output within appropriate time.
* **Ease of use:** The software should be user friendly. Then the customers can use easily, so it doesn’t require much training time.
* **Reliability:** The rate of failures should be less then only the system is more reliable
* **Portability**: It should be easy to implement in any system.

**3.1.1 Specific Requirements**

The specific requirements are:

* **User Interfaces:** The external users are the clients. All the clients can use this software for indexing and searching.
* **Hardware Interfaces:** The external hardware interface used for indexing and searching is personal computers of the clients. The PC’s may be laptops with wireless LAN as the internet connections provided will be wireless.
* **Software Interfaces:** The Operating Systems can be any version of Windows.
* **Performance Requirements:** The PC’s used must be atleast Pentium 4 machines so that they can give optimum performance of the product.

**3.2 Software requirements**

Software requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application.

These requirements or prerequisites are generally not included in the software installation package and need to be installed separately before the software is installed.

**Operating System** Window/ Ubuntu Linux/MacOS

**Programming Language** R, Python

**IDE/Editor** RStudio, Spyder, Jupyter Notebook

**Required Python Packages** TensorFlow, Keras, Pandas, Numpy, Matplotlib, Sklearn, Pytorch

**3.3 Hardware requirements**

The most common set of requirements defined by any [operating system](http://en.wikipedia.org/wiki/Operating_system) or [software application](http://en.wikipedia.org/wiki/Software_application) is the physical computer resources, also known as [hardware](http://en.wikipedia.org/wiki/Computer_hardware), A hardware requirements list is often accompanied by a [hardware compatibility list](http://en.wikipedia.org/wiki/Hardware_compatibility_list), especially in case of operating systems. An HCL lists tested, compatible, and sometimes incompatible hardware devices for a particular operating system or application. The following sub-sections discuss the various aspects of hardware requirements.

All computer operating systems are designed for a particular computer architecture. Most software applications are limited to particular operating systems running on particular architectures. Although architecture-independent operating systems and applications exist, most need to be recompiled to run on a new architecture.

The power of the central processing unit (CPU) is a fundamental system requirement for any software. Most software running on x86 architecture define processing power as the model and the clock speed of the CPU. Many other features of a CPU that influence its speed and power, like bus speed, cache, and MIPS are often ignored.

This definition of power is often erroneous, as AMD Athlon and Intel Pentium CPUs at similar clock speed often have different throughput speeds.

**System/Processor** Intel Core i3,5,7, 2.4-3.0 GHz

**Hard Disk Space** 500 GB or more

**RAM** 4 GB or more

**Internet Connection** Required to auto-download dataset

**CHAPTER 4**

**SYSTEM DESIGN**

System design is the process of defining the architecture, components, modules, interfaces and [data](http://en.wikipedia.org/wiki/Data) for a [system](http://en.wikipedia.org/wiki/System) to satisfy specified [requirements](http://en.wikipedia.org/wiki/Requirement). One could see it as the application of [systems theory](http://en.wikipedia.org/wiki/Systems_theory) to [product development](http://en.wikipedia.org/wiki/Product_development). There is some overlap with the disciplines of [systems analysis](http://en.wikipedia.org/wiki/Systems_analysis), [systems architecture](http://en.wikipedia.org/wiki/Systems_architecture) and [systems engineering](http://en.wikipedia.org/wiki/Systems_engineering). If the broader topic of [product development](http://en.wikipedia.org/wiki/Product_development) "blends the perspective of marketing, design, and manufacturing into a single approach to product development," then design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing [systems](http://en.wikipedia.org/wiki/System) to satisfy specified [requirements](http://en.wikipedia.org/wiki/Requirement) of the user.

**4.1 Dataflow Diagram**

A data flow diagram is a graphical representation of the "flow" of data through an information system, modeling its process aspects. Often, they are a preliminary step used to create an overview of the system which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).

Figure 4.1 Dataflow diagram of the proposed system

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system.

After initialization the user is asked to select or input a company name using a Ticker. The system fetches the stock data online for a given input date range and plots the, on a graph. The data is sent to the train the model. Stock Predictions for n-day ahead are performed and the model is saved for future use.



**Fig 4.2: The Implementation of Neural Network using Clementine 12.0**



**Figure 4.3.** Alyuda – Network Topology.

**CHAPTER 5**

**IMPLEMENTATION**

**5.1 THE PROPOSED NEURAL NETWORK (NN) BASED APPROACH**

**5.1.1 Artificial Neural Network**

An ANN is a complex network that comprises a large set of simple nodes known as neural cells. ANN was proposed based on advanced biology research concerning human brain tissue and neural system, and can be used to simulate neural activities of information processing in the human brain [16]. ANN has the topological structures of information processing nodes that distribute information in parallel fashion. The mappings of inputs and estimated output responses are obtained via combinations of nonlinear transfer functions. We can make use of self-adaptive information pattern recognition methodology to analyze the training algorithms of the artificial neural networks using past experience, neural cells, memory and association, to process fuzzy, nonlinear, and noise-containing data without developing any mathematical models. The various algorithms of the neural networks training are Hebb, Delta, Kohonen, and BP computation. The mostly used BP computation algorithm is the error back propagation algorithm proposed by the PDP group of Rumelhart in 1985

****

**Fig 5.1.1 Neural Network with One Hidden Layer**

Neural networks can be differentiated into single-layer perception and multilayer perception (MLP) network. The multilayer perception consists of multiple layers of simple, two state, sigmoid transfer function having processing element or neurons that interact by using weighted connections. In addition, the neural network contains one or more several intermediary hidden layers of neurons between the input and output layers. Such intermediary layers are called hidden layers and nodes embedded in these layers are called hidden nodes since these are not taking inputs directly from outside. A typical feed-forward multi- layer perceptron neural network consists of the input layer, output layer, and hidden layer (topological structure is as shown in Fig. 1). The neural networks adopting the error back propagation training algorithm are called BP networks, whose learning process comprises both backward and forward propagation. In forward processing, the sample signals will gradually progress through each layer with the sigmoid function

The neural network cell i.e. neuron of each layer (input or hidden) only affects the status of the next neural cell. If the expected output signals cannot be obtained in the output layer, the weight values of each layer of the neural cells must be modified to keep the error minimum. Erroneous output signals will be backward from the source. Finally, the signal error will arrive in certain areas with repeated propagation thus the weights of the different layer neurons are modified.

The network is set at *n* layers and *yj n*, such that *yj n* indicates the output with *n* layers and *j* nodes. If *yj 0* is equal to *xj*, and *j* indicates the inputs. Let *Wij n* be the connection weight between *yi n-1* and *yj n*, then we get the threshold of *θj n* of *n* layers and *j* nodes. The neural network learning algorithm includes the following steps:

1. Initialize node connection weights with random values and set other parameters.

2. Read in the input signal vector and the desired output. The signals progress at the networks with the following formula



It starts to calculate the *j* nodes of each layer with the output *yj n* from the first layer through to the completion of calculation processing. F(*s*) represents one of the sigmoid transfer functions.

3. Compute the actual output via the calculations, working forward through the layers.

4. Compute the error. The error value of each node for the output layer is obtained from the different values between the real output and the required output (*Dj k* ) and is given by:



The error value of each node for the previous (hidden) layers depends on the backward error propagation of each layer *(n = n, n-1... 1)* and is given by:



5. Change the node connection weights by working backward from the output layer through the hidden layers which is done by the following formula:



Where *p* indicates the iterative times (epoch) of the layers. The constant *η* indicates the learning rate and *α* indicates the momentum constant and their values can be from 0 to 1.

**5.1.2 Dataset**

This paper has used the churn data set from the UCI Repository of Machine Learning Databases at the University of California, Irvine [18]. The churn dataset deals with cellular service p rovider’s customers and the data p ertinent to the voice calls they make. Customers have a choice of service providers, or companies providing them with cellular network services. When these customers change cellular service provider they are said to churn which results in a loss of revenue for the previous cellular service provider. The telecommunications company concerned

here has used all its databases like billing database, customer service database etc. and generated a list of pertinent records. The neural network is implemented on Clementine data mining software package from SPSS, Inc [19]. The data set contains 20 variables worth of information about 2,427 customers, along with an indication of whether or not that customer churned (left the company). The variables are as follows-

* State: categorical variable, for the 50 states and the district of Columbia
* Account length: integer-valued variable for how long account has been active
* Area code: categorical variable
* Phone number: essentially a surrogate key for customer identification
* International Plan: dichotomous categorical having yes or no value
* Voice Mail Plan: dichotomous categorical variable having yes or no value
* Number of voice mail messages: integer-valued variable
* Total day minutes: continuous variable for number of minutes customer has used the service during the day
* Total day calls: integer-valued variable

**5.1.3 Training and Testing of Neural Network**

When using neural networks to perform predictive modeling, the input layer contains all of the input fields or variables used to predict the outcome variable. The output layer contains an output field which is the target of the prediction. The input and output fields can be numeric or symbolic. In Clementine, symbolic fields are transformed into a numeric form (dummy or binary set encoding) before processing by the network. The hidden layer contains a number of neurons at which outputs from the previous layer combine. A network can have any number of hidden layers, although these are usually kept to a minimum to simplify the predictive model. All neurons in one layer of the network are connected to all neurons within the next layer while the neural network is learning the relationships between the data and results, it is said to be training. The Figure 2 provides details of implementation of neural network using Clementine 12.0.

Clementine provides two different classes of supervised neural networks, the Multi-Layer Perceptron (MLP) and the Radial Basis Function Network (RBFN). There are five different algorithms available within the Neural Net node of Clementine but this paper has used the widely applied Quick method. The Quick method use a feed-forward back-propagation network whose topology (number and configuration of nodes in the hidden layer) is based on the number and types of the input and output fields.



**Fig 5.1.3: The Implementation of Neural Network using Clementine 12.0**

**CHAPTER 6**

**PERFORMANCE ANALYSIS**

Now that the system has been set up and the results have been obtained, we can do a performance analysis to observe the limits, reliability and accuracy of the proposed system..

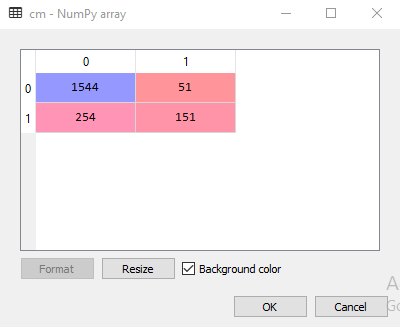


Figure 6.1. Acuracy

**CHAPTER 7**

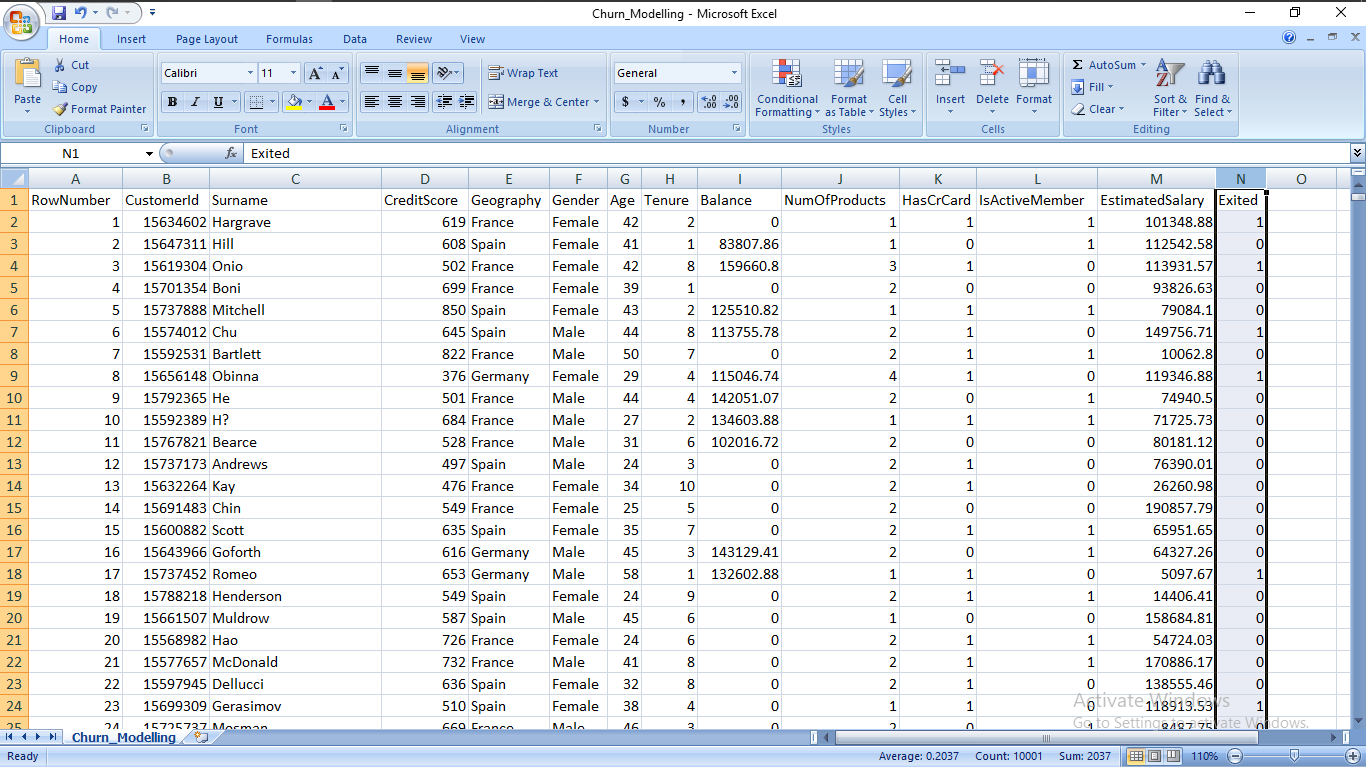
**RESULTS & DISCUSSIONS**

The analysis section of the generated model displays information about the neural network. Figure 3 depicts the final model summary. The predicted accuracy for this neural network is 92.35%, indicating the proportion of the test set correctly predicted. The input layer is made up of one neuron per numeric or flag type field. The input variable named state and phone number were removed from the model since these variables were used for identification only. This study has experimented with multiple hidden layers in the neural network, containing three to seven neurons but the best results were obtained having one hidden layer with three neurons. The output layer contains two neurons corresponding to the two values of the output field (churn being true or false).

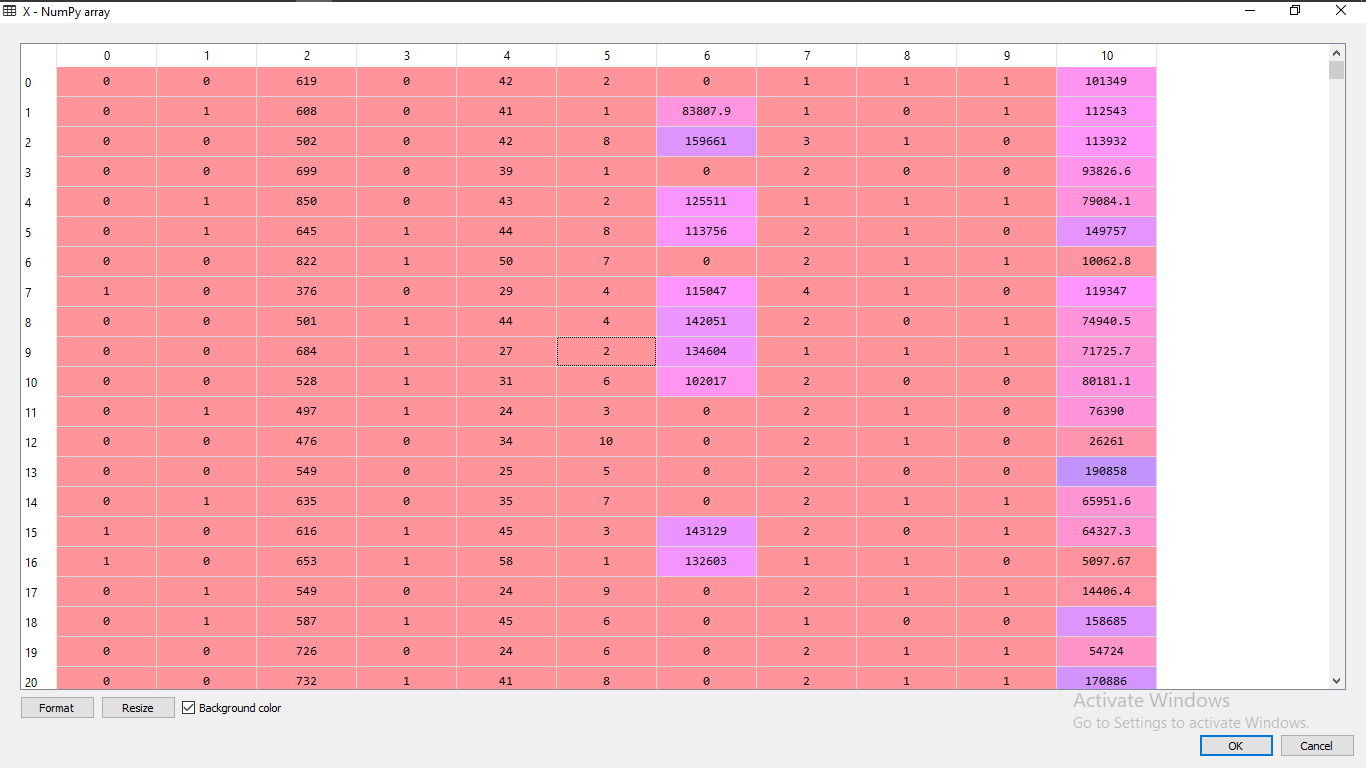
It represents relative importance of the input variables for doing sensitivity analysis of the generated model. The input fields are listed in descending order of relative importance. Importance values can range from 0.0 and 1.0, where 0.0 indicates unimportant and 1.0 indicates extremely important. In practice this figure rarely goes above 0.35. Here we see that Customer Service Calls is the most important field within this current network, and then followed by International Plan, Day Minutes and Day Charge. So we can conclude those customers who are frequently calling to customer service numbers, having international calling plan and spending much time in calling others during day time, may become potential churners. The generated Neural Net calculates two new fields, $N-Churn and $NC- Churn, for every record in the input data base. The first represents the predicted Churn (true or false) and the second a confidence value for the prediction. The latter is only appropriate for symbolic outputs and will be in the range of 0.0 to 1.0, with the more confident predictions having values closer to 1.0. When predicting a symbolic field, it is valuable to produce a data matrix of the predicted values ($N-Churn) and the actual values (Churn) in order to study how they compare.

**CHAPTER 8**

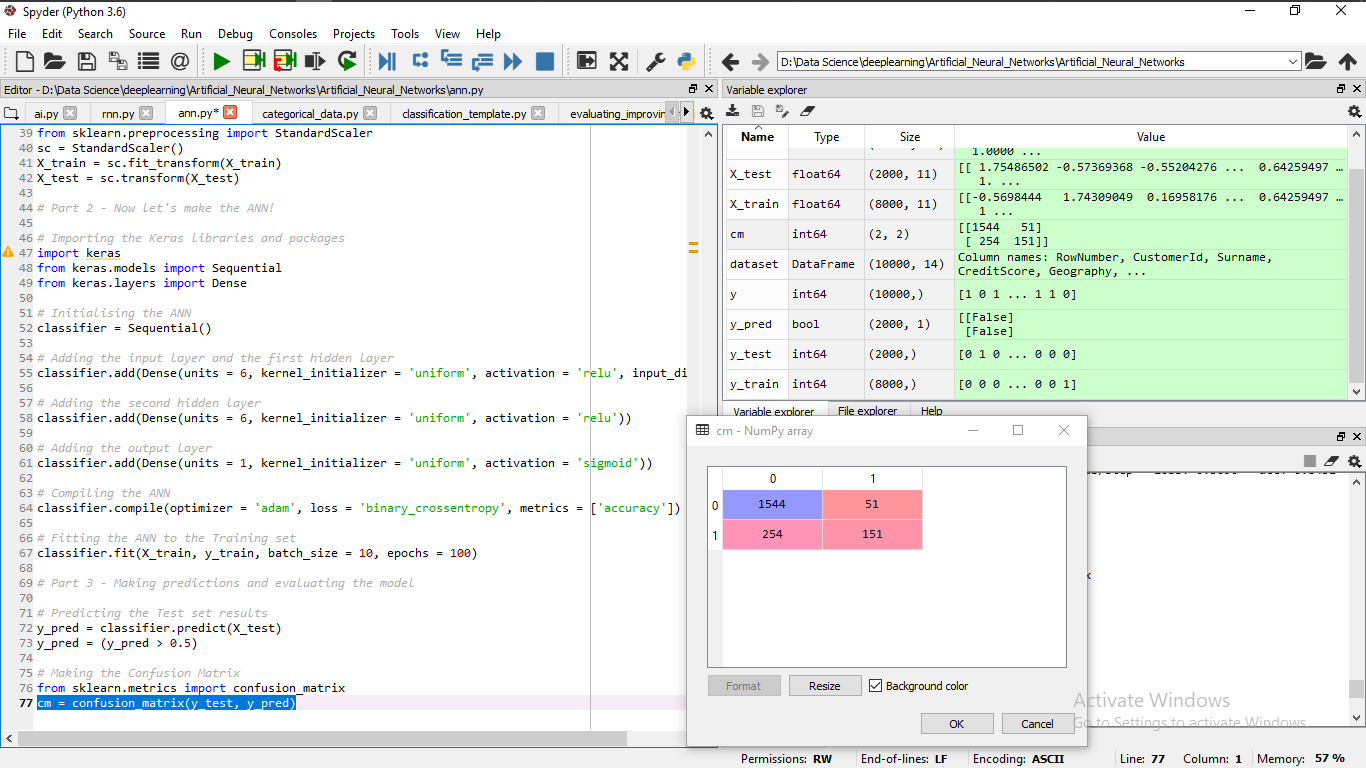
**SCREENSHOTS**



**Figure 8.1:Dataset**

****

**Figure 8.2 Processed Data**

****

**Figure 8.3 Result**

**Chapter- 9**

**Code**

# Artificial Neural Network

# Installing Theano

# pip install --upgrade --no-deps git+git://github.com/Theano/Theano.git

# Installing Tensorflow

# pip install tensorflow

# Installing Keras

# pip install --upgrade keras

# Part 1 - Data Preprocessing

# Importing the libraries

import numpy as np

import matplotlib.pyplot as plt

import pandas as pd

# Importing the dataset

dataset = pd.read\_csv('Churn\_Modelling.csv')

X = dataset.iloc[:, 3:13].values

y = dataset.iloc[:, 13].values

# Encoding categorical data

from sklearn.preprocessing import LabelEncoder, OneHotEncoder

labelencoder\_X\_1 = LabelEncoder()

X[:, 1] = labelencoder\_X\_1.fit\_transform(X[:, 1])

labelencoder\_X\_2 = LabelEncoder()

X[:, 2] = labelencoder\_X\_2.fit\_transform(X[:, 2])

onehotencoder = OneHotEncoder(categorical\_features = [1])

X = onehotencoder.fit\_transform(X).toarray()

X = X[:, 1:]

# Splitting the dataset into the Training set and Test set

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 0)

# Feature Scaling

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

X\_train = sc.fit\_transform(X\_train)

X\_test = sc.transform(X\_test)

# Part 2 - Now let's make the ANN!

# Importing the Keras libraries and packages

import keras

from keras.models import Sequential

from keras.layers import Dense

# Initialising the ANN

classifier = Sequential()

# Adding the input layer and the first hidden layer

classifier.add(Dense(units = 6, kernel\_initializer = 'uniform', activation = 'relu', input\_dim = 11))

# Adding the second hidden layer

classifier.add(Dense(units = 6, kernel\_initializer = 'uniform', activation = 'relu'))

# Adding the output layer

classifier.add(Dense(units = 1, kernel\_initializer = 'uniform', activation = 'sigmoid'))

# Compiling the ANN

classifier.compile(optimizer = 'adam', loss = 'binary\_crossentropy', metrics = ['accuracy'])

# Fitting the ANN to the Training set

classifier.fit(X\_train, y\_train, batch\_size = 10, epochs = 100)

# Part 3 - Making predictions and evaluating the model

# Predicting the Test set results

y\_pred = classifier.predict(X\_test)

y\_pred = (y\_pred > 0.5)

# Making the Confusion Matrix

from sklearn.metrics import confusion\_matrix

cm = confusion\_matrix(y\_test, y\_pred)

**CHAPTER 10**

**FUTURE ENHANCEMENT**

Churn prediction and management is crucial in liberalized cellular mobile telecom markets in developing countries. In order to be competitive in this market, cellular service providers have to be able to predict possible churners and take proactive actions to retain valuable loyal customers. Therefore, to build an effective and accurate customer churn prediction model, has become an important research problem for both academics and practitioners in recent years. This paper suggests that data mining techniques can be a promising solution for the customer churn management and we can establish an early-warning model for this non-steady-state customer system. The final model summary in this paper concludes that the model gives more than 92% overall accuracy for the prediction of the customer churn.

For future work, several issues can be considered. First, as the data pre-processing stage in data mining is a very important step for the final prediction model performance, the dimensionality reduction or feature selection step can be involved in addition to data reduction. Second, along with neural networks, other popular prediction techniques can be applied in combination, such as support vector machines, genetic algorithms, etc to develop hybrid models. Finally, the current methodology of churn prediction can be tested for other sectors like banking, insurance or air line and comparisons can be done for prediction accuracy.

**CHAPTER 11**

**CONCLUSION**

In order to be competitive in this market, banks have to be able to predict possible churners and take proactive actions to retain valuable loyal customers. Building an effective and accurate customer churn prediction model has become an important research problem for both academics and practitioners in recent years. Profiling enables a company to act in order to keep customers may leave (reducing churn or attrition), because it is usually far less expensive to keep a customer than to acquire a new one .

Neural network is a valuable forecast tool in financial economics due to the learning, generalization and nonlinear behaviour properties. It is powerful general-purpose software tool used for a number of data analysis tasks such as prediction, classification and clustering. Neural networks are used in finance such as portfolio management , credit rating and predicting bankruptcy , forecasting exchange rates , predicting stock values , inflation and cash forecasting and others in order to achieve a reliable decision-making process through scientific approaches. The ability of neural networks to discover nonlinear relationships in input data makes them ideal for modelling nonlinear dynamic systems such as banking industry.

The bank must operate on a long term customer strategy, young customers are recognized as being unprofitable in the early stage in lifecycle but will become profitable later on. In this paper we have shown that more and more young people use internet banking and that bank should offer different products/services which could be arranged without the client coming to the bank, such as savings that can be arranged and used only on the internet. It is necessary to develop new products that could be offered to such customers in order to keep them.

Cross-selling is one of the most important ways to increase the profitability of existing customers while increasing their loyalty. By selling additional products to customers we associate with them, thus increasing their loyalty (we have seen that more loyal customers are those who use more than two bank products). Analysing the data available we can determine what the next best offer for a particular client is. For example, bank could offer car insurance together with the car loan.

In this article a customer churn analysis on database of small Croatian bank was presented. The analysis focused on churn prediction based on only one method, Neural network. We could access other important information that could help banks to get competitive advantage by using other methods such as segmentation, decision trees, self-organizing maps. We wanted to show the simple usage of a complex method and to encourage others in similar research. Today, there are many very good software packages for data mining that do not require much pre-knowledge to use, and results can be very useful.

**REFERENCES**

[1] Sumathi, S. and Sivanandam, S.N.: *Introduction to Data Mining Principles*.

Studies in Computational Intelligence **29**(3), 1-20 2013,

[2] Ramageri B.M. and Desai, B.L.: *Role of data mining in retail sector.*

International Journal on Computer Science and Engineering, **5**(1), 47-50, 2013,

[3] Hung, S.Y.; Yen, D.C. and Wang, H.Y.: *Applying data mining to telecomm churn management.*

Expert Systems with Applications **31**(3), 515-524, 2006,

http://dx.doi.org/10.1016/j.eswa.2005.09.080,

[4] Baesens, B.; Viaene, S.; Van den Poel, D.; Vanthienen, J. and Dedene, G.: *Bayesian neural network learning for repeat purchase modeling in direct marketing*.

European Journal of Operational Research **138**(1), 191-211, 2002,

[5] Domingo, R.: *Applying data mining to banking.*

http://www.rtdonline.com, accessed 18th November 2015,

[6] Van den Poel, D. and Larivie‘re, B.: *Customer attrition analysis for financial services using proportional hazard models*.

European Journal of Operational Research **157**(1), 196-217, 2004,

http://dx.doi.org/10.1016/S0377-2217(03)00069-9,