ГОДИШНИК НА СОФИЙСКИЯ УНИВЕРСИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“
ФАКУЛТЕТ ПО МАТЕМАТИКА И ИНФОРМАТИКА
Том 109

ANNUAL OF SOFIA UNIVERSITY “ST. KLIMENT OHRIDSKI”
FACULTY OF MATHEMATICS AND INFORMATICS
Volume 109

eAnalytics: A MODEL OF A LEARNING ANALYTICS VISUALIZATION SYSTEM

DAFINKA MITEVA AND ELIZA STEFANOVA

Collecting big data from e-learning is already a fact. Huge opportunities open up for learning analytics to deeply understand and effectively optimize the process of teaching and learning. At the same time, large amounts of data require more time, efforts and resources to extract useful information from them. This paper presents a research proposing a model of a learning analytics system aiming through advances in big data visualization methods to facilitate data interpretation and assist timely and accurate decision making. The study describes an architectural model called eAnalytics. It consists of three main components: Data sources connection, Data management and Learning analytics visualization. The last component provides a choice of three methods for visualizing analytics: via a ready-to-use virtual dashboard, via a virtual dashboard template, or by composing personalized analytics. The paper depicts the prototyping of the model and its validation by conducting two types of testing: for functional compliance and by applying “Think aloud” protocol. Finally, important conclusions are drawn and some directions for future work are outlined.

Keywords: Learning Analytics, Big Data, visualization, education

CCS Concepts:

- Information Systems∼Information Systems
- Applications∼Decision Support Systems∼Data Analytics

1. Introduction

In today’s world of information technology and big data, learning and teaching are driven by data and benefit from all their advantages. Analyzing big data brings with it great potentials, but in order to unleash them, users need first of all be able to interpret the results, to extract useful information from them and to use it. To enable these data to “speak” quickly and clearly to decision makers, they must be

DOI: 10.60063/GSU.FMI.109.99-119

properly visualized. This research explores how big data visualization can address challenges in modern education.

The focus of traditional education is the average learner, which increases the risk that at some point the choice of learning materials, teaching methods, assessment of knowledge to be designed for a 'non-existent' learner. Modern education therefore faces a number of challenges and issues: Can learning be customized? Can each learner follow his/her own individual curriculum adapted to the starting level, pace of learning, progress and specific needs? Can the learner's performance be predicted and efforts directed to the areas in which the student is struggling? Can the learner's interest be provoked and kept by responding to personal needs? How can learning be more effective and useful? How to evaluate in what direction to develop/improve teaching materials?

These questions can be answered by learning analytics (Learning Analytics, LA), formally defined in 2011 as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs"[1].

According to Siemens [12] the use of LA could (1) reduce the number of dropouts by identifying at-risk students early and notifying teachers, (2) customize the learning process and content so that each learner receives resources according to his/her current knowledge, (3) notify teachers in a timely manner which learners need additional support and which teaching practices have the greatest impact, (4) contribute to the design of high-quality learning, (5) create interactive visualizations of complex information that will help find answers quickly, etc.

Modern learning is served by Learning Management Systems (LMS), which in addition to learning materials provide technological instruments for collecting and storing information about the user's activity. Within an educational institution, more than one learning systems can be used to serve learning process, specific courses, work on projects and tasks. Sometimes learning needs the use of specific tools and environment covering its particular requirements. Often, entire courses are conducted on platforms outside the administrative control of the primary educational institution. In order to provide a meaningful answer to the questions posed at the beginning, learning analytics from all systems supporting learning must be collected and analyzed as a whole.

In a systematic literature review for the period 2000–2017, published in specialized visualization and learning analytics journals [15], the authors conclude that (1) there are relatively few tools for visual analysis of learning analytics, (2) for visualization of learning analytics still mostly uses traditional statistical data representation techniques such as bar and dot plots, and (3) there is a lack of studies that go deeper into educational theories while applying more sophisticated advanced visualizations.

This paper describes the final part of a larger research with the aim to design a LA system in which modern visualization methods outline the ways to increase the

[1] 1st International Conference Learning Analytics & Knowledge (LAK'11), http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=11606

effectiveness of e-learning and support learners, teachers and managers of educational institutions in making the right decisions at the right time. It starts with (1) a short overview of the previous studies and their results; (2) proposes an architectural model of the target LA visualization system; (3) describes the prototyping of the model, and (4) its experimental validation; (5) finally, conclusions and directions for future works are outlined.

2. The previous steps of the study

The main research began with two studies described in this section: (1) to explore and systematize the reports that users (learners, educators, and managers) expect to find and use in LA section of LMS, and (2) to explore and analyze the visualizations of those reports that users find most intuitive and easy to read.

The first study presents experts' opinion on the use of big data to benefit learning by creating learning analytics in the LMS [8]. Its aim is to derive the main requirements on which to concentrate system development efforts.

The Group Concept Mapping [6] method was used to gather expert opinion and analyze the data. The study involved 30 professionals from the Faculty of Mathematics and Informatics, the Faculty of Education and the Centre of Information Society Technologies of Sofia University "St. Kliment Ohridski", philologists, science teachers, PhD students and undergraduates. The focus question in response to which experts had to brainstorming ideas, was "In Learning Analytics (LA) of LMS I would like to have reports for . . . ". As a result, 85 expert suggestions for reports were synthesized for the learning analytics part of the LMS. The collected ideas were then processed by multi-dimensional scaling and hierarchical cluster analysis and were grouped into eight clusters, representing the categories of reports that LMS users expect to find in the LA section: (1) Teacher evaluation (19 reports); (2) Student evaluation (17 reports); (3) Grades (7 reports); (4) Course Feedback (6 reports); (5) LMS Reports (6 reports); (6) Student Support (9 reports); (7) Student Activity (12 reports), and (8) Course Management (9 reports).

Each proposed report is rated on usefulness/significance and applicability/feasibility and visualized in go-zone diagrams that divide the area into four zones according to the average values of both criteria. In addition, these charts allow filtering the results by role: manager, teacher, and student. Thus, on the one hand, it is able to see the reports that each role has prioritized, and on the other hand, the intersections, i.e. the reports that need to be developed for the LMS users in all roles. The first version of the LA visualization system is supposed to implement the reports that are rated above average by the experts on both criteria.

The second study [9] reveals the experts' vision of which visualizations allow these large volumes of data to be read and perceived most quickly and accurately. The results of it form the basis of the learning analytics visualizations that were implemented.

This study involved experts with different background, experience and LMS role, related to Sofia University. Participants filled out an online survey "Data visualization in LMS" sharing their professional opinion on which visualizations present the most appropriately and accurately data in the main types of reports. The list of reports for the LA section of the LMS defined in the previous study was further analyzed in terms of data types and the purpose of their visualization: composition, comparison, linkage or disaggregation. In accordance with Abela's classification [1], 12 main data types were extracted for illustration. For each of them, a section was defined in the questionnaire with a brief description of what type of data is expected to be visualized, for whom and for what purpose. From two to six sample visualizations were proposed for a score from 1 (inappropriate) to 5 (appropriate). Each section ended with an opportunity for comments and new ideas.

The collected data were processed, analyzed and categorized according to the system roles of the respondents in order to propose virtual dashboards with visualizations that would satisfy the largest possible number of LMS users.

3. The eAnalytics model

This section describes an architectural model of a learning analytics visualization system called eAnalytics. The purpose of LA visualization system model is to meet as much as possible the requirements of users using the LMS in the three main roles: student, teacher or manager of an educational institution.

3.1. Users and roles

eAnalytics recognizes several default system roles, which are granted different permissions on the available functions.

Manager. Designed for the institution's management team to derive analytics on both learner and educator participation in a course, overall performance, and student opinion and evaluation of their instructors.

Teacher. Designed for lecturers to monitor the overall conduction of a course, to observe the individual performance of each learner, to create a complex analyses of student performance.

Student. A role for learners to monitor their performance, successes and weaknesses and receive feedback on how to improve their own performance.

Administrator. He is responsible for the installation, configuration and management of the system, is granted full access.

The system allows new roles to be defined if necessary with appropriate rights delegated to them.

3.2. Components of the system

eAnalytics is a web-based system with a hub architecture between connected learning systems that stores data on teaching, learning and learning management processes to give a holistic view of learner progress. Collecting more data for more accurate evaluation is realized by merging data from multiple learning analytics sources. Using APIs, the system connects to different LMSs and databases and uses metadata to describe which analytics they will provide.

The system consists of three main components: (1) Data sources connection component, (2) Data management component, and (3) LA visualization component.

- **Data sources connection component** is responsible for providing connection and access to the sources according to the specific user, his roles and permissions in the respective LMS. The metadata for this access are stored in the system and are used to retrieve data from it (Figure 1). The main unit to which data are associated is a course in the LMS. When a trigger event occurs this component notifies the user and performs an update of connection and metadata. The component, on the other hand, periodically updates the data that are retrieved from LMS and are visualised in the eAnalytics dashboard.

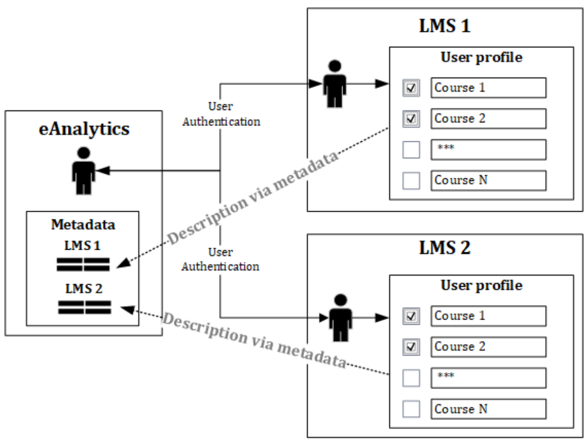

Figure 1. Data sources connection component

- **Data management component** wich includes three modules: (1) metadata module, (2) data processing module, and (3) local repository.

  a) *Metadata module.* Once metadata are extracted and associated relationships described, this module builds a hierarchical metadata model structured by system role, course access, and data provided (Figure 2). The basic unit into which metadata are grouped in eAnalytics is a course, and the access to analytics depends on the user's role in the system.

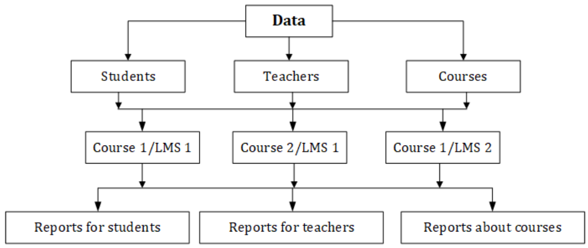

Figure 2. Metadata structure in model

b) *Data processing module.* For some of the analytics, mainly related to combining data from different sources, additional processing is required in the system itself. eAnalytics offers a set of standard formulas, mathematical and statistical functions for data processing and transformation, for text analysis, for data parameterization.

c) *Local repository.* Extractions from big data can be imported and stored in local storage. Such can be used when the data transmission from the source to the system does not need to be implemented in real time. In such cases CSV or XLSX files can be exported from and imported into system.

When the Metadata module receives a report data request, it considers whether to query the local repository or the relevant LMS via Data sources connection component. If the data are sourced externally (the relevant LMS), extracts from them may or may not be copied to the local repository. The data retrieved are passed to the Date Processing module for handling, and then forwarded to the LA visualization component to be displayed in the dashboard (Figure 3).

- **Learning analytics visualization component** provides a choice of several methods for visualizing analytics: (1) via a ready-to-use virtual dashboard, (2) via a virtual dashboard template, or (3) by composing personalized analytics. In the first method, the virtual dashboard contains information about

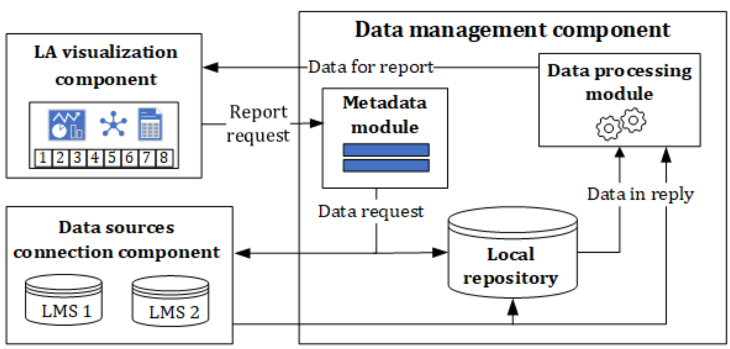

Figure 3. Data management component

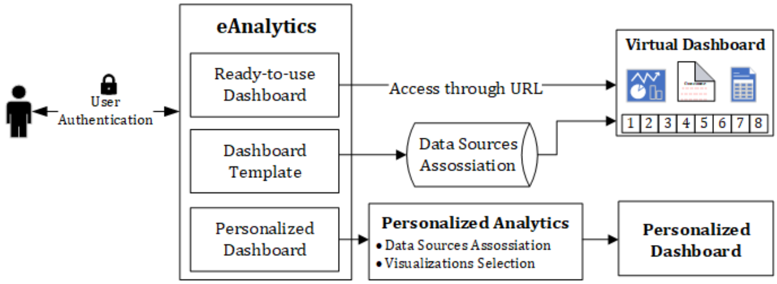

Figure 4. LA visualization component

the data sources and can be shared in a view-only or editable manner. The second option enables the user to select/specify the data sources themselves in order to obtain the corresponding visualizations. The resulting dashboard remains in the user's profile. The third option directly links to a custom analytics composition screen where the data sources must first be selected, then the data itself, and finally the visual objects that will represent the data. Based on the selected data types, "appropriate" visualizations will be recommended, but the final choice is left to the user (Figure 4).

Visualization via a ready-to-use virtual dashboard. The virtual dashboard that the user accesses is completely ready to use. Its main advantage is that it does not require any additional settings, nor knowledge of the data structures and the relationships between them. Its content varies depending on the system role of the user. If a user has more than one role assigned, he/she can choose which one to take at any given time.

The dashboard consists of eight sections corresponding to the LA categories described in part 2: (1) Teacher Evaluation, (2) Learner Activity, (3) Learner Evaluation, (4) Course Management, (5) Grades, (6) Student Support, (7) LMS Report, and (8) Course Feedback. The list of analytics reports and visualizations in each category varies depending on the current role and expert perspective on selecting visualizations.

When the dashboard is opened, the *LA visualization component* calls the *Data management component* and, via the *Metadata module*, extracts the expected data from which LMS they should be extracted. The *Data source connection component* queries the relevant LMSs "on behalf of the given user" and receives the requested data in response. It combines the received data from all the requested LMSs and passes them to the *Data processing module*. If necessary, these data are stored to the local storage. The resulting data are presented with appropriate visualizations and are displayed on the dashboard of the respective user.

The LA visualization component allows switching between the default visualizations developed for different roles, as well as reflecting the settings in the user's profile. Additionally, options are provided to filter the data and export the results in tabular and/or graphical format.

Visualization via virtual dashboard template. The template is structured as a ready-to-use virtual dashboard, contains all the analytics and the appropriate visualizations for them, keeps the metadata model and the relationships between them, but has no connectivity to specific data sources and the local repository is empty. In order to visualize analytics, the template must first be associated with data sources. Then, the data sources connection component and the metadata module complete and update the metadata. This is followed by extracting, processing the data and displaying the relevant analytics in the virtual dashboard.

This visualization method has a number of advantages. The virtual dashboard is versatile, it can work with different sources that support the same data model. For eAnalytics, this means that developed for one specific institutional unit, it can be applied to multiple similar.

Visualization via composing personalized analytics. Creating a personalized virtual dashboard allows the user to define which analytics they want to include and which visualizations to use for them. The personal virtual dashboard is created in the following sequence of actions in the system: (1) selecting data from the metadata module, (2) selecting visualizations, (3) customizing parameters.

The data that are provided to compose analytics are described clearly and intuitively in the system and give a precise idea of what information they include. They can be accessed from different perspectives: student data, faculty data, course data.

Each user can personalize the layout of the system for his profile. The administrator role defines the global layout settings for the entire system.

3.3. Mobility, security and ethics

No matter if users are connecting from a desktop, laptop, tablet or smartphone, the analytics reports are offered/adapted in the best possible format for the respective device. The system is available via Android and iOS apps.

In order to reduce the risk of vulnerability in the storing sensitive data, the model presumes addressing the relevant LMS only upon request and for retrieving only the requested data provided per selected account. When results need to be published to third parties, data are anonymized.

The model envisages the application of a code of ethics, in accordance with the requirements of the organization that will use the relevant learning analytics system and the rules to be published explicitly in the system.

4. Prototype implementation

This section presents the work on prototyping the LA visualization system. An implementation of the model is demonstrated for system role manager and ready-to-use virtual dashboard for the category “Teacher Evaluation” (Figure 5). A dashboard template that can be used repeatedly afterwards is also created in the prototype. The template stores information about the data model, the visualizations created in it, but does not contain specific data.

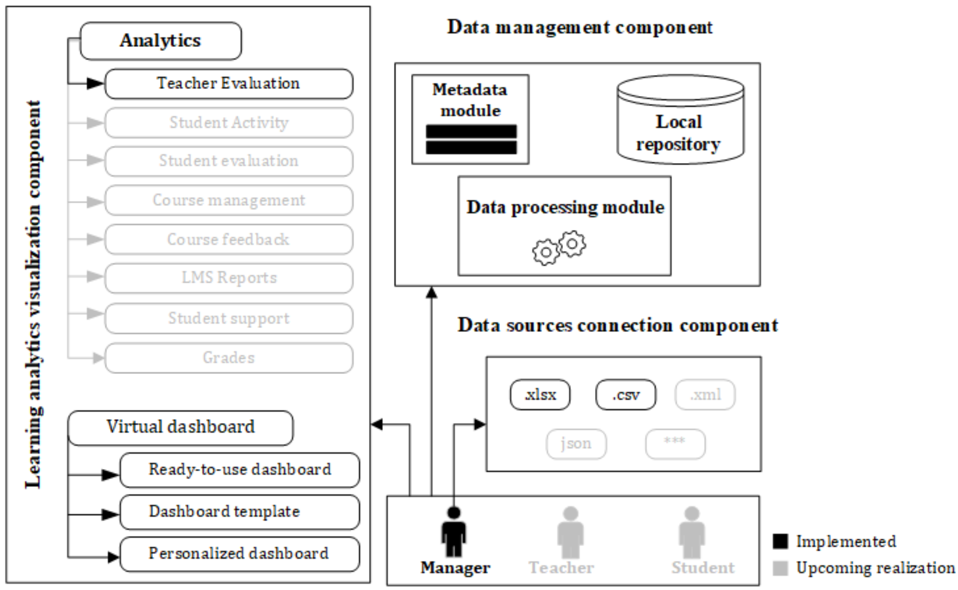

Figure 5. The implementation of model

4.1. Choosing a tool for dashboard implementation

There are a number of modern technologies and tools for visualizing big data increasingly penetrating the field of education. Some of them offer an easy and user-friendly interface, others require considerable technological and programming knowledge. Some of these tools are closely dedicated to particular types of visualizations, e.g., geographic maps, others provide a variety of styles and are open to adding new ones.

For developing an LA visualization system, a good starting point was to choose a relevant off-the-shelf platform that best meets the requirements derived in the previous studies, mentioned in Part II. Available technological environments were analyzed and evaluated according to the following criteria:

- Set of visualizations. On the one hand, the platform should offer a rich set of ready-to-use and customizable visualizations, on the other hand, the user should be able to easily create their own or import ones shared from an external environment;

- Data sources. The platform should allow connection to more than one data source in order to combine, analyze and visualize data from more than one LMS;

- Real-time data visualization;

- User-friendly interface. The user should be able to easily and quickly add new visualizations without requiring special IT skills;
- Mobile version. Users must be able to access the system at any time and from any location;
- Price-quality ratio;
- Reliability;
- Sharing and use of the virtual dashboard.

In 2022, the leading research company Gartner published a report [5] evaluating 20 analytics and business intelligence (BI) platform providers. This type of platform, according to Gartner's definition, provide the ease and convenience for the user to "analyze, explore, share and manage data", enable him to visualize his own and discover others' insights, often based on artificial intelligence. In this study, the platforms were evaluated according to the following criteria: (1) security; (2) management; (3) cloud analytics, (4) connecting data sources; (5) data preparation; (6) catalogue; (7) automated insights; (8) data visualization; (9) natural language query; (10) data storytelling, (11) natural language generation, and (12) sharing. The rating is based on all published reviews, with older ratings halving every 12 months.

Based on these, Gartner defines its "magic quadrant" in which companies are categorized as "leaders", "challengers", "visionaries", and "niche players". For 2022 (Figure 6), the leadership position in combining vision and providing value is occupied by Microsoft. Power BI, combined with Azure and Office 365, offer a visual-based presentation of data, advanced analytics, and interactive virtual dashboards. The next position is occupied by Tableau and Qlik.

Figure 6. Evaluation of analytics and business intelligence platforms

Microsoft Power BI

Microsoft Power BI[2] is one of the leading tools for data visualization and business intelligence. It is a collection of applications, data connectors and software services that connect data from different sources, process it and create visual reports.

Well-known users of Power BI are companies such as Hewlett-Packard, Nokia, RedHat, Heathrow Airport, etc. Among the educational institutions that use Power BI as a primary tool for learning analytics are University of Waterloo (Canada), Southern New Hampshire University (USA), Jacksonville University (USA), University of Surrey (UK), etc. According to InfoCluch[3] statistics, 4.4% of Microsoft Power BI customers are in higher education.

Power BI supports over 500 data sources, including file formats such as XLSX, XML, JSON, PDF, SharePoint, etc., databases such as SQL, MySQL, Oracle, Amazon Redshift, etc., Power Platform products such as Power BI datasets, Power BI dataflows, etc., the Azure cloud platform, online services such as GitHub, Salesforce, Asana, etc.

Power BI provides the ability to build interactive virtual dashboards by updating data in real time. For convenience, it offers pre-defined dashboard templates as Software as a Service (SaaS). Ensures secure and reliable connections to cloud data sources and allows natural language query definition.

Power BI offers ready-to-use almost all popular classic visualizations with customizability. A community has been established where custom visualizations are shared for free. Visualizations can be programmed using Python or R scripts. The local application has an English version.

Tableau

Tableau[4] is one of the most popular visual analytics and business intelligence platforms. Among its users are well-known companies such as Pfizer, BMW Group, Nissan, Splunk, Lenovo, universities such as University of Texas at Austin (USA), University of Kentucky (USA), Florida International University (USA) and others. Some institutions related to higher education management are conducting research on its use in achieving the sustainable development goals [3].

Tableau imports data in various file formats, e.g., CSV, JSON, XML, XLSX, etc., works with relational and non-relational databases such as PostgreSQL, MySQL, SQL Server, MongoDB, etc., accepts data from cloud systems such as AWS, Oracle Cloud, Google BigQuery, Microsoft Azure. It processes data in real time. It can combine two or more heterogeneous big data sources, including data warehouses, data marts, text files, tables, three-dimensional data (data cubes), without requiring prior integration [10]. Interactive dashboards [2] can be created without requiring

[2] Microsoft Power BI, `https://powerbi.microsoft.com`

[3] Microsoft Power BI Customers List. InfoClutch, `https://www.infoclutch.com/installed-base/business-intelligence-software/microsoft-power-bi/#counts-by-industry`, 2022

[4] Tableau Software, `https://public.tableau.com`

significant programming knowledge and effort for fast and compact data viewing. The story telling feature allows the creation of an electronic presentation that can be used to navigate through the data to tell an entire story.

In comparative study of data visualization tools in big data analysis for business intelligence published in 2022 in the international journal IJRASET is reported that Tableau performs "complex table calculations can be performed more rapidly with scripting languages such as R and Python" [12].

Qlick

QlikView and its successor QlikSense[5] are data analytics and visualization tools that allow the user to import data from a variety of sources, and based on business intelligence, discover patterns in the data, create analytics solutions and virtual dashboards [4]. Well-known universities share their experience with QlikView include University of Maryland College (USA), Amsterdam University of Applied Sciences (Netherlands), Saint Joseph University (USA), Xi'an Jiaotong-Liverpool University (XJTLU) (China) and others. An experimental study on the use of Qlik's social analytics products, and in particular Qlik Analytics, shows that Qlik can be successfully applied in educational institutions to collect learner and trainer statistics, success stories, and thus help learners grow, prepare them for better outcomes in the future [4]. Qlik supports connectors for linking to over 32 data sources, such as Amazon S3, Azure, Facebook Insights, Google Analytics, Twitter, and more. Offers traditional interactive charts with customization options. Emoji insertion is possible if programmed. Creating a unique dashboard requires programming. Advantages of QlikView are easy and user-friendly interface, impressive color visualizations, and hassle-free support.

Based on a rigorous analysis of the advantages and disadvantages of the virtual dashboard tools the choice was made to use Microsoft Power BI platform to create the prototype, as it covers the most of the described platform selection criteria.

4.2. Data and metrics used in the prototype

The prototype of the model is implemented on a real problem with the processing and visualization of large volumes of data related to the students' feedback in the courses at Sofia University "St. Kliment Ohridski".

At the end of each academic term at Sofia University (SU) a course evaluation survey is conducted, through which students give feedback on the courses they have studied, the attendance of classes, the work of their lecturers and assistants, share their opinions and recommendations. In this questionnaire, students' rate on a scale of 1 (negative, disagree) to 5 (positive, completely agree) or yes/no/cannot decide the performance of lecturers and teaching assistants – whether they present the teaching material in an accessible way, whether they encourage students and stimulate creative thinking, whether they hold regular classes, whether students

[5] QlikSense, https://www.qlik.com

approve teaching methods, whether students would recommend the course to their colleagues. Each of the questions is evaluated with a certain weight and according to a regulated by SU formula forms part of the teacher's attestation evaluation. So far, the results of this questionnaire have been processed manually and this made it difficult to analyze and extract meaningful information.

The virtual dashboard described in this chapter was developed according to the data model of these surveys.

4.3. Importing the data and binding it to the model

For the purpose of the demonstration, the data were obtained as "Excel Workbook" and "Text/CSV", pre-formed into three different groups simulating three different data sources. Two new tables have been added: "semester" – a list of the codes and full names of each semester for which there are available records, and "instructor-course" – a list of the unique pairs of instructor and course. DAX language[6] was used to clean the data and to calculate average grades for a given period.

A metadata model has been built that specifies the relationships between the different datasets. The possible relations are (1) one-to-one, (2) one-to-many, (3) many-to-one, and (4) many-to-many. The system provides the ability to define relationships visually through drag-and-drop functionality.

When imported, some of the object properties are renamed locally. Thus, complex names and abbreviations get intuitive and easily "understandable" names for the user. Once named in the model, they retain their new names permanently, even if the data source is changed. To avoid confusions some of the key attributes are marked as "hidden" and unavailable for modification. Only attributes accessible to the user for generating personal drag-and-drop reports are left visible.

4.4. Creating a dashboard and selecting visualizations

According to the eAnalytics model, the virtual dashboard for "Teacher Evaluation", accessible for the manager role provides eight reports to the user. The visualizations were grouped into four sections/pages, and a start page was added with instructions what the user should expect from this virtual dashboard and what the evaluation criteria are applied.

4.5. Creating a virtual dashboard template

To make the developed virtual dashboard reusable but with different data sources supporting the same format and data model, the system offers the possibility to create a virtual dashboard template. If the data source is set as a parameter in this template, the next time the template is opened the user will be asked to select what data to associate with the template.

[6] Data Analysis Expressions (DAX) Reference, https://learn.microsoft.com/en-us/dax

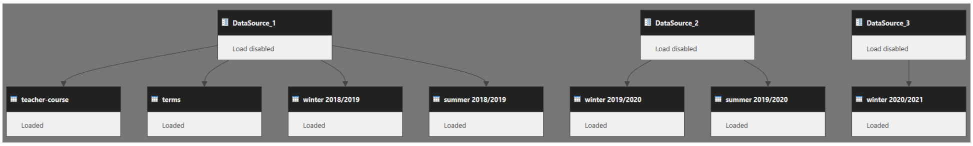

Figure 7. Query dependencies in the prototype

In the prototype, the data sources are set as parameters. On startup, the parameter values are set and all data gets a real path to the corresponding source. The query dependency model is shown in Figure 7 and it reveals that when the DataSource_i (i=1, 2, 3) parameters are set to a value, all the data will get a real path to the corresponding source.

Shared with different system users, this template can also be parameterized by role to display the appropriate data and visualizations depending on the user's role and profile.

4.6. Creating an app and sharing in an online store

The eAnalytics application (eAnalytics_app) was created for the purpose of the prototype and was published in Power BI apps as it is shown in Figure 8. The app was shared with certain stakeholders from Sofia University and is available for downloading and use in their profiles only.

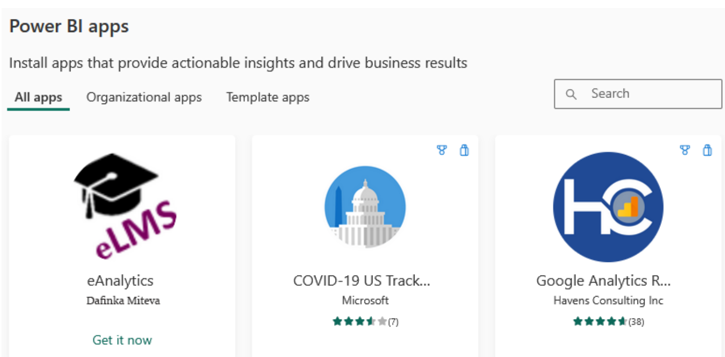

Figure 8. eAnalytics in Power BI apps

After installation, eAnalytics app displays the virtual dashboard with reports in "Teacher Evaluation" category.

5. Testing the prototype

This section describes the experimental testing of the created eAnalytics prototype in response to Task 7. Two types of tests were conducted: functional tests for compliance of the implemented with the set requirements for the prototype and user tests by applying the "Think aloud protocol" [7].

Real data cannot be used for testing because it contains sensitive information that cannot be accessed for research purposes. At the same time, so many "artificial" data corresponding to the real data cannot be generated. For this reason, the prototype was tested in isolation with a real data set that was anonymized. The data are from surveys in the HSE of a specific faculty of Sofia University "St. Kliment Ohridski".

5.1. Preliminary data preparation

Before processing, data were cleaned of incorrect or partially completed records. Reports generated with LMS data contain sensitive information and for the purposes of the demonstration de-identification was applied to mask personal data and minimize the risk of inadvertent disclosure of faculty and their information. The open-source tool Amnesia which follows GDPR guidelines for pseudo-anonymization, was used. Through it, the link to the original dataset is broken. The identifiers first name, lsurname, last name, and academic discipline, which in combination or separately would help to identify the evaluated teacher, were pseudo-anonymized. Strings were masked to three characters, only one of which is a letter, so the full name is represented as "L**J**S**". Courses were masked to five characters, only three of which are letters, e.g. "LEK**". Since by the initials a particular lecturer can again be identified, all visible letters are further replaced, e.g. A is replaced by B, Y is replaced by A. "Name", "Surname" and "Family" hierarchies have been created in which the replacement rules are clearly defined. These hierarchies were applied to all datasets to achieve uniformity in coding all occurrences of the same educator. For teaching disciplines, considering the length of the titles, the option to automatically create a hierarchy was used.

5.2. Functional testing of the prototype

This section describes the tests performed on the developed eAnalytics functionalities: (1) opening a virtual dashboard template and configuring it to work with real data sources, (2) testing a ready-to-use virtual dashboard, and (3) creating a custom virtual dashboard.

Virtual dashboard template testing

When the template was opened, a dialog box first appeared (Figure 9) prompting the user to select (or enter a web address of) real data sources to be visualized in the virtual dashboard.

In the tested template, the user was facilitated to directly select them from a drop-down list. Data prepared for the purpose of the experiment (cleaned, anonymized and completed as described above) were passed, and the virtual dashboard loaded successfully.

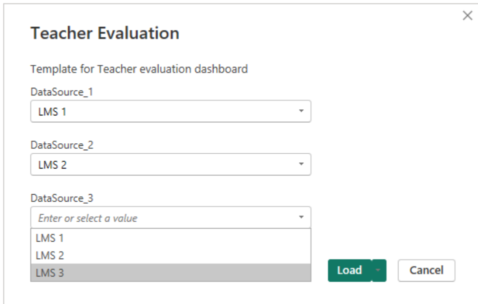

Figure 9. Selecting data sources to display the template

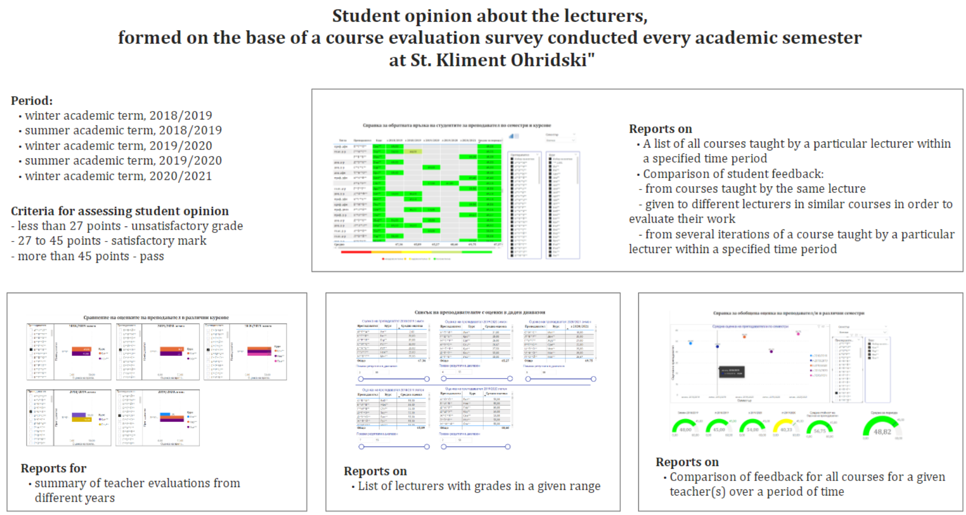

Figure 10. Virtual dashboard "Teacher Evaluation"

The initial screen (Figure 10) was shown first, describing the period under analysis, the criteria for assessing student opinion and the types of reports provided, with links to more detailed presentation.

By following the links or navigation between the pages, all sections of the virtual dashboard were visited and analyzed.

Ready-to-use virtual dashboard testing

The ready-to-use virtual dashboard, published as an application in online store, was tested by a registered Sofia University user of the Power BI service, with whom it had previously been shared. The app was accessed (1) via a URL link provided upon sharing, (2) by scanning a QR code, and (3) by searching for its name "eAnalytics_app" in the Power BI apps online store as it is shown in Figure 3.

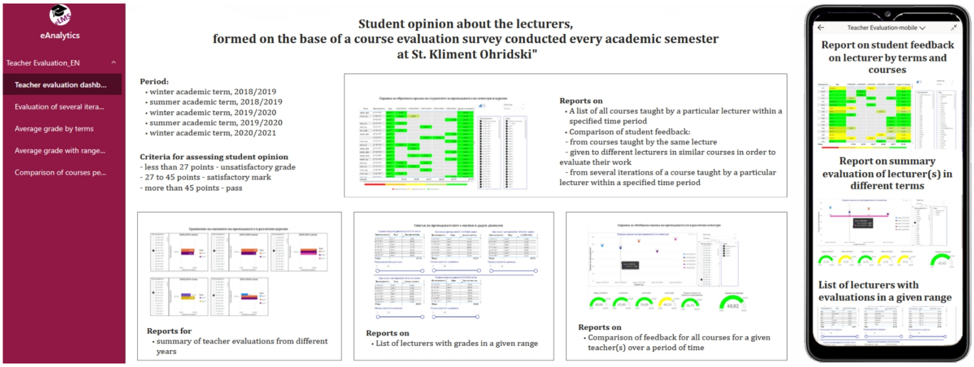

Figure 11. eAnalytics as an application

The application is downloaded to the user's account, installed and available for online use at any time by computer or mobile device (Figure 11).

Personalized dashboard testing

eAnalytics was shared with a test user with editable permissions. This user was able to create own personal analytics with custom visualizations using provided data. The user created a new report in which (1) marked to use data on the number of faculty with different average evaluation scores for all analysed semesters, (2) selected a pie chart as the visual object, (3) customized the visualization settings and illustrated the proportion of all faculty according to their value for the entire analyzed period (Figure 12).

From all the performed tests it can be concluded that the implemented functionalities of the prototype work as intended in the prototype.

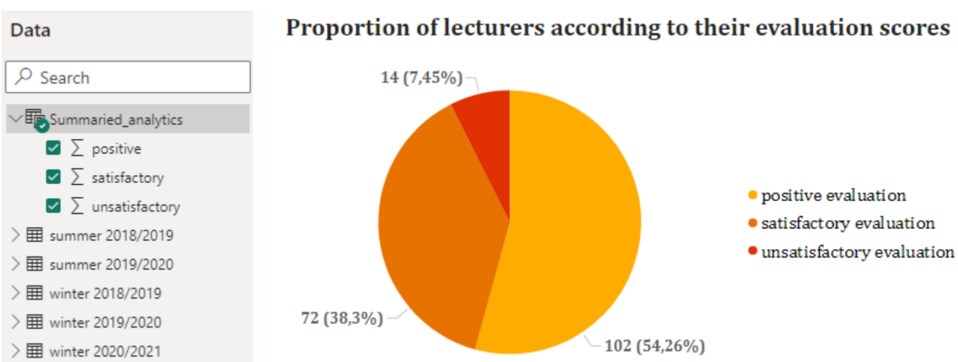

Figure 12. Testing the creation of a personalized dashboard

5.3. User tests with "Think Aloud" protocol

User tests were conducted with a "Think Aloud" protocol. It is a common cognitive method for testing the usability of software in which users speak aloud each thought, feeling, or idea as they perform a series of given tasks. An observer records each opinion expressed or action performed without any intervention. This method has been used successfully in a number of validation and testing studies, including the Share.TEC project [13,14].

In this type of testing of eAnalytics were involved ten experts from the Faculty of Mathematics and Informatics of Sofia University who declared they have been actively using LMS in their work for more than ten years. One of them is a member of the faculty attestation committee. The application was shared from the Power BI online store with them in advance and each participant had to complete the following list of tasks to test the web and mobile version:

1. Find and download the eAnalytics app from the online store.
2. Launch the application.
3. Browse all pages, view the suggested analytics and use the navigation.
4. Switch between the different types of visualizations available for the same analytics.
5. Finding the highest and lowest faculty member's grade for each of the five semesters displayed. (Without specifying which report to search for information.)
6. Retrieve three different lists of lecturers: with unsatisfactory, satisfactory and positive attestation for a randomly selected semester. (Without specifying which report to use for information.)

Each participant was instructed to share aloud any thoughts that occurred to him or her while completing the assigned tasks.

Results of observation

The first and second tasks were done very quickly. Navigation was found easy and intuitive, and switching between the different visualizations was apparently and logical. Tasks 5 and 6 were completed with ease. The only comment was related to the explicit display of the download button, not only when the object is in focus, i.e. the mouse is over it.

First impressions were expressed with positive enthusiasm like "Oh, perfect, this will be very useful for managers!", "Wow, the interface is adapted for users with special needs!", "The idea with the colors is very good, you do not need to look at the numbers to orient yourself!", "Oh, this is very useful too!", etc. After the initial euphoria, a more critical evaluation and list of recommendations was suggested. Some of the proposed interface-related modifications have already been implemented.

One of the experts, using LMS in the role of manager, instinctively started to interpret the data. He found lecturers whose scores had gradually risen and concluded that these lecturers had carefully reviewed the feedback from students and taken on board the recommendations made to them in each subsequent course.

The heat map was identified as the clearest visualization. One of the experts suggested colour differentiation of the grade ranges to be used in all tabular and matrix visualizations.

The mobile version was evaluated as more intuitive, easier to use and with a better design.

Testing of the prototype shows experimentally that big data, statistics and graphic visualization of the data speed up the perception of the data many times and enable the participants in education to answer at least some of the questions asked at the beginning.

6. Conclusions and future works

The research studies the possibilities of increasing the efficiency of learning, teaching and learning management processes by using modern methods for visualization of learning analytics. Its goal is to design a LA visualization system that, through appropriate visualizations, supports learners, teachers and managers of learning institutions to make the right decisions at the right time.

The described LA visualization system will continue to be developed in several directions.

First, it is based on a survey of what learning analytics visualization experts find useful and necessary. For some of these reports, learning systems collect and store enough data to be implemented. For other reports, additional planning is needed on the one hand what data to collect, on the other hand how, by what means.

Another study on the use of advanced data visualization methods is to be conducted. The latest big data visualization methods applied to LA will be presented in a real-world environment so that respondents can better appreciate the bene–fits and challenges of each one. Visualizations programmed in R and Python will also be added, which greatly increases the methods for data presentation and customization. The current version of eAnalytics will be used as a base for this study.

Given the sensitivity of the data in the LMS that is processed and visualized in the learning analytics, one of the directions for further development of eAnalitics is to add a local server for the university to host all the virtual dashboards. Thanks to it, the data processed and visualized by eAnalytics will not leave the Sofia University network. Their sharing will be with users from the university itself or, in the case of external users, secure protocols will be used to ensure data security and protection. A ‘gateway’ application will be added to the server to allow data to be kept on campus even when a virtual dashboard is published to the cloud. In addition, a service will be added to embed the eAnalytics virtual dashboards within the SU systems.

Modern big data processing and image visualization technologies are constantly evolving. The requirements for the LMS are constantly increasing, new functionalities are developed, more and more data are collected. Expectations for learning analytics visualizations will also grow. To be useful, eAnalytics will be keep up-to-date to meet the requirements of its users: educators, students and managers.

Acknowledgements

This study is partially supported by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project No. BG-RRP-2.004-0008-C01.

References

[1] A. V. Abela, The presentation: a story about communicating successfully with very few slides, CreateSpace, 2010.
[2] N. Akhtar, NaziaTabassum, A. Perwej and D. Y. Perwej, Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus), 2020.
[3] A. Amer, Tableau big data visualization tool in the higher education institutions for sustainable development goals, IJCSMC 8(7) (2019) 71–78.
[4] S. Ghosh and K. Neha, Social media analytics using Qlik connectors, IJRTE 8(2S11) (2019) 3527–3530.
[5] J. Heizenberg, G. Herschel, T. Lo, S. Vashisth and A. Jain, Gartner magic quadrant for data and analytics service providers, Gartner, Inc, 2022. [Online]. Available: https://www.gartner.com/en/documents/4011116.
[6] M. A. Kane and S. R. Rosas, Conversations about group concept mapping: Applications, examples, and enhancements, SAGE Publ., Inc, 2017. [Online]. Available: https://lead.to/amazon/com/?op=bt{&}la=en{&}cu=usd{&}key=1506329187.
[7] C. Lewis, Using the "thinking-aloud" method in cognitive interface design, IBM Thomas J. Watson Research Division, 1982. [Online]. Available: https://books.google.bg/books?id=F5AKHQAACAAJ.
[8] D. Miteva and E. Stefanova, Design of learning analytics tool: The experts' eyes view, in: Proc. 12th Int. Conf. on Computer Supported Education, vol. 2, CSEDU, SciTePress, 2020, 307–314, https://doi.org/10.5220/0009395503070314.
[9] D. Miteva and E. Stefanova, Experts' perspective for selecting learning analytics visualizations, in: 2020 Int. Conf. Automatics and Informatics (ICAI), 2020, 1–6, https://doi.org/10.1109/ICAI50593.2020.9311341.
[10] K. Morton, R. Bunker, J. Mackinlay, R. Morton and C. Stolte, Dynamic workload driven data integration in tableau, in: Proc. 2012 ACM SIGMOD Int. Conf. on Management of Data, in SIGMOD '12, New York, ACM, 2012, 807–816, https://doi.org/10.1145/2213836.2213961.
[11] A. Pandey, I. Sharma, A. Sachan and Dr. P. Madhavan, Comparative study of data visualization tools in BigData analysis for business intelligence, IJRASET 10(VI) (2022) 2592–2600.

[12] G. Siemens, D. Gašević, C. Haythornthwaite, S. Dawson, S. Shum, R. Ferguson, E. Duval, K. Verbert and R. S. Baker, Open Learning Analytics: an integrated modularized platform, Working Paper, Jan. 2011.

[13] K. Stefanov, R. Nikolov, P. Boytchev, E. Stefanova, A. Georgiev, I. Koychev, N. Nikolova and A. Grigorov, Emerging models and e-infrastructures for teacher education, in: 2011 Int. Conf. on Information Technology Based Higher Education and Training, 2011, 1–8, https://doi.org/10.1109/ITHET.2011.6018688.

[14] E. Stefanova, N. Nikolova, E. Peltekova, K. Stefanov, T. Zafirova-Malcheva and E. Kovatcheva, SHARE.TEC: An innovative solution for teacher educators, in: ICERI2011 Proc. 4th Int. Conf. of Education, Research and Innovation, IATED, 2011, 1679–1688.

[15] C. Vieira, P. Parsons and V. Byrd, Visual learning analytics of educational data: A systematic literature review and research agenda, Computers & Education 122 (2018) 119–135, https://doi.org/10.1016/j.compedu.2018.03.018.

Received on May 29, 2023
Accepted on June 23, 2023

Dafinka Miteva and Eliza Stefanova

Faculty of Mathematics and Informatics
Sofia University "St. Kliment Ohridski"
5 James Bourchier Blvd.
1164 Sofia
BULGARIA

E-mails: dafinca@fmi.uni-sofia.bg
eliza@fmi.uni-sofia.bg