Learning Analytics Framework for Improving Performance to Students through Educational Virtual Worlds

Rosa Reis, Bertil P. Marques and Isabel Sampaio

Abstract— This paper aims to demonstrate the ongoing work of developing a framework that will allows to improve performance to students. The framework combines use of the open source virtual worlds, the Sloodle module and a learning analytics tool, in order to facilitate the execution of the collaborative learning techniques and improved the performance to students through of analytics learning tool monitorization. This framework is still in the design phase and will later be tested in a classroom context. The target public will be students of the fifth year of basic education, with aim of improve the learning mathematics.

Keywords— Virtual Worlds, Learning Collaborative, Learning Analytics, Instructional Design Model, OpenSim.

I. INTRODUCTION

NOWADAYS, the students increasingly have a heavy contact with technologies, being necessary using the tech in our classes. The using of the tech in our education create new challenges for instructor and learners allowing that learning process becomes more active and improve the student's performance.

The impact these resources are helping to development new environments to teaching and learning, namely, 3D virtual worlds, virtual labs, and games serious, so on. However, the development of these environments must follow some guidelines to produce helpful data that allowing to improve the performance of teachers and students in the educational process. One these guidelines is monitorization the student´s results through tools learning analytics. Teachers can to be able to predict as early as possible, which students are going to require more support and to adapt teaching to better meet individual students' needs.

Rosa Res is a Full Professor at the Department of Informatics Engineering, Instituto Superior de Engenharia do Porto (ISEP), Porto, Portugal and she is a researcher at the Gilt – Games Interaction and Learning Technologies, Porto, Portugal (e-mail: rmr@isep.ipp.pt).

Bertil P. Marques is a Full Professor at the Department of Informatics Engineering, Instituto Superior de Engenharia do Porto (ISEP) Porto, Portugal and she is a researcher at the GILT– Games Interaction and Learning Technologies, Porto, Portugal (e-mail: bmp@isep.ipp.pt).

Isabel Sampaio is a Full Professor at the Department of Informatics Engineering, Instituto Superior de Engenharia do Porto (ISEP) Porto, Portugal and she is a researcher at the ISRC – Interdisciplinary Studies Research Center, Porto, Portugal (e-mail: ais@isep.ipp.pt).

This work presents a proposal to improve teaching using 3D virtual worlds integrating learning analytics techniques, since the data generated by avatars in environments can allow to discover the student behavior in the execution of the learning design.

The paper is divided like following: The section 2 show the concepts involved, the section 3 presents the proposal framework and finally some conclusion are draw.

II. EDUCATION TECHNOLOGIES

According by Huang [1] the educational technology refers to the use of tools, technologies, processes, procedures, resources, and strategies to improve learning experiences in a variety of settings, such as formal learning, informal learning, non-formal learning, lifelong learning, learning on demand, workplace learning, and just-in-time learning.
Based on this assumption in this section we provide a brief description of educational technologies related to work in progress.

A. 3D Virtual Worlds

"A virtual world is a simulated persistent space based on the interaction by computer, inhabited by several users, who are represented by iconic images called avatars, who can communicate with each other's and the world in a synchronized way" [2].

This definition allows to demonstrate that a virtual world is more than a simple virtual environment. The term "virtual world" usually is related with one space similar to the real world where we can live, with details and action of gravity, to move objects and finally, to have the possibility one active communication between the various objects created by the users.

This world must not be interrupted; it should continue to exist even if the user does not connect. One of characteristics of virtual worlds is the persistence. Here that the participant is a member of a dynamic community and so, they communicate and interact with each other and with the environment. In order for communication to be effective, the participants must see the behavior of others, i.e. they must feel presence of others (sharing time). Although the view is the basis for an actual virtual world, you should allow some kind of communication, mediated by the computer, occur among the participants. Such

as the interactivity tends to be dynamic, the communication is in real time.

Moreover, the sense of presence, which is related to the individual and collective, allows when we got in a shared space, each participant becomes a "virtual person", known as avatars, i.e., is a digital representation (graphic or text) in the virtual world. So, we can identify characteristics of the virtual world:

Shared space: the world allows many users to participate at once. All the users have the feeling of being in the same place, room or land. The shared space is a common location where interactions may be occurring. The place can be real or fictional, and this space must have the same characteristics for all participants.

Immersion and Interactivity: The idea of virtual world allows the users in the interaction with the environment, to alter, develop, build or submit contents. The user must feel immersed in the environment and fully engaged with the activities being undertaken. This is normally achieved through representation of the user and environment in-world.

Persistence: the world's existence continues regardless of whether individual users are logged in. The persistence of world leads us to the immediacy feedback and synchronous use of the actions that take place in the word.

Immediacy: An immediate virtual world allows real-time interaction between the user and the world. Interaction between users is also real time. It is a definitional quality of virtual worlds.

Socialization: the world allows and encourages the formation of in-world social groups or communities.

These characteristics have contributed to represent in virtual world the different categories of information by means of three-dimensional objects, and thus, it is possible to develop a world similar to the real life.

Actually, there are several virtual worlds, where it is possible to find a set of scenarios, where each have their avatars and objects inserted into the environment. These scenarios are closed environments, such as the interior of a building or an empty room; or are open environments, like an ocean or a desert. The participants, called by avatars, walk in the different scenarios, making it random navigation. However, when we design a virtual world for education, we must have attention some aspects, for example the interaction. Students must be in full interaction with the system and vice versa. The interaction can occur through problem solving, graphical analysis, simulation, active participation in own environment. "A system of collaborative learning through virtual worlds, allows learners to solve problems and share knowledge. Collaboration among learners is modelled as "socially distributed cognition" [3].

The 3D collaborative virtual environments are mostly fixed environments, which are consisting by graphics filled with the avatars. This leads to some shortcomings, especially in terms of navigation and interaction, leading to disorientation and discouragement of users. So, it is emphasized as factors that most influence the qualities of these environments the Communication, Object Modelling, Navigation, Interaction and Reusing 3D model.

B. Learning Analytics

Different definitions have been proposed for the term Learning Analytics (LA) and analyzing each one we can verify that all emphasize data collection that allows extracting information for instructors to take actions that can improve the teaching-learning process.

Dyckhoff *et al*. [4], suggest that a LA process begins with data collection, followed by a data extraction process. This extraction can be presented as a widget integrated into the learning environment, and from the graphical analysis of the data, teachers must be able to interpret and assess whether the objectives have been achieved.

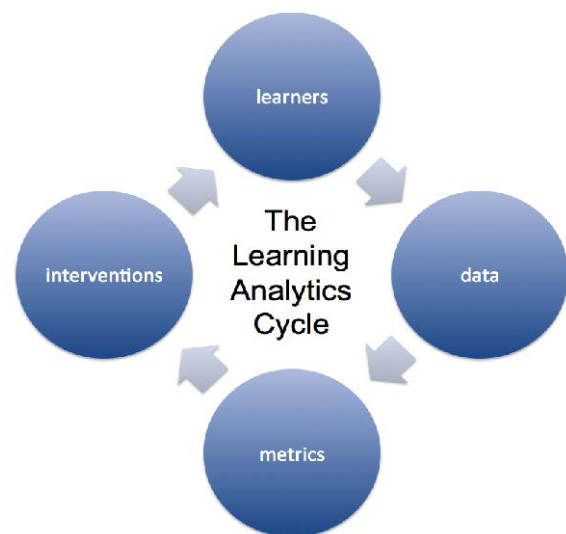The figure 1 shows the process of learning analytics.

Fig. 1. The learning analytics process cycle [5]

The according by Doug Clow [6] the cycle starts with the learners. The next step is the generation and capture of data about learners, for instance, clickstream data generated by learners; postings to a forum; assessment results and etc. The third phase the processing of this data in metrics or analytics, which provide some insight into the learning process. Finally, the last phase is called by intervention. The intervention uses the metrics to drive the instructor n building some alternatives to the learning process that have some effect on learners. However, learning analytics does not necessarily include all fours steps. A project that created reports about learners, but without any mechanism can feed this back into an improved learning experience, but not a very effective one [5].

III. LEARNING ANALYTICS TOOLS

In the last years the area of information technology that has had growing importance with respect to decision support and predictive analysis in the current world is data mining. Research and studies indicate that data mining plays a significant role in helping companies evaluate data and, as such, make informed decisions about different aspects of their

processes and operations, [7]. This awakened the academic community to a new and promising field of study, called of learning analytics. Many studies were developed allowing several tools to emerge that are been applied in education. These tools conducted to clarify benefices and challenges of learning analytics in education.

This section will present some learning analytics tools that are been application in education, namely RapidMiner, Waikato Environment for Knowledge Analysis (Weka) and Tableau.

A. RapidMiner

The software RapidMiner, formerly known as YALE (Yet Another Learning Environment), was developed in 2001 by Ralf Klinkenberg, Ingo Mierswa and Simon Fischer in the Artificial Intelligence Unit of the Technical University of Dortmund[1][8].
The Rapid Miner tool [9] is a tool that provides several data mining and other functionalities (ETL, reporting, etc.). This tool compatible with Windows, Unix and Linux. The data source systems available in Rapid Miner range from excel files, through text files, to the most popular DBMS (e.g., Microsoft SQL Server, Oracle, MySQL, Sybase, etc).
This tool has some characteristic features, of which we highlight [10]: Data Mining Tools; Statistical analysis; Machine learning; Data extraction; connected data management.
This software uses a client/server model with the server offered on-site or in public or private cloud infrastructures.[2]

B. Weka

The Weka software package (Waikato Environment for Knowledge Analysis) started to be written in 1993, using Java, at the University of Waikato, New Zealand and was later acquired by a company in late 2006. Weka is licensed under General Public License therefore it is possible to study and change the respective source code [11].

The Weka is one of the most popular data mining tools. Although well known, it is typically used only for teaching purposes (in data mining courses) or in small experiments. However, the tool provides a very powerful and flexible Java API that allows its integration with any type of Java system. Weka has consolidated itself as the most used data mining tool in the academic environment. It is applied not only in scientific research, but also for didactics [12].

The Weka offers some features like [12] [13]:

⇒ **x Prediction:** Weka provides algorithms that allow the computer to learn, that is, through existing patterns in the data at the time of the analysis, the computer can predict possible future events

⇒ x Classification **and regression**: the tool implements classification and regression algorithms, these, like the forecasting algorithms, aim to predict future occurrences of data through current standards

⇒ x Clustering: by applying clustering algorithms, users are able to identify data that share similar trends and patterns.

C. Tableau

Tableau Software is an American interactive data visualization software company[3] founded in January 2003 by Christian Chabot, Pat Hanrahan and Chris Stolte, in Mountain View, California. The company is currently headquartered in Seattle, Washington, United States[4] focused on business intelligence.[5]. On August 1, 2019, Salesforce acquired Tableau[6]. The Tableau Tool is a powerful data visualization tool used in business intelligence and data analysis. [14].

Tableau products query relational databases, online analytical processing cubes, cloud databases, and spreadsheets to generate graph-type data visualizations. The products can also extract, store, and retrieve data from an in-memory data engine.

The visualization provided by Tableau has completely enhanced the ability to gain more knowledge about the data we are working on and can be used to provide more accurate predictions. Tableau also has functionality for building rich and interactive dashboards, capable of displaying dynamic real-time visualizations to end users. However, Tableau's functionality is limited to this; it does not support predictive analytics or relational data mining.

D. Comparative Analysis

Making now a comparative analysis between that 3 software's, it will help us to choose which tool to use and how it is to be used in our project. All the tools presented were studied. These were subject to comparison and critical analysis, in the end one of them will be chosen for a practical applicability problem. The chosen tool was then used to create a practical proof of concept, aiming to show the quality of the solution produced by this type of tool [12].

The criteria chosen to carry out this analysis are presented in the table 1 and were based on the following assumptions:

1. **Usability** - This feature determines the usability of each tool. This describes which user interface is comparatively easier to use.
2. **Speed** - is an important distinguishing factor among different data mining tools. It helps to understand how system configuration impacts the working of a particular data mining tool.
3. **Visualization -** Visualization is the most important feature of a data mining tool.
4. **Algorithms supported -** This feature classifies data mining tools based on the algorithm implementation supported by them and the descriptor selection choice available.

---

[1]Guido Deutsch, "RapidMiner from Rapid-I at CeBIT 2010," Data Mining Blog, March 18, 2010.
[2]David Norris, "RapidMiner - a potential game changer," IT-Director.com, November 22, 2013.

[3]A Dead-Simple Tool That Lets Anyone Create Interactive Maps
[4]Tableau Software Helping Data Become More Visual
[5]"Tableau Business Intelligence". Tableau Software. Retrieved 2020-02-11.
[6]"Salesforce Completes Acquisition of Tableau". tableau.com. Retrieved 2020-02-20.

5. **Data Set Size - S**maller or larger data set support is another comparable feature among different data mining tools.

6. **Database** - This item is very important because some platforms of virtual world's development only allow you to connect to some databases. So, the LA tools should allow connection to different database, such as Oracle, Microsoft SQL Server, MySQL and access to Excel, Access as well as numerous other data formats. In our case, OpenSim use MySQL.

Based on this study we decide to choose the Weka tool, because:

⇒ Its graphical user interfaces allow the platform to be used more simply and easily;

⇒ It is compatible with almost all modern platforms,

⇒ Not require the ability to manipulate SQL statements and files as the RapidMiner tool;

⇒ Provides access to SQL databases through Java Database Connectivity and can process the results obtained by querying the database.

⇒ Support a comprehensive collection of data preprocessing and modeling techniques.

Table 1. Comparative Features [15]

	RapidMiner	Weka	Tableau
Usability	Easy to use	Most easiest to use	Simple to use
Speed	Requires more memory to operate	Works faster on any machine.	Works fast on any machine
Visualization	More options but less than Tableau	Less options	Many visualization options
Algorithms supported	Classification and Clustering	Classification and Clustering	Not used to implement algorithms
Data Set Size	Supports large and small data set	Supports only small data sets	Supports any data set
Databases	Oracle, SQL Server, MySQL….	Oracle, SQL Server, MySQL….	Oracle, SQL Server, MySQL….

IV. FRAMEWORK PROPOSAL

Our proposed framework is illustrated in figure 2 and is comprised of two key elements:

1- Design of virtual world - this module illustrates how the

virtual world is developed which is explain in detail in section 3.1.

2- Integration of learning analytics tool - this module illustrates how the tool connect to database and how extract the data. Its description is elaborated in detail in section 3.2.
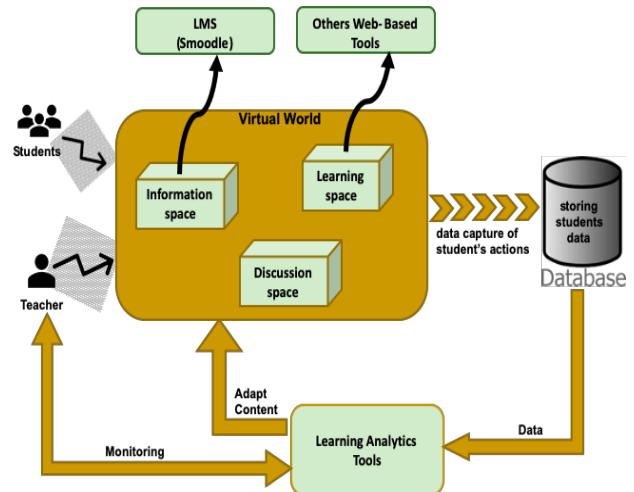

Fig. 2. The proposal framework.

The structure of the framework allows the connection between each of the components. The first component, is the virtual world, developed in an open source platform. We choose the Opensim. The objects of world have scripts that allows to store the actions of avatars in a database. These actions are defined by content experts. The second component is the learning analytics tool which extracts data from the database.

A. Design of Virtual World

The modelling of 3DCVE begin after the identification of the requirements for our application. However, we must have in mind that all educational virtual world must:

⇒ Be a tool which bring the world to class – it is possible access to multiples information sources and online communities from the classroom;

⇒ Support the activities in the classroom for the possibility of evaluations and the realization of innovative activities;

⇒ Open the class to the world for easy access to information, viewing and contact with others.

Also, we must try to identify the essential needs for the development of the system, type of tasks according to the learning objectives and expected interactions that allow obtaining a suitable environment for target audience.

After this phase, the virtual environment should be implement iteratively passing through different phases of development, which we describe.

Analysis phase - begins by focusing on the functionalities, in terms of operations and conditions of the system leaving

aside the details of the platform on which the environment will be develop. We suggest that from the information collected in the initial phase, the space should be divided into three main areas: discussion space, information space and activities space, as shown in figure 3.

Fig. 3. Scenarios of the different spaces of environment

The aim of the discussion area is to allow teachers to clarify some concepts to the students and/or to propose something that is related to the subject, leading the students to reflect, debate and draw conclusions.

The information area presents all the didactic contents to be addressed. Some contents can be accessed by Moodle.

Finally, the area of activities has as main objective to present a set of activities that allow students to consolidate the concepts learned. Some of the activities can be accessed by links to collaboration tools found on the Web, such as Scribblar. Throughout the area of activities students will be able to find indication of some problems to be realized. The activities are directed to be carried out in groups.

Design phase – the aim is to determine how we can facilitate interaction between the user and the environment. The activities are developed based on the personal intrinsic factors, such are curiosity, control, fantasy, and the challenge. "Assessing the factors deemed to support individual student intrinsic motivation may assist in enhancing intrinsically motivated behavior in technology-supported learning environments" [16].

Finally, we suggest that all environment must be evaluated through a set of evaluation requirements, already pre-defined in the initial phase of the project with the aim to analyze whether requirements were being met.

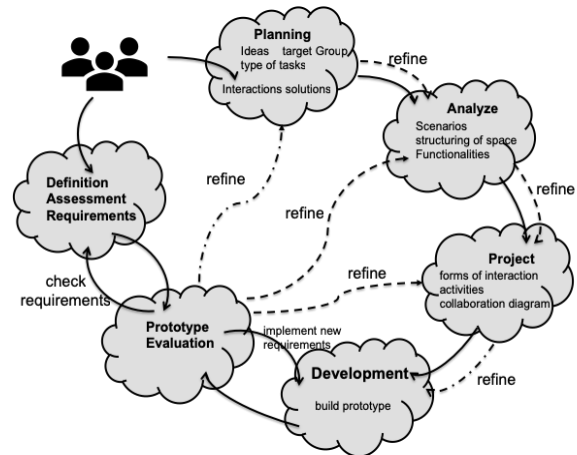Figure 4 shows the entire development process for the any virtual worlds.

Fig. 4. Development process of the virtual world.

B. Integration of Learning analytics Tool

To extract pertinent information about the interactions of students in the virtual world, it is necessary to take into account the coding of some objects. By programming these objects, it is possible in real time to extract data resulting from student interactions with the world or between students. As we intend to test the framework in OpenSim (virtual world platform), the team of development must:

⇒ Create the tables to store the data to collect (database schema), for example, the period during which the students stayed in a particular place; actions which were performed on objects; places they visited, so on.

⇒ Programing the sensors objects to store data in these tables; the data collection from the virtual world is possible through of programming sensor type objects, because OpenSim does have a programming language called OSSL. Thus, the virtual world to be built must have several types of sensors, for example, presence sensors - allows you to control the presence of students in the environment; interaction sensors - allow capturing student interactions (avatars) in the environment and sensors for evaluating actions taken. The scripts of these objects will allow us to store the necessary information in the respective database tables. The communication of the learning analytics tool chosen by us will connect to MySQL database using JDBC Connection object.

⇒ Database schema integration with learning analytics tools.

Thus, we can use the tool which will allows the teacher to analyze the information generated. He can guide and accompany the student in its development because data can be worked visually in tool by teacher, from the generation of graphs. These graphs will help teachers to identify at points where students have performed better or worse and also helping to find strategies to improve results.

To test this integration, we decide to use the Weka because it is used in education and supports all stages of the data mining process, including visualization, validation and optimization results [6]. The support to pre-processing of data, statistical evaluation of learning schemes, as well as methods for processing the results and integration with external

applications through API of java are main advantages in using this tool.

The figure 5 shown the architecture of integration of the learning analytics with the platform server database where was developed the virtual world.
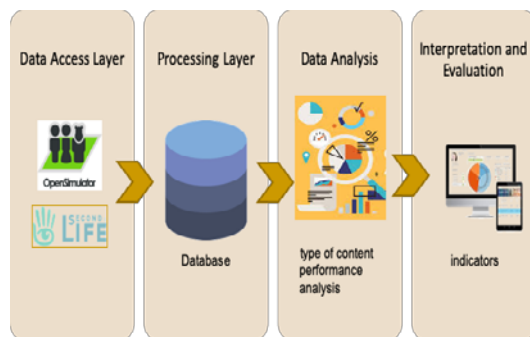

Fig. 5. Architecture Proposal

Data access layer - should be responsible for accessing existing database such as those on the OpenSim platform. Technologies such as Rest API, Open Database Connectivity (ODBC) must be used for the work.

Processing Layer- will be responsible for processing, in parallel, the large volume of data extracted from the data access layer, to then enable the generation of new consolidated databases.

Analysis Layer - here the teacher must analyze the data that are imported through by learning analytics tool, choosing by us. The prediction techniques in LA such as Social Network Analysis (SNA), Sentiment Analysis, Influence Analytics, Learner Success Prediction, so on, will be used if only the tool implement them.

It is expected with the proposal to present information as patterns of student behavior in virtual environments, to bring analyzes and suggestions to improve the teaching-learning process in AVA's environments, providing for predicting and detecting risk of failures, as well as forms of interventions to facilitate the progress and learning of students.

The proposed architecture can be used as a possible solution because we believe that we will be contributing to the unification and formation of a new paradigm of education.

It is also intended, with the result of this work, to help educational institutions to objectively monitor and diagnose, helping to extract some information, such as: the appropriate type of content to improve learning, the correct monitoring of students in the application of educational strategies in online environments.

V. CONCLUSIONS

The increasing use of VLE as an auxiliary tool in the teaching-learning process provides dynamics and modernization in education, as the use of these environments allows the availability of materials, frequency control, use of chats, forums, among others. However, in the context of AVA's, one of the current challenges is directly linked to the treatment of their data in order to improve student's performance. Thus, the purpose of this work is to direct actions to solve problems with the low performance of students, aiming to build a framework capable of integrating LA and the virtual worlds, generating indicators and information to improve the teaching-learning process of educational institutions. A teacher can take advantage of face-to-face classes to complement what has been taught at a distance, to clarify doubts to maximize student learning. This is possible if we improve the design of virtual worlds.

We believe that we are still at the initial stages for exploring further research in how the learning analytics can improve the virtual world design. The directions in which we want to take our research lie not only in the ways of presenting meaningful information to our learners but also, how the usage of learners- produced data contribute to discover learning patterns that are of interest both teachers and students, whose purpose is to achieve educational goals in a given learning context. At the focus of our interest are the learners who design their own learning pathways inside a virtual world, using data in a way that can trace their school success.

REFERENCES

[1] Huang et al. (2019). Introduction to Educational Technology: A Primer for the 21st Century. In book: Educational Technology, pp.3-31. DOI: 10.1007/978-981-13-6643-7_1

[2] Reis, R, Escudeiro, P. et al. (2011) Comparing |Social Virtual Worlds For Education Purposes. Scientific & Academic Publishing. pp. 21-26e-ISSN: 2162-8467; doi:10.5923/j.edu.20110101.04

[3] Sanchez-Segura, A. de Antonio, and A. de Amescua,(2005). Senda: A whole process to develop virtual environments; In M.-I. Sanchez-Segura, editor, Developing Future Interactive Systems, pages 92{114, Idea Group Publishing.

[4] Dyckhoff, Anna Lea et al. (2012) Design and Implementation of a Learning Analytics Toolkit forTeachers. Educational Technology & Society, v. 15, n. 3, p. 58-76.

[5] Clow, D. (2012b). The learning analytics cycle. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12, 134-134. doi:10.1145/2330601.2330636

[6] Jalloule, J., Sallé, J. and Bittencourt, R. RapidMiner Aprenda a Usar. Disponível em: https://prezi.com/-yo8qjamdbbq/rapidminer-aprenda-a-usar/.

[7] Joseph, M. V. 'Significance of data warehousing and data mining in business applications. International Journal of Soft Computing and Engineering, 3(1), 329–333, 2013.

[8] RapidMiner: https://pt.wikipedia.org/wiki/RapidMiner (online) Retrieved 5/1/202

[9] Deutsch G. (2010). RapidMiner from Rapid-I at CeBIT 2010. Data Mining Blog

[10] RapidMiner: https://www.capterra.com.br/software/148220/rapidminer (online) Retrieved 6/1/202

[11] WEKA: https://pt.wikipedia.org/wiki/Weka online) Retrieved 20/1/202

[12] Teses of Master degree: Dissertação TiagoSilva_MSTIO.pdf

[13] Holmes G., Donkin A. (1994). Weka: A Machine Learning Workbench. Proc Second Australia and New Zealand Conference on Intelligent Information Systems

[14] About Tableau: https://en.wikipedia.org/wiki/Tableau_Software online) Retrieved 10/1/202

[15] Rangra et al. "Comparative Study of Data Mining Tools". In Proceedings of International Conference of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 6, pp. 216-223, June 2014..

[16] Shroff, R. H., & Vogel, D. R. (2009). Assessing the factors deemed to support individual student intrinsic motivation in technology supported online and face-to-face discussions. Journal of Information Technology Education, 8, 59-85