

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. Thus propensity modeling helps the business in increasing marketing performance. The objective is to propose a new approach for better customer targeting.

We'll device a method to improve prediction capabilities of existing CRM systems by improving classification performance for propensity modeling.

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

I. INTRODUCTION

CRM requires a big expense in the form of Implementation, updates, and training. Implementing MI on top of crm system will not only increase ROI but will also enable us to derive better results from the huge collection of data available from sales and marketing, customer support. This will enable us to achieve predictive crm which could gather both internal and external data about different prospects

and predict the likelihood of accounts more likely buy. Customer targeting is one of the most important components of the customer relationship management (CRM) systems. Customer targeting helps in identifying promising prospects to generate more revenue. Improving customer targeting is important for reducing overall cost and boosting business performances. Marketing professionals achieve this task using a classification model for buyer targeting.

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 available to target those who have the highest propensity to buy thus improving our chances of success [9]. We will devise a method to exploits the customer data in conjunction with the demographic data from the overall market population containing buyer vs. non-buyer data.

This will increase the predictive accuracy of the classification model and improve Customer Targeting Performance.

II. Related Work

One of the key problems 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 was found to be working best with this methodology. [4] proposed a hybrid algorithm which 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.[12] In 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. Preliminaries

The method proposed here uses the concepts of clustering and classification with class imbalance problem. These concepts are further explained below. 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.

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 with 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 medoids. 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:

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

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 oversampling required, neighbors from the k nearest neighbors are randomly chosen[2].

VI. Proposed Model:

Now, with the advancement in technology it has become easier to manage business relationships and the informative data associated with it. Data like customer contact details, accounts and lead information, sales opportunities in different geographical regions is stored in CRM systems mostly in cloud at one central location, such that the data is available to many at real time with ease and speed. As the amount of data generated from sales and marketing, customer support and product development departments is increasing exponentially, the existing CRM systems potential to generate more revenue has also increased.

Deriving insights from the available data is important not only to get more profit but also to save our efforts and time. After all huge data doesn't necessarily mean better decision making.

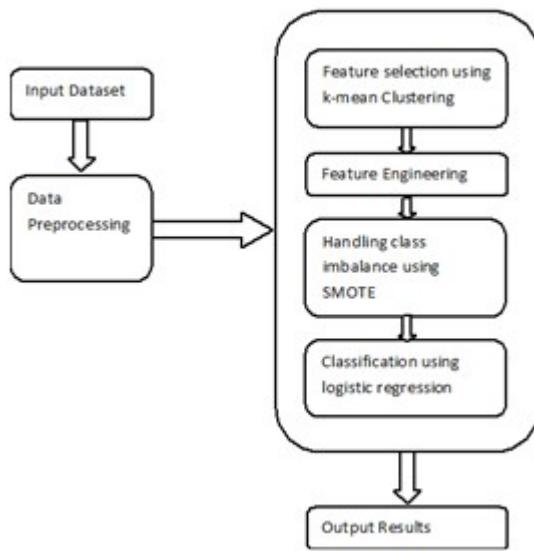


Fig. 1 shows the block diagram representation for the proposed algorithm.

In this paper, we have used algorithms K-means ,Smote and Logistic Regression for customer targeting .Correct logical combination of this algorithms improves the performance metrics of the predictive model.

VI. Case Study

Let us consider a case study involving a marketing manager of a large company selling bikes , must run a mailing campaign to find the best prospective customers which is direct marketing .

To build a buyer propensity model , we require data which must contain:-

- Sales transactions data .The data about our customers including those who bought bikes in the past.
- Data from previous marketing campaigns.
- List of customer who Purchased from third-party vendors.

This dataset has been collected From Azureml gallery at <http://gallery.azureml.net/> The dataset has :

Dataset	Observations	Features
BikeBuyer	10000	18

Missing values = 9% and above.

Class Imbalance = 85.039% Majority class and 14.96% Minority class.

VII. Working

Handling Class imbalance during data cleansing and processing:

This problem occurs in machine learning when one of the class of data is much lesser than the other class of data. This produces a biased classifier .This problem is very common and commonly observable in disciplines like fraud detection.

It is a combination of oversampling and under sampling.

- Firstly we identify the majority class in the given dataset. Then we under sample the majority class by simply removing some of the samples from the majority class .This is done until the minority class becomes some specified percentage of the majority class.
- After having done so , we construct the minority class by oversampling it i.e generating new class data instances using an algorithm SMOTE.
- SMOTE helps in producing synthetic instances of the minority class. Thus SMOTE (Synthetic Minority Over-sampling Technique) enable learning from imbalanced datasets.

SMOTE is widely used because its simple and effective .The new data is given as input to k-means.

Feature selection:-

K-means algorithm is applied for feature selection. Then reduced dataset is given as an input to two class Logistic Regression classifier. The proposed method consists of the following steps:

- The number of clusters k has been determined using average silhouette for the entire range of k as we specify .The mean of all the silhouette values is calculated for distinct clusters. The count of number of clusters having high sill value is picked up to be the value of k. We use euclidian distance as the distance measure of each data points from the centeriod of the cluster.
- After defining the value of k number of clusters, apply k-means on each of the features of the dataset individually .For each cluster we measure the clustering performance using the metric silhouette.
- We include the feature which gives us the best performance and add it to the final set of features. We repeat applying k-means on the remaining features individually and pick the feature which has the highest sill value.
- If we have reached the desired number of features, we stop otherwise we repeat the procedure as mentioned above. This way we have performed feature selection. The data is passed to logistic regression for classification.

	y=+1	y=-1
ŷ=1	TP	FP (Type I)
ŷ=-1	FN (Type II)	TN

Classification:-

We split our preprocessed data into training and test set.

We estimate the coefficients and train our model as follows –

$$f = \underset{\text{model } f}{\arg \min} \frac{1}{n} \sum_{k=0}^n l(y_i, f(x_i)) + \text{Regularization}(f)$$

Where $f(x)$ is chosen to be a linear model.

We use Logistic Regression as a classifier to train our model .We estimate the coefficients as

$$\hat{\beta} = \underset{\beta_1, \beta_2, \beta_3, \dots}{\arg \min} \frac{1}{n} \sum_{k=0}^n \log(1 + e^{-y_i \sum_{j=1}^p \beta_j x_{ij}})$$

Logistic Regression adjusts the value of B.

The predictive accuracy can be increased by regularization. Thus adjusting L2-Norm and L1-norm near about 0.01 gave me the highest predictive power.

- We score our model by computing score for each x_i in test set.

$$f(x_i) = \sum_{j=1}^p \hat{\beta}_j x_{ij}$$

Thus our model can be represented as

$$\text{Bike}_{Buyer} = \frac{1}{e^{-\beta_1 age + \beta_2 n_cars + \beta_3 commute_distance + \dots + n_children}}$$

Where $\beta_1, \beta_2, \beta_3$ etc. represent the coefficients of the independent variables.

e is the error which represents the variability in the data.

In this equation, Bike Buyer is the probability that a customer will buy a bike, given his or her input data. where Age, cars, and Communte_distance are the independent variables.

VIII. Result and Analysis

The method has been implemented in azure machine learning studio using R programming language.

We represent confusion matrix as:-

For evaluation of the quality of results i have used we have used the following metrics-

- Predictive accuracy:-

$$\frac{FP + FN}{n} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[y_i \neq \hat{y}_i]}$$

- Sensitivity:-

$$\frac{TP}{#POS} = \frac{\sum_i^n \mathbf{1}_{[y_i = \hat{y}_i \text{ and } y_i = 1]}}{\sum_i^n \mathbf{1}_{[y_i = 1]}}$$

c. Specificity:-

$$\frac{TN}{\#Neg} = \frac{\sum_i^n 1}{\sum_i^n 1} \quad [y_i = \hat{y}_i \text{ and } y_i = -1]$$

d. False Positivity Rate:-

$$\frac{FP}{\#Neg} = \frac{\sum_i^n 1}{\sum_i^n 1} \quad [y_i \neq \hat{y}_i \text{ and } y_i = -1]$$

e. Precision:-

$$\frac{TP}{\#Predicted\ Positive} = \frac{\sum_i^n 1}{\sum_i^n 1} \quad [y_i = \hat{y}_i \text{ and } y_i = 1]$$

f .F1-Score:-

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

g. Area under ROC Curve which gives us the True Positivity Rate for a particular False Positivity Rate.

True Positive	False Negative
177	5
False Positive	True Negative
4	410

Confusion matrix is as follows:-

Performance Parameter	Results
Accuracy	0.987
Precision	0.989
Threshold	0.6
Recall	0.967
F1-Score	0.978
Positive Label	Yes
Negative Label	No

X. CONCLUSION

The new proposed method resolves the real world buyer propensity modeling problem with noisy and highly imbalanced data .As we demonstrated our proposed method is an effective tool for direct marketing which can improve ROI. One way to subsidize from CRM investment to improve ROI and drive better outcomes is through implementing Machine Learning on top of CRM. In this paper we mainly focused on the research of the customer classification and prediction in Customer Relation

Management. The study will help the company to analyze and forecast customer's pattern of purchase. This method is applicable to industries like banking industry, insurance industry, retail industry, manufacture industries. we outlined a simplified method to guide marketers and managers in making focus their advertising and promotion on those categories of people in order to reduce time and costs. The case study we evaluated shows the practical use and usefulness of the model.

XI. REFERENCES

1. Dimitris 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 CRM:ISSN: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 M odeling 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 Applications Tulsa, 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 Segmentation Perspective. In International Conference on Big Data (Big Data).(2016) 1262-1271

