INTRODUCTION

Data mining and analysis is a set of methods for detecting large quantities of data from previously unknown, practically useful knowledge for decision-making in various spheres of human activity. The peculiarity of the methods of development and analysis of data, distinguishing them from traditional statistical methods are:

- the discovery of non-obvious regularities: these patterns do not

Identified by standard methods of information processing or expert

By way of;

- the detection of objective laws: the knowledge obtained will be

Fully correspond to reality;

- finding practically useful regularities: the found knowledge can be found in concrete application in practice;

- reveal regularities without rigid restrictions to initial data and their distribution.

The most common tasks of data development and analysis are:

- classification,

- forecasting,

- clustering,

- association.

The basis of methods of data development and analysis is various methods of classification, forecasting, clustering, modeling, based on the application of: decision trees; Artificial neural networks; Genetic algorithms; Associative memory; Fuzzy logic, etc. Methods of development and analysis of data also include multidimensional statistical methods: correlation and regression analysis; Factor and component analysis; Variance analysis; Time series analysis; and etc.

Data development and analysis is a multi-stage and very labor-intensive process, which can be divided into three main stages:

- initial research;

- building a model;

- introduction of the model.

The most time-consuming stage of initial data mining involves:

- Data cleaning: removal of duplicate observations from the sample, mistakenly

Entered data with obvious errors, extreme values (emissions),

Checking logical rules and conditions;

- analysis and restoration, if necessary, of missing values in

Data;

- data conversion;

- setting up data properties;

- conducting an exploratory analysis of data using graphic and

Statistical methods;

- selection of the necessary data for building the model.

At the stage of the model construction, various models of data development and analysis are considered, and the best of them is selected.

The introduction of the selected model implies its application to new data in order to obtain forecasts or estimates of expected results, as well as subsequent monitoring of the quality of the model.

The technology of data development and analysis is used practically in all spheres of human activity, where retrospective data are accumulated. The methods of development and analysis of data were most widely spread in the following sectors: retail trade; The banking sector; Insurance; Telecommunications; Industrial production; Stock and currency markets.

In this tutorial, the tasks of forecasting, classifying and clustering on specific examples are discussed in more detail.