

# Big Data e-Learning Analytics for Studying the User Behaviour at Ucha.se

Ivelina Nikolova<sup>1</sup>, Darina Dicheva<sup>2</sup>, Gennady Agre<sup>1</sup>, Zhivko Angelov<sup>3</sup>, Galia Angelova<sup>1</sup>, Christo Dichev<sup>2</sup>, Darin Madjarov<sup>4</sup>

<sup>1</sup> Institute of Information and Communication Technologies – BAS, Sofia, Bulgaria

<sup>2</sup> Winston Salem State University, Winston Salem, NC, USA

<sup>3</sup> ADISS – Sofia, Bulgaria

<sup>4</sup> UCHA.se – Sofia, Bulgaria

**Abstract.** The paper presents a behaviour analysis of users of ucha.se - the most popular e-Learning portal in the high school education in Bulgaria, which contains more than 4 000 online video lessons and more than 1 000 exercises. The portal has more than 200 000 registered users of several categories, which logs exceed 4 millions records. Analysis of such a volume was carried out by applications of some database preprocessing and machine learning algorithms. The main objective of the analysis is to find some patterns in the user choices which will lead to better understanding the user needs and will allow enhancing the portal by new specialized features aiming at increasing the user engagement with the portal. The analysis will also allow assessing the e-Learning resources and drawing a map for improving their quality. The paper describes some preprocessing procedures such as data cleaning, attribute selection and construction, as well as data filtering and data transformation aiming at extracting relevant characteristics of each user. A user description vector consists of 3 types of features: personal information, user activities and user skills - 27 attributes in total. The constructed user description is used for learning symbolic models of user engagement with the portal by application of such Machine Learning algorithms as Decision Trees and CN2. Some approaches to advance the models by improving the highly imbalance structure of the constructed training dataset as well as to analyze the role of gamification elements used for organizing the portal learning materials are discussed.

**Keywords:** E-learning, educational data analytics, machine learning, gamification in education.