Method of Forming the Context of Advertising and Target Audience based on Associative Rules Learning

Khrystyna Lipianina-Honcharenko [1], Taras Lendiuk [1], Anatoliy Sachenko [1,2], Jacek Wołoszyn [2]

[1] *West Ukrainian National University, Lvivska Str., 11, Ternopil, 46000, Ukraine*
[2] *Kazimierz Pulaski University of Technology and Humanities in Radom, ul.Malczewskiego 29, 26-600 Radom, Poland*

Abstract

Nowadays, important mechanisms of study are content and techniques of its creation, the problem of influencing the target audience, which itself seeks to shape communication processes. Internet content occupies a position of powerful communication technology, which continues to grow rapidly and gain influence. Creating a large number of advertisements, especially texts, is extremely expensive. Therefore, it is worth considering how generate these texts automatically. In this regard, it is possible to assume that the development of a method of forming the context of advertising and target audience based on learning associative rules is relevant and can increase the effectiveness of advertising, and thus reduce the cost of online advertising of higher education institutions. The input data used a survey of students majoring in Computer Science, regarding admission. The 152 students took part in the survey and answered 10 questions. The experimental results confirmed, the proposed method enabled to increase the effectiveness of advertising on social networks at least in 23%, and reduce the price in 90%.

Keywords

Data analysis, advertising content, Associative Rules Learning, Apriori algorithm, Facebook.

1. Introduction

The value of advertising is crucial for a company, because only this can make people aware of the company's product and, doing so, can create a good opportunity to sell it to customers [2]. The manual work of a large number of advertisements, especially texts, is extremely expensive. Therefore, it is worth considering how to generate these texts automatically.

Therefore, we can assume that the development of a method of forming the context of advertising and target audience based on learning associative rules is relevant and can increase the effectiveness of advertising, and thus reduce the cost of online advertising of higher education institutions.

This work is devoted to this topic and it is distributed as follows. Section 2 discusses the analysis of related work; section 3 presents a method of forming the context of advertising and the target audience based on the associative rules learning. Section 4 presents the implementation of the method. Section 5 presents the conclusions of the study.

2. Related Work

Given the complexity of the modern digital advertising ecosystem, there are many studies describing the impact of advertising content on social networks on attracting customers [19], using data from Facebook in: medicine [20], psychology [21], sociology [22], politics [23] and others. Research [24,

CMIS-2022: The Fifth International Workshop on Computer Modeling and Intelligent Systems, Zaporizhzhia, Ukraine, May 12, 2022
EMAIL: xrustya.com@gmail.com (Kh. Lipyanina-Goncharenko); tl@wunu.edu.ua (T. Lendiuk); as@wunu.edu.ua (A. Sachenko); jacek.woloszyn@uthrad.pl (J. Wołoszyn)
ORCID: 0000-0002-2441-6292 (Kh. Lipyanina-Goncharenko); 0000-0001-9484-8333 (T. Lendiuk); 0000-0002-0907-3682 (A. Sachenko)


25] provides an understanding of social networks for higher education institutions such as Instagram, Pinterest, Snapchat and WhatsApp.

In [13], a model is proposed that is able to use its targeting strategy in accordance with the received feedback. The model uses the Thompson Sampling algorithm applied to the user space. [1] a set of models of regression, cluster analysis and analysis of association rules to find patterns of user behavior in relation to marketing campaigns, taking into account user characteristics and financially significant indicators.

The aim of the article [12] is to study the analysis of social media data using machine learning tools; new approach to social media marketing strategy uses the Waikato Knowledge Analysis Environment (WEKA). In [5] the article implemented the level of the aspect of mood analysis, based on machine learning algorithms of SVM and NB classification.

In [17] the study analyzed the various needs of customers, focusing on online advertising based on methods of classification, segmentation and clustering. In [16] the model of selection of the optimal amount of advertising on various Internet resources is analyzed in order to achieve the desired coverage of the target audience. In addition, the method of multi-criteria optimization with the definition of the obtained objective function is considered, which allows considering simultaneously the various aspects of media selection issue and optimal budgeting and budget allocation. The proposed approach [2] draws on knowledge that can support several solutions, ranging from marketing campaigns for each customer segment, redesign of the store layout to product recommendations. In [15] the development of advertising art design on the basis of information technologies is mainly investigated.

In a study [4, 26], it is proposed to apply the training of association rules to find influential bloggers in time using the Apriori algorithm. An improved Apriori algorithm has been proposed to identify the relationship between the TECP and the thirty-five factors, which cover four categories of household characteristics, including housing, socio-demographic, household appliances and heating, and energy attitudes [9]. In [10] a new effective system of recommendations based on the Apriori algorithm for user requirements was proposed.

Article [18] proposes the generation of advertising texts based on keywords that take into account product information. Ins [3] was developed a web-based system of recommendations for choosing a property using the method of content-based filtering. The referral system provides information about properties based on user behavior by searching for advertising content that the user previously searched for. [11] presents an intelligent management system for advertising on social networks, based on data analysis techniques to automatically create ads.

The above-mentioned works mostly analyze the actions of users on online advertising. There are also a number of works that present research on methods of associative rules learning. There are also works that use different approaches to the formation of content and target audience of Internet advertising (analogues).

Therefore, goal of this paper is to develop a method of forming the context of advertising and the target audience based on the associative rules learning.

The novelty of the work is the formation of the most profitable (in the economic aspect) text of the Free Economic Zone advertising and the corresponding target group, which will increase the effectiveness of the advertising campaign based on learning the rules of association.

3. Proposed Method

To reduce the time spent on the formation of advertising content for the target audience, the authors have developed a method of forming the context of advertising and the target audience based on the associative rules learning. The proposed method is illustrated schematically (Fig. 1) and is represented by the following steps:

Step 1. Conduct a survey of students (Block 1). It is important to determine the gender characteristics of the respondents, as this will make it possible to determine the target audience in the future. Convert to csv format (Block 2).

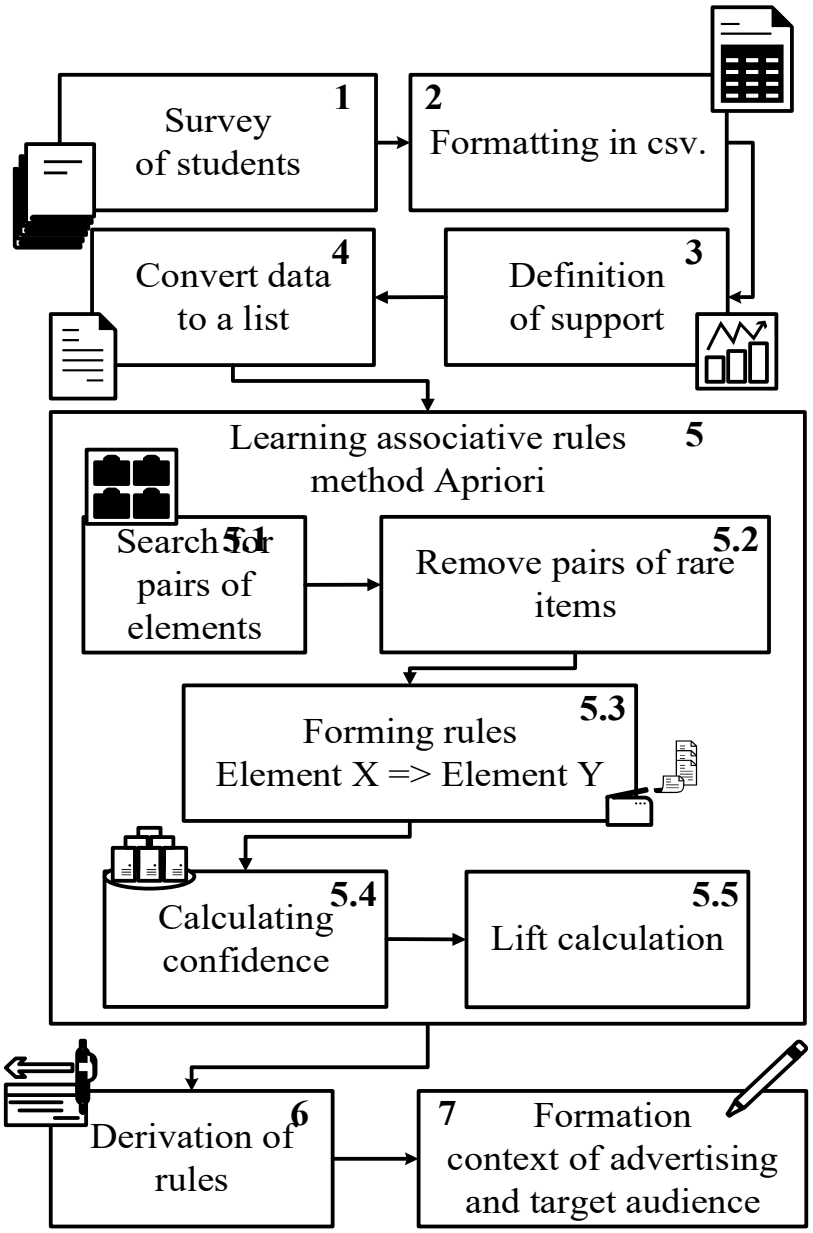


Figure 1: Algorithmic structure of forming the context of advertising and target audience based on learning associative rules

Step 2. Calculation of support for each individual element (Block 3). Support is simply the number of transactions during which a particular product (or combination of products) occurs.

Step 3. Conversion of data into a list (Block 4).

Step 4. Start (Block 5) learning associative rules based on the Apriori method [6].

Step 4.1. Finding support for frequent sets of elements (Block 5.1). Search for pairs of items that appear most often in pairs. The Apriori algorithm ignores all pairs that contain any of the rare elements (Block 5.2).

Step 4.2. Formation of rules (Block 5.3). The most frequent sets of elements, converted into association rules, in the format: Element X => Element Y.

Step 4.3. Calculation of confidence (Block 5.4). Confidence shows the percentage of cases in which this rule applies.

Step 4.3. Elevator calculation (Block 5.5). Raising a rule is an indicator of effectiveness, which indicates the strength of the relationship between the products in the rule. Raising the rule is determined by the following formula:

$$lift = \frac{P(X \cap Y)}{P(X) \times P(Y)}$$

where, P is the probability of the frequency of combination of elements in the generated rule.

Step 5. Output of results-rules (Block 6).

Step 6. Creating the context of advertising and target audience (Block 7), for higher education institutions on the basis of the received rules.

4. Experimental Results

The Python language was chosen to conduct the method of forming the context of advertising and the target audience based on th associative rules learning.

The input data used a survey of students majoring in Computer Science, regarding admission. 152 students took part in the survey and answered 10 questions. All students' feedbacks are in .csv format. The sample is representative, as the survey was conducted among students majoring in Computer Science, who themselves have recently been faced with the choice of where to enter, so their feedback is the most informative for this type of advertising.

Before starting the analysis of the rules of the association, first determine the frequency distribution of the elements (Table 2). The chart shows that the majority of male respondents, as well as the highest number of answers, said that they found information on their own in social networks, majoring in Computer Science.

Table 1

Response rate

Item name	Count
Male	67
independently found information	59
I saw a lot of interesting information on social networks	41
advised familiar relatives	39
Female	37
Interest in computers	35
The presentation of the specialty by the representatives was interesting	33
Interest in design	32
Interest in technology	30
Passed here for public training	29
Augmented reality	29
I got a call and was convinced that it would be interesting to study here	29
Representatives of the specialty came to our school	20
Internet of Things	17
Robotics	15
Internet of Things. Augmented reality	12
Software control of drones	11
The cost of training suited	11
From social networks	10
Due to quarantine, I decided to choose a place closer to home	8
Interest in IT	6
Internet of Things. Augmented reality. Robotics. Software control of drones	5
Augmented reality. Robotics.	4
Internet of Things. Robotics.	4
Augmented reality. Software control of drones	4
Augmented reality. Robotics. Software control of drones	2
Internet of Things. Augmented reality. Robotics.	2

Item name	Count
Internet of Things. Robotics. Software control of drones	2
Quality presentation of the material	2
A friend studied in the same specialty	1
A friend recommended	1
Web	1
I wanted to study to be a programmer	1
Because a large variety of specialties could not be determined	1
Robotics. Software control of drones	1
Chose occasionally	1
I like working with computers and I am interested in the specialty	1
Prospect	1
My chosen specialty is the most relevant	1
Because this specialty covers many areas that I really like	1
Many acquaintances study in Ternopil	1
A relative studied at the university	1
I have wanted to become a programmer for a long time, so I chose this specialty	1
I found information from social networks on my own	1
I really wanted to enroll in FCIT and the most advanced training is on CS	1
I was interested in computer science	1

In addition, it is important to note that even the most frequent response of more than 11% is male (Table 2). Next, we will use this information as a guide when setting a minimum support threshold.

Table 2
The most common answers

Index	Item	Count	Percentage
13	augmented reality	29	4.93
40	passed here for public training	29	4.93
47	interest in technology	30	5.10
45	interest in design	32	5.44
37	interested in the presentation of the specialty by representatives	34	5.78
46	interest in computers	35	5.95
34	female	37	6.29
39	advised familiar relatives	39	6.63
31	saw a lot of interesting information on social networks	41	6.97
41	independently found information	59	10.03
43	male	67	11.39

After converting the dataset into the desired list, we will display the results, namely the rules, which in the future will allow to form the context of advertising and target groups.

Therefore, after starting the algorithm was generated:

1. Counting sets of items of length 1:
- 48 candidates were found for sets of length 1;
- Found 15 large sets of items of length 1.
2. Counting sets of items of length 2:
- 105 candidates were found for sets of length 2;
- 32 large sets of items of length 2 were found.

Based on the experimentally significant parameters of the algorithm, the generated rules are filtered.

Parameters:

min_support = 0.11 – defined as the percentage of most frequent responses;

min_confidence = 0.65 – this probability of determining the rules is sufficient and in experimental studies showed quite good results for this sample.

Rules:

- {male, interest in design} -> {augmented reality} (conf: 0.737, supp: 0.135, lift: 1.965, conv: 2.375);
- {advised by familiar relatives, passed here for public education} -> {male} (conf: 1.000, supp: 0.115, lift: 1.552, conv: 355769230.769);
- {saw a lot of interesting information on social networks, advised familiar relatives} -> {male} (conf: 0.857, supp: 0.115, lift: 1.330, conv: 2.490);
- {saw a lot of interesting information on social networks, interest in technology} -> {found information on his own} (conf: 0.706, supp: 0.115, lift: 1.244, conv: 1.471);
- {independently found information, interest in computers} -> {male} (conf: 0.800, supp: 0.154, lift: 1.242, conv: 1.779);
- {interest in technology} -> {independently found information} (conf: 0.700, supp: 0.202, lift: 1.234, conv: 1.442);
- {passed here for public training} -> {male} (conf: 0.793, supp: 0.221, lift: 1.231, conv: 1.720);
- {advised familiar relatives, interest in design} -> {male} (conf: 0.778, supp: 0.135, lift: 1.207, conv: 1.601);
- {interest in computers} -> {male} (conf: 0.771, supp: 0.260, lift: 1.197, conv: 1.556);
- {Augmented Reality} -> {male} (conf: 0.690, supp: 0.192, lift: 1.071, conv: 1.146);
- {saw a lot of interesting information on social networks} -> {male} (conf: 0.659, supp: 0.260, lift: 1.022, conv: 1.042).

Almost all of the received rules mention the gender of the respondent (male), which indicates the main target audience for the specialty “Computer Science”. Rules that are more than two elements of the answer, allow to create content for advertising. Let's form some examples of content (Table 3).

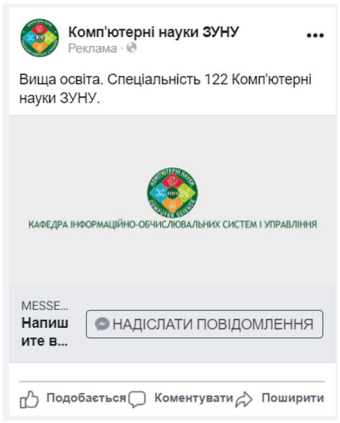
Table 3

Formation of advertising content in relation to the generated rules

# of variant	Rule	Content
1	{male, interest in design} -> {Augmented reality}	Interested in Design, Try Yourself in Augmented Reality
2	{advised by familiar relatives, passed here for public education} -> {male}	We Are Recommended When Public Education Is Important
3	{saw a lot of interesting information on social networks, advised familiar relatives} -> {male}	We are recommended after browsing our social networks
4	{saw a lot of interesting information on social networks, interest in technology} -> {male}	Interested in technology, visit our social networking pages, there's lots of interesting information
5	{self-found information, interest in computers} -> {male}	Interested in Technology, Visit our Social Networking Pages

To compare the effectiveness of the generated advertising content based on the associative rules learning, a comparative experiment was conducted on Facebook on the business page “Computer Science of ZUNU”. The first option (Option 0) advertising (Fig. 2), developed on the basis of the rules identified in previous research, namely:

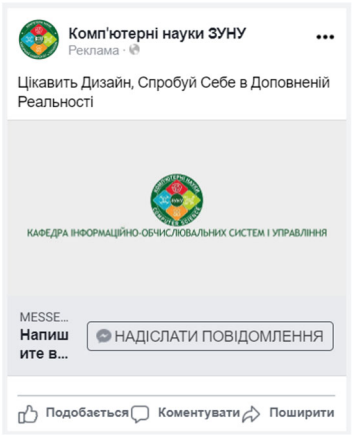
- The greatest interaction with the video advertising of Facebook “Computer Science ZUNU” had males in the age category 18-25, 35-55 [7];
- Male and female clients in the age category of 40-55 had the greatest interaction with the ZUNU Computer Science business page.



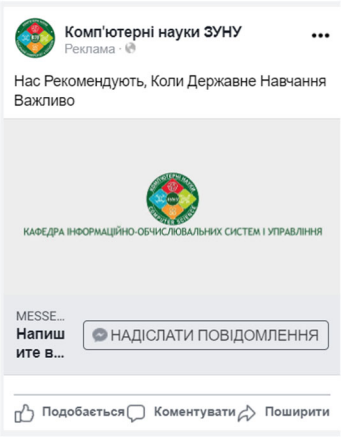
Variant 0

Figure 2: Previous version of the advertisement “Computer Science” of the Western Ukrainian National University on Facebook

Fig. 3 presents an advertisement formed on the basis of learning associative rules, a comparative experiment was conducted on Facebook on the business page “Computer Science of ZUNU”. The content used is from Table 3 and the target audience is male for all age groups.

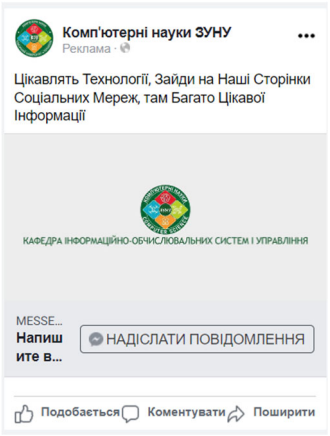


Variant 1

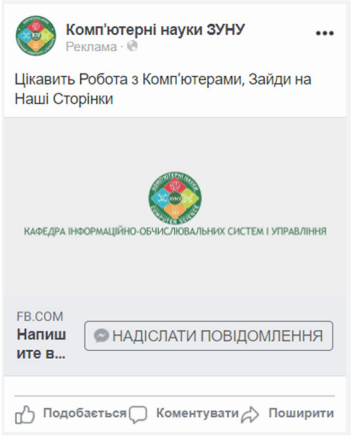


Variant 2

Variant 3



Variant 4



Variant 5

Figure 3: New variants of advertising "Computer Science" of the Western Ukrainian National University on Facebook

Table 4 presents a comparison of the effectiveness of the generated advertising content based on the associative rules learning, in the period from May 4, 2021 – May 31, 2021 with all content options, including the old.

Table 4
Comparison of the effectiveness of the generated advertising content

Advertising variant	Results		Coverage		Readings		Price for the result	
	Index	Changes	Index	Changes	Index	Changes	Index	Changes
Variant 0	120	100%	4498	100%	5395	100%	0,23	100%
Variant 1	147	123%	6561	146%	8776	163%	0,08	37%
Variant 2	197	164%	6364	141%	6560	122%	0,06	25%
Variant 3	240	200%	8561	190%	8192	152%	0,02	10%
Variant 4	160	133%	5442	121%	6671	124%	0,12	53%
Variant 5	196	163%	7024	156%	8867	164%	0,03	14%

Table 4 shows that all variants of the new ad gave improved results. The result indicator shows how many times customers have been in contact with the ad. Advertising performed best in Option 3 (Fig. 4), as it performed 100% better than Option 0. Option 3 performed best in all indicators. Namely, the coverage is 90% better than option 0 and 52% better in terms of the number of impressions. It also reduced the price for the result by 90% in option 3.

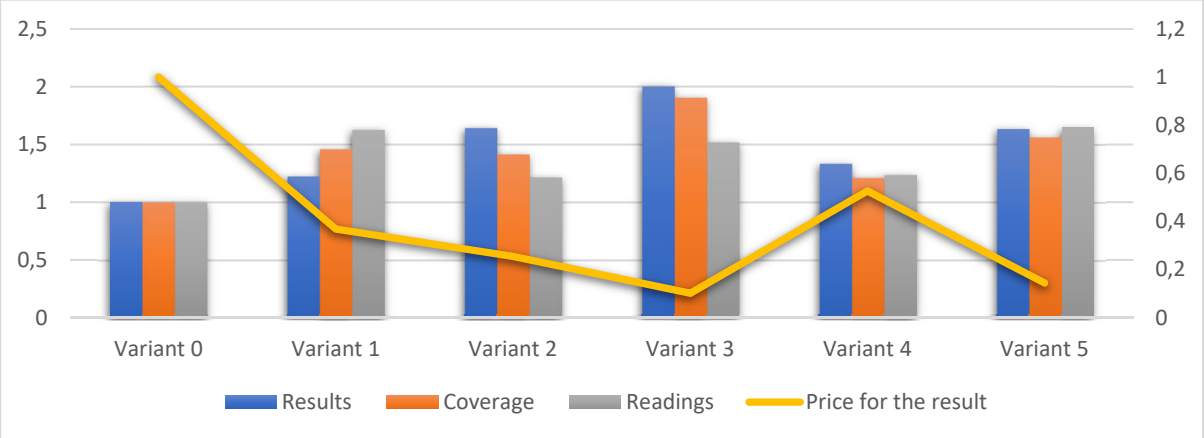


Figure 4: Comparison of the effectiveness of the generated advertising content

Thus, the proposed method enabled to increase the effectiveness of advertising on social networks at least in 23%, and reduce the price in 90%. Moreover, the authors believe that growing the number of student surveys can increase the quality of associative rules learning as well as get better keywords for content formation. That will lead to increasing the effectiveness of advertising and reducing its costs respectively.

5. Conclusions

Developed method of forming the context of advertising and target audience based on the associative rules learning makes it possible to increase the effectiveness of advertising, and thus reduce the cost of online advertising of higher education institutions. Also, the developed method will allow to form rules between the answers of respondents for the formation of advertising content and the definition of the target group.

To implement the developed method, we used a survey of students majoring in Computer Science, regarding admission. 152 students took part in the survey and answered 10 questions.

The proposed method enabled to increase the effectiveness of advertising on social networks at least in 23%, and reduce the price in 90%. Moreover, the authors believe that growing the number of student surveys can increase the quality of associative rules learning as well as get better keywords for content formation. That will lead to increasing the effectiveness of advertising and reducing its costs respectively.

Unlike analogues [3, 11, 18] the developed a method enables to form rules between the answers of respondents, construct the advertising content and determine the target group.

In the future, the authors plan to develop the information system that will automatically generate the content of the advertising message and select the target audience using the deep learning methods [27, 28].

6. References

[1] M. M. Monastyrskaya, V. I. Soloviev, Improving customer relationship management based on intelligent analysis of user behavior patterns, in: Proceedings of the 2020 13th International Conference "Management of Large-Scale System Development" (MLSD), 2020, pp. 1-4. doi:10.1109/mlsd49919.2020.92477 doi: 10.1109/MLSD49919.2020.9247718.

[2] A. Griva, C. Bardaki, K. Pramatari, Katerina, D. Papakyriakopoulos, Retail business analytics: Customer visit segmentation using market basket data, Expert Systems with Applications, 100 (2018), 1-16. doi: 10.1016/j.eswa.2018.01.029.

[3] T. Badriyah, S. Azvy, W. Yuwono, I. Syarif, Recommendation system for property search using content based filtering method, in: Proceedings of the 2018 IEEE International Conference on Information and Communications Technology (ICOIACT), 2018, 25–29. doi: 10.1109/ICOIACT.2018.8350801.

[4] B. Shazad, H. U. Khan, M. Farooq, A. Mahmood, I. Mehmood, S. Rho, Y. Nam, Finding temporal influential users in social media using association rule learning, Intelligent Automation And Soft Computing 26 (2020) 87–98. doi: 10.31209/2019.100000130.

[5] S. Vanaja, M. Belwal, Aspect-level sentiment analysis on e-commerce data, in: Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 1275-1279. doi: 10.1109/ICIRCA.2018.8597286.

[6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases VLDB, September 1994, vol. 1215, pp. 487-499.

[7] H. Lipyanina, S. Sachenko, T. Lendyuk, A. Sachenko, Targeting model of HEI video marketing based on classification tree, in: Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020, Kharkiv, Ukraine, 6-10 October 2020, CEUR Workshop Proceedings, vol. 2732, pp. 487-498. http://ceur-ws.org/Vol-2732/20200487.pdf.

[8] H. Lipyanina, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi, Decision tree based targeting model of customer interaction with business page, in: Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), April 27 – May 1, 2020. CEUR Workshop Proceedings, vol. 2608, pp. 1001–1012. http://ceur-ws.org/Vol-2608/paper75.pdf.

[9] F. Wang. K. Li. N. Duić. Z. Mi. B.-M. Hodge. M. Shafie-khah. J. P. S. Catalão, Association rule mining based quantitative analysis approach of household characteristics impacts on residential electricity consumption patterns, Energy Conversion and Management 171 (2018) 839–854. doi: 10.1016/j.enconman.2018.06.017.

[10] S. Alzu'bi, B. Hawashin, M. Eibes and M. Al-Ayyoub, A novel recommender system based on apriori algorithm for requirements engineering, in: Proceedings of the 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2018, pp. 323-327. doi: 10.1109/SNAMS.2018.8554909.

[11] J. Aguilar, G. Garcia, An adaptive intelligent management system of advertising for social networks: A case study of facebook, IEEE Transactions on Computational Social Systems 5 (2018) 20-32. doi: 10.1109/TCSS.2017.2759188.

[12] B. S. Arasu, B. J. B. Seelan, N. Thamaraiselvan, A machine learning-based approach to enhancing social media marketing, Computers & Electrical Engineering 86 (2020) 106723. doi: 10.1016/j.compeleceng.2020.106723.

[13] A. Popov, D. Iakovleva, Adaptive look-alike targeting in social networks advertising, Procedia computer science 136 (2018) 255-264. doi: 10.1016/j.procs.2018.08.264.

[14] N. Shah, S. Engineer, N. Bhagat, et al., Research trends on the usage of machine learning and artificial intelligence in advertising, Augment Hum Res 5 (2020). doi: 10.1007/s41133-020-00038-8.
[15] Z. Liu, Development of advertising art design based on information technology, in: J. Jansen B., Liang H., Ye J. (eds) International Conference on Cognitive based Information Processing and Applications (CIPA 2021), volume 85 of Lecture Notes on Data Engineering and Communications Technologies, Springer, Singapore, 2021, pp. 3-10. doi: 10.1007/978-981-16-5854-9_1.
[16] O. Barabash, G. Shevchenko, N. Dakhno, O. Neshcheret, A. Musienko, Information technology of targeting: optimization of decision making process in a competitive environment, International Journal of Intelligent Systems and Applications 9 (2017) 1-9. doi: 10.5815/ijisa.2017.12.01.
[17] R. Saito, K. Otake, T. Namatame, Analysis of fashion market trend using advertising data of shopping information site, in: Meiselwitz G. (eds) Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing. HCII 2020, volume 12195 of Lecture Notes in Computer Science, Springer, Cham, 2020, pp. 389-400. doi: 10.1007/978-3-030-49576-3_28.
[18] K. Wakimoto, S. Kawamoto, P. Zhang, Keyword-based text generation for internet advertisement, in: Proceedings of the 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020, pp. 1-4.
[19] D. Lee, K. Hosanagar, H. S. Nair, Advertising content and consumer engagement on social media: Evidence from Facebook, Management Science 64 (2018) 5105-5131. doi: 10.1287/mnsc.2017.2902.
[20] A. M. Jamison, D. A. Broniatowski, M. Dredze, Z. Wood-Doughty, D. A. Khan, S. C. Quinn, Vaccine-related advertising in the Facebook Ad Archive, Vaccine 38, (2019) 512-520. doi: 10.1016/j.vaccine.2019.10.066.
[21] S. Youn, S. Kim, Understanding ad avoidance on Facebook: Antecedents and outcomes of psychological reactance, Computers in Human Behavior 98 (2019) 232-244. doi: 10.1016/j.chb.2019.04.025.
[22] C. L. White, B. Boatwright, Social media ethics in the data economy: Issues of social responsibility for using Facebook for public relations, Public Relations Review 46, (2020) 101980. doi: 10.1016/j.pubrev.2020.101980.
[23] A. Gitomer, P. V. Oleinikov, L. M. Baum, et al., Geographic impressions in Facebook political ads, Appl Netw Sci 6, (2021). doi: 10.1007/s41109-020-00350-7.
[24] A. Peruta, A, B. Shields, Social media in higher education: understanding how colleges and universities use Facebook, Journal of Marketing for Higher Education 27 (2017) 131-143. doi: 10.1080/08841241.2016.1212451.
[25] S. Manca, Snapping, pinning, liking or texting: Investigating social media in higher education beyond Facebook, The Internet and Higher Education 44 (2020) 100707. doi: 10.1016/j.iheduc.2019.100707.
[26] A. Alhegami, H. Alsaeedi, A framework for incremental parallel mining of interesting association patterns for big data, International Journal of Computing 19 (2020) 106-117. URL: http://computingonline.net/computing/article/view/1699
[27] M. Komar, A. Sachenko, V. Golovko and V. Dorosh, Compression of network traffic parameters for detecting cyber attacks based on deep learning, in: Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), 2018, pp. 43-47, doi: 10.1109/DESSERT.2018.8409096.
[28] Zh. Hu, Ye. V. Bodyanskiy, N. Ye. Kulishova, O. K. Tyshchenko, A multidimensional extended neo-fuzzy neuron for facial expression recognition, International Journal of Intelligent Systems and Applications (IJISA) 9 (2017) 29-36. DOI: 10.5815/ijisa.2017.09.04.