Modeling of Effectiveness of Media Investment Based on Data Science Technologies for Ukrainian Bank

Oleksandr Chernyak[1][0000-0002-0453-0063] and Yana Fareniuk[1]

[1] Department of Economic Cybernetics, Taras Shevchenko National University of Kyiv, 90-A, Vasulkivska st., Kiev, 03022 Ukraine
chernyak@univ.kiev.ua, yfareniuk@gmail.com

Abstract. The objective of this paper is to research, modeling and forecast the call-center workload that depends from all media and marketing activities. Data mining approach and machine learning technologies help to clearly identify and distinguish the impact factors on the feedback of potential customers (both positive and negative), determine which communication channels to use to increase inflow of queries. The model for forecasting of effectiveness of media investments and as a result managing of Return of Marketing Investments (ROMI) based on hourly data for all calls to Call Center, media and marketing indicators and macroeconomic factors for banking sector in Ukraine for the period 2013-2018 years was built. Authors used such machine learning technology as econometric modeling (regression analysis) for key metric "Incoming Calls to the Call Center". Data Science technologies help to forecast and manage calls flow with average error that is less than 10%. Article describes how to increase the effectiveness of advertising campaign by 8% in the first 2 months and achieve potential growth of conversion rate by 58%, compared to the standard market level. This article contains the key stages of implementing data mining approach, directly in the process of machine learning and dwell on the important technical aspects of the implementation of forecasting models.

Keywords: data science, marketing, media, machine learning, ROMI.

1 Introduction: market context and business tasks

Businesses today need progressive solutions, and Data Science is a huge, bottomless area to look for. Data analysis is used for both operational and tactical tasks and strategic decisions. The media sphere is no exception. Own data approaches that integrate Data Science, comprehensive expertise and MarTech will be the most valuable resource for optimizing marketing investments and differentiating companies on the market. Predictive analytics is the defining technology of the 21st century and will increasingly be used to solve complex problems, challenges and bring tremendous value to businesses and all humanity.

The researches of the use of machine learning technologies and Data Science for modeling the marketing activity of enterprises were undertaken by such domestic and

foreign scientists as Bazhenov Y., Batra R., Burnet J., Büschken J. [2], Guz M., Luky-anets T., Lysenko Y., Panasenko A. A. [10], Pankratov F., Pargelova A. [9], Romat E., Rositer J. R., Sandage C., Freiburger V., Shakhov D. A. [10], Shapiro S. and others.

A significant amount of research has been conducted on this topic. Marketing (media) mix modeling is the most commonly used method that involves building a regression model on historical data to display business metrics (sales) as a function of marketing and advertising variables, such as media activity, number of impressions, price index, and another variables such as seasonality, weather, market competition.

Mathematical modeling and data analysis open up many opportunities in the implementation of marketing activities of any enterprise. Thus, Chan and Perry (2017) [3] emphasize the importance for businesses to use different approaches to marketing modeling, because advertisers need to understand the effectiveness of their media and marketing spend in driving sales in order to optimize the allocations of marketing budgets. According to their research, the potential of MMM is often limited by the lack of detailed and qualitative data. As a solution they propose to develop better data and models, as well as to test models using simulations as the main areas of improvement for MMM.

Kiygi-Calli et al. (2017) [8] research the case in which data for creating models of advertising response are available every hour, while management decisions can relate to different time intervals (hour, day, week, month). The main conclusion is that models for low-frequency data are much simpler, while models for high-detailed data require to estimate a seasonal component. Using ad-stock approach, Zantedeschi et al. (2016) [16] determine the levels of decay of advertising message for all media, which allows in the process of developing a marketing strategy to forecast and take into account the actual short- and long-term advertising effects of each communication channel.

The contribution of regression analysis to media decision-making is quite significant, but there are alternative methods. Dawes et al. (2018) [5] describe evidence-based methods that have been shown to be useful for forecasting problems. Jin et al. (2017) [7], Zhang and Vaver (2017) [17] suggest using Bayesian hierarchical modelling.

The objective of this paper is to research, modeling and forecast the call-center workload that depends from all media and marketing activities using data mining approach and machine learning technologies for increase inflow of queries to the bank.

7 years ago, a political crisis took place in Ukraine, which caused significant economic crisis, when the Ukrainian national currency has fallen almost 4 times, from 8 to 30 UAH per USD. The financial sector was one of the first to face the economic downturn. Ukrainians have lost trust in banks, as evidenced by the significant decline in the consumer confidence index [12]. For 2014-2015 years credits, deposits in UAH and in foreign currency fell by 38%, 8% and 57% respectively [13].

Media activity of the financial sector is reduced to a lowest level, a lot of companies forced to stop mass media communications in media. Only a few biggest banks trying to maintain the confidence of Ukrainians with image campaigns on air [14, 15].

One of the Ukrainian banks, which is included in TOP 10 banks, had small and short media activity for the last 4 years. 2 years ago, media agency needed to develop a media strategy for the bank. In recent years there has been only one attempt at a media campaign in 2016. But the campaign was soon completed due to low response rate and 2 years the company didn't use media for communication with their audience.

There was a business task in the implementation of the campaign, which will give a maximum response to the calls to the Call Center. The main challenge is to develop the optimal full media mix, which include TV and other channels, and the best budget allocation for the media instruments. The key criterion is to achieve a positive Return of investment (ROI), otherwise current investments can be recognized as ineffective.

2 Data Mining approach as a good instrument to find effective business solution

Project was deploying in accordance with the most widely-used analytics model CRISP-DM [1, 11]. CRISP-DM describes the process through 6 main stages phases: "Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation and Deployment". The process involves the possibility of a flexible transition between phases in any order, going back when the need arises. Data Mining has cyclic nature, as the process of finding solutions continues after the project has been deployed. The key learnings and experience from previous cycle can generate new, more deeper business questions, which have positive influence on future data mining processes [4].

At the beginning of the project there was not enough information to make the best decision, because we had only partial information about the actual results, so the following databases were collected:

1. Detailed all available business indicators from the client for previous advertising campaigns throughout 2013 – 2016 years [6, 14, 15].

2. Open data about socio-economic development of Ukraine and consumer sentiment of population, use of banking products in dynamics [12, 13].

This was one of the key stages that allowed us to assess more accurately the market situation and take this into account when planning media activity.

The key idea was that we should combine the classic approaches of media planning (concentrate on share of voice, frequency of contacts with consumers and coverage of target audience) with completely new approach, which focus only on business indicators (calls, sales, conversion rate). Such approach to work with business client's based on data science, machine learning technologies and deep use of data.

3 How Data Science helps to increase efficiency of media investments

We have changed the traditional approach to media activity planning by focusing on modeling and forecasting key business indicators directly in order to effectively monitor business performance on a daily and weekly basis.

A regression model with a dependent variable called "Incoming calls to the Call Center" was built using Excel and R-Studio software. During the project implementation on the basis of databases of the bank, agency and open sources on business and socio-economic indicators, mathematical methods of data analysis and forecasting were used. Factors that influenced the conversion from media activity to calls, and from them

to orders and sales, were also evaluated, and these factors were improved to obtain the highest conversion rate. Two econometric sub-models were built to control and plan business at different stages of marketing activity:

1) Model for weekly planning. Control the performance of business KPIs and forecasting business results on a weekly basis allows you to quickly react to all changes and make tactical actions. The model allows you to estimate the impact of advertising and other positive or negative factors at any time that determine the level of business result.

2) Model for daily planning. Updating the model on a daily basis allows you to plan the hourly workload of the bank's Call Center accordingly to the amount of advertising activity. The correlation between the volume of realized TV ratings during the day and incoming calls to the Call Center was estimated. We evaluated the effectiveness of TV activities for every day of the week and each hour respectively.

This model optimized the work of the Call Center, as the forecast of calls for the next week was updated on a weekly basis, considering the actual results of TV activity and calls to the Call Center in the previous week.

The optimal model is multiple regression model with more than 30 factors due to daily and hourly specifications and looks like this:

$$Calls\_by\_hours = hours\_coefficient * day\_coefficient * (Constant + a1 * Adstock(TV1) * + a2 * Adstock(TV2) + ... + an * Adstock(TVn) + b * Radio + + ci * billboards\_i + di * Integrated\_economic\_indicator\_i) \quad (1)$$

where *Adstock* is the instant, prolonged and lagged effect of advertising on consumer purchase behavior, which indicate influence of TV activity during a time. $Adstock(TV)_t = TV_t + a * Adstock(TV)_{t-1}$. *Integrated_economic_indicator* include dynamic of GDP, income level and dynamics of the use of banking products.

The model is quite complicated from a technical point of view, because is a combination of patterns for every day and every hour. To determine the technical characteristics of the model below is an example of one of the models (table 1).

Table 1. Technical characteristics of one of the models

Indicator	Coefficient	Stand. Error	t-statistics	P-value
Constant	19,78	5,97	3,31	0,0017
Economic indicator	-3,82	0,08	-50,44	0,0000
Billboard	32,98	0,42	77,77	0,0000
Radio	65,24	4,45	14,67	0,0000
TV1	158,53	0,75	211,77	0,0000
TV2	140,34	1,08	130,09	0,0000
TV3	178,96	1,45	123,61	0,0000
TV4	110,27	7,70	14,32	0,0000
Multiple R^2	0,97		Adjusted R^2	0,97
F-statistics	11894,423		p-value	0,0000

The main criteria of technical model optimization were increasing of R^2, avoidance of problems of autocorrelation, heteroskedasticity and multicollinearity. Results: model estimates of factor's influence with probability at 95% level, $R^2 = 97\%$, homoskedasticity, avoidance of autocorrelation. The main criteria of business optimization were sales increase. Model coefficients have been changed due to data confidentiality.

The creation of a regression model made it possible to evaluate the impact of factors and develop recommendations for maximizing the effectiveness of media activity:

1) The optimal duration of campaign to minimize the wear-out effect.

Exceeding the pressure of the flight at X target rating points (TRPs - the main indicator of television activity) (Y weeks), leads to a decrease in the efficiency of TV activity as a result of the wear-out effect (fig. 1). Recommendation is that you continue to maintain the flight's duration at the necessary level of TRPs to maximize efficiency.
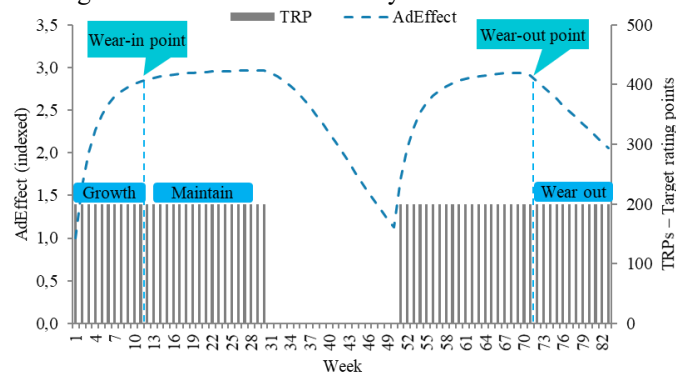

Fig. 1. Wear-out effect (data from bank's internal data base [6], media data bases [14, 15] and authors' calculations)

2) We recommend to rotate the video rollers in the period of campaign for additional calls growth and reduction of the wear-out effect.

Changing the creative allows you to increase incoming calls by 19%, but it doesn't compensate for the wear-out effect. In case of short TV campaigns, our recommendation is to use different creatives for all flights. So, we'll reduce the wear-out effect.

3) We recommend placing only X" roller (fig. 2).

Taking into account the price, placement of X" by the roller has higher efficiency: our recommendation is to use a long video to achieve the business KPIs.

4) We use additional activity on another communication channel at the end of the TV flight to accumulate incremental coverage and increase incoming calls (fig. 3). Another communications channel generates additional calls to calls from TV activity: start of advertising activity provides the incremental calls in every day on air (+ 20% in addition to calls from television).

Also, tactical recommendations on TV placement were developed based on the regression modeling and day by day tracking of business parameters:

1) Placing on weekends and holidays has low efficiency and we don't recommend to use activity in this period. The scenario with media activity at the weekend generates lower number of incoming calls, which reduces the effectiveness of each TRPs.
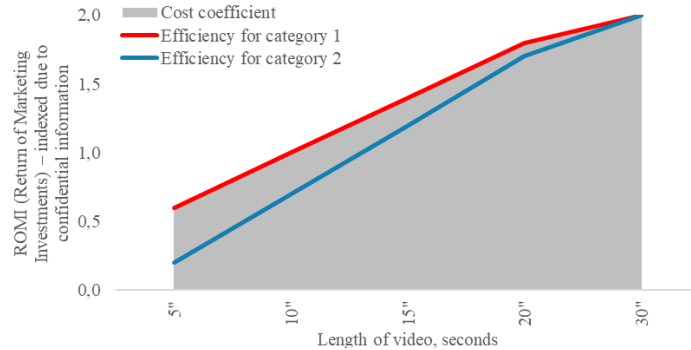
Fig. 2. Efficiency of different duration of creative materials (data from bank's internal data base [6], media data bases [14, 15] and authors' calculations)
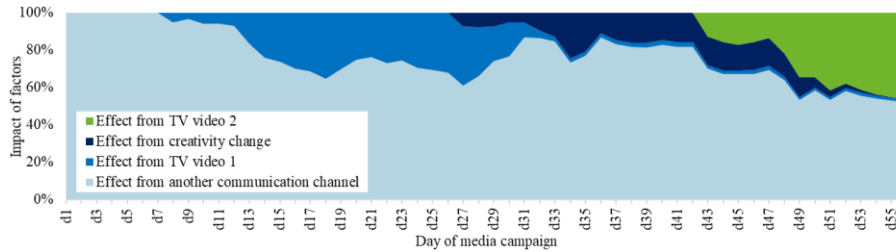

Fig. 3. Model decomposition (data from bank's internal data base [6], media data bases [14, 15] and authors' calculations)
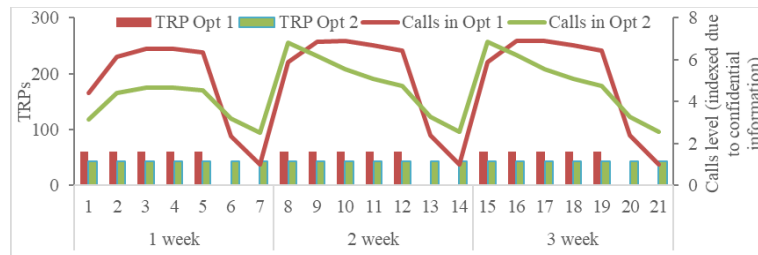

Fig. 4. Calls to the Call Center in different scenario of TV activities' allocation during a week (data from bank's internal [6], media [14, 15] and open data [12, 13] and authors' calculations)

2) Our recommendation is to use uniform distribution of advertising activities during the day, limiting placement in the evening. The mathematical model allowed us to estimate that (Fig. 5): the effect from placement in the evening is lower than during the day and in the morning. The influence of the evening placement to the next day is the same as the daily activity. We do not recommend to increase activity in the evening to stimulate calls the next morning or day.

Such recommendations cannot be taken simultaneously for all companies in the market, as the results are a combination of many factors and conditions that are formed at each time, which requires an individual approach in each case.
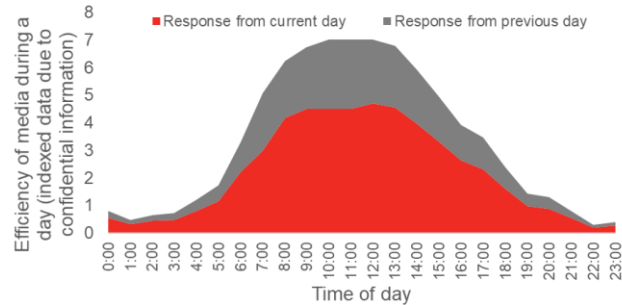
Fig. 5. Number of calls with a uniform distribution of TV ratings throughout a day (effectiveness of TV activity in every hour) (data from bank's internal [6], media [14, 15] and open data [12, 13] and authors' calculations)

4　Conclusions

Therefore, using of machine learning and data science techniques made it possible to make conclusions and develop recommendations for optimization media strategy aimed at the bank's business KPIs maximization: strategic and tactical recommendations for the most effective media mix (incl. TV); necessary volume of media pressure in all communication channels; optimal TV pressure per hour, day and week to maximize the amount of incoming calls to the Call Center.

We changed the approach of media planning for the Ukrainian advertising market, focusing at the first priority on the methodology of data science and machine learning for business data and shifting the traditional parameters of the media campaign's effectiveness on the second priority. Business indicator has become a key.

Working in the conditions of limited information in the market, we step by step collected a unique database, processed it with the help of machine learning methodology and as a result achieved high results. Due to a significant change in the model of planning, we optimized the advertising cost by 14% and achieved a higher conversion rate by 58%, compared to the average level in the market. Also, planning the traffic workload in the Call Center on the basis of the model allowed to properly distribute the workload of the Call Center and minimize possible loss of clients due to high calls flow.

By monitoring the actual results and building econometric models using machine learning tools, the optimal combination of factors was obtained on a daily and weekly basis. The models created an opportunity to evaluate and forecast the results of advertising activity (its effectiveness) on a regular basis. The average forecast error didn't exceed 11% and 8% for daily and weekly forecasting respectively, which confirms the high quality of the model and the received recommendations.

It is recommended to consider the following conclusions for future media campaigns:

- Optimal media pressure by target rating points (TRP) and period by weeks on air;
- The necessary period of a break without activity for restoration of response;
- Placement only X" roller;

- Lack of placement in days and time intervals with low efficiency (weekends, holidays, evening prime-time), uniform allocation of TRPs throughout the daytime;

- Add media activity on another communication channel at the end of the TV campaign for incremental growth of incoming calls.

The recommendations provide an opportunity to increase by 58% the conversion rate compared to the average level in the market. We don't use inefficient channels, time and day intervals, we promptly make the necessary changes in the implementation of the advertising campaign on air.

In the future, we plan to deepen the analysis and evaluate the impact of the media on other channels of the bank's marketing activities: traffic to branches, website traffic, etc.

References

1. Brown M.S.: What IT Needs To Know About The Data Mining Process. Forbes (2015).
2. Büschken J.: Determinants of Brand Advertising Efficiency: Evidence from the German Car Market. Journal of Advertising, Vol. 36, No. 3, pp. 51-73 (2007).
3. Chan, D., Perry, M.: Challenges and Opportunities in Media Mix Modeling. Technical report, Google Inc, 2017. URL https://ai.google/research/pubs/pub45998.
4. Chernyak O., Zaharchenko P.: Data mining: Textbook. Znannya, Kyiv (2014).
5. Dawes, J., Kennedy, R., Green, K. Forecasting advertising and media effects on sales: Econometrics and alternatives. International Journal of Market Research, Vol. 60, No. 6, pp. 611-620 (2018). DOI: https://doi.org/10.1177/1470785318782871.
6. Internal database of Ukrainian bank (Confidential data).
7. Jin,, Y., Wang, Y., Sun, Y., Chan, D., Koehler J.: Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. Technical report, Google Inc, 2017. URL https://ai.google.com/research/pubs/pub46001.
8. Kiygi-Calli, M., Weverbergh, M., Franses, P.H.: Modeling intra-seasonal heterogeneity in hourly advertising-response models: Do forecasts improve?. International Journal of Forecasting, Vol. 33, No. 1, pp. 90-101 (2017). DOI: https://doi.org/10.1016/j.ijforecast.2016.06.005.
9. Pergelova, Albena, Prior Diego, Rialp Josef: Assessing advertising efficiency. Journal of Advertising, v. 39/3 (2010)
10. Shakhov D.A., Panasenko A.A.: Evaluating Effectiveness of Bank Advertising in the Internet: Theory and Practice, World Applied Sciences Journal 18 (Special Issue of Economics): pp. 83-90 (2012).
11. Shearer C.: The CRISP-DM model: the new blueprint for data mining, J Data Warehousing, 5:13-22 (2000).
12. Website of GFK Ukraine, https://www.gfk.com/uk-ua/.
13. Website of National Bank of Ukraine, https://bank.gov.ua/.
14. Website of Nielsen Ukraine, https://www.nielsen.com/ua/uk/.
15. Website of Television Industry Committee, http://itk.ua/en.
16. Zantedeschi, D., Feit, E., Bradlow, E.T.: Measuring Multichannel Advertising Response. Management Science, Vol. 63, No. 8 (2016). DOI: https://doi.org/10.1287/mnsc.2016.2451.
17. Zhang, S., Vaver J.: Introduction to the Aggregate Marketing System Simulator. Technical report, Google Inc, 2017. URL https://ai.google/research/ pubs/pub45996.