

**INFORMATION COMMUNICATION TECHNOLOGY FIELD
COMPUTER LINGUISTICS AND DATA MINING SYSTEM**

Bozorov Sunnatilla Muhiddin o'g'li

**2-year student of Tashkent State University of Uzbek language and
Literature named after Alisher Navoi**

ANNOTATION

The modern computer term Data Mining translates as “information extraction” or “data mining”. Along with Data Mining, the terms Knowledge Discovery and Data Warehouse are often used. The emergence of these terms, which are an integral part of Data Mining, is associated with a new round in the development of tools and methods for processing and storing data. So, the purpose of Data Mining is to identify hidden rules and patterns in large (very large) amounts of data.

Key words: Data Mining, rules, computer, main tool.

АННОТАЦИЯ

Современный компьютерный термин Data Mining переводится как “извлечение информации” или “интеллектуальный анализ данных”. Наряду с интеллектуальным анализом данных часто используются термины "Обнаружение знаний" и "Хранилище данных". Появление этих терминов, которые являются неотъемлемой частью интеллектуального анализа данных, связано с новым витком в разработке инструментов и методов обработки и хранения данных. Итак, целью интеллектуального анализа данных является выявление скрытых правил и закономерностей в больших (очень больших) объемах данных.

Ключевые слова: интеллектуальный анализ данных, правила, компьютер, основной инструмент.

**ОБЛАСТЬ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ
ТЕХНОЛОГИЙ КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И СИСТЕМА
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

Бозоров Суннатилла Мухиддин угли

студент 2 курса Ташкентского государственного университета узбекского
языка и литературы имени Алишера Навои

INTRODUCTION

The fact is that the human mind itself is not adapted to the perception of huge arrays of diverse information. On average, a person, with the exception of some individuals, is not able to capture more than two or three relationships, even in small samples. But traditional statistics, which for a long time claimed to be the main tool for data analysis, also often fail in solving problems from real life. It operates with averaged characteristics of the sample, which are often fictitious values (average client solvency, when depending on the risk function or loss function you need to be able to predict the consistency and intentions of the client; average signal intensity, while you are interested in the characteristic features and background of signal peaks and t . d.).

Therefore, methods of mathematical statistics are useful mainly for testing pre-formulated hypotheses, while determining a hypothesis is sometimes quite complex and time-consuming task. Modern Data Mining technologies process information in order to automatically search for patterns (patterns) characteristic of any fragments of heterogeneous multidimensional data. In contrast to operational analytical data processing (OLAP) in Data Mining, the burden of formulating hypotheses and identifying unusual (unexpected) patterns is transferred from a person to a computer. Data mining is not one, but a combination of a large number of different methods for discovering knowledge. The choice of method often depends on the type of data available and what information you are trying to get. Here, for example, are some methods: association (association), classification, clustering, time series analysis and forecasting, neural networks, etc.

Consider the properties of the detected knowledge, the data in the definition, in more detail.

Knowledge must be new, previously unknown. The expended efforts to discover knowledge that are already known to the user do not pay off. Therefore, it is new, previously unknown knowledge that is of value.

Knowledge should be nontrivial. The results of the analysis should reflect non-obvious, unexpected patterns in the data that make up the so-called hidden knowledge. Results that could be obtained in simpler ways (for example, by visual viewing) do not justify the use of powerful Data Mining methods.

Knowledge should be practically useful. The knowledge found should be applicable, including on new data, with a fairly high degree of reliability. The usefulness lies in the fact that this knowledge can bring certain benefits in their application.

Knowledge should be accessible to human understanding. The found patterns must be logically explainable; otherwise there is a possibility that they are random. In addition, the discovered knowledge should be presented in a human-readable form.

In Data Mining, models are used to represent the knowledge gained. The types of models depend on the methods of their creation. The most common are: rules, decision trees, clusters, and mathematical functions.

The scope of Data Mining is unlimited - Data Mining is needed wherever there is any data. The experience of many such enterprises shows that the return on the use of Data Mining can reach 1000%. For example, there are reports of an economic effect that is 10-70 times higher than the initial costs from 350 to 750 thousand dollars. Provides information about the project of 20 million dollars., Which paid off in just 4 months. Another example is the annual savings of 700 thousand dollars.

Through the introduction of Data Mining in a supermarket chain in the UK. Data mining is of great value to managers and analysts in their daily activities. Business people have realized that with the help of Data Mining methods they can get tangible competitive advantages.

Classification of Data Mining Tasks

Data mining methods allow you to solve many problems that the analyst faces. The main ones are: classification, regression, search for associative rules and clustering. The following is a brief description of the main tasks of data analysis.

* The classification task is reduced to determining the class of an object according to its characteristics. It should be noted that in this problem the set of classes to which the object can be assigned is known in advance.

* The regression problem, like the classification problem, allows us to determine the value of some of its parameters from the known characteristics of the object. In contrast to the classification problem, the parameter value is not a finite set of classes, but a set of real numbers.

* The task of the association. When looking for associative rules, the goal is to find frequent dependencies (or associations) between objects or events. The found dependencies are presented in the form of rules and can be used both for a better understanding of the nature of the analyzed data and for predicting the occurrence of events.

* The task of clustering is to search for independent groups (clusters) and their characteristics in the entire set of analyzed data. Solving this problem helps to better understand the data. In addition, the grouping of homogeneous objects allows us to reduce their number, and therefore, to facilitate analysis.

* Sequential patterns - establishing patterns between time-related events, i.e. detecting the dependency that if event X occurs, then after a specified time, event Y will occur.

* Analysis of deviations - the identification of the most uncharacteristic patterns. The listed tasks for the purpose are divided into descriptive and predictive.

Descriptive tasks focus on improving understanding of the data being analyzed. The key point in such models is the ease and transparency of the results for human perception. It is possible that the discovered patterns will be a specific feature of specific research data and will not be found anywhere else, but this can still be useful and therefore should be known. This type of task includes clustering and searching for associative rules.

Knowledge should be practically useful. The knowledge found should be applicable, including on new data, with a fairly high degree of reliability. The

usefulness lies in the fact that this knowledge can bring certain benefits in their application.

Knowledge should be accessible to human understanding. The found patterns must be logically explainable; otherwise there is a possibility that they are random. In addition, the discovered knowledge should be presented in a human-readable form.

REFERENCE

1. Develop students' knowledge, skills and competencies in the organizational and technical aspects of essay Akhmedovna Yusupova Tursunay Associate Professor, Candidate of Pedagogical Sciences, Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, Uzbekistan Online published on 3 April, 2021

2. Performance in small groups ensures the achievement of efficiency in mastering T. Yusupova 2016 journal Methods of teaching General Education Sciences № 9 (57), 26-29-p. Ministry of Public Education