

ABSTRACT

The thesis is presented in 80 pages. It contains 3 appendixes and bibliography of 52 references. 16 figures and 4 tables are given in the thesis.

Topic relevance. The profit of online stores in the world reached 1.18 trillion US dollars in 2016 [1]. Every year, the number of online stores increases, and competition between them intensifies. Leading online stores use machine learning algorithms to predict user behavior and thus increase the number of users and purchases. To create such algorithms, online stores publish anonymized datasets about the history of user behavior on their sites on the platforms of machine learning tasks [2] and involve independent experts to work with these datasets. Most of the algorithms are designed to offer the user goods that he has not yet seen, but which he may be interested in, thus inducing the user to buy more during one session. However, such algorithms are not designed to make it easier for the user to reuse the website of the online store or facilitate the return of the user to the site for further purchases. That is why it is relevant to develop a system based on machine learning algorithms, the purpose of which is to predict user behavior to facilitate its work with an online store site, which focuses on goods that the user has already bought and can re-select. This approach will ensure a greater number of user returns to this store, and, thus, increase the profit of the store due to this, so in the system's results the most interested online store owners

Thesis connection to scientific programs, plans, and topics. The thesis was prepared according to the scientific research plan of the Applied Mathematics Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute."

Research goal and objectives. The aim of the dissertation work is to improve the ergonomic properties of the e-shopping service by reducing the number of transactions and the time needed to select and search products on the website.

To accomplish this goal, the following objectives were reached:

- systematize existent methods of affinity analysis;
- systematize the existing methods of classification;

- conduct a comparative analysis of existing systems of forecasting purchases of users based on statistical data;
- develop components of the system for forecasting purchases of users based on statistical data;
- develop interfaces for components of the system for forecasting purchases of users based on statistical data;
- get dataset;
- perform preliminary data processing;
- develop the model of forecasting algorithm;
- develop software for forecasting purchases of users based on statistical data;
- conduct experimental research using different methods of classification;
- formulate recommendations for further use and improvement of the system.

Object of research is Market basket analysis problem, affinity analysis, associative rules, algorithms for finding associative rules: Apriori, Ectat, FP-growth, AprioriDP, AIS, OPUS search, data mining, and the concepts of support, reliability, lift, conviction, machine learning algorithms, classification algorithms, namely, the method of k-nearest neighbors, the random forest method, the naive bias classifier, the method of potential functions, the construction of artificial neural networks, software for processing and storage of large volumes of data (Big data), database management systems of categories of NoSQL and SQL, such as Cassandra and PostgreSQL, algorithm for processing large amounts of data - MapReduce.

Subject of research is the algorithm of forecasting of repeated purchases of users with the use of machine learning algorithms for forecasting purchases of users based on statistical data.

Methods of research. To solve the task, the following methods were used: methods of affinity analysis (for the search for associative rules); methods of machine learning (for the development of forecasting algorithm); methods of algorithm theory and programming (for software implementation of developed components), methods of optimization (for choosing system settings to achieve the best results of forecasting).

Scientific contribution consists of the following:

- for the first time, the task of forecasting purchases of users is set, which, differ

from the existing ones in that it focuses on re-selecting a user's product;

- improved methods of predicting the choice of user, which, in differ from the existing ones in that it takes into account the frequency of user selection of goods which makes it possible to make predictions for unpopular goods;

- heuristic data processing strategies are proposed for the processing of available statistical data;

- proposed advanced approaches to forecasting for some data which in differ from the existing ones in that they operate on a schedule rather than in real time, which allows to reduce the load on the hardware system.

Practical value of obtained results. Proposed a system that can be used to improve the ergonomic properties of the online store site by integrating the system with the website of the online store. The system makes prediction of the user's choice at the next stage of purchasing and returns a list of products that may be of interest to the user to display them on the online store site. The system can be customized to meet the needs of the online store. The forecasting algorithm for the available statistical data is developed.

Approbation of the thesis results. The algorithm of preliminary data processing for forecasting was presented at the scientific and technical conference "Applied Mathematics and Computing"(2018).

Publications. Part of the dissertation materials were presented in theses to the conference "Applied Mathematics and Computing"(2018).

Keywords: machine learning, classification, affine analysis, data processing, forecasting.