International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

Direct Marketing Campaign Response Analysis using Logistic Regression, CART and Support Vector Machines

Ankit Bansal

Leander, Texas, USA

Abstract: ***Direct marketing remains a pivotal strategy for businesses aiming to engage customers with personalized content. This article explores how analytics can enhance the effectiveness of direct marketing campaigns by predicting customer responses. By leveraging statistical models and machine learning, businesses can optimize their marketing efforts, reduce costs, and improve conversion rates. In this paper we will focus on comparison of multiple machine learning techniques and learn why SVM (Support vector machine) can be helpful in these type of use cases.***

Keywords: Direct Marketing, Customer Response, Analytics, Machine Learning, Predictive Modeling, Conversion Rate, Data Science, SVM, Support vector machine

1. Introduction

In the rapidly evolving landscape of digital marketing, predicting customer responses to direct marketing campaigns has become a critical focus for businesses aiming to optimize their outreach strategies and maximize return on investment. By leveraging machine learning techniques, organizations can analyze vast amounts of consumer data to identify patterns and predict which customers are most likely to engage with a campaign. This predictive capability not only enhances targeting precision but also helps in personalizing marketing efforts, reducing costs, and increasing conversion rates. As a result, businesses can make data - driven decisions that align closely with consumer preferences and behavior, ultimately leading to more effective marketing strategies and improved customer satisfaction.

2. Problem Statement

In the business landscape of competition and challenges companies grapple with the task of effectively managing their marketing budgets while maximizing the effectiveness of their advertising campaigns. A significant issue lies in the difficulty of predicting which customers will react positively to marketing initiatives resulting in resources being wasted on disinterested or non - targeted customer segments. This lack of precision does not drive - up marketing expenses. Also poses a risk of turning away potential customers with irrelevant promotions. By tackling this challenge head on businesses can refine their targeting strategies. Ensure that promotional messages reach the intended audience at the opportune moments. Using modeling with machine learning provides a way to analyze data in order to predict how customers will respond. This helps businesses enhance their marketing campaigns and boost conversion rates resulting in a return, on investment.

3. Solution

Solution steps are laid down in a flowchart diagram as below:

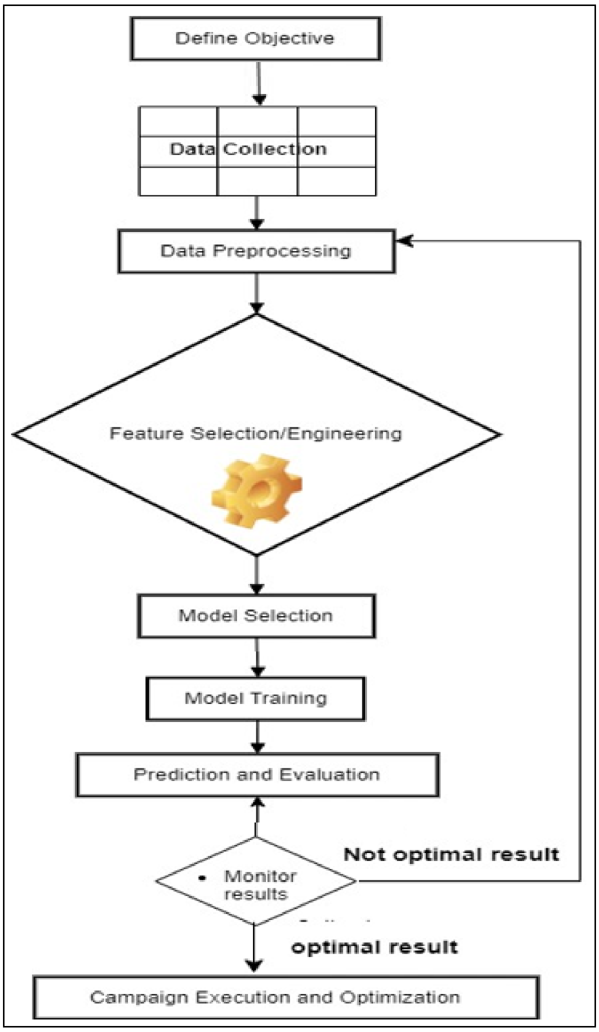

Figure 1: Model building Solution steps

4. Uses

Analytics - driven direct marketing can be utilized in various ways:

1) **Targeted Marketing:** Identify and focus on customers most likely to respond, improving the efficiency of marketing efforts.
2) **Personalized Content:** Tailor marketing messages to individual preferences, increasing engagement and conversion rates.
3) **Resource Optimization:** Allocate budget and resources more effectively by concentrating on high - response segments.
4) **Churn Reduction:** Predict and address potential customer churn through targeted retention strategies.
5) **Improved ROI:** Enhance return on investment by reducing spend on ineffective campaigns.
6) **Dynamic Pricing Strategies:** Implement personalized pricing models based on predicted customer responsiveness.
7) **Cross - Selling and Upselling:** Identify opportunities for additional sales by predicting customer interest in related products.
8) **Customer Segmentation:** Create more accurate customer segments for tailored marketing strategies.
9) **Campaign Timing:** Optimize the timing of campaigns to reach customers when they are most likely to respond.
10) **Feedback Loop for Product Development:** Use response data to inform product improvements and new product development based on customer interests and feedback.
    - **Customer Retention**: Identify at - risk customers and target them with specific retention strategies.
    - **Product Recommendations**: Suggest products based on past behavior, increasing cross - sell and up - sell opportunities.
    - **Churn Prediction**: Anticipate customer churn and intervene with targeted offers.
    - **Campaign Optimization**: Continuously refine marketing efforts based on response data.

5. Impact

Effective use of analytics in direct marketing can significantly improve business outcomes by:
- Increasing response rates and conversion rates.
- Reducing marketing costs through targeted strategies.
- Enhancing customer satisfaction with relevant content.
- Providing measurable insights for ongoing improvement.
- Improved ROI with optimize marketing strategies to achieve a better return on investment.
- Scalability by applying predictive models across various campaigns and channels, scaling marketing efforts efficiently.

6. Scope

1) **Retail:** Predict which customers are likely to respond to a seasonal sale, allowing for targeted promotions and personalized discounts.
2) **Telecommunications:** Identify subscribers who might upgrade their plans or purchase additional services, enabling targeted upsell campaigns.
3) **Banking and Finance:** Predict which clients are likely to respond to offers for new credit cards or loans, optimizing marketing efforts and reducing acquisition costs.
4) **E - commerce:** Forecast customer response to email marketing campaigns about new product launches or exclusive online deals.
5) **Healthcare:** Identify patients who are likely to respond to wellness program enrollments or vaccination reminders, improving patient engagement.
6) **Automotive:** Predict which car owners are likely to respond to service reminders or new vehicle promotions, enhancing customer retention and sales.
7) **Hospitality:** Target previous guests with personalized offers for upcoming holiday packages or loyalty programs.
8) **Insurance:** Determine which policyholders might be interested in additional coverage or new insurance products, facilitating targeted policy upgrades.
9) **Travel and Tourism:** Predict traveler interest in promotional offers for flights or vacation packages, increasing booking rates.
10) **Media and Entertainment:** Identify subscribers likely to respond to offers for premium content or event tickets, boosting subscription and ticket sales.

Real time case study was performed, the aim of this project was to build statistical models for predicting whether the client will subscribe to a term deposit or not using three modelling techniques

Application of Support vector machine

SVMs are supervised learning methods used for classification and regression tasks that originated from statistical learning theory . SVM is a global classification model that generates non - overlapping partitions and usually employs all attributes. The entity space is partitioned in a single pass, so that flat and linear partitions are generated. SVMs are based on maximum margin linear discriminants, and are similar to probabilistic approaches, but do not consider the dependencies among attributes.

One of the main concepts that will help us understand the functioning of Support Vector Machine is the Class separation. This is done by an optimal separating hyperplane between the two classes by maximizing the margin between the classes' closest points. A hyperplane is the extension of the two or three dimensional concepts of lines and planes to higher dimensions. This is shown in the figure below. Here, the points lying on the boundaries are called support vectors, and the middle of the margin is our optimal separating hyperplane.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

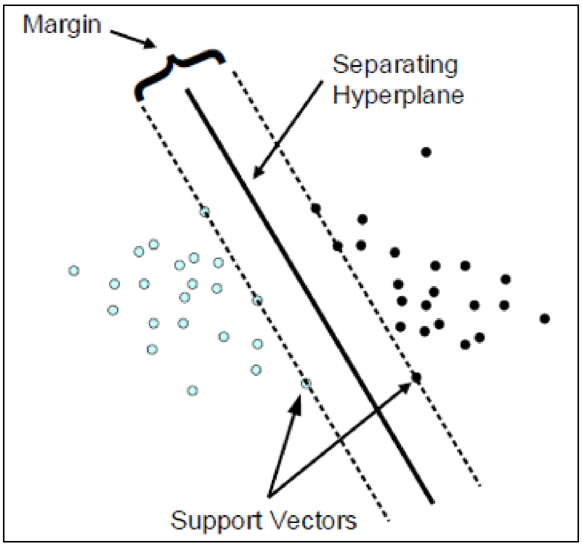

Figure 2: SVM structural understanding

Overlapping classes and non - linearity are other important concepts in SVMs. These are nothing but the data points on the wrong side of the discriminant margin which are eventually are weighted down to reduce their influence. In the case where a linear separator cannot be found, data points are projected into a higher - dimensional space where the data points effectively become linearly separable. This projection into higher dimension space is done via Kernel technique.

A kernel is a function that transforms the input data to a high - dimensional space where the problem is solved. They can be linear or nonlinear (Gaussian). The linear kernel function reduces to a linear equation on the original attributes in the training data. They works well when there are many variable in the training data. The Gaussian kernel transforms each case in the training data to a point in an n - dimensional space, where n is the number of cases. The algorithm attempts to separate the points into subsets with homogeneous target values. The Gaussian kernel uses nonlinear separators, but within the kernel space it constructs a linear equation.

Advantages of SVM

SVM models have similar functional form to well - known data mining techniques like neural networks and radial basis functions but neither of these algorithms has the well - founded theoretical approach to regularization that forms the basis of SVM. The quality of generalization and ease of training of SVM is far beyond the capacities of the traditional data mining methods.

Furthermore, SVM performs well on data sets that have many attributes/variables, even when there are very few observations to train the model. There is no upper limit on the number of variables as far as the hardware can handle. On the other hand traditional neural nets do not perform well under these circumstances.

Histograms of numerical variables

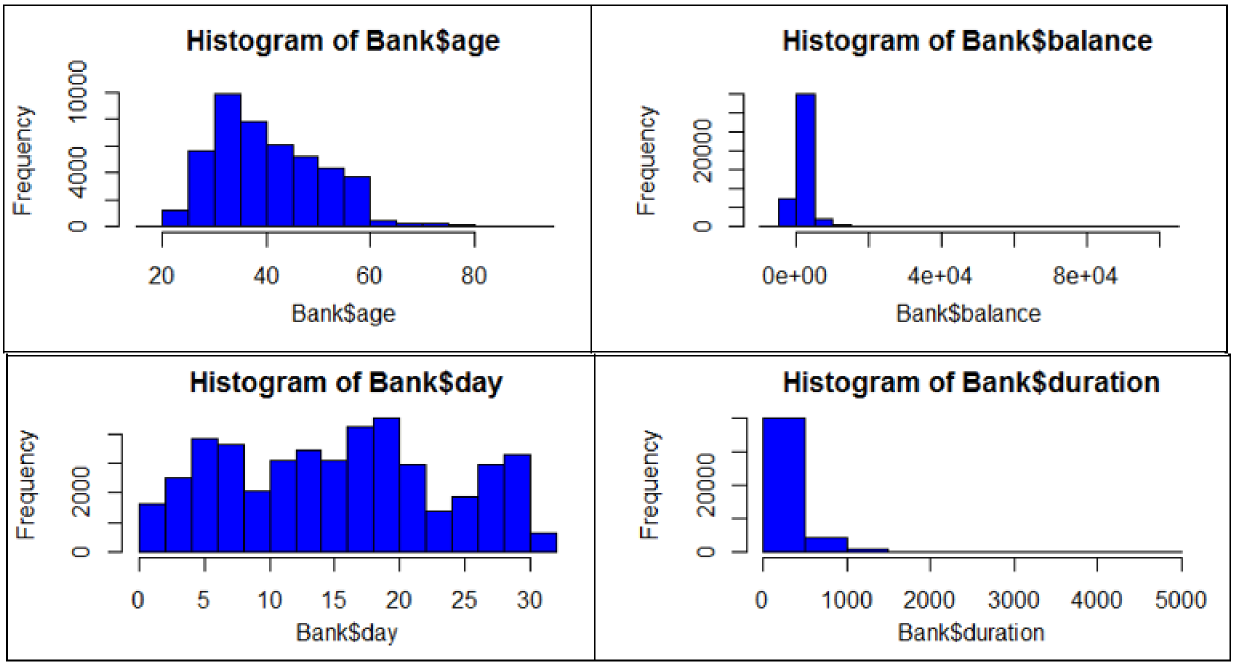

Figure 3: Exploratory data analysis

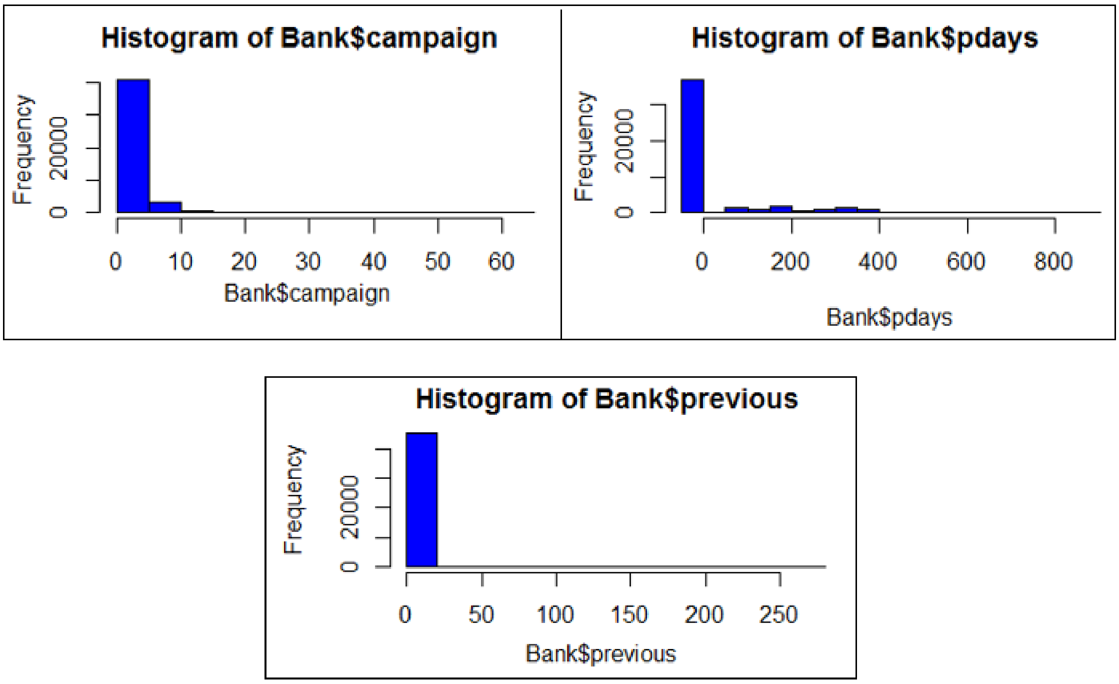

Figure 4: EDA continued

7. Methodology

Logistic Regression:

We split the data set into 80% training set and 20% testing set. We'll then run a step wise logistic regression model and identify the significant predictors for this model.

R code: glm (y~., family=binomial, data=bank_train)

In order to calculate optimal probability and corresponding misclassification rate, we'll assign costs using the following assumptions:

a) Cost of incorrectly identifying customers that will subscribe to term deposit (Predicting '1' as '0') - Business stands to lose a profit of $100 over the life time.
b) Cost of incorrectly identifying customers that will NOT subscribe to the term deposit (Predicting '0' as '1') - Business will not make any profit from such customers and will end up spending $10 as direct mail cost.

So, we'll use the ratio 10: 100 (or 1: 10) in our asymmetric cost function. Now we'll use 'optimal search grid' function to calculate optimal cut off probability. Since this is a marketing campaign or targeting problem, we want to check if the model is successful in identifying customers with high probability to subscribe to term deposit. This is achieved via KS statistics and gains chart. Also, we can plot the ROC curve to find the AUC measure.

Example of Gain Chart and KS statistics

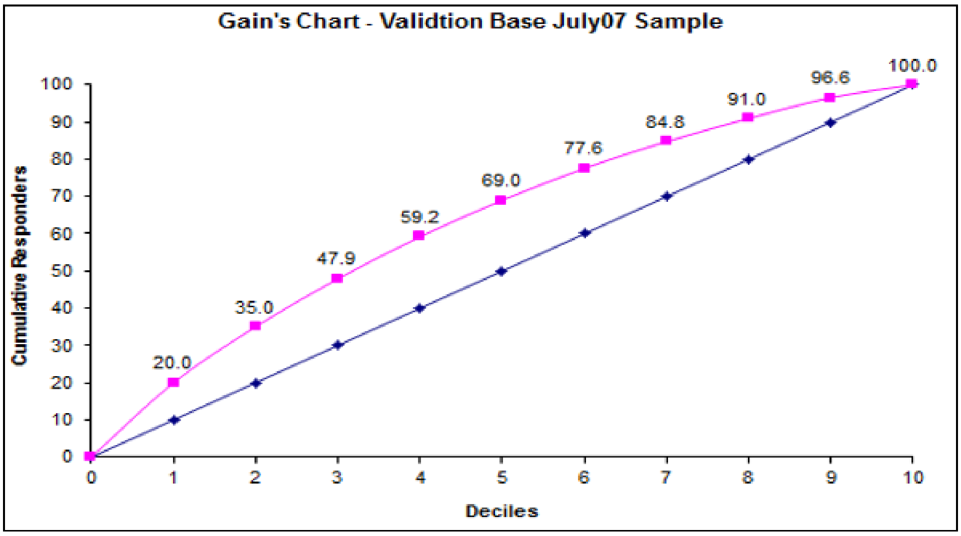

Figure 5: Gain chart

predgroup	obs	resp	nonresp	minpred	maxpred	avgpred	avgact	resp_r	nonresp_r	cumact	cumpred	cumresp_r	cumnonresp_r	ks	predrank	actrank	
												0			0		0
0	2543	1650	893	0.57001	0.95866	0.68293	0.64884	20	5.2	0.64884	0.682928283	19.99515269	5.1966946	14.79845809	1	1	10
1	2544	1239	1305	0.46449	0.56996	0.51184	0.48703	15.01	7.59	1.13587	1.194771787	35.00969462	12.79096834	22.21872628	2	2	20
2	2544	1065	1479	0.39701	0.46447	0.42944	0.41863	12.91	8.61	1.5545	1.624213631	47.91565681	21.39781192	26.51784489	3	3	30
3	2543	931	1612	0.3443	0.397	0.36977	0.3661	11.28	9.38	1.9206	1.993982076	59.19777024	30.77863129	28.41913895	4	4	40
4	2544	808	1736	0.30124	0.34426	0.32172	0.31761	9.79	10.1	2.23821	2.31570667	68.98933592	40.88105214	28.10828378	5	5	50
5	2544	710	1834	0.26331	0.3012	0.28185	0.27909	8.6	10.67	2.5173	2.597561294	77.59331071	51.55377095	26.03953976	6	6	60
6	2543	591	1952	0.22749	0.26331	0.24553	0.2324	7.16	11.36	2.7497	2.843092917	84.75521086	62.91317505	21.84203581	7	7	70
7	2544	514	2030	0.19224	0.22748	0.21022	0.20204	6.23	11.81	2.95175	3.05331563	90.98400388	74.72648976	16.25751412	8	8	80
8	2544	461	2083	0.1534	0.19218	0.17344	0.18121	5.59	12.12	3.13296	3.226757386	96.57052836	86.84823091	9.722297444	9	9	90
9	2543	283	2260	0.03218	0.1534	0.12308	0.11129	3.43	13.15	3.24424	3.349836535	100	100	-1.42109E-14	10	10	100

Figure 6: Result Logistic regression

CART Methodology

The response variable is a categorical variable (Yes/ No) and hence a classification tree will be appropriate in this case. Classification trees are used to predict membership of outcomes in the classes of a categorical dependent variable from the measurements on one or more predictor variables.

To build the classification tree model we use the rpart function and the CP plot is used for finding the optimal number of nodes.

R code: bank_rpart< - rpart (formula = Y ~. - id, data =bank_train, method = "class")
plotcp (bank_rpart)

Similar to the logistic regression methodology above, we use the asymmetric cost in the classification tree model in order to calculate optimal probability and corresponding misclassification rate and we use the ROC curve to find the AUC.

SVM Methodology

We use the same split that we used for the Logistic regression. We'll then run SVM as a classification problem. Now, we will use the tune () function to do a grid search over the supplied parameter ranges (C - cost, gamma), using the train set. The range to gamma parameter is between 0.000001 and 0.1. For cost parameter the range is from 0.1 until 10.

R Code: svm (y ~., data=bank, cost=10, gamma=0.01)

It's important to understanding the influence of this two parameters, because the accuracy of an SVM model is largely dependent on the selection them. For example, if C is too large, we have a high penalty for non - separable points and we may store many support vectors and over fit the model. If it is too small, we may have under fitting.

For classification tasks, we will most likely use C - classification with the RBF kernel (default) because of its good general performance and few number of parameters (only C and gamma). However, better results are obtained by using a grid search over all parameters. For this, we will use the tune. svm () function.

R Code: tune. svm (y~., data = bank, gamma = 10^ (- 2: - 1), cost = 10^ (1: 2))

8. Conclusion

Analytics has revolutionized direct marketing by enabling businesses to predict customer responses with precision. By adopting data - driven approaches, companies can enhance their marketing effectiveness, optimize resource allocation, and achieve higher conversion rates. As the analytics space continues to evolve, direct marketing will become increasingly personalized and efficient.

References

[1] Moro, S., Cortez, P., & Rita, P. (2014). "A data - driven approach to predict the success of bank telemarketing. " *Decision Support Systems, 62*, 22 - 31.

[2] Buckinx, W., & Van den Poel, D. (2005). "Customer base analysis: Partial defection of behaviourally loyal clients in a non - contractual FMCG retail setting. " *European Journal of Operational Research, 164* (1), 252 - 268.

[3] Kumar, V., & Reinartz, W. (2018). "Customer Relationship Management: Concept, Strategy, and Tools. " Springer.

[4] Provost, F., & Fawcett, T. (2013). "Data Science for Business: What You Need to Know About Data Mining and Data - Analytic Thinking. " O'Reilly Media.

[5] Neslin, S. A., & Blattberg, R. C. (1990). "Sales Promotion: Concepts, Methods, and Strategies. "

[6] Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification. " *Expert Systems with Applications, 36* (2), 2592 - 2602.

[7] Hootsuite & We Are Social. (2021). "Digital 2021: Global Overview Report. " Available at: wearesocial.com

[8] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). "Building comprehensible customer churn prediction models with advanced rule induction techniques. " *Expert Systems with Applications, 38* (3), 2354 - 2364.

[9] Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). "Database Marketing: Analyzing and Managing Customers. " Springer.

[10] Burez, J., & Van den Poel, D. (2008). "Separating financial from commercial customer churn: A modeling step towards resolving the conflict between

the sales and credit department. " *Expert Systems with Applications*, *35* (1 - 2), 497 - 514.

[11] Chen, Y. L., Chiu, A. L., & Chang, H. H. (2005). "Mining changes in customer behavior in retail marketing. " *Expert Systems with Applications*, *28* (4), 773 - 781.

[12] Glady, N., Baesens, B., & Croux, C. (2009). "Modeling churn using customer lifetime value. " *European Journal of Operational Research*, *197* (1), 402 - 411.

[13] Lemmens, A., & Croux, C. (2006). "Bagging and boosting classification trees to predict churn. " *Journal of Marketing Research*, *43* (2), 276 - 286.

[14] Lariviere, B., & Van den Poel, D. (2005). "Predicting customer retention and profitability by using random forests and regression forests techniques. " *Expert Systems with Applications*, *29* (2), 472 - 484.

[15] Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006). "Defection detection: Measuring and understanding the predictive accuracy of customer churn models. " *Journal of Marketing Research*, *43* (2), 204 - 211.

[16] Shaffer, G. (1991). "A customer marketing strategy: Combining customer lifetime value and marketing tactics. " *Journal of Business Research*, *22* (3), 177 - 191.

[17] Reinartz, W., & Kumar, V. (2000). "On the profitability of long - life customers in a noncontractual setting: An empirical investigation and implications for marketing. " *Journal of Marketing*, *64* (4), 17 - 35.

[18] Tsiptsis, K., & Chorianopoulos, A. (2011). "Data Mining Techniques in CRM: Inside Customer Segmentation. " Wiley.

[19] Xu, Y., & Walton, J. (2005). "Gaining customer knowledge through analytical CRM. " *Industrial Management & Data Systems*, *105* (7), 955 - 971.

[20] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). "Data Mining: Practical Machine Learning Tools and Techniques. " Morgan Kaufmann.