Software & Systems Received 07.02.19
DOI: 10.15827/0236-235X.126.221-226 2019, vol. 32, no. 2, pp. 221–226

Item-based recommender system with statistical learning for unauthorized customers

A.V. Filipyev [1], Assistant, avfilipev@gmail.com

[1] Dubna State University, Institute of the system analysis and management,
Dubna, 141980, Russian Federation

Abstract. The paper aims to reveal that using statistical learning approaches for recommender systems makes personal communication with customers better than the expert opinion regarding this question does. The author uses a cosine similarity distance as a basis for developing a machine learning recommendation model. However, this distance has high calculation costs, therefore the paper considers the ways of solving this problem. The probability matrix of purchasing one item with another was calculated in order to weight cosine similarity and to avoid the situation when unpopular products are put at the top of a recommendation list. A weighted sum model joins cosine similarity and probability matrices and buildes recommendation sequences. User-based collaborative filtering is the most popular algorithm to build personal recommendation. However, it is useless when it is impossible to identify a user in the system. The developed algorithm based on cosine similarity distances, probability matrix and weighted sums allows building an item-to-item recommendation model. The main idea of this approach is to offer additional products to clients when only products in a cart are known. The item-to-item recommendation algorithm has shown advantages of using statistical machine learning approaches in order to improve communication with clients through a mobile application and a website. An integrated recommendation module has revealed that developing a data-driven culture is a right way of many modern companies.

Keywords: *recommender system, upsell, cosine similarity distance, probabilities, weighted sum model, data-driven culture, statistical learning, machine learning.*

The informatization process in most organizations leads to data gathering. However, the way of using this data does not guarantee a leading position on the market. ually this is due to wrong data management processes. Many organizations are slow in data compilation, classification, and organization. However, every year shows that companies start recognizing that it is importance to handle data correctly in order to become a leader in various industries [1].

Dodo Pizza operates in a very competitive market. Only after a few years of developing, it has become a leader in the Russian market and opened its stores in more than ten other countries. One of the key reasons of becoming a leader is informatization of most business processes by developing its own software. The information system gathers all data about every generated order. Developing software to automatize business processes gives an advantage to the company for fast-growing in markets of different countries. Developing data-based features improves the researching of market preferences and personalizes offers for clients.

A lot of data about clients, orders, geodata, time, etc. makes the company look for new approaches to data analysis. Statistical machine learning algorithms allow finding insights from big data, while regular tools and manual analyzing by looking through billions of data rows cannot explain anything to an analyst. Artificial intelligence is a conventional technology in every modern IT company. It transforms many industries [2]. However, the transformation process is not an obvious task for many companies, and it is very important to choose the right way to implement changes. Fortunately, there is a number of guides on how to resolve this problem, that are written by people who have big experience in this field. For example, one guide offers to focus on five general steps [3], and start from developing a pilot project.

This paper aims to show the results of the developed pilot project based on fundamental statistical approaches, which help to resolve the problem of lacking history about previous orders of unauthorized customers. When it is impossible to identify a user in the system, it is necessary to use other characteristics in order to maximize offer personalization and value for a company.

A machine learning approach for unauthorized clients

Research goal. The research task is to prove that statistical learning algorithms can work better with a huge audience than an expert recommendation system. Moreover, recommendations based on item-to-item collaborative filtering model (I2ICF) for unauthorized users allow personalizing offering and expanding various selling products through the upselling module. First, there are developed null and alternative hypotheses:

– Null Hypothesis (H0): Russian customers who receive a recommendation sequence from I2ICF model do not have higher conversion rates from a recommended purchase compared to customers who receive a recommendation based on expert logic;

– Alternative hypothesis (H1): Russian customers who receive a recommendation sequence from I2ICF model have higher conversion rates from a recommended purchase compared to customers who receive a recommendation based on expert logic.

Recommender system. A recommender system is a way of offering item personalization to a customer based on accumulated knowledge about previously sold products, a composition of products and orders. The recommender system issue is very diverse because it enables the ability to use various types of user-preference and user-requirements data to make recommendations [4]. Many existing recommendation systems use collaborative filtering approaches which are neighborhood-based, and computing similarity between users or items [5–8]. The research result of this paper is a developed algorithm based on a cosine similarity between items.

Upsell recommendation is a form of offering an additional vector R of products consisting of some set of existing goods I $\{i_1, i_1, \ldots, i_M\}$, where M is a menu size of a certain store, there is a vector C of added items in the cart and R must not contain items from the vector C:

when $C \in I \to R \in I \wedge R \notin C.$

It is possible to represent a dataset of orders like a set of N vectors $(x_1, x_1, \ldots, x_n)$, where each vector x consists of information about purchased products $\{x_{11}, x_{12}, \ldots, x_{IM}\}$: x_{ij} is equal to 1 when j^{th} product was sold in the i^{th} order, and 0 when this product was not sold in a certain order. The purpose is to recommend additional products for a user, when we know only the information about chosen products in a cart. This type of upsell products is called item-to-item recommendation [9]. One of the most difficult problems is to choose products without information about customer preferences and to reveal a recommendation sequence of top-N products.

In contrast to the user-based approach where matrix factorization algorithms and collaborative filtering are the basic approach [10, 11], it is reasonable to research the data in terms of orders instead of exploring it as a history of certain users and their preferences. This means that we can operate only with orders and sold items, not with users. Although this approach may lead to the inability of building a recommendation sequence based on personal preferences, it helps to resolve the problem of the cold start.

A regular approach of recommendations for unauthorized clients is based on an expert logic and offers all products from a category without any ranging:

1. $C \in$ [pizzas] $\Rightarrow R \in$ [drinks]
2. $C \in$ [snacks] $\Rightarrow R \in$ [drinks]
3. $C \in$ [desserts] $\Rightarrow R \in$ [drinks]
4. $C \in$ [pizzas, snacks] $\Rightarrow R \in$ [drinks]
5. $C \in$ [pizzas, desserts] $\Rightarrow R \in$ [drinks]
6. $C \in$ [pizzas, drinks] $\Rightarrow R \in$ [snacks]
7. $C \in$ [pizzas, snacks, drinks] $\Rightarrow R \in$ [desserts]
8. $C \in$ [drinks, desserts] $\Rightarrow R \in$ [snacks]
9. $C \in$ [desserts, snacks] $\Rightarrow R \in$ [drinks]
10. $C \in$ [drinks, snacks] $\Rightarrow R \in$ [desserts]

Cosine similarity. Empirical experiments show that using data for ranging products allows improving purchasing of additional products on the stage of cart approving [12]. In contrast to previous experiments, there is no opportunity to limit data set of user orders or to transform it into a smaller matrix. It is also necessary to use a big matrix for all orders to build recommendation sequences for clients. In order to build these recommendations, we can calculate a similarity between certain products. To resolve this task we can represent a dataset of vectors N of sold products as a matrix x_{ij}:

$$\begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix},$$

where x_{ij} equals 1 when j^{th} product was sold in i^{th} order, and 0 when this product was not sold. Previous experiments have shown that using cosine similarity is the right way to calculate the distance between items for our datasets [12]. If we want to calculate a similarity distance S between items we have to select two columns of products, transpose them to vectors (a, b) and add them by the formula:

$$s = \cos(\theta) = \frac{a \times b}{\|a\| \times \|b\|} = \frac{\sum_{i=1}^{m} a_i b_i}{\sqrt{\sum_{i=1}^{m} a_i^2}\sqrt{\sum_{i=1}^{m} b_i^2}}.$$

But cosine similarity has high calculation costs and it is necessary to keep all items in memory at the same time. The matrix with monthly sales data has more than 3 000 000 of rows. The solution was found after checking sparsity of the dataset:

$$sparsity = \frac{count\ of\ non-zero\ items}{count\ of\ all\ items}.$$

This parameter equals 0.9695. It means that vectors of our products have too many zero items, so they are not useful for calculating. It is possible to isolate co-purchased pairs of products that are sold in certain orders like the isolation of co-rated user rates [13].

Figure 1 illustrates the isolation process, where the matrix columns represent products and the rows represent orders. After the algorithm selects orders with *a* and *b* products, vectors of these items become much shorter than before. Now it is reasonable to determine the calculation of similarity as $s = \cos\ (\theta) = \frac{a'\times b'}{\|a'\|\times\|b'\|} = \frac{\sum_{i\in P} a_i b_i}{\sqrt{\sum_{i\in P} a_i^2}\sqrt{\sum_{i\in P} b_i^2}}$,

where vectors *a′* and *b′* consist of co-purchased products and *P* is the index of these orders. This approach significantly reduces the cost of cosine similarity calculation and it can be computed within short time without specific hardware. The calculated matrix of similarities *S*: $\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix}$ reveals how similar products are if computing is based on orders. In this case, we can face another problem when unpopular products exist in a low amount of orders and it leads to a high value of distance. When a client chooses one popular and one unpopular product, the top of recommendation sequence will consist of unpopular products as the value of rare products is higher than other ones.

Weighted sum model. The *weighted sum model* (WSM) is the most recognizable method and a simple multi-criteria of decision making for evaluating a number of alternatives in terms of a number of decision criteria [14]. We suppose that a given *multiple-criteria decision analysis* (MCDA) problem is defined by the alternative option *m* and decision criteria *n*. Next, let's assume that the higher the values, the better all benefit criteria. Further, we suppose that w_j shows the relative importance of the criterion C_j and a_{ij} is the performance indicator of the alternative A_i when evaluated in terms of the criterion C_j. Then, we sum (when all criteria are considered simultaneously) the importance of alternative A_i, denoted as $A_i^{WSM-Score}$, which is defined as follows [15]:

$$A_i^{WSM-Score} = \sum_{j=1}^{n} w_j a_{ij},\ for\ i = 1, 2, 3, ..., m.$$

In this case, the weight might be the probability of purchasing one product with another.

$$p_{ij} = \frac{\sum orders\ with\ i^{th}\ product\ among\ orders\ with\ j^{th}\ product}{\sum orders\ with\ j^{th}\ product}.$$

After obtaining an asymmetric probability matrix, we can calculate prediction values for every product from the menu and range them in order to build a recommendation sequence:

$$r_j = \sum_{j\in C} (1-(1-p_{ij_1})...(1-p_{ij_n}))s_{ij},$$

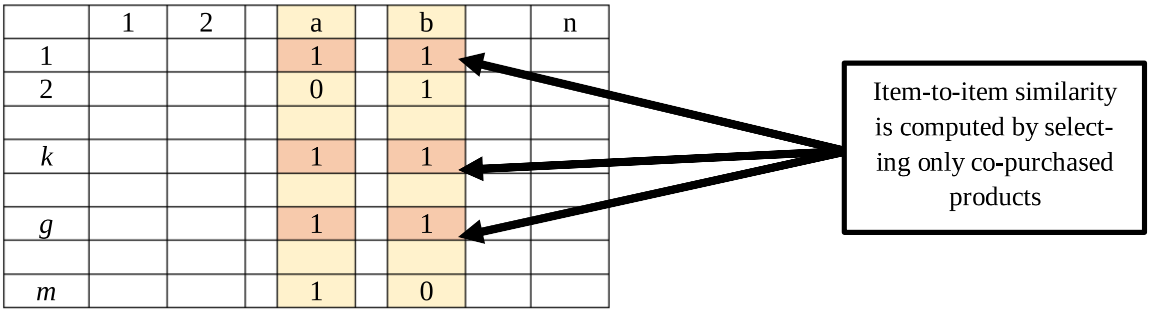
where *C* is an array of products in a shopping cart and *n* equals the size of this array. When sorting items of recommendation values, we get top *N* products to offer to clients. Some examples of testing a shopping cart:

- *C* ∈ [veggie pizza] ⇒ *R* (greek salad, fruit-drink sallow thorn, fries, …)
- *C* ∈ [veggie pizza, dodster] ⇒ *R* (fries, syrnyky, caesar salad, Greek salad, chiken pasta, …)
- *C* ∈ [meat pizza, pepperoni pizza] ⇒ *R* (coca-cola, BBQ wings, corn, vanilla muffin, …)

Above examples illustrate that the algorithm builds sequences with a much more expensive product variety and offers, for example, vegetarian products for pizza products without meat.

A recommender system structure. The main idea of statistical learning is to calculate a cosine similarity and probability matrix before launching a recommender system module:

– Select co-purchased products;



	1	2		a		b		n
1				1		1		
2				0		1		
k				1		1		
g				1		1		
m				1		0		

Item-to-item similarity is computed by selecting only co-purchased products

Fig. 1. Isolation of co-purchased products

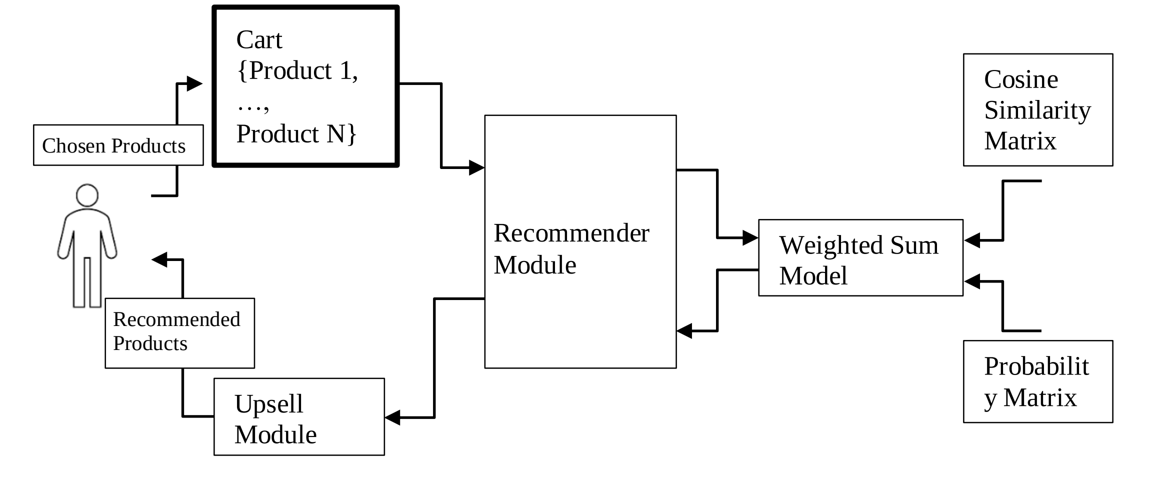


Fig. 2. A recommendation module scheme

– Calculate the Cosine Similarity matrix on selected data;

– Calculate the Probability matrix;

– Upload matrices into a database;

– Launch the recommender module based on the weighted sum module.

Figure 2 illustrates the recommendation process scheme:

1. A customer chooses products from the menu.

2. The list of products goes to the recommender module.

3. The weighted sum model ranges other products from the menu based on the cosine similarity distance and probabilities.

4. The upsell module shows ranged products to clients.

This is the common algorithm of work for a website and a mobile application.

Results

After launching this recommendation module based on cosine similarity, probabilities, and the weighted sum model, gathered statistics data has shown increased sales of additional products. Pictures below reveal sales of additional different goods through a mobile application and a web site.

Before launching cart-based recommendations the most part of additional sales through the upsell module consisted of drinks only. Customers would buy different products like snacks and desserts: fries, BBQ wings, mini rolls cowberry, etc. (ref. http://www.swsys.ru/uploaded/image/2019-2/2019-2-dop/5.jpg). This is the result of recommending a wide variety of products.

Opposed to sales through the web-site, the purchases via a mobile application do not consist of drinks (ref. http://www.swsys.ru/uploaded/image/2019-2/2019-2-dop/6.jpg). The reason for this might be the fact that it is possible to show just one product in a mobile recommendation moduleю However, it is clear that it is possible to offer products based on selected items in a cart.

Figure 3 reveals increasing revenue from selling recommended products since the start of analyzing clients' preferences. There are three important steps which show the difference between various algorithms of offering additional items.

1. Step 1 is the time when customers' orders are clustered, and recommendation sequences are prebuilt to offer products [9].

2. Step 2 is the time of launching cart-based recommendations. Cosine similarity and probability matrices for the weighted sum model are calculated based on data from four months ago.

3. Step 3 is the time of retraining matrix on data from one month ago.

The results given on Figures (ref. http://www.swsys.ru/uploaded/image/2019-2/2019-2-dop/5.jpg, http://www.swsys.ru/uploaded/image/2019-2/2019-2-dop/6.jpg) and Figure 3 show that the statistical machine learning algorithm allows offering a great variety of products to customers, and this offering is more relevant than the one offered by the expert recommender system. After launching a model based on weighted sum, customers started buying additional products through the upsell module more often. So the purchased product range has become wider.

To summarize the results of launching upsell module based on simple statistics algorithms and

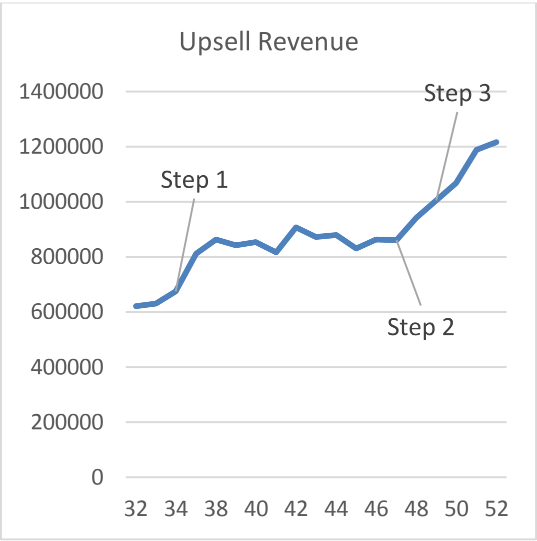


Fig. 3. Upsell revenue by weeks

gathered data, we can observe obvious advantages of applying the machine learning approach for analyzing and making decisions. One more important insight is that people are interested in purchasing various products and it is possible to develop recommendation systems in order to introduce customers to a wide range of products to offer.

Conclusion

The first pilot project for improving offering additional products for customers has shown that a data-driven approach and using artificial intelligence are useful tools for developing communication with customers. The results of improving the upsell module reveal the advantages of applying simple statistical learning. The right way of developing using gathered data of any industry can give undeniable business value. Machine learning algorithms allow companies to become leaders of their markets.

References

1. Kiron D. *Lessons from Becoming a Data-Driven Organization*. MIT, 2016. Available at: https://learning.oreilly.com/library/view/lessons-from-becoming/53863MIT58215/ (accessed February 05, 2019).

2. Filipyev A. *AI Conf. Artificial Intelligence Capabilities. System Administrator*. 2018, no. 12. Available at: http://samag.ru/archive/article/3771 (accessed February 05, 2019).

3. Ng A. *AI Transformation Playbook: How to Lead Your Company into the AI Era*. 2018. Available at: https://landing.ai/ai-transformation-playbook/ (accessed February 05, 2019).

4. Aggarwal Ch.C. *Recommender Systems: The Textbook*. Springer Publ., NY, USA, 2016, 498 p.

5. Ferraro A., Choi K.M., Bogdanov D., Serra X. *Using Offline Metrics and User Behavior Analysis to Combine Multiple Systems for Music Recommendation*. Arxiv:1901.02296v1 [cs.IR]. 2018.

6. Aiolli F. Efficient top-n recommendation for very large scale binary rated datasets. *Proc. 7th ACM Conf. on Recommender Systems*. ACM, 2013, pp. 273–280.

7. Celma O. *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Spinger Publ., 2010, 194 p.

8. Slaney M., White W. Similarity based on rating data. *Proc. 8th Intern. Conf. on Music Information Retrieval (ISMIR)*. 2007.

9. Linden G., Smith B., York J. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*. 2003, no. 7, pp. 76–80.

10. Bokde Dh., Girase Sh., Mukhopadhyay D. Matrix factorization model in collaborative filtering algorithms: A survey. *Procedia Computer Science*. 2015, no. 49, pp. 136–146.

11. Grover P. *Various Implementations of Collaborative Filtering*. 2017. Available at: http://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0 (accessed February 05, 2019).

12. Filipyev A. *Improving Upsell by Clustering Customers' Orders: A Machine Learning Approach*. 2018, no. 3.

13. Sarwar B., Karypis G., Konstan J., Riedl J. *Item-based Collaborative Filtering Recommendation Algorithms*. 2001, Hong Kong.

14. Mesran A., Ginting G., Suginam G., Rahim R. Implementation of Elimination and Choice Expressing Reality (ELECTRE) Method in Selecting the Best Lecturer (Case Study STMIK BUDI DARMA). *IJERT*. 2017, vol. 6, no. 2, pp. 141–144.

15. Handoko D., Mesran M., Darma Nasution S., Yuhandri Yu., Nurdiyanto H. Application of weight sum model (WSM) In Determining Special Allocation Funds Recipients. *The IJICS*. 2017, vol. 1, no. 2. ISSN 2548-8384. Available at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/artcle/view/528/513 (accessed February 05, 2019).

УДК 004.021 Дата подачи статьи: 07.02.19
DOI: 10.15827/0236-235X.126.221-226 2019. Т. 32. № 2. С. 221–226

Рекомендательная система на основе статистического обучения для неавторизованных пользователей

А.В. Филипьев [1]*, ассистент, avfilipev@gmail.com*

[1] *Университет «Дубна», Институт системного анализа и управления, г. Дубна, 141980, Россия*

Цель данного исследования – показать, что использование статистического обучения как основы рекомендательной системы позволяет выстроить персональное взаимодействие с клиентами лучше, чем система, построенная на экспертной логике.

За основу разработки рекомендательной системы была взята косинусная мера сходства. Расчет этой меры имеет высокую вычислительную сложность, и автор статьи предлагает возможный путь решения данной проблемы. Матрица вероятности покупки одного продукта с другим была использована в модели взвешенных сумм с целью избежать ситуации, когда непопулярный продукт может попасть в высокий приоритет рекомендации.

В разработанном модуле модель взвешенных сумм является основой объединения матрицы косинусных мер сходства и вероятностей. Одним из самых популярных алгоритмов для построения персональных рекомендаций является алгоритм коллаборативной фильтрации, но он неэффективен, когда невозможно идентифицировать пользователя в системе. Разработанный алгоритм, основанный на косинусной мере сходства, вероятностях и модели взвешенных сумм, позволил построить рекомендательную систему, работающую на основе выбранных в корзине продуктов.

Рекомендательный алгоритм на основе элементов выявил преимущества использования подходов статистического обучения в задаче улучшения эффективности коммуникации с клиентами через мобильное приложение и веб-сайт.

Интегрированный модуль рекомендаций продемонстрировал наиболее перспективный подход для современных компаний, который заключается в развитии культуры, основанной на данных.

Ключевые слова: *рекомендательная система, кросс-продажи, косинусная мера сходства, вероятности, модель взвешенных сумм, статистическое обучение, машинное обучение.*

Литература

1. Kiron D. Lessons from Becoming a Data-Driven Organization. MIT, 2016. URL: https://learning.oreilly.com/library/view/lessons-from-becoming/53863MIT58215/ (дата обращения: 05.02.2019).

2. Филипьев А. AI Conference. Возможности искусственного интеллекта // Системный администратор. 2018. № 12. URL: http://samag.ru/archive/article/3771 (дата обращения: 05.02.2019).

3. Ng A. AI Transformation Playbook: How to Lead Your Company into the AI Era. 2018. URL: https://landing.ai/ai-transformation-playbook/ (дата обращения: 05.02.2019).

4. Aggarwal Ch.C. Recommender Systems: The Textbook. Springer Publ., NY, USA, 2016, 498 p.

5. Ferraro A., Choi K.M., Bogdanov D., Serra X. Using Offline Metrics and User Behavior Analysis to Combine Multiple Systems for Music Recommendation. Arxiv:1901.02296v1 [cs.IR]. 2018.

6. Aiolli F. Efficient top-n recommendation for very large scale binary rated datasets. Proc. 7th ACM Conf. on Recommender Systems. ACM, 2013, pp. 273–280.

7. Celma O. Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space. Spinger Publ., 2010, 194 p.

8. Slaney M., White W. Similarity based on rating data. Proc. 8th Intern. Conf. on Music Information Retrieval (ISMIR). 2007.

9. Linden G., Smith B., York J. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing. 2003, no. 7, pp. 76–80.

10. Bokde Dh., Girase Sh., Mukhopadhyay D. Matrix factorization model in collaborative filtering algorithms: A survey. Procedia Computer Science. 2015, no. 49, pp. 136–146.

11. Grover P. Various Implementations of Collaborative Filtering. 2017. URL: http://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0 (дата обращения: 05.02.2019).

12. Filipyev A. Improving Upsell by Clustering Customers' Orders: A Machine Learning Approach. 2018, no. 3.

13. Sarwar B., Karypis G., Konstan J., Riedl J. Item-based Collaborative Filtering Recommendation Algorithms. 2001, Hong Kong.

14. Mesran A., Ginting G., Suginam G., Rahim R. Implementation of Elimination and Choice Expressing Reality (ELECTRE) Method in Selecting the Best Lecturer (Case Study STMIK BUDI DARMA). IJERT. 2017, vol. 6, no. 2, pp. 141–144.

15. Handoko D., Mesran M., Darma Nasution S., Yuhandri Yu., Nurdiyanto H. Application of weight sum model (WSM) In Determining Special Allocation Funds Recipients. The IJICS. 2017, vol. 1, no. 2. ISSN 2548-8384. URL: https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/artcle/view/528/513 (дата обращения: 05.02.2019).