ACTA SCIENTIFIC COMPUTER SCIENCES

Volume 2 Issue 7 July 2020

Research Article

Implementation of the Machine Learning Tools on Improving the Decision Support System for Pakistan's Education

Sohaib Latif[1]*, Fang Xianwen[1] and Li-Li Wang[2]

[1]*Anhui University of Science and Technology, Huainan, China*

[2]*The Key Laboratory of Embedded System and Service Computing Ministry of Education (Tongji University), Shanghai, China*

***Corresponding Author:** Sohaib Latif, Anhui University of Science and Technology, Huainan, China.

Received: June 15, 2020

Published: June30, 2020

Abstract

The unknown factors always prevail in developing countries like Pakistan, due to these factors the annual education performance could not show the expected results. The proper implementation of advanced machine learning tools can help to highlight the cause of low performance in the education field. The study aims to address this issue by proposing a new 3-level classifier model for decision support that can identify the reasons for failure by training the neural network which provides reliable results. For experiment purpose, a dataset of 1011 respondents of graduate-level students of "English" and "Physics", used to analyze where the C4.5 algorithm provides maximum accuracy of 83.2%, 88.8%, 83.1% and 89.8%. The study also presents a prototype of the java-based software to implement the proposed model of a 3-level classifier. The Java-based software uses the Waikato Environment for Knowledge Analysis (WEKA) machine learning toolkits with Java Virtual Machine (JVM) to produce reliable results. The enhanced and fully featured java-based software would provide 88.8% accuracy in the decision support system to identify the gray areas of the education sector of Pakistan.

Keywords: Neural Network; Real-World Data Mining; Decision Tree; Education System of Pakistan

Introduction

Higher education institutions of Pakistan overwhelmed with a large amount of data of students, containing the information of students' enrollment, number of courses introduced in the educational institutions, students' achievement in each of the courses, and annual result statistics [1]. The growing use of computerized systems for record-keeping brings a new term, "big-data" [2]. The extensively high amount of inter-related data about any entity that can be used for analytical analysis is called big-data [3]. It is difficult to perform an analytical process on big data for making decisions about curricula reforms and restructuring the education system. Implementation of intelligent methods is essential for extracting data patterns and analytical information to discover hidden knowledge from students' databases [4]. This new research area has become popular and grown exponentially in the new era of modern education, the reason behind its potential and capacity building for familiarizing improvement in the quality of education systems [5].

Conventional education systems are different from the newly emerged machine learning tools assisted systems where availability of multi-dimensional data present in abundance [6]. Intelligent decision support systems (IDSSs) are widely used in various computer science applications for intelligent decision-making [7]. The application of a decision support system is mainly concentrated on improving the learning process by the development of accurate models that predict students' characteristics and performance [8]. The importance of this system is founded on the fact that it allows educators and researchers to extract useful conclusions from sophisticated and complicated questions such as "find the students who 'will' exhibit poor performance" in which traditional database queries cannot be applied [9].

The secondary education system in Pakistan is a two-tier system in which the first two years cover the general topics and introduction of the majors of degree [10]. The remaining two years contain focused learning on the selected subjects, this period is called higher education system [11]. So, the last two years of higher secondary education have immense significance and decisive factors for the life of any student, it acts as a connecting bridge from school learning to higher education, provided by different universities and higher educational institutes [12]. Therefore, the capacity to monitor the students' academic performance and achievement is considered a highly important factor for the identification of possible bad performance that could lead to the decay of education performance [13]. During the last couple of scores, the researchers

aim to develop an efficient and precise decision support system to how the projected students' academic performance [14]. More analytically, an academic decision support system is a knowledge-based information system to capture, handle and analyze information that affects or is intended to affect decision making performed by people in the scope of a professional task appointed by a user [15]. The development and implementation of an academic decision support system is significant to students, educators and educational organizations [16]. Management of education must continue to implement and evaluate over ongoing basis to improve the quality of institutions [17].

Aim of the Study

This study aims to implement the machine learning tools and framework for decision support systems for improving the education system of Pakistan that provides decision support for evaluating students' performance in the final examinations.

Methodology

Descriptive research is conducted to test machine learning tools for developing the decision support system for the improvement of Pakistan's education system. This research has been conducted by collecting data on education institutes of Pakistan. The proposed design model of the decision support system creates new possibilities for the collection, analysis and presentation of data to provide precise and accurate results. Test series like the chi-squared test have been performed under SPSS version 21.0. The proposed model contains three stages by which the machine leaning aided system provides results against the dataset provided to the system. The data collection and processing are the initial approaches to the suggested model. The data is provided to the system as the whole class data contains the successful and unsuccessful student's results. At the same stage, the pass students and fail students are separated which will be used in the next stage as the input values for highly complex classifiers. The second stage consists of implementing the proposed scheme of 3-level classification. The values-driven by previous operation is used in this stage for the decision-oriented classifier. The classifier starts operation on all those students who could stand good in the result. The data of all students who were in the category of fail use as an input value for the b-level classifier. This classifier determines the understanding of students with a curriculum; this data is collected by questionnaires before the examination. This classifier determines whether the succeeded candidate has the understanding of curriculum or pass by cramming. This classifier will later help to improve the curriculum. In the last step, the machine learning algorithms are used to process the dataset to get final results. This classifier determines the overall performance of all passed candidates over failed candidate: and determines by using a neural network that either passed candidates also face issues in understanding the curriculum or only fail candidate has. If the pass candidates also have the same issue then the system decides by multiple other machine learning operations and artificial intelligence (AI) algorithms to determine the actual reason for failure, which gives the result that either reform in curriculum is required or in the teaching method. This proposed model could also determine that the failure is caused by individual hardworking issues or medical issues of failed students. This is how it is very easy to get reasons for the failure of students and help to take necessary actions for the improvement of that particular area

Dataset

The study used a set of data collected from 1011 respondents in graduate-level students having a result of "English" and "Physics" courses. Table 1 enlists the set of parameters used throughout analysis which relate to evaluation details of the participants such as marks obtained in percentage, attendance, class participation, number of students failing, and cumulative comprehension of the curriculum. During each course, the students are assessed by verbal communication and written examination. This investigation lasts for three hours. The 10 interviews with 10 questions were included to determine the perception of every student from the program. The 15-minute tests contain verbal questions and thorough problem-solving. The 3-hour tests lead to multiple theoretical and advanced mathematical problems requiring methods to be solved and evaluated critically. Finally, each student's overall 5-semester grade discusses the student's interest and its success.

The students were categorized into four-level classification scheme which helps in the evaluation of students in the Pakistani schools e.g. Fail (0 - 3), Good (4 - 5), Very good (6 - 8) and excellent (9 - 10).

Nevertheless, since it is significant to an instructor to apprehend poor students within the midst of the instructional cycle, datasets have been created in this context primarily based on the parameters furnished in table 1 and the elegance allocation:

- $DATA_A$: This includes the characteristics relating to the success of the first semester graduates.
- $DATA_{AB}$: This includes the characteristics relating to 1st and 2nd semester results of the graduates.

3-level classifier

The invention of new classification scheme has a great challenge to deal with more precision. To that end, we are implementing a

Attributes	Type	Value
Number obtained in percentage in 1st semester	Real	[0-9]
First semester attendance	Real	[0-9]
Class Participation in 1st semester	Real	[0-9]
Number of student who fails	Real	[0-9]
Curriculum Understanding	Real	[0-9]
Number obtained in percentage in 2nd semester	Real	[0-9]
Attendance in 2nd semester	Real	[0-9]
Class Participation in 2nd semester	Real	[0-9]
Number of student who fails	Real	[0-9]
Curriculum Understanding	Real	[0-9]

Table 1: Collection of functions used in the analysis.

scheme for the classification of architecture at three stages. Three-level identification strategies are deterministic machine learning devices intended to provide precision than one-level tools at the cost of some classification scheme complication.

The study uses the proposed classification scheme on first level to classify the students either they are pass or fail. This performance was judged in the entire class for a particular semester. Furthermore, this algorithm explicitly determines whether the student performs between 0 and 3 (Fail) or between 4 and 10 (Pass). To assess the passing and failing classifier used for student A-Levels. In the case, final examinations have declared Fail at A-Level classifier then B and C classifiers are the best option to determine the most likely reason for failure.

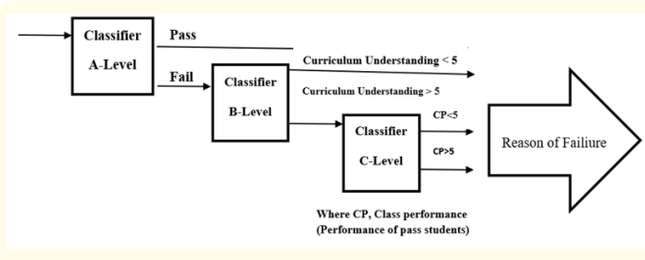

Figure 1: Description of the classifier on 3 rates.

It is worth bearing in mind that the A-Level classifier's judgment is coarser and broader, while the B-Level and C-Level's decision is finer and portrays the explanation for an ineffective student so that either the problem was with the interpretation of the program, the person disinterest of the teaching method.

Experimental Results

Results were made by a series of tests. These tests were also used to analyze the reliability of proposed 3-Level computational model (most successful and widely deployed machine learning classification algorithms). The neural artificial networks (learning algorithm) are the reflection of the Back-Propagation (BP) algorithm. The neural network is constructed and educated by these algorithms. The Sequential Minimal Optimization (SMO) algorithm (simplest training regimes) was also preferred by researcher as a support vector machine. The Bayesian networks were also represented by the Naïve Bayes (NB) algorithm. This C4.5 algorithm was the protagonist in the analysis. For the standard rule-learning strategy, the RIPPER (JRip) algorithm (most widely used methods of producing rules for classification) was chosen. The 3-NN algorithm (instance based learner) was configured as distance metric. Additionally, in the experimental results, the RIPPER, 3-NN, BP and SMO were referred for voting stands (known as quick voting scheme). The WEKA Computing Toolkit contains all these algorithms. The stratified 10-fold cross validation was used to validate the accurate classification. The split of dataset into folds is the best example and the each fold has same grade distribution. It presents the whole data. The accuracy of each individual neural network is illustrated in table 2 and also shows the relation of both the 3-Level clustering algorithm and databases. The each individual classifier's efficiency (3.8% to 9.5%) is improved by 3-Level proposed scheme.

Classifier	$DATA_A$		$DATA_{AB}$	
	Individual	3-Level	Individual	3-Level
Back-Propagation (BP) algorithm	80.1%	88.1%	75.2%	88.1%
Sequential Minimal Optimization (SMO) algorithm	83.1%	89.9%	81.7%	88.9%
Naïve Bayes (NB) algorithm	74.8%	80.2%	70.7%	78.0%
C4.5 algorithm	83.2%	88.8%	83.1%	89.8%
RIPPER (JRip) algorithm	83.7%	81.0%	85.1%	88.6%
3-NN algorithm	84.2%	87.1%	77.7%	84.3%
Voting	85.8%	88.2%	84.4%	87.0%

Table 2: Specific classification system and 3-level classification precision (percentage) for each dataset.

Therefore, it is analyzed that the proposed 3-level scheme using dissimilar categorization algorithms at each level swayed by the previous findings. The goal is to find out which of these classifiers deliver the highest performance. It is best suited for A-Level, B-Level and C-Levels. SMO is also best C-Level classifier to give the absolute best performance, followed by C4.5. Eventually, it is worth mentioning that the C4.5 algorithm was selected as A-Level classifier, SMO as a B-Level, and for the C-Level the best classification performance of the 3-level classifier was described. The above results clearly depict the efficiency of newly proposed model by using tools

of machine learning that give better results to improve the education system specifically of Pakistan. The experimentally proved most salient feature is its three way prediction of most likely reason of failure. The main three strong stockholders are, individuals, teacher and the learning material, which may increase or decrease the standard of education, the proposed model critically evaluate all three concerns of education system to predict accurate results. Moreover, this model has extensibility for future studies because as the new problem emerges the number of complexes classifiers will be changed or increased in order to entertain new problems in the way of education system. Thus, the education system and its all upcoming issues would be entertained by utilizing this proposed model.

Implementation

For this purpose, the newly proposed model can be implemented on the various platform but this study selects the JAVA based prototype for the implementation. The software will use the WEKA machine learning Toolkit which will easily work with Java Virtual Machine (JVM). It is penitence to mention that the prototype for implementation of proposed model with the (i.e. Livieris., *et al*. 2015) proposed decision support system. The prototype will be able to get the data of the entire class in the standard format of comma separated value (CSV), data must be contained prefixed fields such as roll_no, student_name, student_total_mark, student_CUindex, student_CPindex. The comma separated values (CSV) file parse with the OpenCSV a library for Java. The OpenCSV support all basic comma separated value (CSV) format file to perform operation. Currently Java 7 has the compatibility of the OpenCSV. The prototype will utilize basic classes of java with extended toolkits of machine learning to train the neural network for making the decision. The result will be displayed on the interface designed by Java swing/Fx which will show the result about the reason of failure of student according to the data provided.

Discussion

Implementation of intelligent approaches is essential to retrieve data structure and analytical knowledge to explore secret knowledge from the databases of the students. In the global educational age, the application of smart to build accurate and efficient decision support systems for evaluating the output of learners is becoming very common [1]. Outstanding feedback have been received on how machine learning tools (MLT) strives to explore different perspectives into learning through new tools and techniques to influence professional activity at all educational levels, as well as corporate learning and systematic methods of the machine learning tools and how to implement the techniques, however. Alleghany M., *et al*. [2] Illustrated the usefulness of the academic decision support system in evaluating enormous quantities of student-course data. Besides, authors discussed the basic principles used in the development and design of a new set of decision support system tools and discussed various methods of evaluating and presenting student performance data for academic decision making. Beam A., *et al*. [3] studied the Precision of six rising machine learning algorithms in projecting participants who appear to drop out of a Hellenic Open University distance learning course. Based on previous works and a conceptual decision supporting system was implemented to forecast the academic success of students based on key demographic characteristics, attendance, and their marks in written assignments. Sabahi S., *et al*. [5] Proposed a knowledge-driven decision support system (DSS) for semester reward learning using Educational Data analysis [11]. Their suggested guidance framework for educational decision-making is supportive for educational executives to make more effective and fair assumptions about student research and provide more encouragement for graduation to learners. Kostopoulos G., *et al*. [12] studied the implementation of the decision support system focuses mainly on enhancing the learning process by designing detailed models that predict the characteristics and results of the students. Besides, they developed a unique computing method that takes into consideration the active participation of the student in a Moodle forum and their regular utilization. Niet YV., *et al*. [14] proposed a "Student Advisory System" system to create a smart framework. The framework is used to provide a first-year high school student with pieces of consultation to follow a certain educational trajectory in which he/she is likely to excel, intending to reduce the high rate of poor academic performance among such pupils. The system possesses records from the data-sets that maintain students' academic achievements before enrolling in higher education along with their grade of the first year after enrolling in a certain school [16]. After obtaining all the relevant data, the smart system includes both identification and aggregation techniques to make predictions for a particular department. In particular, they submitted a scenario study to prove the feasibility of the theoretical model. Data from the Cairo Higher Institute for Engineering, Computer Science and Management for the period 2000 to 2012 were collected from the students. Géron A [19] based on early detection of high-risk secondary school students while allowing the instructor to act accordingly to enhance the outcome of the students through additional coaching and counseling. Besides, they listed the important qualities that influenced the performance of the third-semester students and calculated the effects of emotional quotient parameters that influenced the influenced placement. In more recent works, Wójcik K and Piekarczyk M. in 2020 introduced a computer tool to forecast student success in the first year of Lyceum's "Mathematics" school. The proposed software is based on a classifier of the neural network which exhibits more consistent behavior and shows better accuracy than other

classifiers. Along the whole line, a user-friendly decision support program was introduced in the authors to predict the success of participants, along with a case study on the Mathematics final exams [20-22]. Their proposed method is based on a system of hybrid predictions that incorporates four learning algorithms using a simple voting scheme. Beam AL., *et al*. [19] Studied the significant issue of early school drop-out prediction. We suggest a technique and a basic classification algorithm for finding intelligible student dropout predictive models as rapidly as possible. In the last two millennia, a new approach for improving the output of single classifiers has been proposed in the field of artificial intelligence by integrating the prediction of a variety of classifiers. Chen and Do (2014) addressed a comprehensive analysis because of its primary objective of investigating the predictability of the output prediction of neural networks for students using preceding exam results, students 'gender and other statistical attributes as input variables. In Newer Plays, Pandey and Taruna (2014) the output of several classifiers was analyzed to efficiently classify poor students, and a multilevel classification model was formulated [23]. They also implemented pre-processing techniques such as resample filtering as well as eliminating the misclassified circumstances from the initial classifier to improve the model's classification accuracy.

Conclusion

This study shows the need for the hour to have a decision support system that may assist to determine the reason behind the failure of students in academic years. The work firstly identifies the dataset for experimental analysis about the working of the proposed three-level classifier for the decision support system. After supposing and experimentally proving the author introduces a prototype of the software to implement this newly proposed model. The software provides the required results as per the experimental analysis proved about the training of neural networks. The Java-based prototype will be enhanced by the addition of custom features for use in the education sector of Pakistan. The objective and expectation will introduce a new paradigm to identify the reasons for the students who are unable to perform well in the academic years and resulted as a fail candidate. The proper utilization of the proposed model can show up to 88.8% (from the table 2) accurate results to identify the real reason and gray area of the Pakistani education sector.

Bibliography

1. Niet YV., *et al*. "Academic decisionmaking model for higher education institutions using learning analytics". In 2016 4th International Symposium on Computational and Business Intelligence (ISCBI) (2016): 27-32.
2. Ala'M AZ., *et al*. "Salp Chain-Based Optimization of Support Vector Machines and Feature Weighting for Medical Diagnostic Information Systems". In Evolutionary Machine Learning Techniques (2020): 11-34.
3. Mothilal RK., *et al*. "Explaining machine learning classifiers through diverse counterfactual explanations". In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (2020): 607-617.
4. Dasgupta J. "Imparting Hands-on Industry 4.0 Education at Low Cost Using Open Source Tools and Python Eco-System". *New Paradigm of Industry* 4 (2020): 37-47.
5. Alexeyev A and Solianyk T. "Decision-Making Support System for Experts of Penal Law". *In Data-Centric Business and Applications* (2020): 163-182.
6. Livieris I., *et al*. "A decision support system for predicting students' performance". *Themes in Science and Technology Education* 9.1 (2016): 43-57.
7. Maxwell AE., *et al*. "Implementation of machine-learning classification in remote sensing: An applied review". *International Journal of Remote Sensing* 39.9 (2018): 2784-2817.
8. Fiebrink R. "Machine learning education for artists, musicians, and other creative practitioners". *ACM Transactions on Computing Education (TOCE)* 19.4 (2019): 1-32.
9. Romeo L., *et al*. "Machine learning-based design support system for the prediction of heterogeneous machine parameters in industry 4.0". *Expert Systems with Applications* 140 (2020): 112869.
10. Hormigo IG., *et al*. "Design and Implementation of Dashboards to Support Teachers Decision-Making Process in e-Assessment Systems". Engineering Data-Driven Adaptive Trust-based e-Assessment Systems (2020): 109-132.
11. Naidu VR., *et al*. "Machine Learning for Flipped Teaching in Higher Education-A Reflection". *In Sustainable Development and Social Responsibility* 2 (2020): 129-132.
12. Soliman HASA and Tabak FATEMA. "Deep learning framework for RDF and knowledge graphs using fuzzy maps to support medical decision". *Journal of International Research in Medical and Pharmaceutical Sciences* 14.3 (2020): 92-97.
13. Scarpazza C., *et al*. "Applications of machine learning to brain disorders". *Machine Learning* (2020): 45-65.
14. Díaz JJS. "Artificial intelligence in cardiovascular medicine: Applications in the diagnosis of infarction and prognosis of heart failure". *Artificial Intelligence in Precision Health* (2020): 313-328.

15. Mosavi A., *et al*. "Reviewing the novel machine learning tools for materials design". *In International Conference on Global Research and Education* (2017): 50-58.

16. Kostopoulos G., *et al*. "Predicting student performance in distance higher education using active learning". International Conference on Engineering Applications of Neural Networks (2017): 75-86.

17. Alloghani M., *et al*. "Implementation of Machine Learning and Data Mining to Improve Cybersecurity and Limit Vulnerabilities to Cyber Attacks". Nature-Inspired Computation in Data Mining and Machine Learning (2020): 47-76.

18. Asthana P and Hazela B. "Applications of Machine Learning in Improving Learning Environment". Multimedia Big Data Computing for IoT Applications (2020): 417-433.

19. Beam AL., *et al*. "Challenges to the Reproducibility of Machine Learning Models in Health Care". *The Journal of the American Medical Association* (2020).

20. Wójcik K and Piekarczyk M. "Machine Learning Methodology in a System Applying the Adaptive Strategy for Teaching Human Motions". *Sensors* 20.1 (2020): 314.

21. Sabahi S and Parast MM. "The Impact of Entrepreneurship Orientation on Project Performance: A Machine Learning Approach". *International Journal of Production Economics* (2020): 107621.

22. Kostopoulos G., *et al*. "Predicting student performance in distance higher education using active learning". International Conference on Engineering Applications of Neural Networks (2017): 75-86.

23. Géron A. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems". O'Reilly Media (2019).