

УДК 004.021

ITEM-BASED RECOMMENDER SYSTEM WITH STATISTICAL LEARNING FOR UNAUTHORIZED CUSTOMERS

Filipjev Andrey

Assistant;
Dubna State University,
Institute of the system analysis and management;
141980, Dubna, Moscow reg., Universitetskaya str., 19;
e-mail: avfilipev@gmail.com.

This article aims to reveal that using statistical learning approaches for recommender systems better build personal communication with consumers than does expert opinion regarding this question. Cosine similarity distance was used as a basis for developing machine learning recommendation model. However, this distance has a high cost of calculation, and ways of resolving this problem were considered. The matrix of the probability of purchasing one item with another was calculated in order to weight cosine similarity and avoid the situation when unpopular products are put in the top of recommendation. Weighted sum model was applied in order to join cosine similarity and probability matrices and build recommendation sequences. User-based collaborative filtering is the most popular algorithm to build personal recommendation, but it useless when it is impossible to identify a user in the system. A developed algorithm based on cosine similarity distances, probability matrix and weighted sums allows building item-to-item recommendation model. The main idea of this approach is to offer additional products to clients when only products in a basket are known. Item-to-item recommendation algorithm has shown advantages of using statistical machine learning approaches in order to improve communication with clients through mobile application and website. An integrated recommendation module has revealed that developing a data-driven culture is the right way of many modern companies.

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

РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА НА ОСНОВЕ СТАТИСТИЧЕСКОГО ОБУЧЕНИЯ ДЛЯ НЕАВТОРИЗОВАННЫХ ПОЛЬЗОВАТЕЛЕЙ

Филипьев Андрей Владимирович

Ассистент;
ГБОУ ВО МО «Университет «Дубна»;
Институт системного анализа и управления;
141980, г. Дубна, Московская область, ул. Университетская 19;
e-mail: avfilipev@gmail.com.

Цель данного исследования показать, что использование статистического обучения как основы рекомендательной системы позволяет лучше выстроить персональное взаимодействие с клиентами, чем система, построенная на экспертной логике. Косинусная мера сходства была взята за основу разработки рекомендательной системы. Так как расчет этой меры имеет высокую вычислительную сложность, в статье был рассмотрен 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 не эффективен 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 показал преимущества использования подходов статистического обучения в задаче улучшения эффективности коммуникации с клиентами через

мобильное приложение и веб-сайт. Интегрированный модуль рекомендаций показал, что развитие культуры, основанной на данных, это правильный путь для современных компаний.

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

Introduction

The informatization process leads to data gathering in most organizations. However, the way this data is used does not guarantee to become a market leader for those organizations. The reason for that is usually wrong data management processes. Many organizations have been slow in compiling, classifying, and organizing the data sitting in silos and dark corners. However, every year shows that companies have awakened the importance of the right communication with data become the leaders in various industries [1].

Dodo Pizza operates in a very competitive market. Only after a few years of developing, it became 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 the way of informatization most of the business processes by developing its own software. It lets to gather all data about every order that was generated by the information system. Developing software for automatization business processes gives an advantage to the company for fast-growing on most countries' markets. Developing data-based features can improve the researching of market preferences and personalize offers for clients.

A lot of data about clients, orders, geodata, time, etc. makes the company seek 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 rows of data cannot explain anything to an analyst. Artificial intelligence is a conventional technology in every modern IT company and it is transforming many industries [2]. However, the transformation process is not an obvious task for many organizations, and it is very important to choose the right way to implement changes. Fortunately, there are a number of guides on how to resolve this problem, and they are written by people who actually have strong experience in this sphere. For example, this guide offers to focus on five general steps [3], and start with developing pilot project.

The present article 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 needed to use other characteristics in order to maximize offer personalization and value for the company.

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 allows personalize offering and expand various of selling products through the upselling module. First, there were developed null and alternative hypothesis':

- Null Hypothesis (H0): Russian customers who receive recommendation sequence from I2ICF model will not have higher from recommendation to buy conversion rates compare to customers receive a recommendation based on expert logic.
- Alternative hypothesis (H1): Russian customers who receive recommendation sequence from I2ICF model will have higher conversion rates compared to customers who receive a recommendation based on expert logic.

Recommender system

A recommender system is a way of offering personalization of items to a consumer based on accumulated knowledge about previously sold products, a composition of products and orders. The topic of recommender systems 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 collabo-

rative filtering approaches which are neighborhood-based, computing similarity between users or items [5, 6, 7, 8, 9]. The result of research of this article is developed algorithm based on a cosine similarity between items.

Upsell recommendation is a form of offering additional vector R of products consists of some set of existing goods $I \{i_1, i_2, \dots, i_M\}$ 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 vector C :

$$\text{where } C \in I \Rightarrow R \in I \wedge R \notin C.$$

It is possible to represent a dataset of orders like a set of N vectors (x_1, x_2, \dots, x_N) , where each vector x consists of information about purchased products $\{x_{11}, x_{12}, \dots, x_{1M}\}$: x_{ij} is equal to 1 when j^{th} product was sold in i^{th} order, and 0 when this product was not sold in the certain order. The goal is to recommend additional products for a user, when we only know the information about chosen products in a cart. This type of upsell products is called Item-to-Item recommendation [10]. One of the heaviest problems is to range products without information about clients preferences and reveal the recommendation sequence of top- N products.

On the opposite of User-Based approach where matrix factorization algorithms and collaborative filtering are the basic approach [11, 12], 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 recommendation sequence based on personal tastes, 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. $C \in [\text{pizzas}] \Rightarrow R \in [\text{drinks}]$;
2. $C \in [\text{snacks}] \Rightarrow R \in [\text{drinks}]$;
3. $C \in [\text{desserts}] \Rightarrow R \in [\text{drinks}]$;
4. $C \in [\text{pizzas, snacks}] \Rightarrow R \in [\text{drinks}]$;
5. $C \in [\text{pizzas, desserts}] \Rightarrow R \in [\text{drinks}]$;
6. $C \in [\text{pizzas, drinks}] \Rightarrow R \in [\text{snacks}]$;
7. $C \in [\text{pizzas, snacks, drinks}] \Rightarrow R \in [\text{desserts}]$;
8. $C \in [\text{drinks, desserts}] \Rightarrow R \in [\text{snacks}]$;
9. $C \in [\text{desserts, snacks}] \Rightarrow R \in [\text{drinks}]$;
10. $C \in [\text{drinks, snacks}] \Rightarrow R \in [\text{desserts}]$.

Cosine similarity

Empirical experiments have shown that using data for ranging products allows improving purchasing of additional products on the stage of basket approving [13]. In contrast to previous experiments, there is no opportunity to limit data set of users orders or to pivot it to a smaller matrix, and it is necessary to use huge matrix about 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 dataset of vectors N of sold products like a matrix X_{ij} :

$$\begin{bmatrix} x_{11} & \dots & x_{N1} \\ \vdots & \ddots & \vdots \\ x_{M1} & \dots & x_{MN} \end{bmatrix},$$

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 [13]. If we want to calculate 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:

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

But cosine similarity has a high cost of calculation and it is necessary to keep all items in memory in one moment. Our matrix with data about monthly sales has more than 3,000,000 of rows. The solution was found after checking sparsity of our dataset:

$$sparsity = \frac{countofnonzeroitems}{countofallitems},$$

and this parameter equals 0.9695. It means that vectors of our products have too many zero items and they are not useful for calculating. It is possible to isolate co-purchased pairs of products are sold in certain orders like the isolation of co-rated users rates [14].

	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

Picture 1. Isolation of co-purchased products

Picture 1 illustrates the process of isolation, where the matrix columns represent products and the rows represent orders. After the algorithm select orders where two products a and b were purchased, vectors of these items become much shorter than before. Now it is reasonable to determine the calculation of similarity as

$$s = \cos(\theta) = \frac{a' * b'}{\|a'\| * \|b'\|} = \frac{\sum_{i \in P} a_i b_i}{\sqrt{\sum_{i \in P} a_i^2} \sqrt{\sum_{i \in P} b_i^2}},$$

where vectors a' and b' consist of co-purchased products and P is an index of these orders. This approach significantly reduces the cost of cosine similarity calculation and it can be computed with short time without specific hardware. The calculated matrix of similarities S

$$\begin{bmatrix} S_{11} & \dots & S_{1N} \\ \vdots & \ddots & \vdots \\ S_{N1} & \dots & S_{NN} \end{bmatrix}$$

reveals how similar products are if the computing is done 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 products for unpopular because of the value of rare products will much higher than the other.

Weighted Sum Model

Weighted Sum Model (WSM) is the most recognizable method and decision making Simple multi-criteria for evaluating a number of alternatives in terms of a number of decision criteria [15]. In general, suppose a given multiple-criteria decision analysis (MCDA) problem is defined on alternative m and decision criteria n . Next, let's assume that all the benefit criteria are, the higher the values, the better. Further, suppose that w_j shows the relative importance of the criterion C_j and a_{ij} is the performance of alternative A_i when evaluated in terms of criterion C_j . Then, total (that is when all criteria are considered simultaneously) the importance of alternative A_i , denoted as $A_i^{WSM-Score}$, is defined as follows [16]:

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

In our case, the weight can be the probability of purchasing one product with another.

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

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

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

where C is an array of products in a basket and n equals the size of this array. When we sort items of recommendation values we get top N products for offering to the client. Some examples of baskets testing:

- $C \in [\text{veggie pizza}] \Rightarrow R(\text{greek salad, fruit-drink, shallot, thorn, fries, ...})$;
- $C \in [\text{veggie pizza, dodster}] \Rightarrow R(\text{fries, syrnyky, caesar salad, Greek salad, chicken pasta, ...})$;
- $C \in [\text{meat pizza, pepperoni pizza}] \Rightarrow R(\text{coca-cola, BBQ wings, corn, vanilla muffin, ...})$.

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

Structure of recommender system

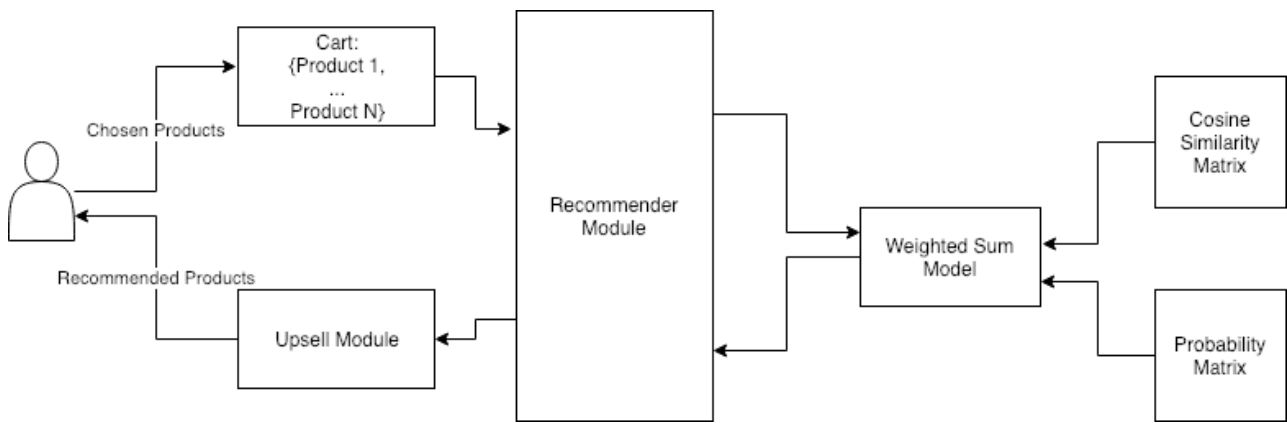
The main idea of statistical learning is to calculate cosine similarity matrix and probability before launch recommender system module:

1. Select co-purchased products.
2. Calculate Cosine Similarity matrix on selected data.
3. Calculate Probability matrix.
4. Upload matrices into database.
5. Launch recommender module based on weighted sum module.

Picture 2 illustrates the scheme of recommendation process:

1. Client choose products from the menu
2. List of products pass to recommender module
3. Weighted Sum Model range other products from the menu based on cosine similarity distance and probabilities
4. Upsell module show ranged products to clients

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

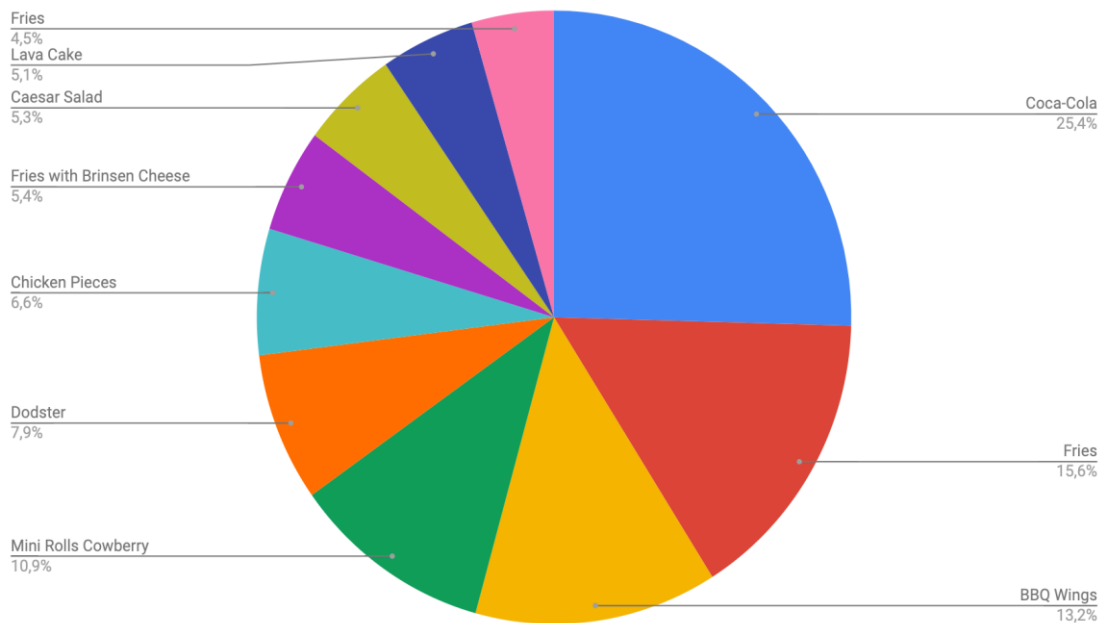


Picture 2. Scheme of recommendation module

Results

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

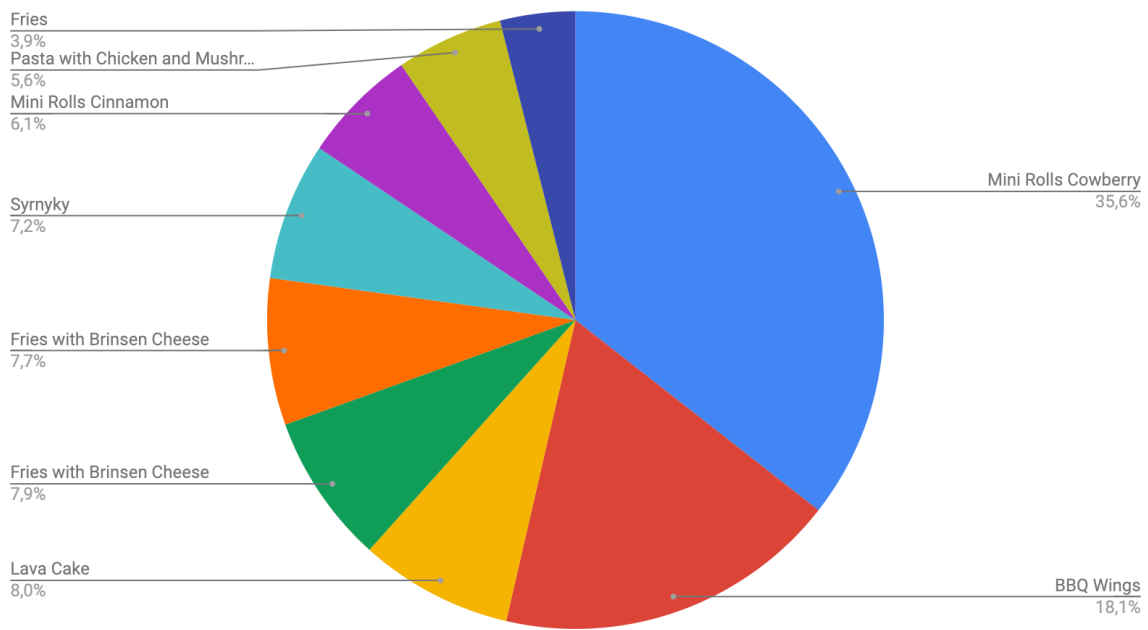
Top 10 sold products through a website



Picture 3. Statistic of top 10 sold products through the web site

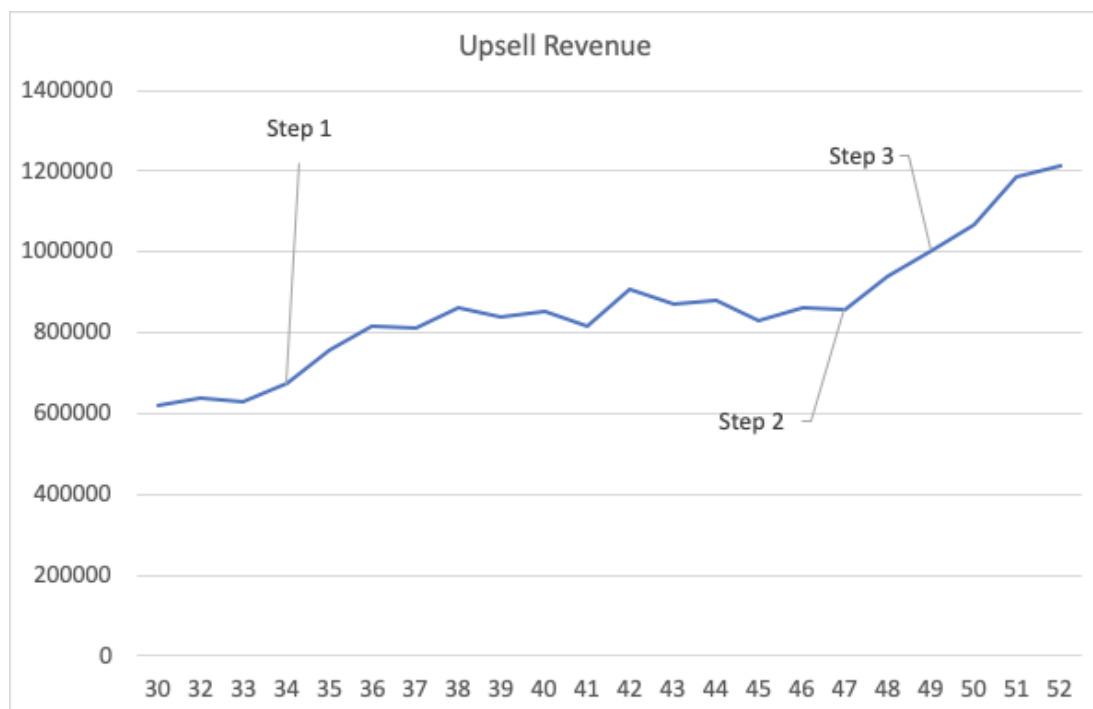
Before the time of launching basket-based recommendations the most of additional sales through the upsell module consisted of only drinks. Picture 3 shows that consumers would buy different products like snacks and desserts: fries, BBQ wings, mini rolls cowberry, etc. This is the result of recommendation wide variety of products.

Top 10 sold products through a mobile app



Picture 4. Statistic of top 10 sold products through the mobile application

Picture 4 shows that as opposed to sales through the web-site purchases via mobile application do not consist of drinks. The reason for this can be an opportunity to show just one product in a mobile recommendation module, but it is clear that it is possible to offer products based on selected items in a cart.



Picture 5. Upsell revenue by weeks

Picture 5 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 clients orders were clustered and sequences of recommendations were prebuilt for offering products [10].
2. Step 2 is the time of launching basket-based recommendations. Cosine similarity and probability matrices for the weighted sum model were calculated on data from four months ago.
3. Step 3 is the time of retraining matrices on data from one month ago.

Results shown on Pictures 3-5 show that statistical machine learning algorithm allows the offer to consumers a great variety of products and this offering is more relevant than expert recommender system. After launching model based on weighted sum customers started buy more often additional products through upsell module and more different items.

Summarize results of launching upsell module based on simple statistics algorithms and gathered data we can observe obvious advantages of applying 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 clients to a wide offering of products.

Conclusion

The first pilot project for improving offering additional products for clients has shown that a data-driven approach and using artificial intelligence are useful tools for developing communication with consumers. Results of improving 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 and machine learning algorithms allow companies to become leaders of their markets.

Literature

1. David Kiron. Lessons from Becoming a Data-Driven Organization., October 2016, Massachusetts Institute of Technology. – [Электронный ресурс]. URL: <https://learning.oreilly.com/library/view/lessons-from-becoming/53863MIT58215/> (date of request: 2019-02-05).
2. Андрей Филипьев. AI Conference. Возможности искусственного интеллекта. «Системный администратор». – 2018. – № 12. – С. 10.
3. Andrew Ng& AI Transformation Playbook: How to lead your company into the AI era. December 2018. – [Электронный ресурс]. URL: <https://landing.ai/ai-transformation-playbook/> (date of request: 2019-02-05).
4. Charu C. Aggarwal. Recommender Systems: The Textbook. 2016.
5. Andres Ferraro, Kyumin Choi, Dmitriy Bogdanov, Xavier Serra. Using offline metrics and user behavior analysis to combine multiple systems for music recommendation. Arxiv:1901.02296v1 [cs.IR]. 8 January 2018.
6. Fabio Aiolli. 2013. Efficient top-n recommendation for very large scale binary rated datasets. In Proceedings of the 7th ACM conference on Recommender systems. ACM, 273-280.
7. Oscar Celma. 2010. Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space. Springer.
8. Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web. ACM, 285-295.
9. Malcolm Slaney, William White. 2007. Similarity Based on Rating Data. In 8th International Conference on Music Information Retrieval. ISMIR.
10. G. Linden, B. Smith, J. York. Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing, Jan/Feb 2003.

11. Dheeraj Bokde, Sheetal Girase, Debajyoti Mukhopadhyay. Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey. *Procedia Computer Science* 49. – 2015. – Pp. 136-146.
12. Prince Grover. Various Implementations of Collaborative Filtering. [December 2017]. – [Электронный ресурс]. URL: <http://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0> (date of request: 2019-02-05).
13. Andrey Filipyev. Improving upsell by clustering customers' orders: machine learning approach. – 2018. – № 3.
14. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. *Item-based Collaborative Filtering Recommendation Algorithms*. 2001, Hong Kong.
15. Mesran, G. Ginting, Suginam, and R. Rahim. Implementation of Elimination and Choice Expressing Reality (ELECTRE) Method in Selecting the Best Lecturer (Case Study STMIK BUDI DARMA). *Int. J. Eng. Res. Technol. – IJERT*, 2017. – Vol. 6. – № 2. – Pp. 141-144.
16. Dikki Handoko, Mesran, Surya Darma Nasution, Yuhandri, Heri Nurdiyanto. Application Of Weight Sum Model (WSM) In Determining Special Allocation Funds Recipients. *International Journal of Informatics and Computer Science (The IJICS) ISSN 2548-8384 (online)*. – 2017. – Vol. 1. – № 2.