Predictive Analytics for High Business Performance through Effective Marketing

Supriya V. Pawar
Department of Computer
Engineering, Sinhgad Academy of
Engineering (SAE), Pune, India

Gireesh Kumar
Department of Computer
Engineering, Sinhgad Academy of
Engineering (SAE), Pune, India

Eashan Deshmukh
Department of Computer
Engineering, Sinhgad Academy of
Engineering (SAE), Pune, India

ABSTRACT

With economic globalization and continuous development of e-commerce, customer relationship management (CRM) has become an important factor in growth of a company.CRM requires huge expenses. One way to profit from your CRM investment and drive better results, is through machine learning. Machine learning helps business to manage, understand and provide services to customers at individual level Both customer segmentation and buyer targeting help the business to increase marketing performances.

The objective is to propose a new approach for better customer targeting.

Keywords: Customer Relationship Management (CRM) , Machine Learning , Customer Segmentation ,Customer sTargeting ,K-means algorithm ,Smote, Logistic Regression, Classification , Clustering

I.   INTRODUCTION

CRM requires a big expense in the form of Implementation, updates, and training. One way to improve ROI and drive better outcomes from the huge data available from sales and marketing, customer support, is through implementing machine learning on top of existing CRM systems. Thus Predictive CRM is a system that gathers both internal and external data about prospects to predict which accounts are more likely to buy. Customer targeting is one of the most important components of the customer relationship management (CRM) systems. Customer targeting identifies promising prospects to increase revenue. Improving customer targeting is important for reducing overall cost and boost business performances. Marketing professionals achieve this tasks using a classification method for buyer targeting. We will propose a hybrid algorithm that will improve Customer Targeting Performance.

1.1 Analysis Scenario:

For identifying prospective customers It is important to measure a subject's "propensity to buy" a particular product . We can take advantage of the large amount of demographic data to target only those who have the highest propensity to buy thus increasing our chance of success[9].

We will devise a method that exploits the customer data in conjunction with the demographic data from the overall market population that contains buyer vs. non-buyer data, using hybrid algorithms to increase customer targeting by improving classification performance.

II. Related Work

One of the key problem in CRM is buyer targeting, that is, to identify the prospects that are most likely to become customers. Marketers are applying data

mining tools to solve the problem, such as in [1] the authors focused on classification of online customers based on their online website behaviors, and [2] applied neural networks guided by genetic algorithms to target households. [3] proposed a new feature selection technique .In this work the classification performance of C4.5 Decision Tree , NaiveBayes classifier, SVM classifier and KNN classifier was compared, SVM classifier is found working best with this methodology.[4] proposed a hybrid algorithm that uses the concept of clustering and decision tree induction to classify the data samples. This approach solves issues of burdening decision tree with large datasets by dividing the data samples into clusters. In [5] the author suggested a customer classification and prediction model in commercial bank that uses collected information of customers as inputs to make a prediction for credit card proposing. She implemented Naive Bayesian classifier algorithm[8] developed individually tailored predictive models for each segment to maximize targeting accuracy in the direct-mail industry. In such a step-by-step approach, the buyer targeting (the second step) becomes dependent on the results of customer segmentation (the first step). However, the customer segmentation has to be implemented independently.[9] proposed to first use K-Means clustering to segment customers and then build the segment-wise predictive models for better targeting the promising customers. In [12] customer segmentation and buyer targeting as a unified optimization problem was formulated as a single problem. The integrated approach not only improves the buyer targeting performances but also provides a new perspective of segmentation based on the buying decision preferences of the customers. A new K-Classifiers Segmentation algorithm was developed to solve the unified optimization problem.

III. Algorithms

The algorithm k-means was used for customer targeting[12] and for feature selection in [4].SMOTE algorithm was used for handling class imbalance in [11].Logistic Regression was used as a benchmark for the comparative analysis of RFM and FRAC methods in [13].This algorithms are elaborated in detail.

A. K means Clustering:-

The real life datasets has multiple number of features[4].Grouping these features on the basis of similarity is required. Clustering is an unsupervised method of separating a large number of data into subsets of similar characteristics. Different clustering methods can generate different groupings for same set of data samples. Clustering can be broadly classified as partition based and hierarchical based. Some examples of the techniques used for partition based clustering are k-means and k medeoids. The algorithm proposed in this paper uses k-means algorithm for feature selection.

B. Logistic Regression:-

In the logistic regression model, the predicted values for the dependent variable will always be greater than (or equal to) 0, or less than (or equal to) 1. [10].

The name logistic stems from the fact that one can easily linearize this model via the logistic transformation. Suppose we think of the binary dependent variable y in terms of an underlying continuous probability p , ranging from 0 to 1. We can then transform that probability p as:

$$ p' = \log_e \left(\frac{p}{1-p} \right) $$

Logistic regression is very useful for several reasons: (1) logistic modeling is conceptually simple; (2) easy to interpret as compared to other methods like ANN (3) logistic modeling has been shown to provide good and robust results in comparison studies[6].For database marketing applications, it has been shown by several authors [7]that logistic modeling may outperform more sophisticated methods.

C. SMOTE:-

An over-sampling approach in which the minority class is over-sampled by creating "synthetic" examples rather than by over-sampling with replacement[2,11].The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any or all of the k minority class nearest neighbors[11,2]. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen[2].

VI. CONCLUSION

Previous research mainly focus on providing a general predictive model for the total customer base, .Our work in this paper is an attempt to unify the supervised and the unsupervised learning methods for

better customer targeting.Is it possible to get maximum ROI ,by using hybrid algorithms to exploit buyer vs non-buyer customer data from the overall market population for customer targeting by improving the classification performance? This is the research question we would like to answer. **A case study on a real world marketing data will be used for evaluating the the performance of the proposed approach .**

REFERENCES

1. imitris Bertsimas, Adam J Mersereau, and Nitin R Patel. Dynamic classification of online customers. In SDM, pages 107–118. SIAM, 2003.

2. Yong Ge, Hui Xiong, Wenjun Zhou, Siming Li, and Ramendra Sahoo. Multifocal learning for customer problem analysis. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3):24, 2011.

3. S.Ummugulthum Natchiar. Customer Relationship Management Classification Using Data Mining Techniques .In ICSEMR ,2014.Page(s):1 - 5 , ISBN: 978-1-4799-7613-**3**

4. Akanksha Ahlawat , Bharti Suri, Improving Classification in Data mining using Hybrid algorithm . In IICIP-2016 –DTU .Pages: 1-4

5. Gao Hua, Customer Relationship Management Based on Data Mining Technique. IN CRMISSN:2051-0845, Vol.47, Issue. Page(s):1416-1420

6. SCOTT A. NESLIN, SUNIL GUPTA, WAGNER KAMAKURA, JUNXIANG LU, and CHARLOTTE H. MASON ,Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models.,Journal of Marketing ResearchVol. XLIII (May 2006), 204–211.

7. NissanLevinJacobZahaviContinuous,Predictive Modeling a Comparative Analysis.JOURNAL OF INTERACTIVE MARKETINGVOLUME 15 / NUMBER 2 / SPRING 2001.

8. Chidanand Apte, E Bibelnieks, Ramesh Natarajan,E Pednault, Fateh Tipu, Deb Campbell, and Bryan Nelson. Segmentation-based modeling for advanced targeted marketing. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 408–413. ACM,2001

9. Paul B Chou, Edna Grossman, Dimitrios Gunopulos,and Pasumarti Kamesam. Identifying prospective customers. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 447–456. ACM, 2000.

10. Hill T and Lewicki P (2007 STATISTICS Methods and ApplicationsTulsa, OK: StatSoft)

11. NiteshV.Chawla, KevinW. Bowyer, LawrenceO.Hall, W.Philip Kegelmeyer, SMOTE: Synthetic Minority Over-sampling Technique,Journal of Artificial Intelligence Research 16 (2002) 321–357.

12. Jingyuan Yang,Chuanren Liu,Mingfei Teng,Hui Xiong,March Liao,Buyer Targeting Optimization: A Unified Customer sSegmentation Perspective. In International Conference on Big Data (Big Data).(2016) 1262-1271