PRODUCT AND CHANNEL PREDICTION FOR DIRECT MARKETING IN BANKING SECTOR

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

MERVE MITIK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

SEPTEMBER 2017

Approval of the thesis:

PRODUCT AND CHANNEL PREDICTION FOR DIRECT MARKETING IN BANKING SECTOR

submitted by **MERVE MITIK** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** ______________

Assoc. Prof. Dr. Pınar Karagöz
Supervisor, **Computer Engineering Department** ______________

Prof. Dr. İsmail Hakkı Toroslu
Co-supervisor, **Computer Engineering Department** ______________

Examining Committee Members:

Prof. Dr. Ahmet Coşar
Computer Engineering Department, METU ______________

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering Department, METU ______________

Prof. Dr. İsmail Hakkı Toroslu
Computer Engineering Department, METU ______________

Assist. Prof. Dr. Şeyda Ertekin
Computer Engineering Department, METU ______________

Assoc. Prof. Dr. Osman Abul
Computer Engineering Department, TOBB ETU ______________

Date: Sept 08, 2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MERVE MITIK

Signature :

ABSTRACT

PRODUCT AND CHANNEL PREDICTION FOR DIRECT MARKETING IN BANKING SECTOR

Mıtık, Merve
M.S., Department of Computer Engineering
Supervisor : Assoc. Prof. Dr. Pınar Karagöz
Co-Supervisor : Prof. Dr. İsmail Hakkı Toroslu

September 2017, 52 pages

Direct marketing is an advertisement method in which customers are directly informed for product offers through one-to-one communication channels. With the advancements in technology, customer databases of businesses began to grow well, therefore, detecting the needs of each customer and offering the optimal product becomes harder. Large customer dataset needs to be analyzed to make the best product offerings to the potential customers over the most proper channels and to increase the return rate of a marketing campaign. However, this goal is not very easy to accomplish, since, negative returns in the dataset usually outnumber the positive ones. Imbalanced datasets cause data mining algorithms reveal poor performance. This thesis studies the similar problem for bank product marketing. Two data mining solutions are proposed, partitioning based method and model based method for bank product and channel prediction. First proposed approach depends on unsupervised learning method, and uses clustering to predict product and channel for new customers. Second one presents a hybrid approach with unsupervised and supervised learning methods, which first constructs a classification model to detect if the customer is buyer or non-buyer and then clusters customers for product and communication channel offers. Experimental analysis on real life banking campaign dataset shows promising results.

Keywords: Direct Marketing, Bank Product Marketing, Supervised Learning, Unsupervised Learning, Customer Relationship Management

ÖZ

BANKACILIK SEKTÖRÜNDE DOĞRUDAN PAZARLAMA İÇİN ÜRÜN VE KANAL TAHMİNİ

Mıtık, Merve
Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Doç. Dr. Pınar Karagöz
Ortak Tez Yöneticisi : Prof. Dr. İsmail Hakkı Toroslu

Eylül 2017 , 52 sayfa

Doğrudan pazarlama, müşterilerin bire bir iletişim kanalları aracılığıyla ürün teklifleri konusunda bilgilendirildiği bir reklam metodudur. Teknolojideki ilerlemelerle birlikte, işletmelerin müşteri veritabanları iyice büyümeye başladı, bu nedenle, her müşterinin ihtiyacını tespit etmek ve en uygun ürünü sunmak zorlaştı. Potansiyel müşterilere en uygun kanallar üzerinden en iyi ürün tekliflerini yapmak ve bir pazarlama kampanyasının getiri oranını artırmak için geniş müşteri veri setinin analiz edilmesi gerekmektedir. Bununla birlikte, bu hedefin gerçekleştirilmesi kolay değildir, çünkü veri setindeki negatif dönüşler genelde pozitif olanlardan daha fazladır. Dengesiz veri setleri, veri madenciliği algoritmalarının düşük performans göstermesine neden olur. Bu tez, bankacılık ürün pazarlaması için benzer bir problemi araştırmaktadır. Bankacılık ürünü ve kanal tahmini için bölümleme tabanlı yöntem ve model tabanlı yöntem olmak üzere iki veri madenciliği çözümü önerilmiştir. İlk önerilen yaklaşım gözetimsiz öğrenme yöntemine dayanır ve yeni müşterilere yönelik ürün ve kanal tahmininde küme-

leme kullanır. İkincisi, önce müşterinin alıcı olup olmadığını tespit etmek için bir sınıflandırma modeli oluşturur ve daha sonra ürün ve iletişim kanalı teklifleri için müşterileri kümeleyen, gözetimsiz ve gözetimli öğrenme yöntemlerine sahip hibrid bir yaklaşım sunar. Gerçek bir bankacılık kampanyası veri seti üzerinde yapılan deneysel analiz umut verici sonuçlar göstermektedir.

Anahtar Kelimeler: Doğrudan Pazarlama, Bankacılık Ürün Pazarlaması, Gözetimli Öğrenme, Gözetimsiz Öğrenme, Müşteri İlişkileri Yönetimi

To the loving memory of my father...

ACKNOWLEDGMENTS

I would like to express my sincere gratitude and appreciation to my supervisor Assoc. Prof. Dr. Pınar Karagöz and my co-supervisor Prof. Dr. İsmail Hakkı Toroslu for their guidance, encouragement, motivation and continuous support throughout this study.

I would like to express my deepest gratitude to my dear husband, Ozan Korkmaz for always being with me in this long period.

I am deeply grateful to my mom for her love and support over all these years.

I also thank to my company MilSOFT Software Technologies for allowing me to continue my education and conduct my research while I am working there.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CRM	Customer Relationship Management
PCL	Product Channel Label
PCY	Product Channel Yes (Label)
SVM	Support Vector Machine
MLPNN	Multilayer Perceptron Neural Network
RBFNN	Radial Basis Function Neural Network
TAN	Tree Augmented Naïve Bayes
ROC	Receied Operating Characteristic
IVN	Interactive Voice Network
CC	Call Center
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

CHAPTER 1

INTRODUCTION

1.1 Motivation & Problem Definition

Marketing is defined as "total of activities involved in the transfer of goods from the producer or seller to the consumer or buyer, including advertising, shipping, storing, and selling" [12]. In simple terms, marketing is providing products from business to customer. There are some concepts that organizations should follow in order to perform effective marketing campaigns. They are called "marketing mix" [6]. It includes four items referred as 4Ps of marketing which are product, price, place and promotion [26]. Choosing the product to promote, determining the price of selected product, deciding the communication channel with the place of promotion and executing the promotion are four fundamental elements of marketing.

Two different techniques are used in marketing, which are mass and direct marketing [29]. The objective of mass marketing is reaching as many people as possible with a single shot without considering the specific needs of the audience. This is accomplished through broadcasting channels like television, radio, billboards and newspaper. On the other hand, direct marketing method is about informing people on specialized products and reaching them directly by one-to-one communication channels like phone calls, emails and text messages. Comparing these two methods, mass marketing is more expensive than direct marketing due to the high cost of commercials in public places and low ratio of people responding positively to campaigns.

With the advancements in technology and widespread usage of Internet, people are more aware of wide variety of products being offered and they pursue the best. Mass marketing does not help handling this situation. Businesses should tailor their products and try to give the customer what they need. This can be done with customer relationship management tools. CRM allows businesses to manage their relationship with customers and make more profitable campaigns. CRM systems generally have large customer information and it is infeasible to analyze the data and make reasonable suggestions for products to customers manually. Data mining techniques are very useful to extract hidden information, thus they can be utilized to group customers according to their characteristics and needs.

Bank product marketing is a process of promoting banking services and products to customers. It is one of the many sectors where CRM, therefore data mining techniques are very practical. In this thesis, I am concerned about the 3Ps of marketing which are product, place and promotion excluding the price. Main aim is to determine whether a customer is willing to buy a product proposed through a communication channel or not.

The problem under interest can be described as follows: Given a set of banking products P, and a set of communication channels C, for a given user u, predict whether the user u will accept any product offer, if s/he will accept, predict the product $p \in P$ and communication channel $c \in C$. I propose two solutions for this problem; first is partitioning based approach and second is model based approach which is a hybrid solution. In the partitioning method, a set of clusters are created with respect to the products, channels and sold or not sold labels, and then, using these clusters, predicting customer decision is aimed. In the model based method, models generated from the logs of product offer to customers are used to predict the right product and the channel if they are likely to accept the offer.

1.2 Contributions

Competitive environment and increasing volume of customer information force businesses to find solutions for optimizing direct marketing campaigns. Since data mining techniques are very effective in this topic, there are many approaches proposed in scope of direct marketing [5] [14] [21] [27] [28] [38] [39]. There are also some hybrid approaches that combines two or more data mining algorithms to solve problems due to imbalanced data distribution [2] [10] [20] [32] [36]. However, all of these works either use one of the data mining methods or they are concerned about one product to propose and one communication channel only.

Cohen and Nobibon et al. proposed solutions considering multiple product and multiple channel constraints [9] [11] [31]. However, these approaches are not data mining based but they offered integer programming solutions. While Cohen used the data from marketing department of a large international commercial bank, Nobibon et al. did not use real life data but generated instead.

My focus in this thesis is providing a solution for the problem of selecting product and communication channel for potential customers. To achieve my goal,

- a real life Turkish bank dataset is processed,
- a baseline technique is implemented to compare proposed methods,
- two new methods are proposed for predicting product and communication channel for potential customers among multiple products and communication channels.

1.3 Organization of the Thesis

This thesis is organized as follows:

Chapter 2- Literature Survey summarizes related work conducted about direct marketing.

Chapter 3- Background gives background information about the topics discussed in this thesis.

Chapter 4- Proposed Methods for Bank Product and Channel Prediction describes the dataset properties and explains the details of the methods that are proposed.

Chapter 5- Experimental Analysis Results and Discussions shows the results of the experiments conducted on proposed methods and discusses the results.

Chapter 6- Conclusion and Future Works is where summary about thesis is given. Opinions about future work are also discussed in this chapter.

CHAPTER 2

LITERATURE SURVEY

Marketing has a crucial role in banking industry like in any other business sector. It is important to choose right marketing strategy to promote products or services to potential customers. Increasing the return rate of the marketing while decreasing the cost of it is the main approach. There are mainly two marketing strategies to consider, mass marketing and direct marketing [29].

2.1 Mass Marketing

Mass marketing is a strategy that uses broadcasting media as communication channels to promote products or services. These broadcasting channels can be television, radio, billboard or newspaper. The advertisement is published without considering the different characteristics and needs of the audience. Main aim in this strategy is to reach as many people as possible to increase the return rate of the campaign. Promoted products are not customized for specific customers but they are generic to appeal standard people. There was a time after World War II when the advertisements of electronic goods in television were influential on promoting these products [19]. However, mass marketing is not an effective way of selling as it is used to be. With the increasing types of products and services, large number of companies and many similar advertisements make mass marketing lose its effect on people. They see various kinds of products advertised and desire the best [35].

Return rate of campaigns with mass advertisements is very low although the cost for it is high. Ling and Li compared the results of mass mailing and direct mailing campaigns [24]. Results show that, even though the number of customers mailed with mass mailing is five times more than direct mailing, net profit from the promotion with direct mailing is much more than mass mailing. Beside the uneconomic side of mass marketing, it also disturbs most people who do not want any product by communicating with them redundantly. Therefore, because the profit/cost rate of mass marketing is very low and unnecessary contacts with customers annoy them, bank marketing systems are trying to shift from mass marketing to direct marketing strategy [14].

2.2 Direct Marketing

Direct marketing is a strategy that potential customers are contacted directly for promotion of specific products. Banks try to promote their credit card, loan, deposit account type of banking products through a communication channel like SMS, e-mail and phone call. However, banking management systems have large customer information and it is costly to make telephone calls, send messages or mails to contact all the customers directly since the return rate is generally very low [4]. Moreover, it is also infeasible to analyze the big data manually and make reasonable suggestions for products to customers. Therefore, data mining techniques can be utilized to group customers according to their characteristics or needs and offer specific products to them.

General overview of direct marketing is shown in Figure 2.1. As it is seen in the figure, direct marketing processing is divided into two main phases as dataset generation and data mining. Data collection is the first step of dataset generation. Bank databases already have massive customer information about former campaign offerings. If a new product to be offered, then a pilot promotion can be made and the results of it can be used as existing data [24]. In the data preprocessing part, the collected data is prepared for data mining step according to the implementation. Attribute selection can be applied to find the attributes that provides more information. Data cleaning can be applied for the missing or duplicate records. Ultimately, dataset gets ready for the data mining phase. Then, using the information in the dataset, patterns for likely buyers

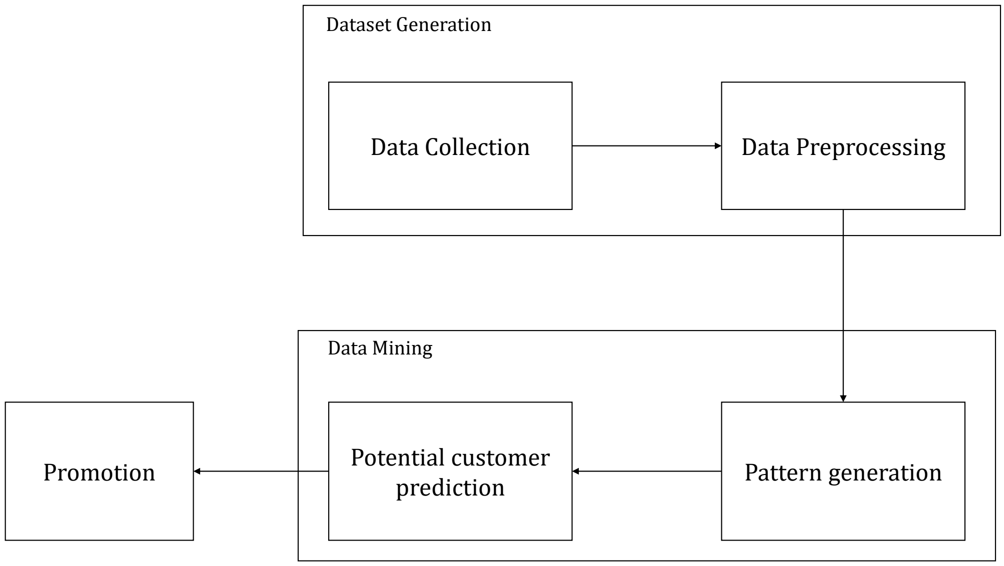

Figure 2.1: General overview of direct marketing

and non-buyers are generated using data mining techniques like clustering and classification. After that, a new customer can be predicted whether s/he will respond positively to the offer or not by using the generated patterns.

2.2.1 Applications of Direct Marketing

Cohen declares the ultimate motto of bank marketing as "delivering the right product to the right customer at the right time" [9]. However, selecting one of many products to offer to a specific customer using one of the several contact channels considering the business constraints is very complicated. Budget, channel capacity, number of products in a campaign are mentioned as some of the business constraints in the paper. The solution of Cohen does not include data mining approaches, however it provides broad information about bank direct marketing. Cohen converts the database marketing problem into optimization problem and uses mathematical programming models to maximize profit with satisfying the business constraints. He does the optimization by choosing the customer and the product with highest profit instead of the greedy approach that chooses the customer with the highest expected return.

Another integer programming approach is proposed by Nobibon et al. [31]. They consider the optimization for selection of target customers to offer one or more products is strongly NP-hard and they propose a branch and price algorithm and eight heuristics for optimal solution. Some business constraints are also handled in this paper as Cohen but also two new constraints are added [9].

2.2.2 Data Mining Applications of Direct Marketing

In the domain of direct marketing, the biggest issue is to market the right product to the right customer through the right channel. Companies and industries keep large databases of their customers with their characteristic information. This information mostly contain personal information such as age, marital status, education, address and specific information about sector like the previous products purchased, last contact time, last contact duration. It is impossible to generate patterns manually from this huge data. Therefore, data mining techniques are very helpful for making sense of the hidden data and discarding the unnecessary data which does not provide any information.

The most used data mining technique applied in direct marketing is classification for predicting behaviors of customers [30]. There are many different classification techniques applied in scope of data mining for direct marketing domain. Large majority of the studies use the bank direct marketing dataset provided by University of California at Irvine (UCI) Machine Learning Repository for evaluation [23]. The dataset is about the campaigns of a Portuguese bank and the records represent if the customers subscribed to bank term deposit or not. There is only one product and the customers are contacted through a single communication channel which is phone call. This dataset was collected and preprocessed by Moro et al. and Naïve Bayes, decision tree and SVM methods were applied [27]. Results show that SVM has the best performance among three methods applied. Wiaseng applied J48-graft, LAD tree, radial basis function network and SVM over this dataset and the results show that SVM achieved the best results in terms of accuracy, sensitivity and specificity [39]. MLPNN, Naïve Bayes, tree augmented Naïve Bayes, Logistic Regression, C4.5 and C5.0 are other classification algorithms used on the same topic and the same dataset [5] [14] [21] [38]. Nachev extended the methodology in the previous studies by eliminating the dummy attributes and testing 300 times to reduce the effect of lucky set composition using the classification algorithms Neural Network, Logistic Regression, Naïve Bayes and Linear and Quadratic Discriminant Analysis [28].

Although classification is widely used in direct marketing applications, there are some problems of the classification algorithms due to the characteristics of databases related to direct marketing campaigns [24]. First and the most significant issue is the imbalanced class distribution. Positive return rates for this kind of campaigns are usually very low which results in %1-2 positive and %99-98 negative responded customers. Classification algorithms do not perform well on this type of datasets. To increase the accuracy of the prediction, they tend to classify all instances as negative and simply ignore the positive responded minority. The second problem emerges due to the class imbalance. Since the classification algorithms detect negative responses mostly, the accuracy of predictions is very high. Actually, its highness is not because of the good prediction of the algorithm but it is because of the majority of negative responses. Therefore, accuracy is not always a proper evaluation criteria for classification algorithms.

There exist several solutions to the problems that emerge due to imbalanced data distribution. One of them is using different performance measuring metrics instead of accuracy. Sensitivity and specificity are proper choices along with accuracy [5] [14] [39]. Lift is another approach preferred for performance measurement [24] [27] [38]. In addition to all these metrics, area under the receiver operating characteristic is one of the most widely used [5] [13] [21] [28] [29].

Re-sampling the dataset is the other solution for imbalanced datasets [13]. Oversampling the minority class or undersampling the majority class is the mostly used two techniques. Oversampling the minority class includes increasing the number of the instances with low ratio. For instance, since the positive responded customers are usually very low in direct marketing examples, they are copied to increase their ratio. Undersampling the majority class means decreasing the number of instances in the class which outnumbers the other class by removing random instances. Shih et al. used 1 to 1 oversampling method to overcome the imbalanced dataset problem [37] while Ling and Li combined oversampling with undersampling [24]. There are some works applied SMOTE technique [1] [22] which is an oversampling technique used for balancing that creates synthetic positive instances by taking into consideration of nearest neighbours [8]. Li et al. applied four different sampling techniques and observed that the performance is mostly dataset dependent [22].

Since oversampling and undersampling methods have some drawbacks as creating non-existent data and removing existent data, ensemble methods are proposed as another technique for coping with the problems of direct marketing applications. Ensemble methods are generated by combining several supervised or unsupervised data mining techniques. Bagging and boosting are two well known and much used ensemble methods which both include multiple classifiers [7] [15]. In bagging method, random subsets are generated by resampling and classifiers are trained on these subsets. On the other hand, boosting trains the classifiers iteratively. Pan et al. applied bagging and boosting techniques to Portuguese bank dataset and results show that the best performing classifier in terms of ROC curve is bagged neural network model [32]. Instead of using multiple classifiers, mixing clustering and classification methods is another approach in hybrid techniques. Santos et al. proposed a method that first clusters the dataset into 25 partitions by Self Organizing Maps and then applies C5.0 decision tree to all the clusters [36]. Another ensemble approach was proposed which includes k-means clustering and bagged neural networks and it was compared with traditional subsampling approach [10]. Partitioning the data with k-means and then applying the bagged neural network gave better accuracy, sensitivity and specificity than subsampling the dataset and applying bagged neural network alone. Kang et al. proposed a novel approach, CUE, based on clustering, undersampling and ensemble using k-means clustering and different classifiers: logistic regression, multi-layer perceptron neural network, k-nearest neighbor classification, and SVM [20]. They generated multiple training sets by using all the positive responded instances and adding the undersampled negative responded instances. Result are promising in terms of true positive and true negative ratios. A very similar solution was proposed for bank direct marketing [2]. The bank marketing dataset from UCI machine learning repository was used in this ensemble approach. SVM, MLPNN, RBFNN, decision tree and logistic regression classification algorithms were used in this solution. Results were compared with using single classification algorithms and using these classification algorithms in ensemble approach. Results show that ensemble method increases both accuracy and true positive prediction rate.

CHAPTER 3

BACKGROUND

Two important data mining techniques namely, clustering and classification are used in development of the algorithms proposed in this thesis.

Clustering is an unsupervised learning method. Its purpose is to divide dataset into partitions that data in each partition become more similar than others. Similarity or dissimilarity of data is measured by calculating distances. Minkowski distance, Manhattan distance, Euclidean distance are some distance functions that are used in finding the similar clusters.

Classification is a supervised learning method. It is a two step process including learning and prediction. In the learning step, a model is constructed using training dataset and its class labels. Rules are generated based on the characteristics of the dataset to predict the class label. In the prediction step, test data is used to calculate accuracy of the generated model.

In this thesis, for clustering, k-means clustering algorithm is used and for classification, C4.5, Naïve Bayes and Tree Augmented Naïve Bayes classification algorithms are used. WEKA Data Mining Software is used in implementation of proposed algorithms [17].

3.1 K-Means Clustering

The objective is partitioning the dataset with N objects to k clusters so that sum of squares in each cluster is minimized. K-means clustering partitions data to k clusters where each datum belongs to the closest cluster. This algorithm requires a k value which is not known and best value is found experimentally for different datasets. K-means clustering is the most basic therefore the most popular clustering technique among the others [25]. It is selected as number 2 data mining algorithm in the Top 10 Algorithms in Data Mining [40].

The main idea is to define k centroids beforehand for each cluster and assign each datum to the cluster with the minimum distance to the centroid. It is a good idea to choose these centroids as far from each other as possible since the selection of centroids results in different cluster formations. The centroids of the clusters are re-calculated by finding the mean of the assigned data. After, distances of each datum to each centroid are calculated and they are assigned to the closest clusters again. This is done until no change occurs in the locations of the centroids. As a result, N data are assigned to k clusters with k centroids, in which the data are more similar than the data in other clusters.

WEKA uses k-means++ instead of original k-means algorithm which is a minor extension over the original one. Since selecting initial centroids is random in k-means algorithm, it is a bit of luck to get right clusters all the time. K-means++ is proposed to overcome this situation [3]. Instead of choosing all initial centroids arbitrarily, it choses only one initial centroid and then evaluates the other *k-1* centroids by weighting the sum of squares of each point to closest centroid. The results show that k-means++ is better than k-means in terms of speed and accuracy [3].

3.2 C4.5 Decision Tree Classification

C4.5 decision tree algorithm is developed by Quinlan which is an extension over his ID3 algorithm [33] [34]. ID3 builds a decision tree from top to bottom using the information gain (entropy) of the training instances. Then, it uses this tree to predict unknown data. However, there are some drawbacks of this algorithm. Firstly, it only handles the nominal attributes, but does not handle numeric ones. Besides, it also cannot handle the missing values in the dataset. ID3 is a greedy algorithm that it finds the best decision by information gain, and does not look for alternative choices, it does not have pruning.

C4.5 is an improvement over ID3 algorithm. Different from ID3, C4.5 handles both numeric and nominal attributes. It also handles the missing values in the dataset. It uses gain ratio to select the best attributes for decision making. Moreover, from bottom to top it goes through the created tree and removes the branches that are unnecessary, that is it uses error based pruning. It is selected as number 1 data mining algorithm in the Top 10 Algorithms in Data Mining [40].

Applying top-down recursive divide and conquer manner, root dataset is divided into leaf subsets repeatedly until the stopping criterion is satisfied to improve the prediction accuracy. Root and internal nodes are known as decision nodes. Leaf node is the result of purification process. In resultant tree, leaf nodes represent class label and internal nodes show conjunctions that lead to leaf nodes with class labels. At each node of the tree, algorithm selects the attribute to find information gain and splits itself into minor subsets by using the attribute with highest gain ratio to provide more homogeneity. All attributes are expected to be categorical. If there exists any continuous attributes, then they are discretized in preprocessing time. Lastly, from bottom to top pruning process is applied to the tree to avoid overfitting.

3.3 Naïve Bayes Classification

Naïve Bayes is a statistical classifier based on conditional probabilities using Bayes' Theorem. Bayes' Theorem is for calculating the probability of event A will occur knowing probability of event B occurred. Posterior probability of event A is calculated as,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{3.1}$$

This probability is calculated for each class of the dataset. For example, in this thesis' scope, probability of new customer accepting the campaign offer is calculated for all attributes and their values. The probability of customer accepting a product offer via a channel with education level high school is known from the training dataset. Therefore, a new customer's probability of accepting the product offer via the channel can be calculated. The highest calculated probability with all attributes and their values is chosen as the result of prediction.

Naïve Bayes is based on strong independence assumptions between the attributes. It assumes that each attribute is independent of the other attributes [18]. The relationship graph of the dataset used in this thesis is shown in Figure 3.1. Only the class label is sold is parent of each attribute and no other relation exists between attributes. Although it is naive and easy to implement and requires small set of data to train, it still outperforms other classification algorithms most of the time. However, as it assumes probabilistic independence of attributes this method is clearly unrealistic in real life examples in this case the bank product marketing problem. It is not very clever to ignore the relationship between the attributes like age, education, asset and loan.

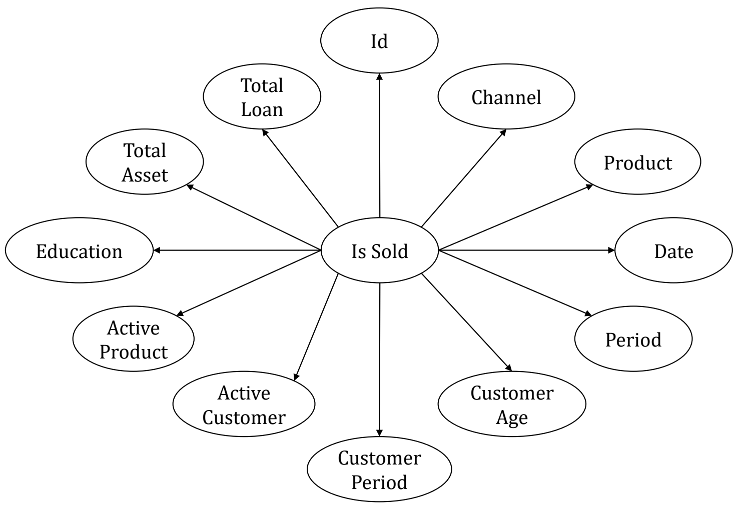

Figure 3.1: Naïve Bayes relationship graph of dataset

3.4 Tree Augmented Naïve Bayes Classification

The independence assumption of attributes in Naïve Bayes Classification does not fit the real life problems always. Tree augmented Naïve Bayes is an extension to the original Naïve Bayes classification algorithm. Unlike Naïve Bayes, in TAN the attributes are not all independent but there are edges between them. Thus, TAN is more life like structure than Naïve Bayes. Each attribute has at most one parent attribute along with class attribute, however class attribute has no parent. It is important that the edges between attributes do not result in a loop [16].

Dependence of the attributes of the dataset used in this thesis is shown in Figure 3.2. As in Naïve Bayes, class label is sold is parent of each attribute while date attribute is also parent of some other attributes. For example, the edge from active customer to active product means that, effect of attribute active product to class label also depends on attribute active customer.

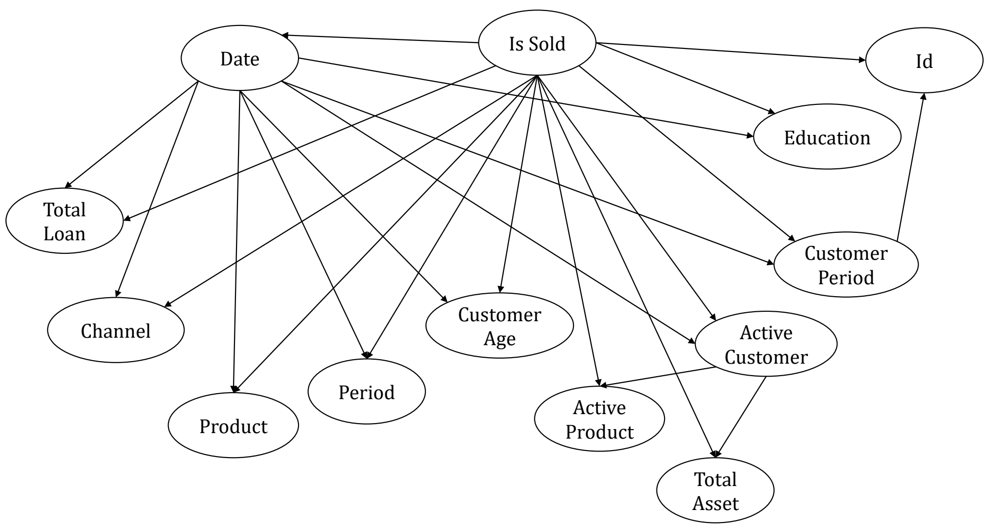

Figure 3.2: Tree Augmented Naïve Bayes relationship graph of dataset

CHAPTER 4

PROPOSED METHODS FOR BANK PRODUCT AND CHANNEL PREDICTION

In this chapter, firstly the dataset used in this thesis is introduced and the attributes are explained in detail. Then, a baseline technique and proposed methods are described.

4.1 Dataset

The dataset related to direct marketing campaigns used in this study is provided by an anonymous Turkish Bank. There are 81915 customer records with 13 attributes. Each entry in the dataset represents whether the customer with the specified id and some other personal information accepted the product offer via the offered channel or not. Table 4.1 shows the list of attributes and their domain values.

The attributes in the dataset are as follows:

- *Id*; unique identification number of the customer,
- *Date*; date of the contact,
- *Period*; period of the year of contact,
- *Customer Age*; age of the customer,
- *Customer Period*; number of weeks that the customer works with the bank,
- *Active Customer*; activeness of the customer,

- *Active Product*; number of active products the customer have,
- *Total Asset*; total asset of customer,
- *Total Loan*; total loan of customer,
- *Channel*; the communication channel used to contact with the customer,
- *Product*; offered product to the customer,
- *Education*; education level of customer,
- *IsSold*; response of the customer to the campaign.

Table 4.1: Attributes of Turkish Bank Dataset

ID	Attribute Name	Attribute Type	Domain Values
1	Id	Numeric	[6477, 60051815]
2	Channel	Nominal	{IVN, CC, EMAIL, SMS}
3	Product	Nominal	{Credit Card, Overdraft Account, Loan, Deposit Account}
4	Date	Numeric	[2015.01.13, 2015.05.31]
5	Period	Numeric	[120, 124]
6	Customer Age	Numeric	[16, 90]
7	Customer Period	Numeric	[1, 1384]
8	Active Customer	Nominal	{0, 1}
9	Active Product	Numeric	[0, 17]
10	Education	Nominal	{Bachelor's Degree, Academy, Secondary School, Primary School, Master's Degree, High School, Uneducated, Doctoral Degree, Null}
11	Total Asset	Numeric	[0, 569667.025]
12	Total Loan	Numeric	[0, 450743.725]
13	IsSold	Nominal	{Yes, No}

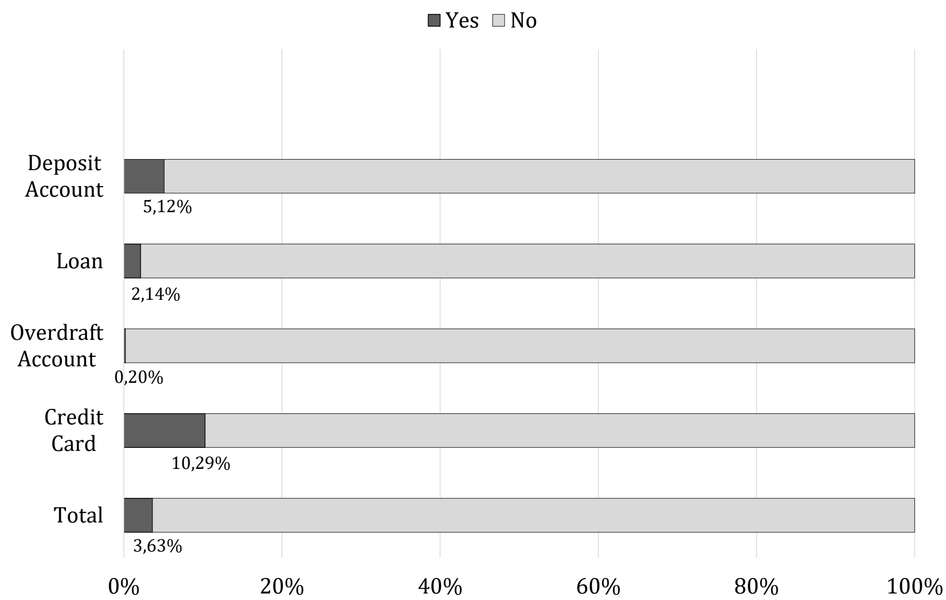

Figure 4.1: Product vs IsSold quantity

There are 2975 positive (*Yes*) and 78940 negative (*No*) instances according to the class label *IsSold* attribute. Positive responded records are 3.63% of the total data. The ratio of positive and negative responses for each product and channel are shown in Figure 4.1 and Figure 4.2 respectively. The figures show how negative responses outnumber the positive responses.

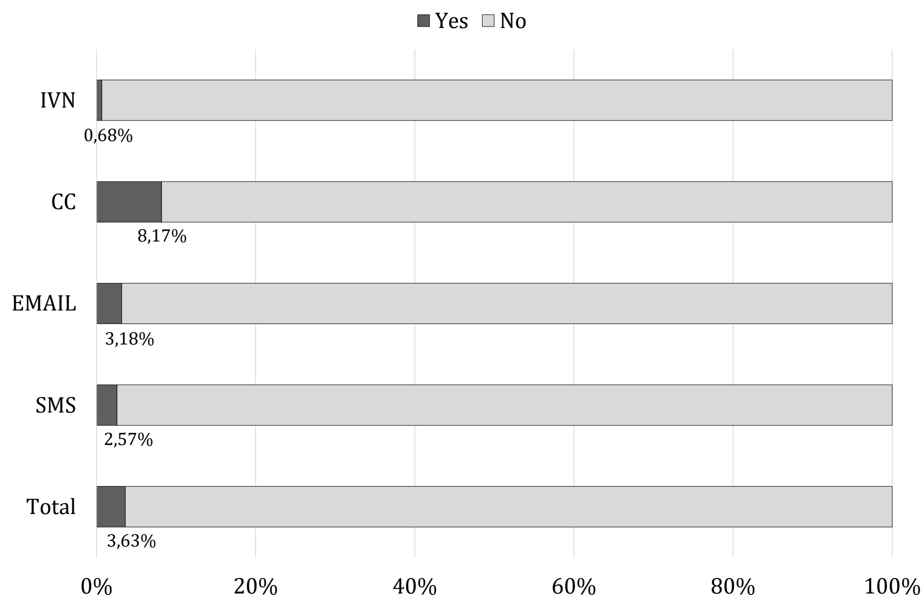

Figure 4.2: Channel vs IsSold quantity

4.1.1 Data Preprocessing

The dataset file given from a Turkish Bank was in .txt format. However, WEKA needs .arff file format in order to process data. The dataset was converted from .txt to .arff file format.

There are some faulty records and missing attribute values, so some preprocessing operations on the dataset were applied. First of all, out of 81939 entries in the dataset, 24 of them were removed because of being duplicate. Therefore, the dataset of 81915 records is used in this thesis.

Then, the dataset had some blank values in the three attributes namely *Education*, *Total Asset* and *Total Loan*. In *Education* attribute 33202 values, in *Total Asset* attribute 7604 values and in *Total Loan* attribute 7604 values were empty. These values were replaced with *Null* values. Since *Education* is a nominal attribute, another nominal attribute *Null* for missing values was added. However, *Total Asset* and *Total Loan* are numeric attributes, their *Null* values were assigned as 0.

Also, in some of the records *Product* and *Channel* attributes were mixed. For example, in *Product* attribute of one record there was *CC* and in its *Channel* attribute there was *Credit Card*. This values were exchanged with each other.

Moreover, attribute selection was made to find the most effective attributes on the dataset.

Finally, dataset is divided into two parts as training and test data. 90% of the dataset is used as training purposes and 10% of the dataset is used for testing purposes in the experiments.

4.2 Baseline Technique

Baseline method is a simple prediction model developed to compare proposed methods with a similar base method without clustering. The idea behind the baseline method is checking the new customer's similarity with the positive responded customers and negative ones. This similarity comparison is done with Euclidean distance. A cluster is said to be closest if the distance with the new customer and that cluster centroid is smaller than all other cluster centroids. If the new customer is closer to the positive ones then a product to offer with a proper channel is tried to be found. By partitioning the *Yes* labeled instances with same channels and same products as 4 partitions for products and 4 partitions for channels, closest product and channel clusters to the customer are found. Then, these closest product and closest channel are assigned to the customer. However, if the new customer is more inclined to *No* responded customers than *Yes* responded ones, then the customer label can be predicted as *No* and there is no need to try to find a suitable product and channel. For accuracy and sensitivity evaluation, the customer's real product, channel and label are checked if they match with the predictions. Experimental results of baseline technique are shown in Table 5.2.

4.3 Partitioning Based Bank Product and Channel Prediction Technique

This method is an enhanced version of the baseline method with the addition of clustering mechanism. Instead of partitioning the dataset for each channel and each product separately as in baseline method, the dataset is divided for each product, channel and label combinations (PCL). Detailed drawing of this division is shown in Figure 4.3. By doing this, the aim is to group customers who buy a specific product with a specific channel and checking the similarity of a new customer with this groupings separately.

Since the dataset contains 4 product (Credit Card, Overdraft Account, Loan, Deposit Account), 4 channel (IVN, CC, EMAIL, SMS) and 2 label (Yes, No) values, in this method 32 PCLs are formed. For instance, one of the PCLs includes all the customers with product *Credit Card*, channel *EMAIL* and label *Yes*, another PCL includes all customers with product *Overdraft Account*, channel *SMS* and label *No*. Then, each PCL is clustered to make partitions more similar inside. For clustering k-means algorithm is used and optimal value of k

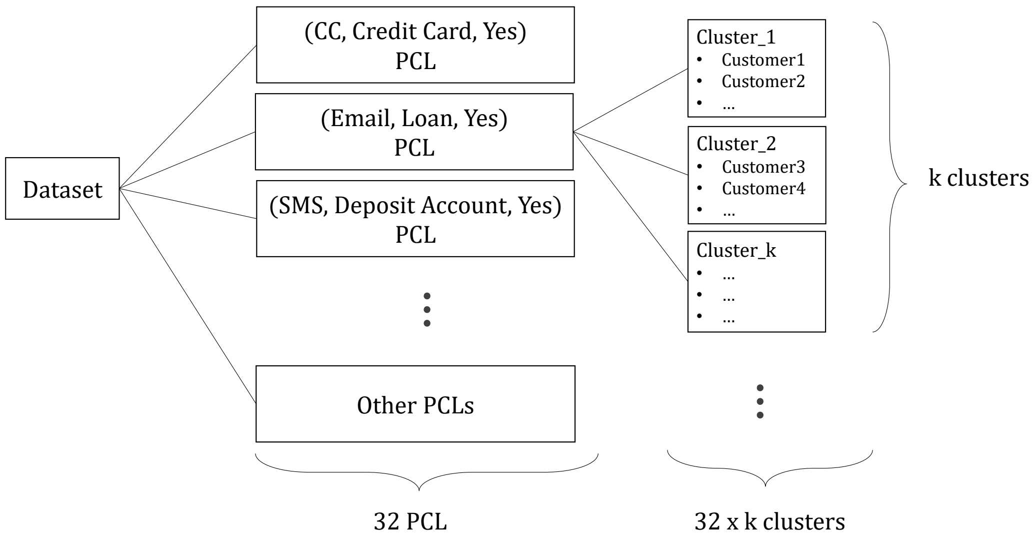

Figure 4.3: Partitioning based technique PCL formation

is found experimentally. To assign a product, a channel and a label for each customer, the nearest *Yes* and *No* clusters to the customer are found. The nearest clusters are found by comparing the customer's data with the centroids of the clusters. If the distance from a new customer to a *No* cluster is smaller than the distance to a *Yes* cluster, then the label of his/her is decided to be *No* and no further processing is applied. On the other hand, if a *Yes* cluster is closer, then the label of the customer is decided to be *Yes* and that cluster's product and channel are assigned to him/her. The algorithm of this technique is shown in Algorithm 1 and experimental results are discussed in Section 5.2.

Four variations of this algorithm are implemented to find the best accuracy, true positive and true negative ratios. These four variations are as follows:

- *Attribute Selection Approach* is explained in Section 4.3.1 and it is about using different attributes in Euclidean distance function for finding closest clusters,
- *Heuristic Approach* is about altering the rule used to find class label and is explained in Section 4.3.2,
- *Neighbor Approach* is declared in Section 4.3.3 and is not just looking one

Yes and one *No* clusters, but finding multiple closest clusters and averaging them,

- *Multiple Offering Approach* is offering multiple products through multiple channels to a customer which is described in Section 4.3.4.

Algorithm 1 Partitioning Based Bank Product and Channel Prediction Technique's Algorithm

procedure CONSTRUCTMODEL($trainingData$, n)
  partition the $trainingData$ into same PCL instances
  for each PCL instances **do**
    construct n clusters for each P_i, C_j, L^k as $P_iC_jL^k$
procedure PREDICTINSTANCE($data$, $P_{(*)}C_{(*)}L^{(*)}$)
  $P_iC_jY^{(k)}$ = closestCentroid($P_{(*)}C_{(*)}L_Y^{(*)}$, $data$)
  $P_iC_jN^{(k)}$ = closestCentroid($P_{(*)}C_{(*)}L_N^{(*)}$, $data$)
  if distance($data$,$P_iC_jY^{(k)}$) $<$ distance($data$,$P_iC_jN^{(k)}$) **then**
    $data$.Label = *Yes*
    $data$.Product = P_i
    $data$.Channel = C_j
  else
    $data$.Label = *No*

4.3.1 Attribute Selection Approach

Finding closest clusters includes using Euclidean distance function to find the minimum distance. Figure 4.4 shows a simple explanation of finding closest cluster. The closest distance to a *No* cluster is *d1* and the closest distance to a *Yes* cluster is *d3*. Since *d3* $<$ *d1*, the cluster with distance *d3* is chosen as the closest cluster and new customer is decided to be offered *Credit Card* through *CC* and expected to answer *Yes*.

The Euclidean function takes dataset attributes as parameter to be used in distance calculation. It means that the distance is calculated by using only the attributes given as parameter. Since channel(2), product(3) and isSold(13)

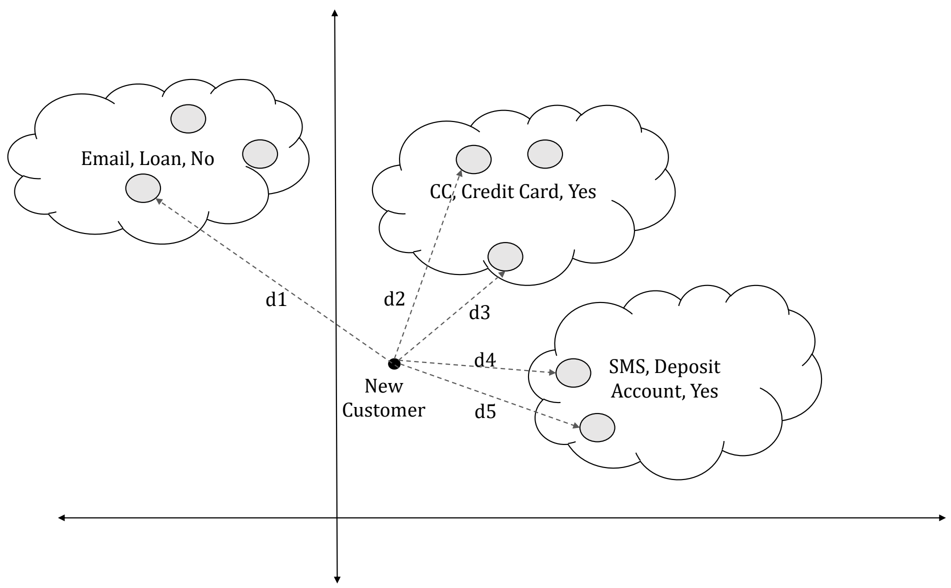

Figure 4.4: Finding closest cluster centroid

attributes of the new customer are tried to be found, the attributes with ids 2, 3 and 13 are not added to calculation. The results of experiments with the whole attributes except channel(2), product(3) and isSold(13) are shown in Table 5.3.

Except trying all attributes, to see the effect of attribute selection in accuracy, true positive and true negative ratios, the aim is to select different attributes for similarity calculation. Information gain of the attributes to the class label are calculated and Table 4.2 shows the information gain ratios of the attributes in decreasing order. The table means that contribution of date attribute to class label is highest and it is lowest in active customer attribute.

By excluding channel(2), product(3) and isSold(13) attributes, the results of selecting different attributes are discussed in Section 5.2.1.

4.3.2 Heuristic Approach

After examining results of the experiments of partitioning based technique, it is found out that there are some instances which are a bit closer to a *No* centroid than a *Yes* centroid but their real label is *Yes*. However, since the rule for deciding label is $distanceYES < distanceNO$, they are decided as *No* and no offer

Table 4.2: Ranked attributes by information gain

Attribute ID	Attribute Name	Information Gain
4	Date	0.031812
7	Customer Period	0.018425
1	Id	0.015250
10	Education	0.008634
12	Total Loan	0.002817
5	Period	0.001817
11	Total Asset	0.001659
6	Customer Age	0.001514
9	Active Product	0.001035
8	Active Customer	0.000577

is made. Some distance values of false negatives, which are actually positive but falsely classified as negative, to closest *Yes* and closest *No* clusters are shown in Table 4.3. As it can be seen from the table the values are very close to each other however, the customer is decided as *No*.

Table 4.3: Samples of distance values of false negatives

DistanceYes	DistanceNo
1.03442765	1.03417839
1.08960464	1.08871269
1.03933533	1.03763921
1.02150705	1.01916096
0.28502250	0.28240719
1.04201695	1.03931048
1.04733341	1.04455979

In the partitioning based technique the rule for deciding class label is simple which is:

$$\begin{aligned} &\textit{if distance(ClosestYesCentroid, NewCustomer)} < \\ &\quad \textit{distance(ClosestNoCentroid, NewCustomer)} \\ &\quad\quad \textit{then LABEL=YES else LABEL=NO.} \end{aligned} \tag{4.1}$$

However, in this approach to overcome the problem of false negatives that are very close to a *Yes* cluster centroid, different coefficients to distance functions are added to see the change in accuracy and sensitivity compared to the partitioning based method. By using different coefficients the aim is to expand the threshold of *Yes* values to predict more positive instances. The generic rule applied in this method is,

$$\begin{aligned} &\textit{if } (\boldsymbol{k} \times \textit{distance(ClosestYesCentroid, NewCustomer)}) < \\ &\quad (\boldsymbol{t} \times \textit{distance(ClosestNoCentroid, NewCustomer)}) \\ &\quad\quad \textit{then LABEL=YES else LABEL=NO}, \end{aligned} \tag{4.2}$$

in which $\boldsymbol{k}$ and $\boldsymbol{t}$ coefficients are found experimentally. The experiments made on this approach are shown and the results are discussed in Section 5.2.2.

4.3.3 Neighbor Approach

In this method, neighbor effect is tried to increase the accuracy of the prediction result. In clustering method, finding distance between clusters and test data instances and suggesting the closest one as a result is the main idea. In this method, on the contrary, the aim is not finding only the closest cluster centroid but also neighbor ones. By doing so, a new customer is tried to be found whether s/he is around *Yes* or *No* responded customers. Therefore, distance function does multiple search to find a set of cluster centroids with minimal distance among them. This method minimizes some irrelevant matches, therefore, it provides an approximation to the optimal distance between train and test data.

The motivation behind the neighboring effect is same as in Section 4.3.2. There are some test instances far close to the result *Yes* after comparing distance *Yes* and *No*, however, the final decision is unexpectedly *No* in some cases. This is concluded that irrelevant test data instances may cause unexpected behavior on decision. In neighborhood concept, although there are some irrelevant cluster centroids which are selected exists, because of multiple centroid selection, dis-contiguous ones are tried to amortize. The prediction phase depends on summation of more than one distance of *Yes* and *No* cluster centroids. The experimental results of this approach are shown in Section 5.2.3.

4.3.4 Multiple Offering Approach

This approach is about offering multiple products to a customer using multiple channels. This means a customer is contacted more than once for campaign offerings. In this method, if the new customer is decided to answer *Yes*, then not just the closest cluster's product and channel is offered, but second and third closest clusters' products and channels are also offered. If the new customer's real product and channel matches with any of these products and channels, then prediction is done correctly. This method definitely increases the cost of a campaign, however the aim is to find how much it increases the return rate. Results are discussed in Section 5.2.4.

4.4 Model Based Bank Product and Channel Prediction Technique

This is a hybrid method, which both uses clustering and classification, implemented over partitioning based method explained in Section 4.3 which makes the label selection based on the rule (4.1). In model based method, this label prediction rule is replaced with a classification model. Training dataset is used to generate models for predicting labels and test dataset instances use these models to predict their own labels. The algorithm of this method is presented in Algorithm 2. Naïve Bayes, C4.5 decision tree and Tree Augmented Naïve Bayes classification algorithms are applied in the classification phase of the algorithm.

In the dataset only 3.63% of the data is positive labeled. This big difference between *Yes* and *No* responses makes the dataset very imbalanced and generated models are more inclined to *No* response rather than *Yes* response to increase accuracy. To prevent this, a new train dataset is created for generating models using equal number of *Yes* and *No* instances from the original training dataset. If a customer's label is decided as *No* from the model, then no further processing is needed and the customer is labeled as *No*. On the other hand, if the customer's label is decided as *Yes* from the model, then the same approach in

Algorithm 2 Model Based Bank Product and Channel Prediction Technique's Algorithm

```
procedure ConstructModel(trainingData, n)
    partition the trainingData into same PCY instances
    for each PCY instances do
        construct n clusters for each P_i, C_j as P_iC_jY^(k)
    classificationModel = classifier(trainingData)
    //train data contains same amount of Yes and No instances
procedure PredictInstance(classificationModel, data, P_(*)C_(*)Y^(*))
    if classificationModel(data) == YES then
        P_iC_jY^(k) = closestCentroid(P_(*)C_(*)Y^(*), data)
        data.Label = Yes
        data.Product = P_i
        data.Channel = C_j
    else
        data.Label = No
```

partitioning based method is applied. Only difference is that PCYs are formed in this method rather than PCLs. PCYs include only *Yes* labeled PCLs since the label of the new customer is decided by classification, *No* labeled PCLs are ignored. Therefore, from 4 product (Credit Card, Overdraft Account, Loan, Deposit Account), 4 channel (IVN, CC, EMAIL, SMS) and 1 label (Yes) combination 16 PCYs are formed in this case. Detailed drawing of this division is shown in Figure 4.5. After dividing the PCYs into clusters, closest *Yes* cluster is found and its product and channel is selected to offer to the customer. Results of this method with different classification algorithms are discussed in Section 5.3.

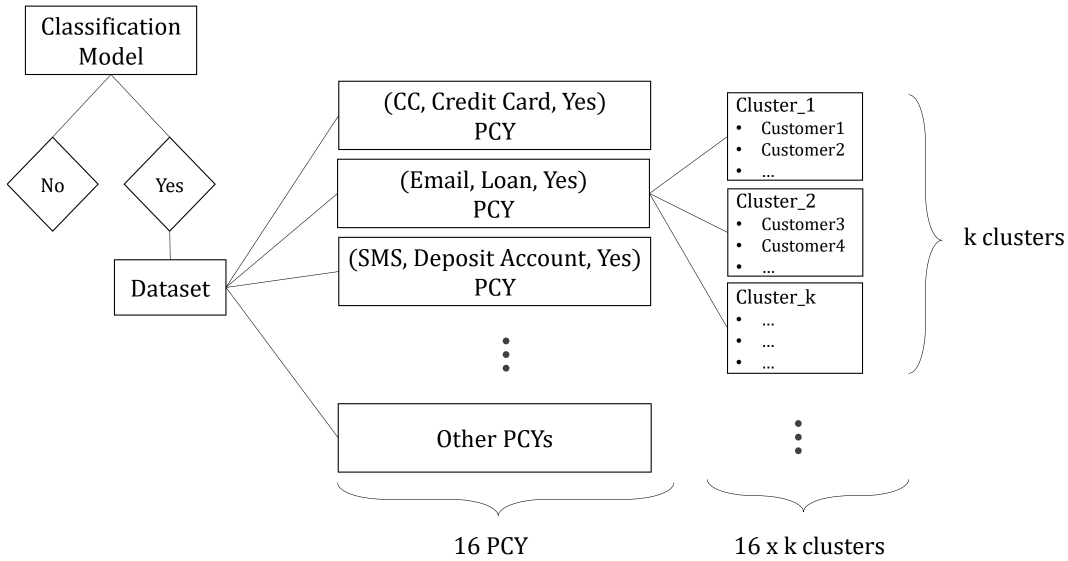

Figure 4.5: Model based method PCY formation

CHAPTER 5

EXPERIMENTAL ANALYSIS RESULTS AND DISCUSSIONS

In the experimental analysis, 10% of the dataset is used for testing. WEKA Data Mining Software is used in implementation and testing of proposed methods on a computer with Intel® Core™ i5-3317U, 2 physical 1.70 GHz CPUs and 4 GB physical memory [17].

As the responses of customers are known, for evaluation, the real responses of them are compared with the predicted ones. Confusion matrix is used to evaluate the performance of proposed methods. Table 5.1 shows the confusion matrix.

Table 5.1: Confusion matrix

	Predicted *Yes*	**Predicted *No***
Actual *Yes*	True Positive (TP)	False Negative (FN)
Actual *No*	False Positive (FP)	True Negative (TN)

True positives are the instances which are correctly classified as positive while true negatives are the ones which are correctly classified as negative. False positive instances are actually negative, however, incorrectly classified as positive and false negatives are actually positive mistakenly predicted as negative. However, since predicting not only the label but also product and channel of the customer is the motivation of this thesis, in the experiments true positives are the instances whose product, channel and label are correctly predicted. Besides, false positives may also include the positive instances whose product or channel are incorrectly predicted. Accuracy of the methods are evaluated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (5.1)$$

However, accuracy is not a reliable metric because of the problems mentioned in Chapter 2. Therefore, true positive, true negative, false positive, false negative ratios and F1-score are also used for evaluation of the results in the experiments which are defined as:

$$TP(\%) = \frac{TP}{TP + FN}, \quad (5.2)$$

$$TN(\%) = \frac{TN}{TN + FP}, \quad (5.3)$$

$$FP(\%) = \frac{FP}{FP + TN}, \quad (5.4)$$

$$FN(\%) = \frac{FN}{FN + TP}. \quad (5.5)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \quad (5.6)$$

5.1 Baseline Technique

Accuracy rate is one important measure for calculating performance of the solution. As it can be seen from the Table 5.2 the experimental results of the baseline method, accuracy of this method is 50%. This means that with the baseline method, half of the overall responses overlap with the predicted responses. However, the table also shows that true positive rate is 32.96% while true negative rate is 50.38%. Moreover, f-score is very low due to lowness of precision. The results indicate that most of the accurate predictions comes from the predicted negative responses which includes no prediction of product and channel attributes. As the dataset comprises nearly 97% negative responses, these results are quite expected. However, from the bank marketing point of view true positive ratio may even be more important measurement for the effectiveness of the solution to increase the return of investment. Therefore, in this problem the aim is to enhance the true positive rate as well as accuracy.

Table 5.2: Checking Product, Channel and Label with Baseline Technique

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	50.01	32.96	50.38	49.62	67.04	2.8

5.2 Partitioning Based Bank Product and Channel Prediction Technique

Partitioning based technique is based on clustering, so in the experiments this method's performance is tested by increasing the number of clusters. The results are shown in Table 5.3. In addition to predict all channel, product and label attributes, the results of predicting just label, both product and label, both channel and label are also added. For example, in the label part of Table 5.3, just the customer's *Yes* or *No* response is checked with proposed method's predicted response.

To obtain the best results, different number of clusters are tried. Because of the different characteristics of the customers, monotonic behavior cannot be expected. In general, accuracy tend to increase with the increasing number of clusters. This form is seen in all the four versions of the problem; checking label, channel-label, product-label, product-channel-label. Also, sensitivity, i.e. true positive rate, seem to achieve its maximum with cluster count 100. In general, the results show that clustering has positive effect on increasing accuracy, true positive and true negative ratios to a number. After that, no significant change occurs.

In the following approaches the results are compared with the last part of Table 5.3, that is *Product, Channel and Label* part. The experiments in the approaches shows whether an enhancement is achieved over partitioning based technique or not.

Table 5.3: Changing cluster count with partitioning based technique

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score	
1	50.48	69.23	49.84	50.16	30.77	8.53	Label
3	64.29	57.51	64.52	35.48	42.49	9.69	
5	65.43	62.27	65.53	34.47	37.73	10.72	
10	68.61	63.00	68.81	31.19	37.00	11.80	
20	70.30	63.00	70.55	29.45	37.00	12.39	
40	70.26	63.00	70.51	29.49	37.00	12.37	
100	72.79	61.54	73.18	26.82	38.46	13.10	
1	49.21	50.30	49.19	50.81	49.70	3.93	Channel and Label
3	63.37	41.41	63.92	36.08	58.59	5.18	
5	64.85	54.42	65.15	34.85	45.58	7.87	
10	67.84	51.90	68.26	31.74	48.10	7.64	
20	69.59	53.02	70.04	29.96	46.98	8.39	
40	69.60	53.88	70.03	29.97	46.12	8.66	
100	72.37	56.07	72.86	27.14	43.93	10.59	
1	48.85	39.57	49.01	50.99	60.43	2.56	Product and Label
3	63.53	45.02	64.02	35.98	54.98	5.98	
5	64.77	52.97	65.09	34.91	47.03	7.44	
10	68.03	55.11	68.39	31.61	44.89	8.65	
20	69.59	53.02	70.04	29.96	46.98	8.39	
40	69.69	55.31	70.09	29.91	44.69	9.15	
100	72.34	55.51	72.83	27.17	44.49	10.36	
1	48.76	36.36	48.96	51.04	63.64	2.24	Product, Channel and Label
3	63.26	38.62	63.85	36.15	61.38	4.63	
5	64.63	50.48	65.00	35.00	49.52	6.76	
10	67.72	49.50	68.18	31.82	50.50	6.97	
20	69.41	49.50	69.90	30.10	50.50	7.32	
40	69.45	51.21	69.93	30.07	48.79	7.81	
100	72.18	52.91	72.72	27.28	47.09	9.38	

5.2.1 Attribute Selection Approach

Attribute selection is performed based on information gain calculation of the attributes. The sorting of the attributes in descending order is shown in Table

4.2. The attributes are chosen based on this sorting. Since there is a visible gap between id(1) and education(10) attributes, first, 3 attributes date(4), customer period(7) and id(1) from the start of the table are tried. The results of this selection are shown in the Table 5.4. These results are for comparing all predicted product-channel-label attributes with the real data. As it can be seen accuracy and true negative ratios decrease with respect to choosing all attributes as in the previous section, however true positive ratio increases significantly.

Table 5.4: Changing cluster count with selected attributes 1, 4, 7

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	28.49	44.56	28.31	71.69	55.44	1.38
3	70.85	37.71	71.80	28.20	62.29	6.72
5	58.40	43.31	58.81	41.19	56.69	5.23
10	55.60	50.88	55.74	44.26	49.12	5.95
20	58.34	62.44	58.22	41.78	37.56	7.98
40	58.22	60.66	58.14	41.86	39.34	7.81
100	52.35	69.20	51.79	48.21	30.80	8.53

Then, more attributes are added to see if there is an improvement on the results or not. Attributes education(10), total loan(12) and period(5) are added depending on the information gain of the attributes. Results are shown in Table 5.5. Although there is a slight increase in true positive ratio with respect to using attributes 1, 4, 7, accuracy and true negative ratios show dramatic rise. However, accuracy and true negative ratios are not as good as using all of the attributes. This is a trade-off between two choices. One choice is to increase the number of customers who will return positive to the offer but needlessly contacting more customers who will reject the offer. This preference will increase the cost of the campaign as well as profit of it. Other choice is to decrease the customers who will buy a product if the offer is made but increasing the number of customers who will not be contacted needlessly. This choice decreases the cost of the campaign however it also decreases its profit.

Table 5.5: Changing cluster count with selected attributes 1, 4, 5, 7, 10, 12

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	47.10	35.64	47.24	52.76	64.36	1.63
3	63.87	48.78	64.26	35.74	51.22	6.33
5	59.46	58.25	59.49	40.51	41.75	6.74
10	64.78	50.49	65.15	34.85	49.51	6.72
20	65.07	48.37	65.52	34.48	51.63	6.78
40	62.81	60.09	62.89	37.11	39.91	8.42
100	61.21	71.31	60.89	39.11	28.69	10.13

5.2.2 Heuristic Approach

Experiments are done by changing k and t coefficients in the Equation 4.2 to see the increase in the prediction of *Yes* instances. The results are shown in Table 5.6. The experiments are done with cluster count 100 and the second line of the table with $k = 1.0$ and $t = 1.0$ is the result of *Product, Channel and Label* part of Table 5.3 with cluster count 100. As the coefficient k decreases and coefficient t increases, true positive ratio increases as expected. However, while there is an increase in the correctly predicted positive results, true negative ones are lost evenly.

Table 5.6: Changing k and t with 100 cluster

k	t	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1.05	0.95	76.65	49.35	77.44	22.56	50.65	10.65
1.00	1.00	72.18	52.91	72.72	27.28	47.09	9.38
0.95	1.00	69.64	55.20	70.04	29.96	44.80	8.93
0.90	1.05	65.24	59.35	65.4	34.6	40.65	8.19
0.85	1.10	61.15	62.74	61.11	38.89	37.26	7.71

5.2.3 Neighbor Approach

Neighbor count is increased from 1 to 20 to see the difference of the prediction results while the number of cluster count is fixed, 100. The result with 1 neighbor in Table 5.7 is same with the result of last part of Table 5.3 with 100 clusters whose accuracy is 72.18%. However, in neighborhood effect, even with

2 neighbor, total accuracy increased to 75.84%. The table shows that increasing neighbor count enhances the result of the accuracy. Note that, there is a limit for total neighbor count and after that point, the concept of neighborhood starts not to enhance results. Accuracy ratio increases with neighbor count, however, there arises a trade-off between true positives and true negatives. The difference between the experiments show that adding neighborhood concept cause to decrease on sensitivity, true positive prediction rate and increase on specificity, true negative prediction rate. Selection between two methods depends on a decision whether to lose some true positive instances or to gain true negative ones.

Table 5.7: Changing neighbor count with 100 cluster

Neighbor #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	72.18	52.91	72.72	27.28	47.09	9.38
2	75.84	48.67	76.61	23.39	51.33	10.00
3	77.44	46.02	78.33	21.67	53.98	10.12
4	78.84	46.22	79.76	20.24	53.78	10.72
5	79.62	47.98	80.51	19.49	52.02	11.36
6	79.84	47.32	80.76	19.24	52.68	11.38
7	80.41	47.56	81.33	18.67	52.44	11.76
8	80.53	45.81	81.52	18.48	54.19	11.54
9	80.63	44.54	81.66	18.34	55.46	11.39
10	80.86	44.54	81.90	18.10	55.46	11.51
15	81.78	41.85	82.92	17.08	58.15	11.30
20	81.69	40.79	82.86	17.14	59.21	11.03

5.2.4 Multiple Offering Approach

The difference of this approach from partitioning based technique is that not only 1 but also 2 and 3 offering to a customer is applied. Table 5.8 and Table 5.9 shows the results of 2 and 3 offers for a customer, respectively. The results show that accuracy, true positive and true negative ratios increase as expected. It can be seen that while the rate of change in accuracy and true negative ratios are small, it is significant in true positive ratio. With 100 cluster and 1 offer the true positive ratio is 52.91%, while it is 57.49% with 2 offers and 58.90% with 3 offers. Although, making more offers increases the cost of the campaign, increasing number of true positives makes the profit of the campaign increase also.

Table 5.8: Changing cluster count with making 2 offers

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	49.13	48.15	49.15	50.85	51.85	3.61
3	63.66	47.51	64.10	35.90	52.49	6.59
5	64.80	53.60	65.11	34.89	46.40	7.63
10	68.15	57.02	68.48	31.52	42.98	9.32
20	69.80	56.47	70.18	29.82	43.53	9.58
40	69.78	56.84	70.16	29.84	43.16	9.70
100	72.47	57.49	72.94	27.06	42.51	11.19

Table 5.9: Changing cluster count with making 3 offers

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score
1	49.59	58.00	49.38	50.62	42.00	5.32
3	63.86	51.26	64.24	35.76	48.74	7.62
5	64.97	56.36	65.23	34.77	43.64	8.48
10	68.39	60.39	68.65	31.35	39.61	10.63
20	69.93	58.44	70.28	29.72	41.56	10.34
40	69.88	58.26	70.24	29.76	41.74	10.26
100	72.54	58.50	72.99	27.01	41.50	11.63

5.3 Model Based Bank Product and Channel Prediction Technique

In this experiment, among the classification methods applied, Tree Augmented Naïve Bayes performed better than the others. Naïve Bayes, C4.5 Decision Tree and TAN results are shown in Table 5.10, Table 5.11 and Table 5.12, respectively. The results of Naïve Bayes classifier shows that there is improvement on true positive predictions compared to partitioning based method, however true negative ones are decreased. Therefore, the accuracy is 43.27% for 100 clusters, which is less than accuracy ratio of the partitioning based method. As seen in the results in Table 5.11, with C4.5 Decision Tree, 74.47% accuracy rate is obtained for predicting product, channel and label of the customer. True positive rates and f-score are also higher than partitioning based method. TAN results are even better than C4.5 Decision Tree. With TAN accuracy result is 77.12% while true positive ratio is 75.97% for product, channel and label prediction.

For predicting only label the results are all same with 1 to 100 clusters. The reason of this is the class label is decided by the result of classification model. After deciding label, clustering mechanism is applied if the label is decided to be *Yes*. Therefore, clustering does not have any effect on label part of the tables. Clustering have positive effect on deciding channel-label, product-label and product-channel-label as in the partitioning based method.

Table 5.10: Applying Naïve Bayes while cluster count changes with checking Product, Channel, Label

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score	
1-100	43.48	77.66	42.31	57.69	22.34	8.39	Label
1	41.87	56.74	41.61	58.39	43.26	3.25	Channel and Label
3	42.66	70.39	41.95	58.05	29.61	5.82	
5	42.80	71.89	42.01	57.99	28.11	6.24	
10	42.90	72.89	42.05	57.95	27.11	6.55	
20	43.21	75.70	42.19	57.81	24.30	7.55	
40	43.27	76.17	42.22	57.78	23.83	7.74	
100	43.31	76.45	42.23	57.77	23.55	7.86	
1	41.48	44.04	41.45	58.55	55.96	1.96	Product and Label
3	42.66	79.39	41.95	58.05	29.61	5.82	
5	42.76	71.50	42.00	58.00	28.50	6.13	
10	42.74	71.23	41.99	58.01	28.77	6.05	
20	43.04	74.26	42.12	57.88	25.74	7.02	
40	43.09	74.69	42.14	57.86	25.31	7.17	
100	43.31	76.45	42.23	57.77	23.55	7.86	
1	41.28	34.40	41.37	58.63	65.60	1.31	Product, Channel and Label
3	42.47	67.89	41.87	58.13	32.11	5.19	
5	42.58	69.35	41.92	58.08	30.65	5.54	
10	42.61	69.80	41.93	58.07	30.20	5.66	
20	42.98	73.71	42.09	57.91	26.29	6.82	
40	43.03	74.15	42.11	57.89	25.85	6.98	
100	43.27	76.17	42.22	57.78	23.83	7.74	

Table 5.11: Applying C4.5 Decision Tree while cluster count changes with checking Product, Channel, Label

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score	
1-100	74.61	71.43	74.72	25.27	28.57	15.80	Label
1	73.21	50.64	73.66	26.34	49.36	6.80	Channel and Label
3	73.92	63.89	74.19	25.81	36.11	11.44	
5	74.06	65.79	74.31	25.69	34.21	12.38	
10	74.05	65.64	74.30	25.70	34.36	12.30	
20	74.31	68.55	74.49	25.51	31.45	13.91	
40	74.44	69.89	74.60	25.40	30.11	14.75	
100	74.49	70.35	74.63	25.37	29.65	15.05	
1	72.78	36.59	73.34	26.66	63.41	3.88	Product and Label
3	73.91	63.73	74.19	25.81	36.27	11.36	
5	73.99	64.87	74.25	25.75	35.13	11.91	
10	73.91	63.73	74.19	25.81	36.27	11.36	
20	74.13	66.53	74.35	25.65	33.47	12.76	
40	74.22	67.64	74.43	25.57	32.36	13.38	
100	74.49	70.35	74.63	25.37	29.65	15.05	
1	72.61	28.44	73.21	26.79	71.56	2.69	Product, Channel and Label
3	73.76	61.58	74.07	25.93	38.42	10.42	
5	73.87	63.21	74.16	25.84	36.79	11.13	
10	73.82	62.50	74.12	25.88	37.50	10.82	
20	74.09	66.09	74.32	25.68	33.91	12.53	
40	74.19	67.23	74.40	25.60	32.77	13.15	
100	74.47	70.11	74.62	25.38	29.89	14.90	

Table 5.12: Applying TAN while cluster count changes with checking Product, Channel, Label

Cluster #	Acc. (%)	TP (%)	TN (%)	FP (%)	FN (%)	F1-score	
1-100	77.30	77.29	77.30	22.70	22.71	18.50	Label
1	75.69	56.03	76.04	23.96	43.97	7.35	Channel and Label
3	76.56	70.75	76.71	23.29	29.25	13.51	
5	76.69	72.20	76.82	23.18	27.80	14.43	
10	76.65	71.69	76.78	23.22	28.31	14.10	
20	76.96	74.69	77.03	22.97	25.31	16.25	
40	77.05	75.40	77.10	22.90	24.60	16.81	
100	77.16	76.25	77.19	22.81	23.75	17.54	
1	75.31	43.64	75.75	24.25	56.36	4.53	Product and Label
3	76.61	71.30	76.75	23.25	28.70	13.85	
5	76.66	71.82	76.79	23.21	28.18	14.18	
10	76.52	70.33	76.69	23.31	29.67	13.26	
20	76.79	73.16	76.90	23.10	26.84	15.10	
40	76.84	73.62	76.94	23.06	26.38	15.43	
100	77.16	76.25	77.19	22.81	23.75	17.54	
1	75.11	33.33	75.59	24.41	66.67	2.95	Product, Channel and Label
3	76.40	68.84	76.59	23.41	31.16	12.42	
5	76.50	70.05	76.67	23.33	29.95	13.09	
10	76.41	69.00	76.60	23.40	31.00	12.50	
20	76.74	72.69	76.86	23.14	27.31	14.77	
40	76.79	73.16	76.90	23.10	26.84	15.10	
100	77.12	75.97	77.16	22.84	24.03	17.30	

CHAPTER 6

CONCLUSION AND FUTURE WORK

In this chapter, firstly, proposed methods and experiment results of them are summarized. Then, enhancements that can be performed on this area are expressed.

6.1 Conclusion

Data mining methods provide practical solutions for direct marketing applications and they are widely used in this area. However, some problems emerge due to the characteristic structure of direct marketing databases. Databases are very large with very low positive response rate and very high negative response rate. Since negative responded customers outnumber the positive ones, the datasets are significantly imbalanced. Imbalanced data distribution makes learning algorithms more prone to predicting negative responses and discarding positive ones. Moreover, accuracy alone is no longer a proper evaluation criteria for imbalanced datasets.

I propose two methods for product selection and communication channel detection for a potential customer, partitioning based method and model based method. Partitioning based approach uses unsupervised learning method clustering, and model based approach uses both unsupervised and supervised learning methods, clustering and classification. In the partitioning based method, firstly the dataset is divided into PCLs, product-channel-label combinations. Then, each PCL is divided into clusters. After that, the closest cluster centroid to a new customer is found, and that cluster's product, channel and label are assigned to the new customer. Moreover, to increase the accuracy, true positive and true negative ratios I try four different approaches on partitioning based method; attribute selection approach, heuristic approach, neighbor approach and multiple offering approach. In the model based approach, firstly a classification is performed with equal sized positive and negative instances in order to find if the new customer is positive or negative respondent. Then, if the result is positive, the positive responded data is divided into product-channel combinations and divided into clusters. The closest cluster's product and channel are assigned to new customer.

Experimental results show that clustering makes positive increase on accuracy, true positive and true negative ratios. In the partitioning based method accuracy is 72.18%, true positive is 52.91% and true negative is 72.72% with 100 clusters and correct prediction of product, channel and label attributes. The approaches extended upon partitioning based method lead to a trade-off, increasing the accuracy or increasing the true positive ratio. Attribute selection increases true positive, however it decreases accuracy and true negative ratios, while neighbor approach does vice versa. Heuristic approach is also, decreases accuracy and true negative or increases true positive ratio depending on different k and t selection. Only multiple offering increases all accuracy, true positive and true negative ratios, however in this case the cost of the campaign increases. Model based method results in higher accuracy, true positive and true negative ratios. The best result for true positive ratio for predicting all product, channel and label attributes is achieved with Naïve Bayes which is 76.17%, and best accuracy and true negative ratios are achieved with TAN which are 77.12% and 77.16% respectively. F-score is low in all cases in both solutions as expected since the dataset is very imbalanced and number of true positives are very low compared to false positives. Although, f-score is not a proper evaluation criteria for this problem, clustering makes positive increase on it also.

6.2 Future Work

Data mining application in direct marketing provides very promising solutions in increasing the effectiveness of campaigns. The ensemble solution in this thesis and in many other works show that combining data mining approaches gives better results than using them alone. These clustering and classification combinations can be extended and new ensemble methods can be generated.

In this thesis, I am interested in finding the correct product and correct communication channel to propose to a customer if s/he is likely to buy any. However, there is a profit/cost constraint of campaigns which is not in the scope of this thesis. For example, multiple product offering in partitioning based approach increases accuracy, true positive and true negative ratios, however it also increases the cost of the campaign. The problem discussed can be extended by considering the cost of the channels used to reach to the customers and the expected profits for selling the products. In that case, the problem will be trying to reach to the customers through the most appropriate channel to minimize the cost while maximizing the profit obtained by selling the products. To solve the problem, besides data mining techniques, heuristic approaches can also be helpful.

REFERENCES

[1] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. *Machine learning: ECML 2004*, pages 39–50, 2004.

[2] M. Amini, J. Rezaeenour, and E. Hadavandi. A cluster-based data balancing ensemble classifier for response modeling in bank direct marketing. *International Journal of Computational Intelligence and Applications*, 14(04):1550022, 2015.

[3] D. Arthur and S. Vassilvitskii. k-means++: the advantages of carefull seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035, 2007.

[4] E. F. Ayetiran and A. B. Adeyemo. A data mining-based response model for target selection in direct marketing. *International Journal of Information Technology and Computer Science (IJITCS)*, 4(1):9, 2012.

[5] T. F. Bahari and M. S. Elayidom. An efficient crm-data mining framework for the prediction of customer behaviour. *Procedia Computer Science*, 46:725–731, 2015.

[6] N. H. Borden. The concept of the marketing mix. *Journal of advertising research*, 4(2):2–7, 1964.

[7] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[9] M.-D. Cohen. Exploiting response models—optimizing cross-sell and up-sell opportunities in banking. *Information Systems*, 29(4):327–341, 2004.

[10] M. Daneshmandi and M. Ahmadzadeh. A hybrid data mining model to improve customer response modeling in direct marketing. *Indian Journal of Computer Science and Engineering*, 3(6):844–855, 2013.

[11] S. Delanote, R. Leus, and F. T. Nobibon. Optimization of the annual planning of targeted offers in direct marketing. *Journal of the Operational Research Society*, 64(12):1770–1779, 2013.

[12] Dictionary. Marketing. http://www.dictionary.com/browse/marketing, 2017. [Online; accessed 5-August-2017].

[13] E. Duman, Y. Ekinci, and A. Tanrıverdi. Comparing alternative classifiers for database marketing: The case of imbalanced datasets. *Expert Systems with Applications*, 39(1):48–53, 2012.

[14] H. A. Elsalamony. Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*, 85(7), 2014.

[15] Y. Freund, R. E. Schapire, et al. Experiments with a new boosting algorithm. In *Icml*, volume 96, pages 148–156, 1996.

[16] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[18] D. J. Hand and K. Yu. Idiot's bayes—not so stupid after all? *International statistical review*, 69(3):385–398, 2001.

[19] A. M. Hughes. *The complete database marketer: second-generation strategies and techniques for tapping the power of your customer database*. McGraw-Hill, 1996.

[20] P. Kang, S. Cho, and D. L. MacLachlan. Improved response modeling based on clustering, under-sampling, and ensemble. *Expert Systems with Applications*, 39(8):6738–6753, 2012.

[21] M. Karim and R. M. Rahman. Decision tree and naive bayes algorithm for classification and generation of actionable knowledge for direct marketing. *Journal of Software Engineering and Applications*, 6(04):196, 2013.

[22] Y. Li, P. Murali, N. Shao, and A. Sheopuri. Applying data mining techniques to direct marketing: Challenges and solutions. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, pages 319–327. IEEE, 2015.

[23] M. Lichman. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml, 2013. [Online; accessed 29-July-2017].

[24] C. X. Ling and C. Li. Data mining for direct marketing: Problems and solutions. In *KDD*, volume 98, pages 73–79, 1998.

[25] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on*

mathematical statistics and probability, pages 281–297. Oakland, CA, USA., 1967.

[26] E. McCarthy. Basic marketing, a managerial approach. *Homewood, Ill.: RD Irwin*, 1960.

[27] S. Moro, R. Laureano, and P. Cortez. Using data mining for bank direct marketing: An application of the crisp-dm methodology. In *Proceedings of European Simulation and Modelling Conference-ESM'2011*, pages 117–121. Eurosis, 2011.

[28] A. Nachev. Application of data mining techniques for direct marketing. *Computational Models for Business and Engineering Domains*, 2015.

[29] A. Nachev and M. Hogan. Application of multilayer perceptrons for response modeling. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, page 1. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2014.

[30] E. W. Ngai, L. Xiu, and D. C. Chau. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602, 2009.

[31] F. T. Nobibon, R. Leus, and F. C. Spieksma. Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms. *European Journal of Operational Research*, 210(3):670–683, 2011.

[32] Y. Pan and Z. Tang. Ensemble methods in bank direct marketing. In *Service Systems and Service Management (ICSSSM), 2014 11th International Conference on*, pages 1–5. IEEE, 2014.

[33] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[34] J. R. Quinlan. *C4. 5: programs for machine learning.* Elsevier, 2014.

[35] C. Rygielski, J.-C. Wang, and D. C. Yen. Data mining techniques for customer relationship management. *Technology in society*, 24(4):483–502, 2002.

[36] M. F. Santos, P. Cortez, H. Quintela, and F. Pinto. A clustering approach for knowledge discovery in database marketing. *WIT Transactions on Information and Communication Technologies*, 35, 2005.

[37] J.-Y. Shih, W.-H. Chen, and Y.-J. Chang. Developing target marketing models for personal loans. In *Industrial Engineering and Engineering Management (IEEM), 2014 IEEE International Conference on*, pages 1347–1351. IEEE, 2014.

[38] L. Sing'oei and J. Wang. Data mining framework for direct marketing: A case study of bank marketing. *International Journal of Computer Science Issues (IJCSI)*, 10(2):198–203, 2013.

[39] K. Wisaeng. A comparison of different classification techniques for bank direct marketing. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(4):116–119, 2013.

[40] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.