

A FRAMEWORK TO INCREASE THE PERFORMANCE ANALYSIS OF THE STUDENT USING MACHINE LEARNING

Prof. S.B.Zanke¹, Vaishnavi Choudhari², Ayusha Nemade³, Rutuja Dhoke⁴, Chaitali Kolte⁵

¹Assistant Professor, Computer Science and Engineering, Padm. Dr.VBKCOE, Malkapur ,Maharashtra, INDIA

²Student, Computer Science and Engineering, Padm. Dr.VBKCOE, Malkapur ,Maharashtra, INDIA ³Student,

Computer Science and Engineering, Padm. Dr.VBKCOE, Malkapur ,Maharashtra, INDIA ⁴Student, Computer

Science and Engineering, Padm. Dr.VBKCOE, Malkapur ,Maharashtra, INDIA ⁵Student, Computer Science and

Engineering, Padm. Dr.VBKCOE, Malkapur ,Maharashtra, INDIA

Abstract

For scholars, mining data and extracting information from huge databases has emerged as an intriguing field of study. Since a few decades ago, the concept of using data mining techniques to extract information has been around. The dataset was originally intended to be partitioned and the inherent features examined using classification and clustering algorithms. They base their predictions on these characteristics. These forecasts have been made in the area of educational data mining for a variety of reasons, including to predict student success based on personal characteristics and help students find the right professors and courses. These goals have been drawn from the attrition and retention of students. These objectives are the focus of our research on student attrition and retention. Additionally, we have found exciting variables that aid in predicting students' success, suggesting the most qualified instructors, and assisting them in course selection.

Keywords: Mining, databases, information, dataset, predictions, performance

1. INTRODUCTION

Data mining and information extraction from massive databases have become fascinating academic fields of study. The idea of employing data mining techniques to extract information has existed for a few decades. The first plan was to divide up the dataset and use algorithms for classification and clustering to analyse the innate traits. They use these traits as the foundation for their predictions. These predictions have been made in the field of educational data mining for a number of purposes, such as predicting student success based on individual traits and assisting students in selecting the best instructors and courses. These objectives were derived from student retention and attrition rates. Our study on student attrition and retention is centred on these goals. Additionally, we have discovered intriguing characteristics that support success prediction, instructor recommendation, and course recommendation for students.

2. PROBLEM DEFINITION

A major problem that the academic institutions are facing worldwide is poor performance of students in academics. This causes high attrition rate that is a loss to students, parents and institutions. Student's performance prediction can help to reduce attrition rate as it raises an early alarm for students not performing well and are likely to leave the institution/studies. The application of this research is versatile. The experiments of this research have been conducted on the dataset that has been collected from various colleges and university and was available in various dataset links. Such factors can be applied on the data of Government institutions as well as schools and colleges.

3. OBJECTIVES

- To focus on predicting the grades of low performance students by applying various machine learning techniques.
- To compare the techniques on the basis of mean absolute error.
- To make prediction made by machine learning algorithms will help the teachers and mentors to decide the course of education to be applied to specific students.
- To make the university management system efficient
- To enable Course recommender system pro-active
- To build such recommender system that helps students in selection and registration of the courses will be quite beneficial for the institutions.

4. LITERATURE SURVEY

The initiative of online course learning system is based on e-commerce venture. With the growth of this venture, data from web resources started collecting and storing in the excel that contains customer and product information and order information (Kokina, J, 2017). E-commerce is a term that refers to use as online business through internet. There are various websites that are working for this purpose such as e- bay, Alibaba, and Amazon etc. It can be said that data storage in excel from the resources has been derived from e-commerce. Predicting student's performance by using data mining techniques to extract information from the academic dataset of universities has become state of the art research in the scientific society. Universities are confronting with some challenges now a day to analyze the performance of their students. That's why researchers are focusing on student's profiles and characteristics to make the university management aware of student's performance and overall academic result (Kabakchieva, D, 2013). There is another dimension of student's performance that is the dependence of student retention upon student student's performance. To minimize the problem of student retention cases in the universities, different researchers have proposed different methods to predict the performance of students in their future semester based on the performance of previous one. To predict the courses of next term grades, four parameters have been considered in this study such as; admission records, High school scores, SAT/ACT scores and grades of previously completed courses. Based upon these parameters, recommender system can be trained to predict the grades of students accurately in any of the educational institution. Historical information about the course has also been considered in this study such as which course is taught by which teacher and information about contents of the course. Many researchers have used LMS and Moore to predict the successive chances of success and failure of students. In this research, regression based methods such as course specific regression (CSPR) and personalized linear multi regression 10 (PLMR) has been used. Another method known as matrix factorization based methods in which standard matrix factorization (MF) has been used for the grade prediction of students (Elbadrawy, A, et al, 2015).

Data mining is a process of sorting data and extracting information from existing databases. With the help of pattern mining and data analysis, hidden information can be obtained from huge datasets. The strategy of data mining is now applied in the field of education by researchers. They are busy in exploiting a lot of dimensions in education sector. This is now known as educational data mining. Data mining is applying in educational sector by considering the performance of students and finding the position of students by using their academic records. Educational dataset is being collected from various resources such as interactive learning systems, computer-supported collaborative systems, and administrative datasets of school, colleges and universities. Data mining methods are now implemented in well known universities to analyze the patterns of student performance from the dataset through which information can be extract and decision making may become easier for the management of institutions (Kabakchieva,

D. 2013). With the incremental growth in the use of technology everywhere, educational institutions are now busy in finding hidden trends and patterns in their larger datasets (Scheuer, O. 2012). (Merceron, A. 2005).

Datasets that are used for experimental purposes in educational institutions have become possibly available because of web based educational systems in which LMS, Moore and Portal system have become common. With the help of these sources, dataset can easily be collected if authorization is accessed. One purpose of extracting information from its own dataset is to make its prestige among other educational institutions stronger. Another purpose is to build the student career by covering its each and every aspect such as improvement of their grades if lacking, overall performance booster, support them financially, make them enable to select the courses suggested by course recommender system, assign them appropriate teachers based on their inclination of interest in course selection.

Web based learning environments have become common right now. It has become powerful medium to act as a bridge between learner and instructor and provides interesting learning mechanisms to both (Nam, C. S, 2007) (Osmar, R, 2002) has elaborated the concept and working of web based learning which is also known as e-learning. A recommender system helps to recommend the actions to a learner by using his previous actions with the help of intelligent software agents. The use of such recommender systems were first initiated in ecommerce and now carrying out by researchers in the domain of e- learning. Osmar has implemented this recommender system in assisting the offline web miners who are responsible to find hidden trends in the data and online course navigation. With the help of recommender system, educational institutions are now strengthening their management and departmental progress. Researchers have proposed several recommender systems among which course recommender system and teacher recommender system are commonly known. students who enrolled (Kizilcec, R. F, 2017) (Pérez-Sanagustín, M., 2017) (Maldonado, J. J. 2017).

5. USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor.

Roles of the actors in the system can be depicted.

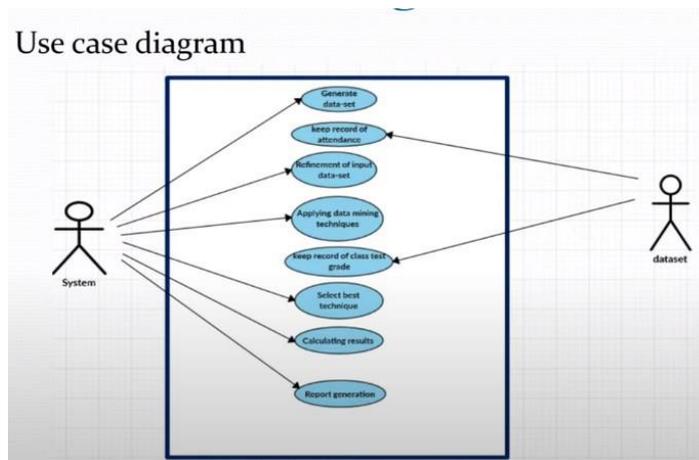


Fig. 1 Use case diagram

6 .ACTIVITY DIAGRAM

Activity diagrams are visual depictions of workflows with choice, iteration, and concurrency supported by activities and actions. Activity diagrams can be used to depict the operational and business workflows of system components in the Unified Modelling Language. An activity diagram demonstrates the total control flow.

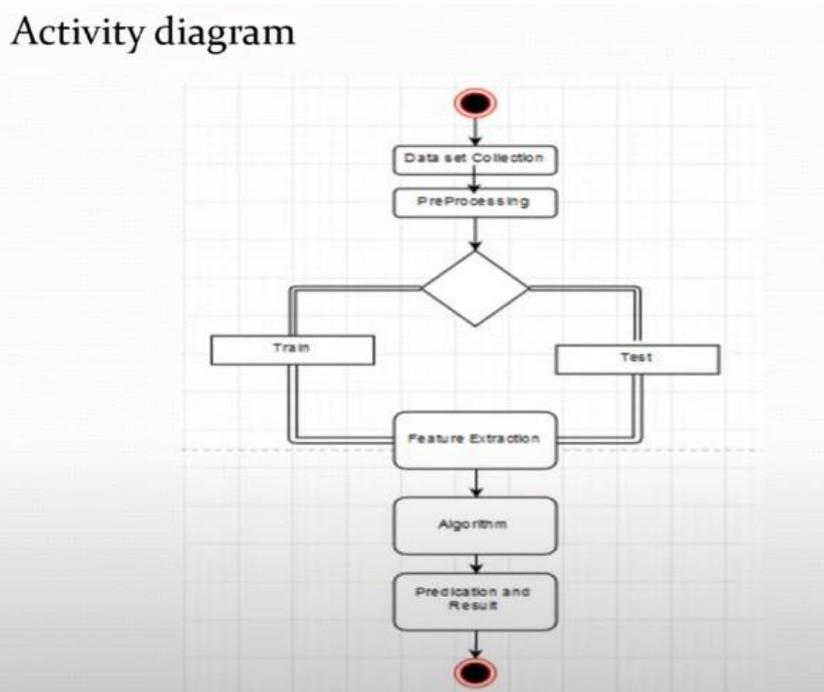


Fig. 2. Activity Diagram

7. ALGORITHMS AND TECHNIQUES

Nave Based and J48 classification algorithms have been compared in order to assess our research. These classifiers were chosen for a variety of factors, including the fact that they support categorical and nominal data. The programme is open source and accessible online; it is typically used for data training and testing, classification, and machine learning algorithms. The data can then be shown graphically as well. The created attribute combinations have been assessed in the current study to address our research questions. In order to enter the data into the dataset for this purpose, it was converted to csv format, and the output from various filters and classifiers was compiled.

a) Naïve Based

Naïve based algorithm is comparatively fast algorithm in terms of classification. It works faster on huge datasets by using Bayes algorithm of probability. Bayes algorithm generally used to predict the class of unknown

dataset4. Naïve based algorithm works on assumptions to label an item whose features are known but name is unknown. For example; a fruit is labeled as an apple if it is round and red in color and its size is 3 inches in diameter. These features of apple will raise the probability of this fruit that it is an apple.

b)J48

J48 decision tree is used to predict the target variable of new dataset. If dataset contains predictors or independent variables and set of target or dependent variables, then this algorithm is applied to extract the target variable of new dataset5.

c)Linear Regression

Generally linear regression classification algorithm is an approach to identify the relationship between dependent and independent variable. It is generally used for predictive analysis and has two main points. One is to check whether predictor variable does a good job in predicting the expected outcome variable. Second main thing that linear regression does is the identification of variable that are significant predictors of dependent variables. At the end, the regression equation is used which helps to determine the set of predictor which are used to predict the outcome. In this research, algorithms are being used to compare the trend and pattern of the factors with other approaches like non-linear regression and SMO.

8. DATA FIELD

Entering various field into the system which helps to predict the current scenario of the grade of the students. It helps the students to judge the grade and study accordingly with due to respect to increase or decrease in the parameters schedule

8.1 Study time from terminal

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:-
```

Fig.3 Study time from terminal

8.2 Age from terminal

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- █
```

Fig. 4 Age from terminal

8.3 Travel time to and fro from terminal

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- █
```

Fig. 5 Travel time to and fro from terminal

8.4 Absentee from terminal

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- 20
Enter school absences:-
```

Fig. 6 Absentee from terminal

8.5 First grade marks

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- 20
Enter school absences:- 23
Enter first period grade:-
```

Fig. 7 First grade marks

8.6 Second grade marks

```
PS G:\BE PROJECTS\VBKOLTE\Student_Performance\n7\Student-Grade-Prediction-main> &
N/AppData/Local/Programs/Python/Python311/python.exe" "g:/BE PROJECTS/VBKOLTE/Stu
Student-Grade-Prediction-main/Student_Performance.py"
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- 20
Enter school absences:- 23
Enter first period grade:- 17
Enter second period grade:-
```

Fig.8 Second grade marks

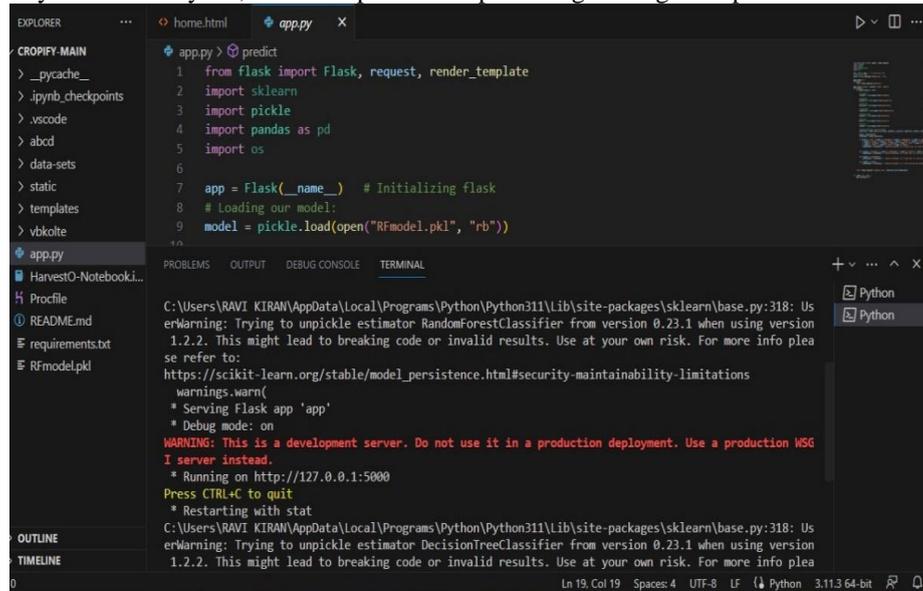
8.7 Predicted Grade according to system

```
Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995 0.10044543 0.0527645 0.11366621 0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- 20
Enter school absences:- 23
Enter first period grade:- 17
Enter second period grade:- 12
C:\Users\RAVI KIRAN\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\b
erWarning: X does not have valid feature names, but LinearRegression was fitted with fe
warnings.warn(
Predicted Grade is: 13.960941463768478
```

Fig. 9 Predicted Grade according to system

9. RESULTS

Researchers have made major contributions to the field of educational data mining by introducing elements that can be used by educational institutions to uncover hidden patterns in academic databases. Datasets from physical institutions as well as online learning management systems are employed in the expanded field of educational data mining. One drawback of this strategy is that in order to perform analysis and carry out research, one must obtain access to the dataset from the administration. The subsequent sections of this chapter will provide a detailed overview of the results that were attained. The first phases in the analysis of the research are data gathering and pre-processing. The registrar's office provided the demographic information for this project, including the city and gender of the students. Data on preuniversity and institutional elements have been gathered with the use of dataset administrative resources. Information about midterm, term, GPA, intermediate, and matriculation marks can be found in institutional and pre-university variables. Access to this student data was made possible by the university site, which helped us compile a large enough sample size for our research.



```

EXPLORER  ...  home.html  app.py  X
CROFY.MAIN
> _pycache_
> .ipynb_checkpoints
> .vscode
> abcd
> data-sets
> static
> templates
> vbkolte
app.py
HarvestIO-Notebooki...
Profile
README.md
requirements.txt
RFmodel.pkl
OUTLINE
TIMELINE

app.py > predict
1 from flask import Flask, request, render_template
2 import sklearn
3 import pickle
4 import pandas as pd
5 import os
6
7 app = Flask(__name__) # Initializing flask
8 # Loading our model:
9 model = pickle.load(open("RFmodel.pkl", "rb"))
10

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL
Python
Python

C:\Users\RAVI KIRAN\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\base.py:318: Us
erWarning: Trying to unpickle estimator RandomForestClassifier from version 0.23.1 when using version
1.2.2. This might lead to breaking code or invalid results. Use at your own risk. For more info plea
se refer to:
https://scikit-learn.org/stable/model_persistence.html#security-maintainability-limitations
warnings.warn(
  * Serving Flask app 'app'
  * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSG
I server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
C:\Users\RAVI KIRAN\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\base.py:318: Us
erWarning: Trying to unpickle estimator DecisionTreeClassifier from version 0.23.1 when using version
1.2.2. This might lead to breaking code or invalid results. Use at your own risk. For more info plea

```

Fig. 10 Dashboard of the local host system

```

Mean squared error is: 5.524221095174995
Weights: [-0.00482364 -0.24366995  0.10044543  0.0527645  0.11366621  0.99885866]
intersect: 2.2046770814240126
Enter Study Time:- 5
Enter student's age:- 22
Enter travel time:- 20
Enter school absences:- 23
Enter first period grade:- 17
Enter second period grade:- 12
C:\Users\RAVI KIRAN\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\b
erWarning: X does not have valid feature names, but LinearRegression was fitted with fe
warnings.warn(
Predicted Grade is: 13.960941463768478

```

Fig. 11 Entering all values and predicting the result

10.CONCLUSION

In order to improve the accuracy of the student performance system, we employ an improved platform that can access information intelligently to supply and carry out a number of tasks using the online educational system. The primary goals are to enhance the real-world experience for both students and teachers and to develop understanding. Making students' grasping abilities stronger will help them improvise more broadly and attain desirable results without stress. This study puts someone on the list of candidates for those who will devote enough time and effort to finish the course successfully.

11.FUTURE SCOPE

Future research should focus on many methods to enhance the system's functionality, such as: Since many variables affect students' academic performance, the prediction system should be improved to take into account the effects of various human factor categories. Emotional qualities, study schedule, familial attributes, and institutional attributes are a few examples of these factor categories. Multiple elements in each category have an effect on students' performance, either favourably or unfavourably.

REFERENCES

- [1] Abaya, S. A., & Gerardo, B. D. (2013, September). An education data mining tool for marketing based on C4. 5 classification technique. In *e- Learning and e-Technologies in Education (ICEEE), 2013 Second International Conference on* (pp. 289-293). IEEE.
- [2] Aher, S. B., & Lobo, L. M. R. J. (2012). A comparative study of association rule algorithms for course recommender system in e-learning. *International Journal of Computer Applications*, 39(1),48-52.
- [3] Al-Barrak, M. A., & Al-Razgan, M. (2016). Predicting student's final GPA using decision trees: a case study. *International Journal of Information and Education Technology*, 6(7), 528.
- [4] Anand, S. S., Bell, D. A., & Hughes, J. G. (1996). EDM: A general framework for data mining based on evidence theory. *Data & Knowledge Engineering*, 18(3), 189-223.
- [5] Azarcon Jr, D. E., Gallardo, C. D., Anacin, C. G., & Velasco, E. (2014).
- [6] Attrition and retention in higher education institution: A conjoint analysis of consumer behavior in higher education. *Asia Pacific Journal of Education, Arts and SCIENCE*, 1(5), 107-118.
- [7] Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *learning analytics* (pp. 61-75). Springer New York.
- [8] Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyse students' performance. arXiv preprint arXiv: 1201.3417.
- [9] Bydžovská, H. (2013). Course Enrolment Recommender System (Doctoral dissertation, Masarykova univerzita, Fakulta informatiky).
- [10] Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *The journal of higher education*, 64(2),123-139.
- [11] Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. Drea, C. (2004). Student Attrition and Retention in Ontario's Colleges. *College Quarterly*, 7(2), n2.
- [12] Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting Student Performance Using Personalized Analytics. *Computer*, 49(4), 61-69.
- [13] Garcia-Saiz, D., & Zorrilla, M. E. (2011, November). Comparing classification methods for predicting distance students' performance. In *WAPA* (pp. 26-32).
- [14] Guarín, C. E. L., Guzmán, E. L., & González, F. A. (2015). A model to predict low academic performance at a specific enrollment using data mining. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 10(3), 119-125.
- [15] Hung, J. L., & Zhang, K. (2008). Revealing online learning behaviors and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*.
- [16] Inoue, S., Rodgers, P. A., Tennant, A., & Spencer, N. (2017). Reducing Information to Stimulate Design Imagination. In *Design Computing and Cognition'16* (pp. 3-21). Springer, Cham.
- [17] Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72.
- [18] Kaminski, J. (2005). Moodle—a User-Friendly, Open Source Course Management System. *Online Journal of Nursing Informatics*, 9(1).
- [19] Kaur, G., & Singh, W. (2016). Prediction Of Student Performance Using Weka Tool.
- [20] Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Selfregulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104, 18-33.
- [21] Kokina, J., Pachamanova, D., & Corbett, A. (2017). The role of data visualization and analytics in performance management: Guiding entrepreneurial growth decisions. *Journal of Accounting Education*.