Handling Class Imbalance In Direct Marketing Dataset Using A Hybrid Data and Algorithmic Level Solutions

Haya Abdullah Alhakbani
Department of Computing
Goldsmiths College
University of London
London SE14 6NW,UK
Email: halha001@gold.ac.uk

Mohammad Majid al-Rifaie
Department of Computing
Goldsmiths College
University of London
London SE14 6NW,UK
Email: m.majid@gold.ac.uk

Abstract—Class imbalance is a major problem in machine learning. It occurs when the number of instances in the majority class is significantly more than the number of instances in the minority class. This is a common problem which is recurring in most datasets, including the one used in this paper (i.e. direct marketing dataset). In direct marketing, businesses are interested in identifying potential buyers, or charities wish to identify potential givers. Several solutions have been suggested in the literature to address this problem, amongst which are data-level techniques, algorithmic-level techniques and a combination of both.

In this paper, a model is proposed to solve imbalanced data using a Hybrid of Data-level and Algorithmic-level solutions (HybridDA), which involves oversampling the minority class, undersampling the majority class, and additionally, optimising the cost parameter, the gamma and the kernel type of Support Vector Machines (SVM) using a grid search. The proposed model perfomed competitively compared with other models on the same dataset. The dataset used in this work are real-world data collected from a Portuguese marketing campaign for bank-deposit subscriptions and are available from the University of California, Irvine (UCI) Machine Learning Repository.

Keywords—*Imbalance data; minority class; grid search; sampling; SMOTE; ; classification; SVM*

I. Introduction

Customer relationship management (CRM) is the term given to a set of activities and processes (closely allied with marketing) that are aimed at identifying valuable customers, building relationships with them, and encouraging them to make repeated purchases. This relationship built between a business and its preferred customers is known as the customer life cycle.

In general, there are four stages to the customer life cycle, a customer can be a prospect, a responder, an active customer, or a former customer. Prospects are people who are not yet customers but are in the demographic that is a target market for the product or service, and can therefore be enticed to try the company's offerings [1]. Responders are people who respond to promotional or communication campaigns and show interest in trying the product. When responders try the product, they become active customers. A company's

objective is generally to continue to provide services and promotions to active customers, so that they remain loyal to the company and continue to purchase products and services while sparing the company the costs of finding new customers. The bulk of CRM activities are targeted toward active customers. A company aims to increase profits for minimal investment by offering a higher diversity of products and services of greater value, which includes upselling, cross-selling and deep-selling. Lastly, people who are no longer active customers are classified as former customers. Former customers may have left the company to switch to the products of another company. However, a company may also classify certain customers as former customers if they are not responsive to promotions and campaigns, thereby leading to losses to the company for no value in return. If a customer delays or defaults on payments, or participates in fraud, including unfair abuse of the company's policies, the company may choose to stop approaching the customer for further sales and to classify them as former customers, which is the last stage in the customer life cycle. Former customers are rarely targeted for CRM campaigns [2].

As a customer's perception of the company is essential to building relationships, and thus ensuring customer retention, CRM involves a range of inbound and outbound processes, which are often managed by dedicated software packages and accompanying processes, collectively known as operational CRM systems. Operational CRM systems support the management of customers, sales and marketing and customer service, and build a consistent relationship involving all customer touchpoints [1].

Data analysis is central to customer-retention activities, such as choosing a target demographic of customers, managing appropriate interaction at various customer touchpoints, and effective management of all inbound and outbound customers. Such analysis uses data collected by operational CRM systems to fulfil CRM objectives by delivering the right message to the right customer [3].

II. Imbalanced Data

Over the past years, class imbalance issue has been widely discussed in the literature [4], [5]. This is because data mining

models usually tend to be influenced with the majority class, therefore the minority class usually is misclassified leading to bad performance and low predictive accuracy. Solving this problem is a major challenge especially in customer related data such churn prediction where the class of interest is the minority and customer response in direct marketing for a product or a service. This section will investigate the suggested solutions for this major issue at the data level and the algorithmic level.

A. Major problem in direct-marketing dataset

Companies that choose direct marketing as a way to promote their products or services are interested in classifying potential buyers. In direct marketing, potential buyers are usually classified as a minority class. Data analysis techniques and machine learning algorithms tend to be biased toward the majority class; therefore, the minority class is usually misclassified, leading to a bad performance and low predictive accuracy [6]. Solving the problem of class imbalance is a major challenge, especially in cases where companies are interested in the minority class of customers, such as in churn prediction [7], which is one of the well-known big data application in business, where the aim is detecting potential customers who are likely to cancel a subscription to an already used service. The next section will explore suggested solutions for overcoming class imbalance.

1) Data Level Solution: The issue of class imbalance has received attention in the literature[8], [9], [10], [11], [6], [12]. Various solutions have been suggested to overcome the problem. At the data level, solutions work by applying various sampling techniques to balance the dataset. Different researchers have suggested varying techniques to overcome data imbalance, these include: oversampling the minority class, undersampling the majority class and a combination of both [8], [11], [13].

This process of randomly oversampling or undersampling has its disadvantages. Random oversampling increases the minority examples. The process has performed well in some studies when compared to other sampling methods [14], however, as the process only replicates existing examples, it has been argued that random oversampling does not add *actual data* to the training set for the classifier to learn from. This makes the decision region for the minority class very specific, which can cause over-fit. Another way is to randomly undersample the dataset by decreasing the instances in the majority class. Similar to random oversampling, the process has performed well in some cases, but it exhibited the disadvantage of removing useful information from the training dataset.

Nevertheless, despite the above-mentioned issues, sampling is still used to address class imbalance issues. Various researchers have made valuable progress in finding better ways to undersample and oversample. In Chawla [12], a model has been proposed using an approach where the minority class is oversampled by creating a synthetic example. The approach, called Synthetic Minority Oversampling Technique (SMOTE), takes each minority class sample and creates a synthetic example along the line segments using minority class *k*-Nearest Neighbour (*k*-NN). The suggested approach can be combined with random undersampling to increase the model's overall efficiency.

2) Algorithmic Level Solution: Algorithmic level solutions include adjusting costs to training examples [15]. Some learning algorithms assume that the misclassification costs are the same, but that is not true in all cases; For example, in applications for direct marketing, the cost of not calling buyers is more than the cost of calling non-buyers. This problem can be solved by assigning different costs to the samples of the classes or adjusting the probabilistic estimate at the tree leaf when using decision tree, which makes the minority samples more important than the majority samples when performing the training phase. Assigning different penalty constants for the positive samples and negative samples has proven to be effective [16]. Moreover, numerous classification methods have been proven to perform well with unbalanced datasets. These include adjusted *k*-NN, Support Vector Machine (SVM) and genetic programming [17].

SVM is a widely used machine-learning classifier. Many applications have reported using SVM models such as bio-informatics [18], churn prediction [7], bankruptcy prediction [19] and many other real-world applications. SVM has advantages that gives it unique features, including speed and high accuracy [20]. It has also been found that SVM produces optimal results with imbalanced datasets. Generally, an SVM classifier, when trained on an imbalanced dataset, produces a model that is biased towards the majority class and has low predictive accuracy on the minority class. There are different data preprocessing and algorithmic solutions to deal with this problem. Batuwita and Palade [21] present the existing techniques proposed in the literature to handle the class imbalance problem for SVM such as data preprocessing methods and algorithmic methods.Data preprocessing methods include re-sampling that can be applied to balance the data and algorithmic methods include adjusting the missclassification cost for the positive and the negative examples There are also some papers that applied both methods on SVM such as in Akbani, Kwek and Japkowicz [22], where the authors combined SMOTE sampling to balance the data with the Different Error Cost (DEC) method in SVM to overcome the class imbalance problem.

In summary, there is no clear and general answer as what the best method might be to apply to imbalanced datasets. Weiss, McCarthy and Zabar [23] compared cost-sensitive learning and sampling to find which is better at handling class imbalance. The authors used three models to deal with the issue on 14 datasets with different minority class percentages. The first model adjusted the misclassification cost in the learning algorithm and the other two models incorporated over-sampling and undersampling to balance the data. It was found that the choice depended on the data; for example, in datasets with $10,000$ instances, cost-sensitive learning outperformed sampling. Maloof [24] also compared cost-sensitive learning with sampling on one dataset and found that they all performed nearly the same, but this assumption cannot be generalised because only cases of class distribution were compared.

3) Performance Measurement: In terms of the model's performance measurement, the predictive accuracy rate (Acc) (see Eq. 1) is not an effective evaluation tool for a model on imbalanced datasets because it does not show how the

model correctly classified the minority-class instances. As for churn-prediction models [7], because the class of interest was small, accuracy was not the best evaluation metrics. Hit rate, churn rate and lift rate were considered by the authors when evaluating the model against other machine-learning approaches:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

where TP is true positive, TN is true negative, FN is false negative and FP is false positive.

Suggested performance measurements are sensitivity and specificity in [25]. Sensitivity (see Eq. 2), also known as the true positive rate (TPR) or recall rate, refers to the portion of positive instances correctly classified and calculated using the following equation:

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

Specificity (see Eq. 3), also known as true negative rate (TNR), refers to the portion of negative instances that are correctly classified using the following equation:

$$TNR = \frac{TN}{TN + FP} \quad (3)$$

Another important evaluation metric that can be used is the geometric mean (G-mean) of sensitivity and specificity (see Eq. 4) [33]:

$$G - Mean = \sqrt{TPR \times TNR} \quad (4)$$

Other suggested performance metrics are the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) [12]. The ROC curve is created by plotting the true positive rate (y-axis) against the false positive rate (x-axis) that is can be calculated as $1 - TNR$ which shows the portion of the misclassified examples. It is an ideal performance measurement to compare more than one classifier on the dataset. The perfect point on the ROC curve would be (0,100), that is when all positive instances are classified correctly and no negative instances are misclassified as positive. AUC is a performance metric generated from the ROC curve,

III. EXPERIMENTS AND RESULTS

Some studies have found that a combination of data level and algorithmic level solutions can lead to better [26]. The proposed HybridDA model uses SVM algorithm for classification purposes Additionally, to balance the data, this will be combined with both SMOTE to create a synthetic example of the minority class and random undersampling for the majority class. For a better result, parameters like C (misclassification cost) , gamma and the kernel type will be optimised using a grid search, which is a standard way to optimise parameters where a range of parameters is specified manually before running the process.

The data available at the UCI Machine Learning Repository[1] is used to apply the proposed model related to a direct

[1] http://mlr.cs.umass.edu/ml/datasets/Bank+Marketing

marketing campaign of a Portuguese banking institution. This dataset is collected in an attempt to get its clients to subscribe to a term deposit using phone calls as a means of reaching potential clients and imporve the quality of CRM in banks [27]. It contains 21 attributes and 4119 examples (see Table I for the list of the attributes).

TABLE I.   ATTRIBUTES LIST

Attributes	Name	Description
1	Age	In integer
2	Job	Type of job
3	Marital status	Married, single, divorced or unknown
4	Default	Has a credit
5	Balance	Average yearly balance
6	Housing	If there is a housing loan
7	Loan	Has a personal loan
8	Contact	Type of communication
9	Day	Last contact day of the month
10	Month	Last contact month of the year
11	Duration	Duration of last contact in second
12	Campaign	Number of contacts performed during this campaign and for this client
13	Pdays	Number of days passed since last campaign contact
14	Previous	Number of contacts performed for this client from previous campaigns
15	P-outcome	Previous outcome results
16	Emp-Var-rate	Employment-variation rate
17	Con-Price-index	Consumer price index
18	Cons-Conf-index	Consumer confidence index
19	Euribor3m	Euribor 3-month rate
20	No Of employed	Number of employees
21	Subscription	Label class Target: has the client subscribed

In the data, the target variable is subscription, a binomial type of attribute that is set as the label class in the model. From the meta data, it has been found that the number of subscribers is dramatically below that of the non-subscribers, which causes a class imbalance, as shown in table II.

TABLE II.   THE LABEL CLASS BEFORE BALANCING THE INSTANCES

Label	Number of instances
Yes	451
No	3668

A. Data Level Solution

There are several methods to deal with the imbalance in the dataset. The proposed model will focus on balancing the data by undersampling the majority class and oversampling the minority class.

1) SMOTE: To oversample the minority class, the model will implement a technique known as the Synthetic Minority Oversampling Technique (SMOTE) where the parameters are initialized as follows:

- Class: set to zero to detect the minority class automatically.

- Nearest neighbours: set to 5, this will create synthetic instances from the five nearest neighbours.

- The percentage of instances to create is 458.

- The number of seeds used for sampling is 0.

2) Random undersampling of the majority class: After applying the SMOTE algorithm, randomising the instances is required, especially in some cases (e.g. when applying 10-fold cross-validation where some folds will have too many positive instances or too many negative instances). After applying the randomization process, the majority class has to be randomly undersampled to balance the data further. As a result, the data are balanced and the minority class is now only slightly smaller than the majority class (See table III).

TABLE III. THE LABEL CLASS AFTER BALANCING THE INSTANCES

Label	Number of instances
Yes	2516
No	2550

3) Algorithmic Level Solution: At the algorithmic level, solutions to fix imbalanced datasets include adjusting the cost measurement to consider the class imbalance, or adjusting the probabilistic estimate at the tree leaf when using a decision tree. HybridDA model uses SVM and a grid search to optimise the C, gamma and the kernel type. As suggested in [28], the range for C has been defined as $[2^{-5}, 2^{15}]$ and the range of gamma as $[2^{-15}, 2^3]$. After evaluating all possible combinations, the optimised results evaluated by the grid search are shown in Table IV.

TABLE IV. THE OPTIMIZED PARAMETERS COMBINATION USING GRID SEARCH

Parameter	Optimised value
Kernel	Poly
Gamma	4.800
C	3276.828

4) Results: In cases of imbalanced data and the class of interest being the minority class, predictive accuracy is not the best performance indicator. In this work, the true positive rate will be used as a performance measurement as the higher the recall rate, the better the class of interest is classified. To complement TPR, TNR, the accuracy and AUC will also be used. The proposed model presents an improved predictive model with the following results:

- Acc = 96.73%
- TPR = 97.93%
- TNR = 94.82%
- AUC = 0.98

As discussed in the next section, the proposed system outperforms previously reported results conducted on the dataset. In summary, This has been achieved by applying SMOTE, which creates more synthetic examples to learn from, undersampling the majority class to balance the data, and then combining this with the optimised values for C, gamma and the kernel type in SVM. A flowchart of the proposed system is shown in Figure 1.

IV. COMPARISON WITH OTHER TECHNIQUES

There has been some relevant researches on various versions of the dataset used in this paper. For example, Moro, Cortex and Laureano [29] implemented the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology on previous versions of the dataset. The CRISP-DM methodology allows the building of a Data Mining (DM) model in non-grid six phases that can be used in a business environment. The study implemented three classification methods using an R package for data mining: Decision tree, Naïve Bayes and SVM. The algorithms have been evaluated using both ROC and lift curve. According to the evaluation metrics, SVM gave the best results with AUC value and Area Under The Lift Chart (ALIFT) value equal to 0.9 and 0.8. In another research few years later, Moro, Cortez and Rita [27] used four data-mining classification methods: Decision tree(DT), neural networks (NN), SVM and logistic regression (LR). Two evaluation metrics were also used: AUC and ALIFT. The models were tested on the most recent contacts and NN preformed the best with AUC=0.8 and ALIFT=0.7. In terms of simplicity, LR and DT produce a more understandable model while giving good results. However, SVM and NN are more flexible and have better learning capabilities, thus giving a better results. At the beginning of the above-mentioned study a semi-automated approach was used to reduce the attributes from 150 to 22. This was done in two steps. In the first step the attributes were analysed from a business perspective and in the second step the forward selection method was implemented.

Vajiramedhin and Suebsing [30] proposed a model on the same dataset, focusing on using correlation-based feature subset selection algorithm and a dataset balancing technique. The balancing technique is used to make the dataset label equivalent by randomly selecting the dataset of each label equally. For the correlation-based feature subset selection algorithm for feature correlation measurements, the paper proposed a model that implements a C4.5 algorithm. The proposed model scored a high true positive rate of 92% and an ROC rate of 95%, when compared with other methods where balancing or feature selection was excluded.

Another model by G. Feng et al. [31] proposed on the direct marketing dataset combines Bayes networks (BNs). The experiment started by combining two BNs, then three, then more than three. All average accuracies were compared and provided improvement in the results with average accuracies of 0.83.

In another piece of research, Elsalamony [32] focused on increasing the effectiveness of the marketing campaign by finding the major attributes that affected the success of the phone call. The author compared three classification methods: multilayer perception neural network (MLPNN), Bayesian networks, logistic regression (LR), and C5.0 on the direct marketing dataset and found that C5.0 gave the best results, with the testing part scoring:

- Acc = 90.09%
- TPR = 59.06%
- TNR = 93.23%

The author also found that (duration) was the most important

attribute from C5.0, MLPNN and LR, and that (age) was the most effective attribute from BN.

Another model that has been proposed in [33] is an example-dependent cost-sensitive decision-tree algorithm where the direct marketing dataset has been used with two other datasets. In this model, the Cost-Sensitive Decision Tree (CSDT) has been evaluated against the standard decision tree with three different cases: without pruning, with error-based pruning and cost-sensitive pruning training. The three different tree algorithms have been trained using the training, undersampling, cost-proportionate rejection sampling and cost-proportionate sampling dataset. The results show that the proposed learning algorithm is the best performing for all three databases and the direct marketing dataset accuracy of 88.28% and the F1-score of 0.35. This model has been used further to measure the cost savings.

V. Conclusion

This paper proposes a new approach in addressing imbalanced data by using a combination of both Data-level and Algorithmic-level solutions. One such imbalanced dataset is used in the experiment to evaluate the proposed method using this dataset, different researchers have taken different approaches to classifying potential responder. While most researches have compared different classification algorithms on the dataset some focused on finding the most effective attribute by applying different classification algorithms. The experiment reported in this paper provides promising results when evaluated against previous work on the same dataset. This is likely to be caused by focuing on minimising the imbalance effect on the learning algorithm without affecting identification of the class of interest of customers who subscribed to the term deposit. SMOTE has been chosen to oversample the minority class in order to create more synthetic examples, which is better than random oversampling (causing over fitting). Random undersampling has been applied only to the majority class to balance the data further. To complement the sampling process, an algorithmic solution has been applied by adjusting the cost, the kernel type and the gamma. The adjustment has been made by applying the standard grid search to optimise the parameter combination. Therefore, the proposed model, HybridDA, uses a combination of data and algorithmic level solutions to handle the existing class imbalance in the dataset. The result of this hybridisation demonstrates a competitive performance. It is suggested that balancing the data using a more powerful random oversampling technique and combining it with random undersampling to balance the data, then applying a cost-sensitive learning algorithm where parameters are optimised, worked well with imbalanced data. Given the proposed model is not generalised, caution should be exercised when applying the system to the other datasets. As such, future work will focus on testing the suggested model on the large dataset for the Portugues marketing campaign, and other datasets from different fields as well as using other techniques at the data level.

References

[1] H. Edelstein, "Building profitable customer relationships with data mining," in Customer Relationship Management, SCN Education B.V. Eds, Braunschweig: BertelsmannSpringer, 2001, pp. 339- 351.

Fig. 1. The Proposed Model Flowchart
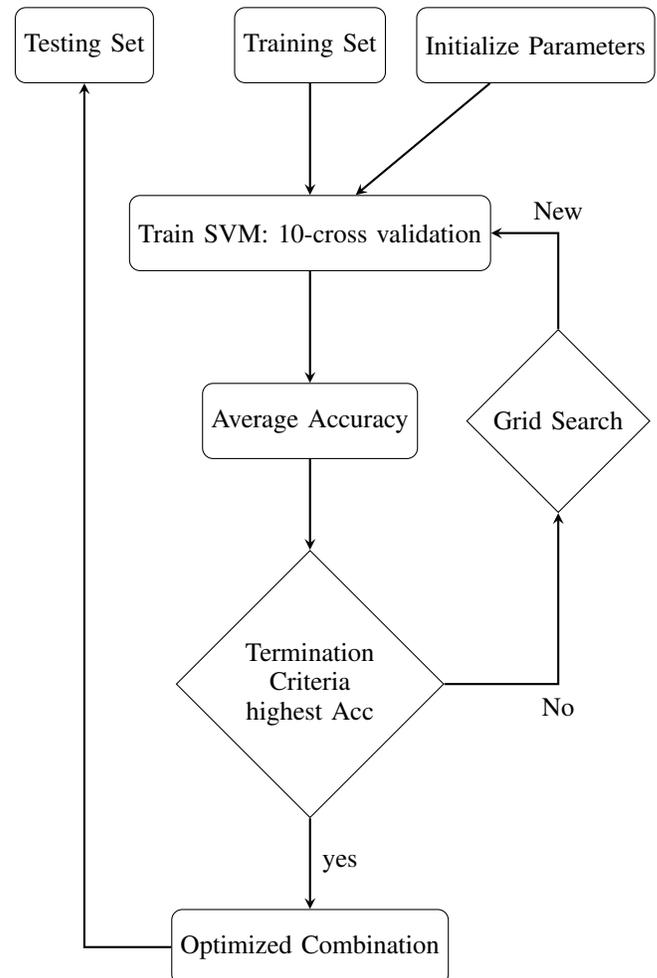
[2] C. W.Rygielski, "Data mining techniques for customer relationship management," Technology in Society, vol. 24, no. 4, pp. 483- 502, 2002.

[3] K. Tsiptsis and A. Chorianopoulos, Data mining techniques in CRM: inside customer segmentation. New YorK: Wiley, 2011.

[4] Garca, Vicente, Javier Salvador Snchez, and Ramn Alberto Mollineda. "On the effectiveness of preprocessing methods when dealing with different levels of class imbalance." Knowledge-Based Systems 25, no. 1 (2012): 13-21.

[5] Drown, Dennis J., Taghi M. Khoshgoftaar, and Ramaswamy Narayanan. "Using evolutionary sampling to mine imbalanced data." In Machine Learning and Applications, 2007. ICMLA 2007. Sixth International Conference on, pp. 363-368. IEEE, 2007.

[6] Ling, C. X and C. Li, "Data mining for direct marketing: problems and solutions," in Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, 1998.

[7] Ali Rodan, H. Faris, J. Alsakran and O. Ali-Kadi, "A support vector machine approach for churn prediction in Telecom industry", The Scientific World Journal, vol. 17, no. 8, pp. 3961- 3970, 2014.

[8] V. Nitesh and K. W. Chawla, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 1, pp. 321- 357, 2002.

[9] H. He and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263- 1284, September 2009.

[10] Hui Han, Wen-Yuan Wang and Bing-Huan Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in

TABLE V.   SUMMARY OF THE RESULTS FOR PREVIOUS MODELS ON THE DIRECT MARKETING DATASET

Models	AUC	ALIFT	Acc	TPR	TNR	ROC
Moro et al. [29]	0.938	0.887	NA	NA	NA	NA
Moro et al. [27]	0.8	0.7	NA	NA	NA	NA
Vajiramedhin and Suebsing [30]	NA	NA	NA	NA	92.14%	95.6%
G. Feng et al. [31]	NA	NA	83%	NA	NA	NA
Elsalamony [32]	NA	NA	90.09%	59.06%	93.23%	NA
Bahnsen et al. [33]	NA	NA	88.28%	NA	NA	NA
HybridDA	0.98	NA	96.73%	97.93%	94.82%	NA

Advances in Intelligent Computing, Berlin: Springer, 2005, pp. 878-887.

[11] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," in Proceedings of the Fourteenth International Conference on Machine Learning, 1997, Morgan Kaufmann, pp. 179-186.

[12] N. V. Chawla, "Data mining for imbalanced datasets: an overview," in Data Mining and Knowledge Discovery Handbook, Oded Maimon and Lior Rokach, Eds. New York: Springer-Verlag, 2005, ch. 40, pp. 853-867.

[13] N. Japkowicz, "Learning from imbalanced datasets: a comparison of various strategies," in Papers from the AAAI Workshop, Technical Report WS-00-05, 2000, pp. 10- 15.

[14] G. E. A. P. A Batistaa, R. C. Prati and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data", SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 20- 29, 2004.

[15] P. Gaspar, J. Carbonell and J. L. Oliveira, "On the parameter optimization of support vector machines for binary classification". Journal of Integrative Bioinformatics, vol. 9, no. 3, 2012.

[16] T. H. Nguyen, Z. Gantner and L. Schmidt-Thieme, "Cost-sensitive learning methods for imbalanced data," in Proceedings of International Joint Conference on Neural Networks (IJCNN), 2010.

[17] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets," International Journal of Emerging Technology and Advanced Engineering, vol. 2, pp. 42- 47, 2012.

[18] B. Schlkopf, K. Tsuda and J. P. Vert , Eds., Kernel Methods in Computational Biology. MIT Press series on Computational Molecular Biology, 2004.

[19] C.H. Wu, G. T. Tzeng, Y. J. Goo and W. C. Fang, "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," Expert Systems with Applications, vol. 32, no. 2, pp. 397- 408, 2007.

[20] G. Q. Zhu, S. R. Liu and J.S. Yu, "Support vector machine and its applications to function approximation," Journal of East China University of Science and Technology, vol. 5, pp. 555- 559, 2002.

[21] R. Batuwita and V. Palade, "Class imbalance learning methods for support vector machines", in H. Hee and Y. Ma, Eds, Imbalanced Learning: Foundations, Algorithms, and Applications. New York: Wiley, 2012, ch. 6.

[22] R. Akbani, S. Kwek and N. Japkowicz, "Applying support vector machines to imbalanced datasets," European Conference on Machine Learning, 2004.

[23] G. M. Weiss, K. McCarthy and B Zabar, "Cost-sensitive learning vs. sampling: which is best for handling unbalanced classes with unequal error costs," in Proceedings of the International Conference on Data Mining, DMIN 2007. Las Vegas, Nevada, USA: CSREA Press, pp. 35-41.

[24] M.A. Maloof, "Learning when data sets are imbalanced and when costs are unequal and unknown," ICML 2003 Workshop on Imbalanced Datasets.

[25] Tang, Yuchun, Yan-Qing Zhang, Nitesh V. Chawla, and Sven Krasser. "SVMs modeling for highly imbalanced classification." Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 39, no. 1 (2009): 281-288.

[26] J. Burez, J. and D. Van den Poel, "Handling class imbalance in customer churn prediction", Expert Systems with Applications, vol. 36, no. 3, pp. 4626- 4636, 2009.

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

[28] Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." (2003).

[29] S. Moro, P. Cortex and R. Laureano, "Using data mining for bank direct marketing; an application of the CRISP-DM methodology," in Proceedings of the 25th European Simulation and Modelling Conference ,ESM'2011, pp. 117- 121, October 2011.

[30] C. Vajiramedhin and A. Suebsing, "Feature selection with data balancing for prediction of bank telemarketing", Applied Mathematical Sciences, vol. 8, no. 114, pp. 5667- 5672, 2014.

[31] G. Feng, J. D. Zhang and S. S. Liao, "A novel method for combining Bayesian networks, theoretical analysis, and its applications," Pattern Recognition, vol. 47, no. 5, May 2014, pp. 2057- 2069.

[32] H. A. Elsalamony, "Bank direct marketing analysis of data mining techniques," International Journal of Computer Applications, vol. 85, no. 7, January 2014.

[33] A. C. Bahnsen, D. Aouada and B. Ottersten, "Ensemble of example-dependent cost-sensitive decision trees," arXiv preprint arXiv:1505.04637y 2015.