DATA MINING AS A SERVICE

by

Artsiom Klimets, 2nd year student, post-graduate student

Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus

Scientific Supervisor: Dmitry Pertsau – Associate Professor

ABSTRACT

Data Mining is a relatively young branch of Data Science, which has taken a special place with evolution of the Internet. This method is widely used in the modern world, in particular for analyzing user behavior and identifying their interests, since classical methods of data analysis are rather trivial and do not allow you to quickly respond to changes in user desires. Data Mining is not a specific technology, but a collection of different methods for solving specific problems. A number of tools make data mining easier, opening up the possibility of applying these approaches to people who are not experts in Data Science. This paper presents the concept of a new tool that implements some of the concepts of existing solutions, but provides a more flexible environment for data analysis.

A large amount of traffic generating every day, for example, according to the latest research conducted by Domo [1], Google conducts 5.7 million searches and Instagram users share 65 thousands photos every minute. All created information flows are analyzed, processed and applied in various fields of activity: advertising, forecast of successful campaigns.

Data Mining is a research and discovery of hidden knowledge that was not previously known, non-trivial, practically useful, available for human interpretation in raw data [2]. In this area, both commercial for example, MatLab, Statistica and free for example, Weka, R, etc. specialized tools have been widely developed. However, they have both advantages designed for specific tasks and disadvantages: complicated interface with many parameters, operation only on a personal computer. Data Mining is not one, but a collection of a large number of different knowledge discovery methods and it is multidisciplinary in nature, since it includes elements of numerical methods, mathematical statistics and probability theory, information theory and mathematical logic, artificial intelligence and machine learning. The scope of Data Mining is quite wide; however, it is limited, despite the use of a large number of previously created methods of analysis and data processing. The special features of data mining require the support of a large variety of data analysis methods in tools under development, and the library of these methods needs to be updated over time. Among already existing tools, the concept of visual programming is widely used, that is a very good solution, simplifying the perception and enhancing the user experience. These features are reflected in the tool under development as part of the research work.

The tool provides a simple and user-friendly interface that does not require any program writing skills or prior knowledge of how to interact with the system. It has a wide range of extensibility possibilities through modularity: any library in Python as the main programming language used in projects related to mathematical statistics, machine learning, neural networks, and data analysis can be plugged in.The system consists of two parts: a server that provides an API to interact with Python libraries and auto generate the final program that will handle the data, and a web interface that can be accessed through the Internet. The interface consists of two areas: the list of available functions to use, presented as a multi-level list, and the area where the order

of operations on the data is lined up, presented as separate tabs for each function. One of the first mock-ups provided below (Figure 1).

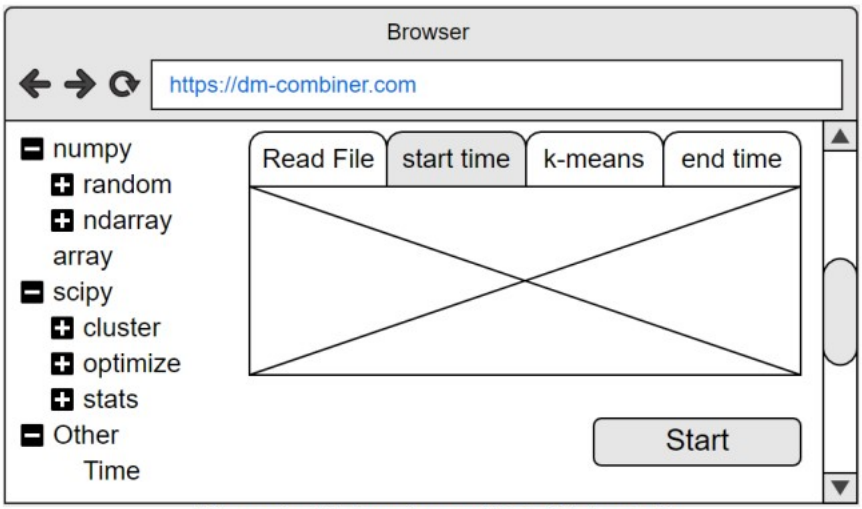

Figure 1 – First mock-up of Data Mining tool

References:

1. Data Never Sleeps 9.0. URL: https://www.domo.com/learn/data-never-sleeps-9 (accessed 10 October 2021)

2. Barsegyan A.A., Kupriyanov M.S., Kholod I.I., Tess M.D., Yelizarov S.I. Analiz dannykh i protsessov [Data and process

analysis]. Saint-Petersburg, BHV-Petersburg Publ., 2009. 512 p. (in Russian)