A Data Mining Approach to Improve Benefit of Telemarketing: A Case Study of E-Commerce Company

Bramantyo Erlangga[1], Indra Budi[2], Mohamad Ivan Fanany[2], Qorib Munajat[2]

[1] Magister of Information Technology, Faculty of Computer Science, University of Indonesia
[2] Faculty of Computer Science, University of Indonesia, Depok, West Java, Indonesia, 16424

***Abstract*—XYZ Company, an e-commerce company focusing on Customer to Customer (C2C) classified ads, have been using telemarketing since June 2016. However, their weekly average conversion rate from telemarketing is only 5,1% which is less than the average industry conversion rate for telemarketing (8,21 %). To improve the conversion rate, we need to understand the customer behavior and characteristic. Therefore, in this study, we conduct a data mining approach to classify the potential customer that will help increase the sales. We aim to find the best class imbalance technique and learning algorithm for e-commerce C2C classified ads company case. CRISP-DM methodology was used as the guideline principle in this research. Learning algorithms experimented are Bayesian Network, Decision Tree, Random forest, Support Vector Machine, Neural Network, Bagging Neural Network, Deep Neural Network, Adaboost Deep Neural Network, Convolutional Neural Network, and Extreme Gradient Boosting Tree. SMOTE and undersampling technique were compared in addressing class imbalance problem. The classification performance are evaluated using cost benefit analysis. The result found that Adaboost Deep Neural Network ensemble learning algorithm trained using the SMOTE training data produced the highest profit value with a potential gain of 3.56 times compared without using the classification model. In general, SMOTE method for balancing classes on training data provides better results than the undersampling method.**

***Index Terms*—data mining, cost-benefit analysis, telemarketing, class imbalance, SMOTE, undersampling, CRISP-DM**

I. Introduction

Many companies have been widely using telemarketing in order to increase their sales. Telemarketing is a type of direct marketing that use telephone and call centers to do marketing activity [1]. Telemarketing can be categorized as outgoing telemarketing (calls originated from company), and incoming telemarketing (calls originated from customer) [2], or also known as outbound telemarketing and inbound telemarketing [1]. Report from The Direct Marketing Association (DMA) shows that most of direct marketing budget on company was spent on telemarketing activity, more than direct mail non-catalog marketing [3]. That report indicates the importance of telemarketing activity for a company.

XYZ Company is one of leading e-commerce platforms in Indonesia with more than three billion monthly page views and more than four million monthly ads listing in 2016. XYZ Company main business is to provide classifieds ads platform towards its customer. Classifieds ads can be considered as customer-to-customer (C2C) type of e-commerce [4] [5]. Although their classifieds ads service is free to use, XYZ Company try to monetize their services through the introduction of premium feature towards customer who place ads in their website. Example of XYZ Company premium feature is top listing feature, which allows customer to pay some amount of money to put their ads on top of search result.

Telemarketing in XYZ Company was used to drive premium feature sales since June 2016. However, the current telemarketing activity in XYZ Company have low conversion rate. From their weekly average telemarketing data in June 2016, it was found that their conversion rate is just 5,10 percent from the total calls. Report from DMA shows that the average industry conversion rate for telemarketing is 8,21 percent [3], higher than XYZ Company conversion rate. In addition, XYZ Company also encounters problem with limited telemarketer resources, which made telemarketing activity cannot follows all provided leads within limited timeframe.

Previous research shows that data mining could help to improve bank telemarketing success rate by predicting the customer's respond to marketing campaign on the product offered [6] [7] [8] [9] [10] [11] [12]. Other research shows that using decision engine to optimize direct marketing campaign could increase profit up to 27 percent and reduce cost up to 6 percent [13]. Moro, Cortez, and Rita [7] claimed that data mining could achieve good result by correctly predicting 79 percent of buyers by only contacting half of total population, which translates as 29 percent improvement compared with traditional practice. Meanwhile, Li, et al. [14] evaluated their telemarketing prediction model with actual field test. Their result indicated that data mining could help to improve direct

This work was supported by International Publication Grants for Student Thesis provided by Universitas Indonesia (PITTA Grant No. 408/UN2.R3.1/HKP.08.00/2017)

marketing response rate up to 2.3 times compared with traditional approach [14].

In this paper, we propose a data mining approach to help XYZ company identify potential premium feature customer in their telemarketing activity. The aim of this research is to find the best learning algorithms and class imbalance technique for the case of e-commerce company specializing in C2C classifieds ads. We contributed by employing more comprehensive approach in C2C classified ads company case study that has not been covered in previous research. First, our research emphasizes on measuring data mining model quality based on its cost and benefit value, not only based on traditional metric such as accuracy. Second, we contributed by comparing two class imbalance problem techniques which are SMOTE and undersampling technique. Third, we compared various learning algorithms, including latest popular method such as Deep Neural Network, Convolutional Network, and XGBOOST. The research is based on real data so the findings contributed not only on theoretical level but also on practical level.

II. LITERATURE REVIEW

A. *CRISP-DM*

CRISP-DM is an iterative methodology consisting of six phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment [15]. CRISP-DM sequence is not rigid; it allows movement between phases [15]. Those phases are Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment phase.

There are other popular data mining methodologies such as SEMMA and KDD [16] [17] [18] [19]. Comparison study between those methodologies agrees that most of the phases between different methodology can be mapped to each other [18] [18] [19]. In those surveys, all agreed that the most widely used data mining methodology is CRISP-DM [17] [18] [19]. In our research, we choose to follow CRISP-DM methodology because of the following reasons: it was developed by practitioner and professional consultant from industries [20]; it is the most widely used data mining methodology in industry [16] [17] [18] [19]; and it starts with business goal in minds [15] which is suitable for real world problem.

B. *Class Imbalance*

The most obvious problem in data mining for direct marketing is extreme class imbalance problems [14] [21], which can go as far as only 1 percent of positive class distribution in population [21]. Previous research in telemarketing domain also encountered class imbalance problems, ranging from 11.7 percent to 12.8 percent positive class [6] [7] [8] [9] [10] [11] [12]. Imbalance class distribution could lead learning algorithm to a major pitfall: If learning algorithm discover a pattern that all data belongs to negative class, it still yields a high accuracy [21].

A study categorized major method to address class imbalance problems as Problem-definition-level method, Data-level method, and Algorithm-level method [22]. Our research focuses on using data-level method to address class imbalance problem since the business problem cannot be redefined to a smaller subset, and not all machine learning algorithm used in this research are capable of addressing class imbalance.

The most basic sampling methods is undersampling and oversampling [22]. Oversampling works by duplicating minority class, and undersampling works by eliminating majority class, so the training set contains balanced class distribution [22] [23]. There are more advance sampling methods, that use intelligence to duplicates or eliminate data, and such method tends to give better performance than random oversampling or random undersampling [22].

Synthetic Minority Oversampling Technique (SMOTE) was introduced by Chawla, et al. on 2002 [24]. The idea behind SMOTE is to generates synthetic examples for minority class using k-nearest neighbor algorithm so that learning algorithm could build larger decision regions using those synthetic minority examples [24]. SMOTE tends gives better result than plain undersampling during experiments [24].

Although class imbalance is a major problem in direct marketing domain [21], not all previous research in telemarketing domain address this problem. For example, some researches [6] [7] [8] [11] [12] that using bank telemarketing data did not address class imbalance problem as part of the solutions. Other researches on bank telemarketing data that address class imbalance problem are [9] and [10], which use SMOTE methods and gives better result than [6] [7] [8] [11] [12].

Our research address class imbalance problem by using SMOTE methods, as it gives promising result on previous research [9] [10] [24]. We analyze the performance of this technique by comparing the model's performance built using SMOTE, random undersampling, and without using any sampling method. In order to understand impact of oversampling percentage in SMOTE, we also analyze SMOTE parameter, ranging from 100 percent to 500 percent.

C. *Customer Lifetime Value*

Moro, Cortez, and Rita [8] used Customer Lifetime Value (LTV) in generating feature to improve model quality. LTV features yields up to 4 pp improvement in cumulative lift curve compared without using LTV in bank telemarketing domain [8]. LTV represents expected benefit from a customer in the future purchase or interactions [25]. Recency, Frequency, and Monetary (RFM) characteristics of a customer can be used as the base to compute LTV [26]. In direct marketing domain, RFM can be described as: 1) R which represents period since the last purchasing activity; 2) F which represents how often customer made a purchase in certain period; and 3) M which represents the total monetary amount of purchase made by a customer on a certain period [27].

RFM and historical telemarketing data are used in Moro, Cortez, and Rita research [8] to represent LTV from a customer. Extracted features on [8] based on LTV are: Months since last purchase (R); Number of times client subscribed to deposit previously (F); Total amount of money client subscribed previously (M); Average value subscribed per successful telemarketing contact (M); Last telemarketing contact result (H); Last telemarketing contact result for the same product sold (H); Total call duration of all previous contact (H); Average call duration of all previous contact (H); Total number of previous contact (H); Total success per total contact (H); and Total success minus total unsuccessful contact (H) [8].

Previous research [8] shows promising result using LTV value in telemarketing domain. Our research use LTV and historical telemarketing data in order to yields a better model. Feature extraction based on RFM and historical telemarketing data are subject to availability of the data itself, which is unique on each data mining problems including this research.

D. *Evaluation Metric*

Another problem in data mining for direct marketing is that simple measurement such as accuracy value is less suitable to measure the quality of a model [21]. Error in classifying needs to be treated differently, such as having false positive (recognize non-buyers amongst buyer) more favorable than having false negative (recognize buyer amongst non-buyer) [21]. Accuracy measurement cannot be used to select proportion of population (such as 20 or 30 percent) that having the most probability of being buyer [21].

Prediction from binary class dataset can be drawn into a confusion matrix that holds four possible classification results: 1) True positive (TP), if predicted class is positive and actual class is positive; 2) False positive (FP), if predicted class is positive but the actual class is negative; 3) False negative (FN), if predicted class is negative but the actual class is positive; and 4) True negative (TN), if predicted class is negative and actual class is negative [23] [28]. Confusion matrix can be used to calculate various measurements such as sensitivity, specificity, precision, and recall [23]. In this study, we use confusion matrix to perform cost benefit analysis.

Cost Benefit analysis was performed by associating the confusion matrix with the cost and benefit [29]. For example, True Positive result can be correlated to benefit since the result of classification shows the number of marketing success which generates income to the company. True Negative result can provide benefit by reducing unnecessary telemarketing cost. False Positive result is correlated to cost since there will be unnecessary cost for unsuccessful marketing. Lastly, False Negative is attributed to cost considering the cost of opportunity lost due to uncontacted potential buyer.

III. METHODOLOGY

This research used experimental approach to find which technique (class imbalance problem technique and learning algorithm) is best suited to classify potential customer in XYZ company case. The research process is based on CRISP-DM methodology with modification to meet the research goal. The research process is presented in Figure 1.

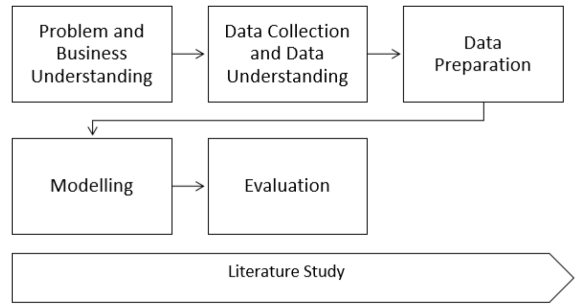

Figure 1 Research Process [29]

Data was provided by XYZ Company which includes customer personal data, customer history of transaction (ads placement), public interest with ads, and premium feature purchase data. The data are formatted using CSV format. The next phase is data preparation which includes several processes such as data consolidation, data cleansing, data transformation, and attribute adjustment.

The raw data obtained from PT XYZ was provided in separate form, such as customer data, historical advertising data, data of interest on advertisements posted by customers, and premium features purchasing data. Customer data, telemarketing data, and premium feature purchasing data are incorporated in the data consolidation process, resulting in customer data describing premium buyer and non-premium buyer. The output class is obtained from premium purchasing data.

In the process of data cleansing, data removal was performed to row having incomplete attribute. Duplicated data and outlier data was also eliminated in this process. After data cleansing, data transformation process will be performed. The LTV method is used to perform data transformation by adding the LTV attributes extracted from the customer's historical data. After that, the dataset will be divided into test data and training data. SMOTE and undersampling method are used to address data imbalance problem. Minority classes with the majority class on training data Output of this stage is data that has been processed as training data and test data.

The classification model was generated using the following algorithms selected based on previous studies: Bayesian Network [9][10], Decision Tree [6][7], Support Vector Machine [6][7][8], Neural Network [7][8], Ensemble Learning (Bagging and Boosting) [11], Deep Convolutional Neural Network [12], and Extreme Gradient Boosting Tree (XGBOOST) [30]. The generated model for each algorithm is then evaluated using confusion matrix and cost-benefit analysis.

IV. EXPERIMENT AND RESULT

The explanation of the experiment and result are structured using the process structure in CRISP-DM Methodology.

A. *Business Understanding*

The main goal in the implementation of data mining for telemarketing in PT XYZ is to classify or predict which customers will buy premium features when contacted by sales telemarketing. Implementation of data mining classification is expected to increase the ratio of success calls (customers who buy premium features after being contacted by telemarketing sales). There are two classes in this classification case which are customer purchasing premium features and not purchasing premium features. The classification model is evaluated based on the result of cost-benefit analysis.

To perform cost-benefit analysis, components of cost and benefit need to be identified. The cost component of telemarketing are the cost of making calls and telemarketing employee salaries. Meanwhile, the benefit is the revenue potential calculated from the average revenue per customer who purchased premium features within a week of telemarketing. Table 1 below shows the result of cost benefit calculation linked with confusion matrix.

Table 1 Cost and Benefit Table

Classification Result	Benefit[1]	Description
True Negative	4,352	Cost saving potential due to unneeded telemarketing contact
True Positive	111,380	Potential benefit
False Negative	-115,732	Cost of opportunity loss due to uncontacted likely buyer
False Positive	-4,352	Cost of unnecessary telemarketing contact

[1] in IDR currency.

The number are obtained based on the calculation of several cost and benefit aspects such as salary, revenue, etc. in XYZ Company.

B. Data Understanding and Preparation

The data obtained from XYZ Company is separated in four (4) tables which are Premium table, Replies table, Call Result table, and Listing table. Premium table contains premium feature purchasing transaction and top-up of premium customer credit data. Listing table contains customer personal data and their ads placement transaction. Replies table records the frequency and detail of certain ads being accessed which shows the level of attractiveness of ads. Call Result table contains data of telemarketing activities though phone call and the response of customer.

The number of data is as follow: (1) Call Result table contains 38,488 records; (2) Premium table has 89,570 records; (3) Listing table has 1,744,333 records and contains only data that correlates to phone attributes in the Call Result table; and (4) Replies table has 6,871,080 records and contains only data that correlates to data in the listing table. All the data are form 1st March 2016 until 14th August 2016.

Data was then cleansed to treat the noise such as inconsistent data format and invalid input. After the cleansing process, features were generated from the data. This study added features based on Customer Lifetime Value (LTV) consisting of recency (R), frequency (F), monetary (M), and history (H) attributes. The features are listed in Table 2.

Table 2 List of Feature

Feature	LTV[1]	Description
count_prv_contact	H	Total number of previous telemarketing contact to customer
count_user_id	-	Total number of user id related to a phone number
count_listing	F	Total number of ads listing by customer in the last 90 days
avg_listing_price	M	Average price in all ads listed by customer in the last 90 days
max_listing_price	M	Maximum price in all ads listed by customer in the last 90 days
min_listing_price	M	Minimum price in all ads listed by customer in the last 90 days
stdev_listing_price	M	Minimum price in all ads listed by customer in the last 90 days
most_l{1,2,3}_cat_id[2]	-	Most category id level {1,2,3} in all ads listed by customer in the last 90 days
most_{province, city}_id[2]	-	Most {province, city} id in all ads listed by customer in the last 90 days
count_distinct_listing_{province, city}[2]	-	Number of distinct {province, city} id in all ads listed by customer in the last 90 days
most_device_id	-	Most device id used to put ads in the last 90 days
count_rejected_listing	F	Number rejected ads listing by customer in the last 90 days
latest_balance	M	Customer latest balance before telemarketing contact
count_distinct_cat_l{1,2,3}[2]	-	Number of distinct category id level {1,2,3} in all ads listed by customer in the last 90 days
total_replies	F	Total number of replies of all customer ads in the last 90 days
avg_replies_per_listing	F	Average number of replies for each ad listed by customer in the last 90 days
days_since_last_{topup, purchase, listing}[2]	R	Number of days since customer last time do a {top-up, purchase, listing}
days_since_last_replies	R	Number of days since an ad of customer got replies
count_listing_device_{android, apple, i2, m, none}	-	Number of ads listed by customer using {android, apple, i2, m, none} device in the last 90 days
count_cat1_{86, 87, 88, 89, 90, 92, 94, 96, 97, 98}	-	Number of ads listed by customer where category level 1 id is {86, 87, 88, 89, 90, 92, 94, 96, 97, 98} in the last 90 days

[1] R: Recency, F: Frequency, M: Monetary, H: History
[2] Parentheses indicates that there were more than one feature based on values on the parentheses.

C. Modelling

In the modeling stages, classification model was developed using the aforementioned classification algorithms with 3 different training datasets: SMOTE training dataset, Undersampling training datasets, and unmodified training datasets. Rapidminer was used as the tool to develop the model. Data were also normalized to accommodate several learning algorithms limitation. Normalization includes mapping continues data in normal distribution using z-transformation, changing categorical data into numerical data, and adjusting data range (i.e for Deep Neural Network, the range is real number from 0,0 until 1,0).

The dataset was split into training data and test data based on the call_date attribute on Call Result table. The training data contains customer data contacted on 1st June 2016 until 1st

August 2016, while the test data contains customer data contacted on 1st to 14 August 2016. This was done to get the model evaluation result to be as close as possible to the actual condition in XYZ Company.

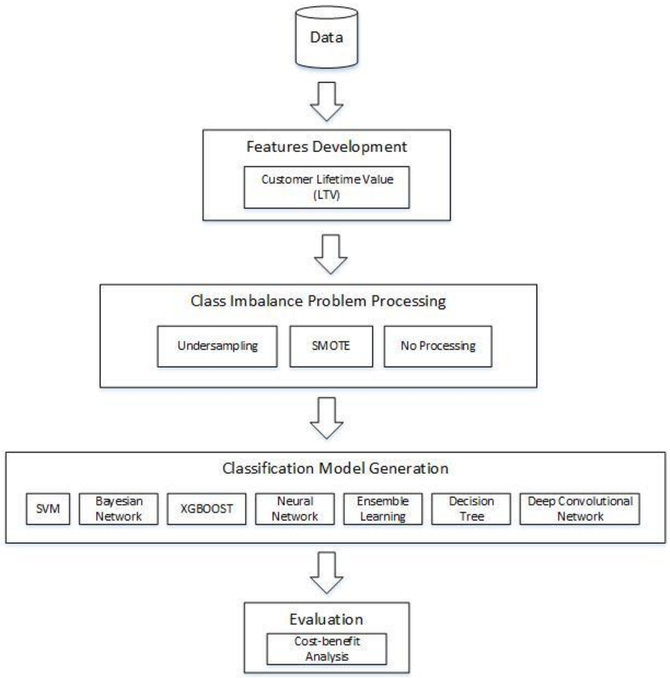

Figure 2 Experiment Process [29]

The training dataset formed with SMOTE contains 6,795 data with negative class 3,171 data and positive class of 3,624 data. Training dataset formed using undersampling method amounted to 1.244 data with positive class of 604 data and 640 class of negative data. While unmodified training dataset amounted to 13,403 data with positive class of 604 data and negative class 12,799 data.

The training process was conducted using 10-fold cross validation method. This was done to ensure that the model can generalize to the data well. The final evaluation of the model was performed using unmodified test data, which is telemarketing data from 1st August to 14th August 2016.

There were three inputs to the process of cross validation. The execution of each process will be repeated as many as the available dataset which are the SMOTE dataset, undersampling dataset, and unmodified dataset. The difference between each modeling process is the algorithm used. The overall experiment process is depicted in Figure 2.

D. Evaluation

In general, the oversampling method with SMOTE gave better results than the undersampling method. However, the undersampling method had much faster modeling time. This is because the amount of data generated by the undersampling method is fewer which is only 10 percent of the total data held (all minority classes plus the percentage of the majority class of minority classes). The result of Cost Benefit evaluation can be observed in Table 3.

Table 3 Cost Benefit Analysis Result

Class Imbalance Technique	Algorithm	Cost	Benefit	Profit
-	Baseline - Without Model	23,522,560	31,966,060	8,443,500
SMOTE	Bayesian Network	19,570,304	36,614,636	17,044,332
	Decision Tree	13,159,412	42,716,536	29,557,124
	Random Forest	13,091,552	42,780,044	29,688,492
	Neural Network	14,070,440	41,844,676	27,774,236
	SVM	13,949,548	41,952,512	28,002,964
	Bagged Neural Network	14,102,520	41,830,004	27,727,484
	Deep NN	13,795,116	42,037,312	28,242,196
	AdaBoost Deep NN	12,814,456	42,978,804	30,164,348
	Convolutional NN	14,097,828	41,760,712	27,662,884
	XGBOOST	14,556,588	41,415,104	26,858,516
UNDERSAMPLING	Bayesian Network	14,274,984	41,640,132	27,365,148
	Decision Tree	14,659,576	41,272,948	26,613,372
	Random Forest	13,883,304	42,031,812	28,148,508
	Neural Network	15,292,884	40,687,512	25,394,628
	SVM	13,949,548	41,952,512	28,002,964
	Bagged Neural Network	14,066,088	41,849,028	27,782,940
	Deep NN	14,758,708	41,186,872	26,428,164
	AdaBoost Deep NN	14,320,120	41,612,404	27,292,284
	Convolutional NN	14,995,488	40,945,740	25,950,252
	XGBOOST	13,501,292	42,400,768	28,899,476
UNMODIFIED	Bayesian Network	15,828,676	40,203,944	24,375,268
	Decision Tree	29,482,968	27,111,060	-2,371,908
	Random Forest	33,215,084	23,522,560	-9,692,524
	Neural Network	29,607,404	26,990,976	-2,616,428
	SVM	13,949,548	41,952,512	28,002,964
	Bagged Neural Network	27,473,776	29,041,916	1,568,140
	Deep NN	14,686,496	41,254,732	26,568,236
	AdaBoost Deep NN	18,142,352	37,990,364	19,848,012
	Convolutional NN	33,215,084	23,522,560	-9,692,524
	XGBOOST	26,427,836	30,048,688	3,620,852

We compared the cost and benefit for each model to the basic value which is the cost-benefit obtained without using the model (all customers contacted). All models trained using SMOTE data and undersampling data yield greater benefits than without modeling. Models trained using unmodified data show poor results, with 6 out of 10 models providing no greater benefit than without modeling.

V. DISCUSSION

This study implemented the data mining approach comprehensively from data preparation to evaluation to demonstrate how data mining can be used in telemarketing. It enriches and extend the previous study in telemarketing domain. Implementation of data mining classification is expected to reduce the number of customers to be contacted by sales to increase the ratio of success calls. However, in business perspectives, the expected value is not about the number of calls but the benefits they can get by telemarketing. Therefore, measurement using cost-benefit analysis is considered more suitable for business organization.

This study also focused on class imbalance problem which is a common problem in telemarketing case. There are several techniques such as undersampling and SMOTE to address class imbalance problem. This study found that SMOTE technique is better than undersampling technique. The result is aligned with previous studies who used SMOTE in their data processing [9][10].

VI. CONCLUSIONS

Based on the result, we found that the classification model that produces the highest profit value based on the cost-benefit analysis is the model developed using the Adaboost Deep Neural Network ensemble learning algorithm trained using the SMOTE training data with a potential gain of 3.56 times more than the current condition (telemarking process without using the classification model). As for class imbalance technique, the SMOTE method for balancing classes on training data provides better results than the undersampling method. The best SMOTE oversampling parameter is 500 percent.

The findings contribute in data mining domain by showing how data mining is applied in e-commerce company. The best model found in this case study can be used as comparison on similar cases of data mining. However, it needs to be noted that the best model in this study is based on cost benefit analysis which may not be the best model in other evaluation methods. Future study need to include more evaluation methods to compare the model in various perspectives.

ACKNOWLEDGMENT

This study was supported by International Publication Grants for Student Thesis provided by Universitas Indonesia (PITTA Grant No. 405/UN2.R3.1/HKP.05.00/2017).

REFERENCES

[1] P. Kotler and K. L. Keller, Marketing Management 14th Edition, New Jersey: Prentice Hall, 2012.

[2] J. J. Marshall and H. Vredenburg, "Successfully Using Telemarketing in Industrial Sales," *Indusrrial Marketing Management 17*, pp. 15-22, 1988.

[3] The Direct Marketing Association, "Statistical Fact Book 2013 The Definitive Source For Direct Marketing Benchmarks," Direct Marketing Association, New York, 2013.

[4] D. Chaffey, E-Business and E-Commerce Management: Strategy, Implementation and Practice 5th Edition, New Jersey: Pearson, 2011.

[5] E. Turban, D. King, J. K. Lee, T.-P. Liang and D. C. Turban, Electronic Commerce A Managerial and Social Networks Perspective 8th Edition, New York: Springer, 2015.

[6] S. Moro, R. M. S. Laureano and P. Cortez, "Using Data Mining for Bank Direct Marketing: an Application of The CRISP-DM Methodology," in *Proceedings of the European Simulation and Modelling Conference - ESM' 2011*, Guimarães, 2011.

[7] S. Moro, P. Cortez and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems Volume 62*, pp. 22-31, 2014.

[8] S. Moro, P. Cortez and P. Rita, "Using Customer Lifetime Value and Neural Networks to Improve The Prediction of Bank Deposit Subscription in Telemarketing Campaigns," *Neural Computing and Applications Volume 26 Issue 1*, pp. 131-139, 2015.

[9] P. Ruangthong and S. Jaiyen, "Bank direct marketing analysis of asymmetric information based on machine learning," in *The 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Songkhla, 2015.

[10] P. Ruangthong and S. Jaiyen, "Hybrid Ensembles of Decision Trees and Bayesian Network for Class Imbalance Problem," in *8th International Conference on Knowledge and Smart Technology (KST)*, Chiangmai, 2016.

[11] Y. Pan and Z. Tang, "Ensemble methods in bank direct marketing," in *11th International Conference on Service Systems and Service Management (ICSSSM)*, Beijing, 2014.

[12] K.-H. Kim, C.-S. Lee, S.-M. Jo and S.-B. Cho, "Predicting the success of bank telemarketing using deep convolutional neural network," in *7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, Fukuoka, 2015.

[13] J. Berry, "The benefits of using a decision engine to optimise campaign planning for direct marketing," *Database Marketing & Customer Strategy Management Vol. 13*, pp. 192-202, 2006.

[14] Y. Li, P. Murali, N. Shao and A. Sheopuri, "Applying Data Mining Techniques to Direct Marketing: Challenges and Solutions," in *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Washington, 2015.

[15] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer and R. Wirth, "CRISP-DM 1.0 Step-by-step data mining guide," CRISP-DM Consortium, 2000.

[16] E. Turban, R. Sharda and D. Delen, Decision Support and Business Intelligence Systems 9th Edition, New Jersey: Prentice Hall, 2010.

[17] G. Mariscal, Ó. Marbán and C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review,* vol. 25, no. 2, pp. 137-166, 2010.

[18] A. Azevedo and M. F. Santos, "KDD, SEMMA and CRISP-DM: A Parallel Overview," in *IADIS European Conference of Data Mining 2008,* Amsterdam, 2008.

[19] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA)," *International Journal of Innovation and Scientific Research,* vol. 12, no. 1, pp. 217-222, 2014.

[20] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *Journal of Data Warehousing,* pp. 13-22, 2000.

[21] C. X. Ling and C. Li, "Data Mining for Direct Marketing: Problems and Solutions," *KDD,* vol. 98, pp. 73-79, 1998.

[22] G. M. Weiss, "Foundations of Imbalanced Learning," in *Imbalanced Learning:Foundations, Algorithms, and Applications*, Wiley-IEEE Press, 2013, pp. 13-41.

[23] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques 3rd Edition, Waltham: Morgan Kaufmann Publishers, 2011.

[24] 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*, pp. 321-357, 2002.

[25] F. R. Dwyer, "Customer lifetime valuation to support marketing decision making," *Journal of interactive marketing,* vol. 11, no. 4, pp. 6-13, 1997.

[26] P. Fader, B. Hardie and K. L. Lee, "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research,* vol. 42, pp. 415-430, 2005.

[27] D. L. Olson and B. Chae, "Direct marketing decision support through predictive customer response modeling," *Decision Support Systems,* vol. 54, no. 1, pp. 443-451, 2012.

[28] D. T. Larose and C. D. Larose, Data Mining and Predictive Analytics, New Jersey: Wiley, 2015.

[29] N. Japkowicz, "Assessment metrics for imbalanced learning," in *Imbalanced Learning:Foundations, Algorithms, and Applications,* Wiley-IEEE Press, 2013, pp. 187-206.

[30] T. Chen and G. C, "A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* pp. 785-794, 2016.