

УДК: 330.4:005.93

DOI:10.5281/zenodo.11920570

M.A. Naumenko, PhD in Economics (*Charitable Foundation "Ukrainians Together", Poltava*)

DATA MINING IN EFFECTIVE ENTERPRISE ADVERTISING

In today's digital environment, online advertising is an essential part of an effective marketing strategy for businesses of all sizes. As advertisers invest heavily in digital advertising, it's important to have tools to help predict the likelihood of an ad campaign's success. Competition between advertisers is extremely tough. Every company is trying to attract the attention of Internet users and turn that attention into clicks, sales and conversions.

Business invests millions in advertising. The leading segment of the advertising industry today is online advertising. Online games, mobile applications, social networks, e-commerce platforms, search portals, online videos, banners and other digital media are just a few examples of the many digital media used in online advertising to reach potential customers. Data Mining has revolutionized the way businesses approach marketing by providing insights into customer behavior, identifying new opportunities and improving marketing strategies. The main components of Data Mining that support Internet marketing are the collection, aggregation, analysis and appropriate display of business intelligence data. After Data Mining, it is business analytics that helps determine the real pattern of consumer behavior. Thus, the emphasis on personalization and customization of content across multiple digital channels is one of the main implications of the impact of Big Data, in any form, on digital marketing.

Above all, it is important to understand consumer behavior, preferences and buying patterns through data analysis, which allows companies to adapt marketing campaigns, develop new products, optimize sales strategies and improve customer satisfaction. Among the key approaches considered in the Data Mining process are: problem identification, data preparation, model development, model use, and model monitoring. A number of techniques are used for classification and regression models, including decision trees, artificial neural networks, evolutionary algorithms, and association rules. Sophisticated Data Mining approaches such as text mining, web mining, and multimedia mining are explored to find hidden patterns and insights from various data sources such as documents, web materials, and multimedia files.

In recent years, there has been a lot of interest in implementing personalized marketing techniques and Data Mining tools in digital

companies as a mixed methods strategy that included both qualitative and quantitative research methods. In terms of qualitative methods, for example, in-depth interviews were conducted in one of the studies. Quantitative methods were used to collect data on consumer behavior from various online sources, including e-commerce Shopee, Instagram, TikTok, WhatsApp and Instagram. Quantitative analysis of this data is performed using methods such as association analysis, which identified relationships between products and purchasing patterns, recommendation systems that match individual customer preferences, and cluster analysis, which identified groups of customers with similar preferences and behaviors. Using these methods, researchers were able to gain a comprehensive understanding of how consumers relate to individual marketing tactics and how they behave in online stores.

Predicting CTR (click-through rate) is a calculation of the probability that a user will click on an advertisement. This indicator is very important in online advertising systems. Under the traditional pay-per-click paradigm, advertisers only get paid when viewers click on their ads, so a higher CTR can increase advertisers' profits. Using modern machine learning approaches based on large volumes of data, you can get a more accurate CTR forecast, which will increase the effectiveness of advertising campaigns and reduce advertising costs. Approaches that allow publishers to choose which ad image to show to a specific user on a web page to maximize CTR are in high demand due to the need to maximize revenue for publishers. Therefore, in most online advertising systems, the main responsibility of advertising placement is to predict CTR.

Decision tree models, deep learning models, factorization machine models, and multivariate statistical models are four modern models available for CTR prediction. Two of the most common multivariate statistical methods for predicting advertising CTR in the literature are logistic regression and second-degree polynomials. One method most commonly used to predict CTR based on multiple factors is logistic regression (LR). It can return a number between 0 and 1 that indicates the probability of a click, as well as a linear mixture of several attributes that represents the correlation between the qualities and the label. However, LR cannot account for interaction effects or non-linear correlations. In the literature, LR is often used as a benchmark to evaluate the performance of recommended CTR prediction models.

Researchers note that factorized parameters and factorization (MF) machines can help model estimation on sparse data to predict CTR. MFs and their extensions are often used to predict clicks in the advertising research literature. At the 2012 KDDCup competition, they demonstrated that MFs can accurately estimate feature interactions in predicting ad CTRs.

In 2015, an online learning strategy for CTR prediction using the Follow-TheRegularized-Leader (FTRL-proximal) algorithm in MF was presented. Although low-order models are easy to build, they do not do a very good job of predicting, especially in complex advertising systems. AUC shows that LSTM and RNN-based CTR prediction models outperform linear (e.g., LR and Naive Bayesian) and nonlinear (e.g., neural networks and random forests) models. However, since LSTMs are learned sequentially, they require more time and memory during the training phase.

Tree models are used in a wide range of computational advertising applications. Gradient Boosting Decision Tree (GBDT) and XGBoost are based on the Gradient Boosting Machine (GBM), which has shown great success in predicting CTR. For Facebook mass advertising, the authors propose a concatenation of GBDT and LR to predict CTR, where the output of GBDT serves as the input of LR. As a result, GBDT significantly improves the performance of LR due to autonomous feature selection.

Another approach to studying user satisfaction and dissatisfaction is using text analysis. For this, the method of latent semantic analysis (LSA) can be used, which allows you to extract hidden semantic connections of words and phrases from the corpus of documents.

In the process of data processing, redundant words are removed, related phrases are combined, and variables indicating user satisfaction and dissatisfaction are determined. The results of the study confirm the importance of certain elements, and also help to determine the offers or services that affect the satisfaction or dissatisfaction of users. Further, this knowledge is used to create phrases on creatives or quizzes, which increases the interest of users, which in turn increases the company's income. Because text analytics is open-ended, it can provide more accurate reports on the customer experience than conventional survey methods such as questionnaires or interviews. Analyzing big data on customer behavior and satisfaction is made much easier by applying LSA and text analytics.

The strategic value of Data Mining in advertising is emphasized, especially as it helps companies stay competitive in a fast-changing digital market, helping them predict future trends, make informed decisions, and so on. Data Mining ensures the adaptability of marketing efforts, improvement of sales forecasts and strengthening of relationships with customers, which as a result contributes to the prosperity and longevity of organizations.