Journal of Advanced Research in Embedded System
Volume 7, Issue 1 - 2020, Pg. No. 1-7
Peer Reviewed Journal

Article

Educational Data Mining in Practice Literature Review

Kamal Bunkar[1], Sanjay Tanwani[2]

[1]Ph.D Scholar, School of Computer Science and IT DAVV, Indore 452001, India.
[2]Professor and Head of Department of Computer Science and IT DAVV, Indore 452001, India.
DOI: https://doi.org/10.24321/2395.3802.202001

I N F O

Corresponding Author:
Kamal Bunkar, School of Computer Science and IT DAVV, Indore 452001, India.
E-mail Id:
kamal.bunkar@gmail.com
Orcid Id:
https://orcid.org/0000-0003-2774-1864
How to cite this article:
Bunkar K, Tanwani S. Educational Data Mining in Practice Literature Review. *J Adv Res Embed Sys* 2020; 7(1): 1-7.

Date of Submission: 2020-02-18
Date of Acceptance: 2020-03-30

A B S T R A C T

Educational Data Mining (EDM) is an evolving field with a suite of computational and psychological methods for understanding how students learn. Applying Data Mining methods to education data help us to resolve educational investigation issues. The growth of education data offers some unique advantages as well as some new challenges for education study. Some of the challenges are an improvement of student models, identify domain structure model, pedagogical support and extend educational theories. The main objective of this paper is to present the capabilities of data mining in the context of the higher educational system and their applications and progress, through a survey of literature and the classification of articles. We observed the works on investigational situation studies showed in the EDM during the recent past, in addition, we have introduced three data models based on descriptive and predictive data mining techniques. This is oriented towards students in order to recommend learners' activities, resources, suggest path pruning and shortening or simply links that would favor and improve their learning or to educators in order to get more objective feedback for instruction.

Keywords: Data Mining, EDM, Classification, Prediction, Presentations, E-Learning

Introduction

Data Mining, often called Knowledge Discovery in Database (KDD) is known for its powerful role in uncovering hidden information from large volumes of data[1] Data Mining is an important data analysis methodology that has been successfully employed in many domains including e-commerce, bioinformatics, industry, business, science and lately, within the educational research. From a statistical point of view, it can be seen as computer automated exploratory data analysis of the huge amount of complex data sets. Despite the strong association between data mining and statistical/mathematical data analysis, a major amount of the techniques used in data mining has so far come from the field of statistics.[2]

Data mining techniques are helping to analyze the data automatically in place of manual data evaluation. In order to recover patterns that are application oriented, data mining algorithms are used. Now, these techniques are acceptable in various other domains for making decisions, forecasting or finding trends and patterns. In this work the data mining approaches are applied in the educational domain for recovering the patterns of students, their growth and future road maps.[3]

Educational Data Mining is an developing discipline,

concerned with developing approaches for exploring the exclusive kind of data that arise in the educational domain and using data mining approaches understand students and patterns which they learn. Researchers in this field concentrate on finding valuable information either to enable the Educational Institute for dealing with their students to better or to help students to upgrade their academic achievement.[4-5]

Literature Survey

There was a comprehensive literature review that covers various significant researches in the area of EDM. Therefore, some reviews are introduced the latest trends in data mining for educational research.

Students presentation is an essential part in higher learning institutions. This academic presentations may be impacted by many factors, subsequently, it will be fundamental to create predictive data mining model for learning outcomes so as to identify the distinction between high learners and moderate learners student. Brijesh Kumar Bhardwaj et al.[6] propose an experimental methodology to generate a database. According to Radek Pelánek et al.[7] mostly historical data used for student demonstrating. Data collection could impact on interpretation of student models. Radek Pelánek et al.[7] Mastery attrition bias and adaptive choice of items are two approaches that show how data gathering can impact on the results. They provide numerous ways for the collection and publication of datasets. Alejandro Peña-Ayala et al.[8] identify twofold objective, the primary objective is to preserve and enhance the chronicles of recent Educational Data Mining (EDM); the secondary objective is to organize review produced by data mining methods in a suitable manner. The finding shows that vast amount of the EDM techniques are grouped into three kinds of educational systems, disciplines, tasks, methods, and algorithms each. This work concludes with an EDM work, their strengths, weakness, opportunities, and threats represent. Raheela Asif et al.[9] show data mining technique will help to identify the presentations of undergraduate students. And further analyze that, it will give better if the prediction result combines with progression. Similarly, Ajay Kumar Pal et al.[10] describes the data mining techniques which is helping to the learner to improve their efficiency in order to improve final grades and identify those students who need special attention or counseling by the tutor at regular interval. Cristóbal Romero et al.[2] classify student based on final marks obtained with the help of numerous classification methods used in data mining and claim that a classifier model is appropriate for the educational domain to take a decision by the instructor.

Summary: According to the listed literature the EDM is an application of data mining technique. In this application, the student records are analyzed for predicting the student's presentation. In addition to that, it is also necessary to keep the attention on the collection and evaluation of data for preparing the accuracy and uses precise data model. In this context, the major contribution is observed in applying the data model or data mining algorithms that help to predict the presentations of students majorly in two main classes' namely poor or effective learner. In addition to that, some of the contributions also available that works on descriptive data modeling for providing the suggestions for achieving the goals.

Issues and Challenges in EDM

According to Rajni Jindal and Malaya Dutta Borah [11] the EDM is performed for the two key objectives (1) Academic Objectives and for (2) Administrative Objectives. When the data are mined for Academic purpose the intention is to find information about the following three criteria

- Person oriented for teaching and learning patterns such as Student learning, modeling, behavior, risk, presentations analysis, predicting decision, etc. for both in a traditional and digital environment that also involve Faculty modeling- presentations and satisfaction. Gwo-Jen Hwang et al [13] works on Identifying students' learning status or problems.
- Department/Institutions oriented that analysis is oriented to Particular departments with respect to time, sequence and demand such as redesign courses according to industry requirements, identify realistic problems.
- Domain Oriented involves the Designing Methods-Tools, Techniques, Knowledge Discovery based Decision System (KDDS) etc.

On the other hand, the Administrative Purposes involve the following outcomes from mining:

Administrator Oriented that include the monitoring and Participation of the ruling classes/Administrator, Infrastructure and Human Resource, Industry academia relationship, marketing. Nicholas A. Ashford et al.[14] addressing the following themes and questions for EDM for administrative point of view:

How can multi-and trans-disciplinary research and teachings coexist in a meaningful way in today's university structures? Does education relevant to sustainable progress require its own protected incubating environment to survive? How can difficulties in linking the needed teaching and research be overcome? Even if there exist technical options to do so, how can it be made safe for courageous students to take educational paths numerous from traditional tracks? What can we learn from the comparative analysis of universities in numerous nations and environments? What roles can national and EU governments have in accelerating the needed changes? How can difficulties in

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001

linking the needed teaching and research be overcome?

To explore educational data from educational systems for decision making. In this field, Gwo-Jen Hwang et al.[13] contributes on enabling policy-makers and administrators to have a global view of current educational policies or problems by analyzing large-scale educational data to identify important indicators for evaluating the educational status from local, regional, national or international perspectives.

In order to achieve the goals they consider the following key issues:

Data nature and Lack of Interoperability: Due to the exponential growth of data, maintaining data is difficult. To monitor data sources, student interest, intentions and impact analysis is the main issue. Secondly, alignment and translation of incremental data which depends on time and context required optimal utilization of computing and human resources. Additionally, a wide range of storage locations, data platform, and social networking make it more complicated to keep track. So need to design a model to classify/cluster data to find relationships.

- **Uncertainty of Data:** Presence of uncertain noise and errors, no model can predict accurate results for students or academic planning.
- **Relation between Student-Teacher:** In Higher Educational Institutions students have a project work relevant to research work. Supervisors are assigned as per availability, but it is not possible to assign all students-supervisor with similar area, hence the result of the project is not predictable.

On the other hand, Rebecca Ferguson[12] works on knowledge analytics and considers three main challenges:

- Technical challenges
- Educational challenges
- Political/economic challenges

Where they find the following key issues in providing the solutions for target challenges.

Build strong connections with learning sciences, Develop approaches of working with a wide range of datasets, Focus on the perspectives of learners, Progress and apply a clear set of ethical guidelines.

During their research Gwo-Jen Hwang et al[13] obtained the following key issues:

Proposing new approaches of learning analytics and educational data mining, Investigating the impacts of providing personalized supports based on educational data mining on students' learning presentationss.

Samira ElAtia et al[15] worked on the similar directions of [11] and aims to address following questions.

Would it be feasible to develop an integrated data acquisition system for collecting and storing data from all departments and units within a university? Can the collected data be converted to useful knowledge and provide new insights into educational research? Can such practices be done in a way that does not infringe on legal issues relating to privacy and confidentiality?

Research using data mining techniques can be useful in three contexts:

- When it is geared towards providing information that enhances learning from the students' perspective
- When it is geared towards providing insight into learning and teaching for university teachers
- When it is geared towards providing the institution and its administrators with valuable knowledge that would be an asset for decision-making

In a global context, using data mining resources for educational research and to improve the education experience on all levels can be grouped into four major categories:

- Focus on the learning
- Focus on the teaching
- Managing the institution from an academic experience
- Managing the institution from an infrastructure perspective

Algorithms Applied in EDM

In traditional data mining and their techniques, a number of numerous kinds of data models are available. Additionally, the utilization of these data models is achieved on the basis of their applicability and usage. This section provides the

Table 1.Clustering Algorithms in EDM

Reference	Problem/objective	Algorithms
[16]	Evaluation of Student's Presentations	K-Means, Hierarchical Clustering
[17]	improve educational process mining	AXOR Algorithm, k-means
[18]	predicting dropouts using outlier detection	Dynamic time warping, k-means++
[19]	Predicting Students' Academic Presentations	Hybrid of K-means and Decision tree
[20]	analyze the meaningful rules among great amount of students' raw information	Apriori association algorithm; KMeans clustering;

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001

overview of literature where how and for which purpose

Table 2.Classification Algorithms in EDM

Reference	Problem/objective	Algorithms
[26]	To understand the knowledge and presentations of students	Apriori algorithm
[27]	guiding the learner's activities and intelligently generate and recommend learning materials	Apriori algorithm
[28]	to Improve Academic Presentations	Apriori algorithm
[29]	to improve quizzes and courses	genetic programming, Apriori algorithm
[30]	to improve the quality of their educational processes	FP-Growth algorithm

numerous data models are applied in EDM applications.

Table 3.Association Rules in EDM

Reference	Problem/objective	Algorithms
[21]	Recognition of Slow Learners	J48, REPTree, Naive Bayes, SMO, Multilayer Perceptron
[22]	prediction for Student's Presentations	ID3 Algorithm
[23]	Predicting Student Presentations	C4.5/J48 decision tree
[24]	predicting the drop out feature of students	J48, NBTree, REPTree, SimpleCart
[25]	Predicting Student Presentations	Support Vector Machine (SVM), C4.5, Classification and Regression Tree (CART), Bayes Network (BN), Naive Bayes (NB)

Clustering

Classification

Association Rules

Summary: Essentially, two kinds of DM models are designed: descriptive and predictive. Usually, the implementation of a model is made by a task. For instance, clustering, association rules, correlation analysis, produce descriptive models; whilst, classification, regression, and categorization generate predictive models. Some new methods, Response Theory, Bayes net and Markov Decision Theory significantly more prominent in recent few years than earlier year. The rise in the camaraderie of these approaches is likely a reflection of the addition of researchers from the psychometrics and student demonstrating communities into the EDM community.

Proposed Work

Conventionally in EDM, the key efforts are made in order to comprehend and predict the following:

- Student presentations
- Student dropouts
- Student failure
- Improvements in educational processes

Additionally, for obtaining their primary goals the classification, clustering, and association rule mining algorithms are frequently utilized. In a similar manner for providing contributions in the EDM, the following key objectives are established.

To identify teaching efficacy and teaching patterns of university teachers with the help of efficient data mining techniques.

- To provide feedback to support educators in decision-making about how to advance students' learning and enable them to take suitable proactive or remedial action.
- To predict student presentation based on knowledge and grade with the help of the proposed framework.
- Create groups of students according to their knowledge indicators, personal characteristics and personal learning data.
- To map academic presentation with future prospects and achievements.

In order to accomplish the proposed objectives the entire work is sub-divided into following three sections:

Data collection: the data mining and their outcomes directly depend on the nature and kind of data. In addition to that, the quality of data is also a key factor in effective data mining process. Thus the unbiased data collection, reduction of noise and outliers from the data is a primary challenge.

Data Analysis: To enhance the quality of data for accurate and effective data analysis. Thus required data models or data mining algorithm selection is also a second key

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001

issue in this work. That is also expected the data analysis models are needed to be efficient and accurate during their applications.

Obtaining the conclusion from data and data models: The generated outcomes from the numerous data modeling approaches and their interpretation are also included the complexity. Thus need to be made, additional efforts for understanding and correlating the outcomes of the data

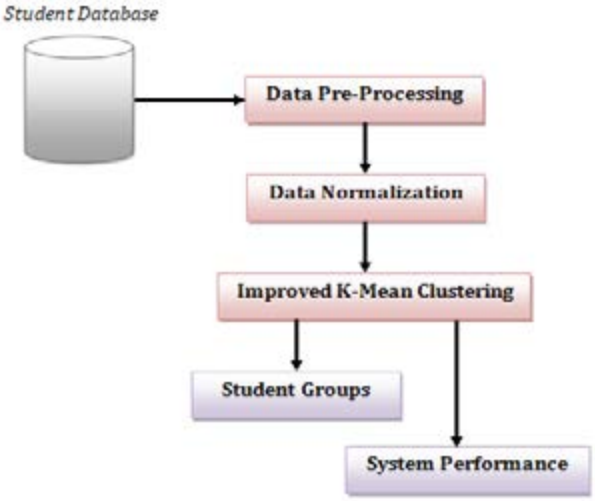

Figure 1.Creating Student Groups

model for obtaining real- world advantage is also a key aim of the proposed work.

The above-listed issues are complicated issues for any data

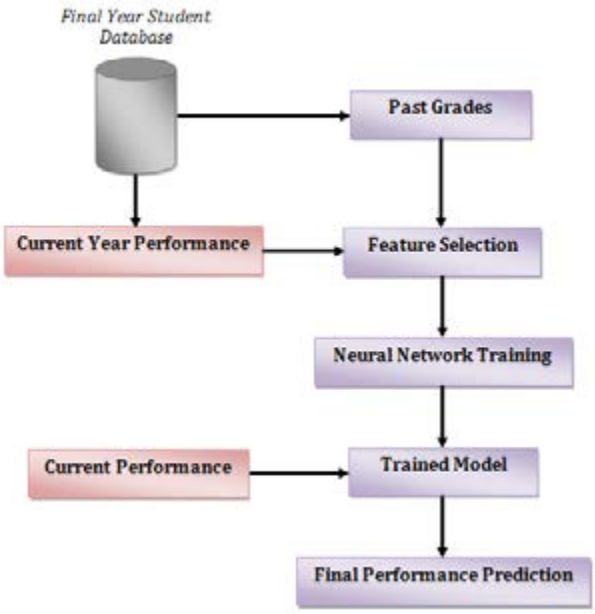

Figure 2.Final Year Presentations Prediction

mining based application design. In order to rectify the key issues the following strategy is proposed for work:

The proposed concept for predicting the student's presentations three data models are provided using figure 2, 3 and 4. Figure 2, shows the data model which is used to predict the learning ability of students.

In this context, the database of students is considered which contains the student's class presentations for any semester according to their class test, and midterm exams, a number of assignments completed and additional achievements list. This dataset is pre-processed for discovering the missing values and other noise complexities in data. Which is further processed with the improved k-means clustering algorithm. The improved k-means clustering algorithm is a fixed point kernel based k-means clustering technique that clusters the student's into three group's namely good learners, average learners, and the weak learners. This data model helps to understand the student's current presentations and also help to provide feedback on their improvements.

In a similar manner, the second data model as described using figure 3 is aimed to predict the final year marks according to their previous patterns and current presentations. In this data model, the pass out student's database is provided as input. That data is a combination of the previous presentations of students and the final year presentations or current year presentations. This dataset is used with the correlation coefficient with the student's final grade or

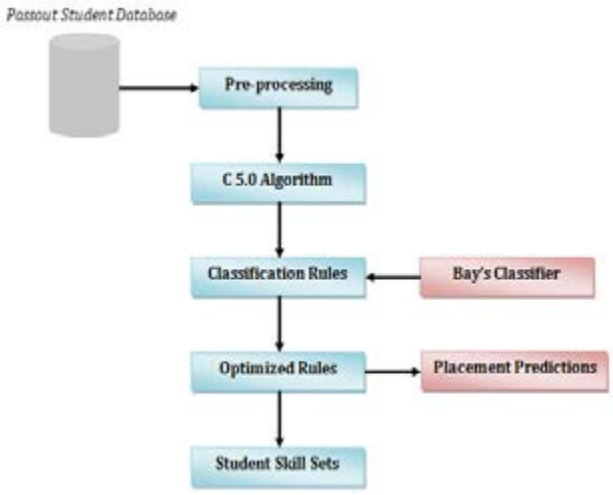

Figure 3.Students Placement Prediction

classes. Using this technique the features are selected from the training dataset which is used with the artificial neural network for performing the training on selected features of student records. After preparing the trained data model the current or pursuing student's current presentations and old records are produced to the trained model for predicting the final marks or grades.

The third and final data model is described in figure 4 which provides the prediction of the student's presentations for the placement in some company. This data model also accepts

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001

the pass out student records as input for performing the training. The data is first preprocessed for cleaning and normalize the dataset. Further, the C5.0 algorithm is applied to the dataset. This process results in the tree form of decision, that is used to extract the cataloging rules. These classification rules are then validated with the help of the bay's classification algorithm. This data model is used for predicting the student's placement chances.

Conclusion and Future Work

EDM has potential mining the aspect that may be helpful for student's learning capability, and also enhance the process of learning. EDM is a sub-domain of data mining for analyzing the academic data relevant to administrative and academic points of view.

One of the most recent and biggest challenges that higher education faces today is making students skillfully employable. Many universities/institutes are not in a position to guide their students because of lack of information and assistance from their teaching-learning systems. To better administer and serve student population, the universities/institutions need better assessment, analysis, and prediction tools. Considerable amount of work is done in analyzing and predicting academic presentations, but all of these works are segregated. There is a clear need for a unified approach. Suitable data mining techniques are required to measure, monitor and infer these factors for prediction. Thus enriching the input vector with qualitative values may rise the accuracy rate of prediction as well.

Integrated Models/Frameworks are required for all the stakeholders of an Institution; hence ensuring sustainable growth for all (Management, Teachers, Students and Parents).

References

1. Romero, Cristobal, Ventura S. Educational data mining: A survey from 1995 to 2005. Expert systems with applications 33.1, 2007; 135-146.
2. Romero, Cristóbal, Ventura S et al. Data mining in course management systems: Moodle case study and tutorial. Computers & Education 51.1, 2008: 368-384.
3. Ahmed, El Din AB, Elaraby IS. Data Mining: A prediction for Student's Presentations Using Classification Method. *World Journal of Computer Application and Technology 2.2*, 2014; 43-47.
4. Saa AA. Educational Data Mining & Students' Presentations Prediction. *International Journal of Advanced Computer Science and Applications* 2016; 7(5).
5. Gautam R, Pahuja D. A Review on Educational Data Mining. *International Journal of Science and Research (IJSR),*2014; 3(11).
6. Bhardwaj BK, Pal S. Data Mining: A prediction for presentations improvement using classification. Data Mining: *A prediction for presentations improvement using classification* 2011; 9(4).
7. Pelánek, Radek, Rihák J et al. Impact of data collection on interpretation and evaluation of student models. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ACM, 2016.
8. Peña-Ayala, Alejandro. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications* 41.4 (2014): 1432-1462.
9. Asif, Raheela, Merceron A. Analyzing undergraduate students' presentations using educational data mining. Computers & Education, 2017.
10. Pal AK, Pal S. Analysis and Mining of Educational Data for Predicting the Presentations of Students", *International Journal of Electronics Communication and Computer Engineering,* 2013 4, Issue 5, , pp.1560-1565
11. Rajni Jindal R, and Malaya Dutta Borah. "A survey on educational data mining and research trends." International Journal of Database Management Systems 5.3 (2013): 53.
12. Ferguson R, Learning analytics: drivers, progresss and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6): 304–317.
13. Hwang GJ, Hui-Chun Chu & Chengjiu Yin (2017) Objectives, methodologies and research issues of learning analytics, Interactive Learning Environments, 25:2, 143-146, DOI: 10.1080/10494820.2017.1287338
14. Nicholas A. Ashford, Professor of Technology and Policy, Massachusetts Institute of Technology. Major Challenges to Engineering Education for Sustainable Progress: What Has to Change to Make it Creative, Effective, and Acceptable to\the Established Disciplines. Samira ElAtia, Donald Ipperciel, Ahmed Hammad, "Implications and Challenges to Using Data Mining in Educational Research in the Canadian Context"
15. Rana S, Garg R. Evaluation of Student's Presentations of an Institute Using Clustering Algorithms. *International Journal of Applied Engineering Research*, 2016;11(5): 3605-3609.
16. Ariouat H, Cairns AH, Barkaoui K et al. A two-step clustering approach for improving educational process model discovery", 25th IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, 13 June 2016 - 15 June 2016.
17. Oeda S, Hashimoto G. Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes", *Procedia Computer Science* 2017; 1126: 14-621. Published by Elsevier B.V
18. Ogwoka TM, Cheruiyot W, Okeyo G. A Model for Predicting Students' Academic Presentations using a Hybrid of K-means and Decision tree Algorithms.

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001

International Journal of Computer Applications Technology and Research 2015; 4(9): 693 - 697.

19. Jiaqu Yi, Sizhe Li, Maomao Wu et at. Cloud-based Educational Big Data Application of Apriori algorithm and K-Means Clustering algorithm based on Students' Information", *Fourth International Conference on Big Data and Cloud Computing,*IEEE 2014.
20. Kumar M, Shambhu S, Aggarwal P. Recognition of Slow Learners Using Classification Data Mining Techniques. *Imperial Journal of Interdisciplinary Research* (IJIR) 2016; 2(12).
21. Ahmed ABED, Elaraby IS. Data Mining: A prediction for Student's Presentations Using Classification Method. *World Journal of Computer Application and Technology* 2014; 2(2): 43-47.
22. Kabakchieva D. Predicting Student Presentations by Using Data Mining Methods for Classification. *Cybernetics and Information Technologies,* Sofia, 2013; 13(1).
23. Gulati H. Predictive Analytics Using Data Mining Technique. 978-9-3805-4416-8/15/$31.00 c 2015 IEEE
24. Daud A, Aljohani NR, Abbas RA et al. Predicting Student Presentations using Advanced Learning Analytics. Companion, April 3-7, 2017, Perth, Australia. ACM.
25. Vijayalakshmi MN, Kumar SA, Kavyashree BN. Investigating Interesting Rules Using Association Mining for Educational Data. *International Journal of Advanced Research in Computer Engineering & Technology* 2014; 3(2).
26. Jha J, Ragha L. Educational Data Mining using Improved Apriori Algorithm. *International Journal of Information and Computation Technology.* 2013; 3(5): 411-418
27. Arora RK, Badal D. Mining Association Rules to Improve Academic Presentations. *International Journal of Computer Science and Mobile Computing*,2014; 3(1): 428-433
28. Romero C, Zafra A, Luna JM et al. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data", Expert Systems, 2013; 30(2)
29. Chalaris M, Gritzalis S, Maragoudakis M. Cleo Sgouropoulou and Anastasios Tsolakidis. Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques. *Procedia - Social and Behavioral Sciences* 2014; 147: 390-397.
30. Ryan SJD, Baker, Yacef K. The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*.

ISSN: 2395-3802
DOI: https://doi.org/10.24321/2395.3802.202001