International Journal for Multidisciplinary Research
International Conference on Multidisciplinary Research & Studies 2023
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Comparison of Applications for Data Mining in Education

Dr. Prafullkumar Ashokrao Ghuge

Shankarlal Khandelwal Arts, Science & Commerce College, Akola

Abstract:

Educational Data Mining focuses on the use of data mining in educational research (EDM). Data collected during the teaching and learning process can be analysed using a variety of methods, including machine learning, statistics, data mining, and data analysis. – An educational data mining technique is the process of extracting useful information from large educational databases.

In recent years, scholars have presented a variety of mining methods. Using appropriate data mining methods for educational datasets is the goal of this paper. Data mining is one of the most popular methods for delivering feedback on the teaching-learning process.Open-source apps for educational data mining have proliferated in recent years. The current tendency in educational research is to leverage powerful statistical learning approaches to extract relevant insights from large data streams.

Keywords:

Educational Data Mining, Student Performance, Data Mining Techniques, Educational Dataset, Teaching, Student, Learning, Classification.

Introduction:

For the most part, educational systems are preoccupied with preparing students to enter the workforce at a fixed moment in time. Social and economic advancement are strongly influenced by educational systems' ability to achieve this goal. Technology in educational systems has produced a significant amount of data that people cannot interpret. In educational contexts, educational data mining (EDM) employs a variety of data mining techniques to analyse student data.When it comes to analyzing and solving educational problems, EDM is the primary tool. As a result, it aims to study educational data in order to resolve education-related challenges. Education-related data mining (EDM) is the process of obtaining meaningful, interpretable information from educational datasets [1-4].

Engineering, science, medicine, business, and education can all benefit from data mining, which is a rapidly growing field of study. When it comes to educational applications, the size of the data base might range from a few thousand to tens of millions of records. The amount of data being collected from many domains is growing at an exponential rate.The combination of machine learning, artificial intelligence, statistics, and database systems has been a beehive of innovation in data mining. Extracting data from a vast database and making it understandable is the ultimate goal of data mining. Data mining techniques are increasingly being used by educational institutions to extract information from massive volumes of data. [5] Experts in educational data analysis are troubled by ambiguity due to the plethora of DM tools now on the market and used in many different data analysis fields [6].In light of the fact that this habitat has unique characteristics that make it distinct from other ecosystems, this is extremely important. It is true that dedicated programmes have been built for university environments [7],However, these programmes do not contain all of the features of the most commonly used DM applications.

Educational Data Mining (EDM)

In many educational institutions, the amount of data being collected and kept has grown to such an extent that manual data analysis is no longer feasible. Educational data mining, often known as EDM, is a relatively new concept that arose from the application of data mining techniques to educational data sets. In 2008, the first global conference on EDM research was held in Montreal, Canada. The Journal of Educational Data Mining was created in 2009, and the International Educational Data Mining Society was founded in 2011. (Romero and Ventura, 2010, pp. 601-618). Since its inception, EDM has spawned areas such as data mining and machine learning, as well as psychology, artificial intelligence, and computer modelling. [8]

The introduction of data mining techniques to educational data has led to the development of academic analytics and learning analytics. Stakeholders who stand to gain from academic analytics vs. learning analytics are the most significant distinctions.Data from educational institutions can be mined for information about how administrators and other high-ranking members of the institution's leadership are carrying out their duties. These stakeholders are looking for information on the institution's retention rate, variables that influence it, and how students use the resources available to them. Academic analytics can help answer these questions.In contrast, the goal of the discipline of learning analytics is to assist both students and teachers. Student performance is examined to discover their shortcomings and the teaching method that appeals to them most. As a result of this research, teachers will be able to make adjustments to their instructional strategies that will benefit pupils.

Reviews of Literature:

In Educational Data Mining: A Survey from 1995 to 2005, a survey of EDM literature from 1995 to 2005 is provided. Written by Romero and Ventura (2007). [9] This study examined the effectiveness of traditional classroom training to online instruction using web mining technologies.

Researchers Al-Razgan and Al-Khalifa conduct a comprehensive evaluation of the EDM literature published between 2006 and 2013 using Google Scholar's highly cited publications. (2014). [10] A bibliographic overview of educational data mining studies, approaches, and contributions to their application was presented in Educational Data Mining Applications and Trends (Peria-Ayala, 2013).

Researchers Kumar and Chadha (2011) conducted an empirical examination into the possible applications of data mining techniques in higher education, in an effort to identify those areas of potential use. As a result of their findings and the data mining techniques employed, they came to the following conclusions: syllabus organization, prediction of student enrollment in educational programmes, student performance forecasting (including cheating detection), abnormal or erroneous value identification (including abnormal or erroneous value clustering), and outlier analysis (which identifies outliers).

According to Ali (2013) [11], educational data mining can uncover students' patterns, preferences, and course requirements; select specialities; forecast students' final results; autonomously explore data; and profile individuals.

Educational Data Mining's Advantages and Uses

There are significant advantages and uses to educational data mining. Numerous studies and publications have been done on data mining in the education sector. Later, we'll talk about a few more.Most commonly, educational data mining is used to enhance the learning experience, increase completion rates, assist students in course selection, create profiles of their academic progress, detect dropouts, better target students, construct curricula, and aid in enrollment decision-making.

According to Romero and Ventura (2010) [8], EDM can be used in a number of different contexts such as: modelling, prediction, data visualisation, social network analysis, feedback in support management, scheduling and student grouping, and behaviour detection. [8]

Baradway and Pal (2012) presented a data mining model for the university system to show the potential of data mining techniques in higher education.When evaluating a student's performance, the decision tree method was applied. This study could aid instructors in the early diagnosis of pupils who are at risk of dropping out and those who require particular care. [12]

Luan (2002) discussed the potential applications of data mining in higher education and how data mining techniques may save resources and boost efficiency. [13]

According to Kovai (2012), a student's socio-demographic traits, such as their gender, ethnicity, education, employment status, and disability, can be used to distinguish successful and unsuccessful pupils. Classifying students based on pre-enrollment information, according to the findings of this study, can assist in identifying students who are at risk of dropping out of the course and providing advice and mentorship programmes to help them succeed. [14]

One data mining technique employed by Kabackijeva (2012) was the nearest neighbour classifier, which was a combination of the rule learner, a decision tree classifier, a neural network, and other data mining techniques. [15]

Using data mining, according to Maqsood (2013), can be utilised to construct marketing programmes for targeted students. [16]

Objectives:

We wanted to compare the performance of three data mining (DM) tools in our classroom experiment (RapidMiner, Knime, and Weka). To find these features, researchers looked at the academic records of three engineering programmes at an Ecuadorian institution using clustering and segmentation techniques.Comparing the various EDM tools can help those who want to work in EDM and learn how to select the appropriate equipment for a specific study.

Research Methodology:

Tools will be compared in terms of technical aspects as they are put to use in the various stages of the DM cycle.For example, outcomes obtained by each tool during process development are contrasted, as are available algorithms for conducting DM operations and the working environment in which each tool is used.

Figure 1 depicts the DM process, which will be explained in depth in the following section.

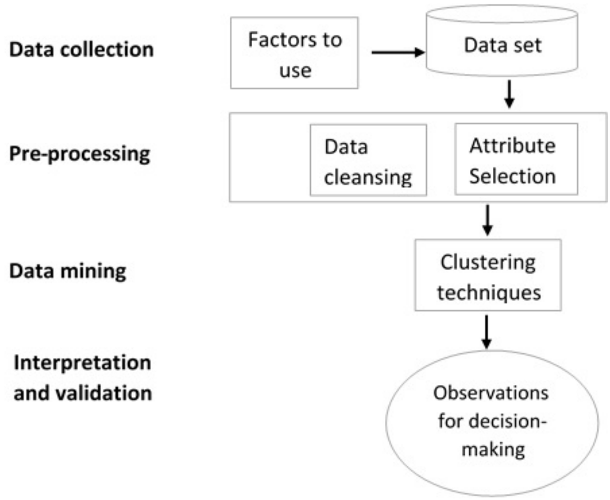

Figure 1 Depicts the Standard Method Used In The MDE Procedure.

Tools of Data Mining:

Commercial and open-source data processing solutions are widely available. However, we have only compared three of them because they are the most often utilised in educational settings [17]. As a result, they are GPL-licensed and open-source in nature. RapidMiner, Knime, and Weka are all part of the same family. RapidMiner is a software suite that focuses on data mining and model building. The implementation and execution of DM procedures are simplified by a user-friendly graphical user interface [18]. The tool enables the chaining of operators to construct data analysis procedures. Operators for data analysis, preprocessing, and visualisation are offered in more than 500 varieties. In addition, the Weka tool's algorithms can be utilized.

DM platforms such as Knime (Konstanz Information Miner) offer model building in a visual setting. It's built on top of the Eclipse platform and is primarily written in Java. A wide range of operating systems are supported. To visualise and create statistical models and representations of DM, it provides a wide range of tools (such as pie charts and point clouds) as well as a variety of additional methods (such as decision trees and regressions) for doing so. It also makes it possible for users to call the Weka tool directly and to add R or Python code in a straightforward manner [19].

It is part of Pentaho's Intelligence Suite, Weka (Waikato Environment for Knowledge Analyses). Java-based machine learning techniques are embedded throughout. WEKA can be incorporated into custom Java programmes or into a graphical user interface on its own. In this package, you'll find a wide range of essential tools for pre-processing, classification, association rules, clustering, regression, and visualisation. Among the various open-source software and DM libraries that have adopted WEKA in recent years are RapidMiner, Knime, and R [20].

Result and Discussion:

Case studies based on three Ecuadorian engineering school transcripts demonstrate the above processes in great detail.

Using a query from the university's academic database, an Excel file of the academic record is to be analysed (March 2016 - July 2016). Information about pupils' academic advancement is included in the file, which is referred to as progress.In addition, the number of times the subject has been seen by the student and the name of the teacher responsible are described. For the study, network and telecommunications engineers as well as those in the fields of computer science and information systems engineering were included in the findings.Table I provides a breakdown of the gathered data.

Table 1 summarizes the records in the files.

Description	Quantity
Number of careers	3
Number of records	3743
Number of students	662
Number of subjects	96
Number of teachers	170

Table II then displays the file's properties along with a brief explanation.

Data File Attribute Descriptions in Table II

International Journal for Multidisciplinary Research
International Conference on Multidisciplinary Research & Studies 2023
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Attributes	Description
Ano_lectivo	Year of the academic calendar.
Periodo	The academic year has two periods
Cod_carrera	Career code.
Matricula	Academic code of the student.
Cod_asignatura	Code of the subject.
Nom_asignara	Name of subject.
Nom_docente	Name of teacher.
Paralelo	Number of courses in each subject.
Nota_progreso_1 Nota_progreso_2 Nota_examen_final	Notes of the three evaluation moments. The maximum evaluation grade of each component is 10.
Nota_examen_final	Final grade obtained in the course. The minimum passing grade is 6
Situacion	Discretized variable: 1 = approve; 0 = does not approve
Repitencia	Number of repetitions of the subject

e

9 records (0.2 % of the total file) were found to be incomplete because of the mining methods used in pre-processing data (missing values). It was decided to eliminate the data set after conducting an effect analysis.In this phase, one of the most significant tasks is to figure out what kind of data each attribute contains and what function it serves in the DM process. Initially, all variables are recognized numerically by the tools utilized.Because of this, the analyst must manually configure these variables according to their needs. As a last step and with the help of the tools discussed, the user can evaluate the basic statistics and also view other graphical alternatives. The visual representation aids in the verification of the facts.

Algorithm execution utilizing each tool is depicted in Figs. 1, 2 and 3. An operator was used in RapidMiner to analyse the performance of each cluster's centroids. However, in Knime, an operator was used to implement a column filter, which was designed to work only with integer variables.For ScatterPlot style charts, Knime also has operators for determining the colour of points in each cluster.

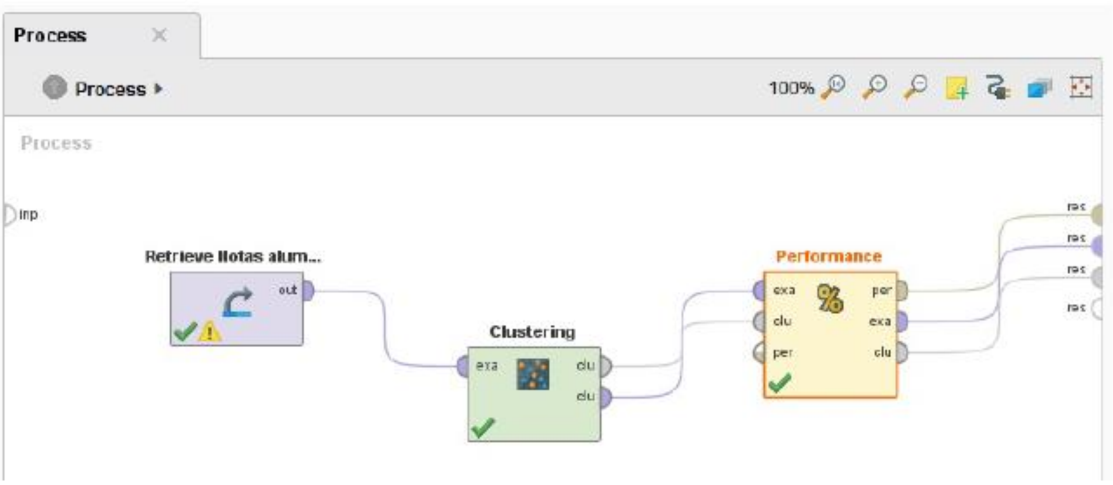

Figure 1 shows the RapidMiner process for running the K-means algorithm.

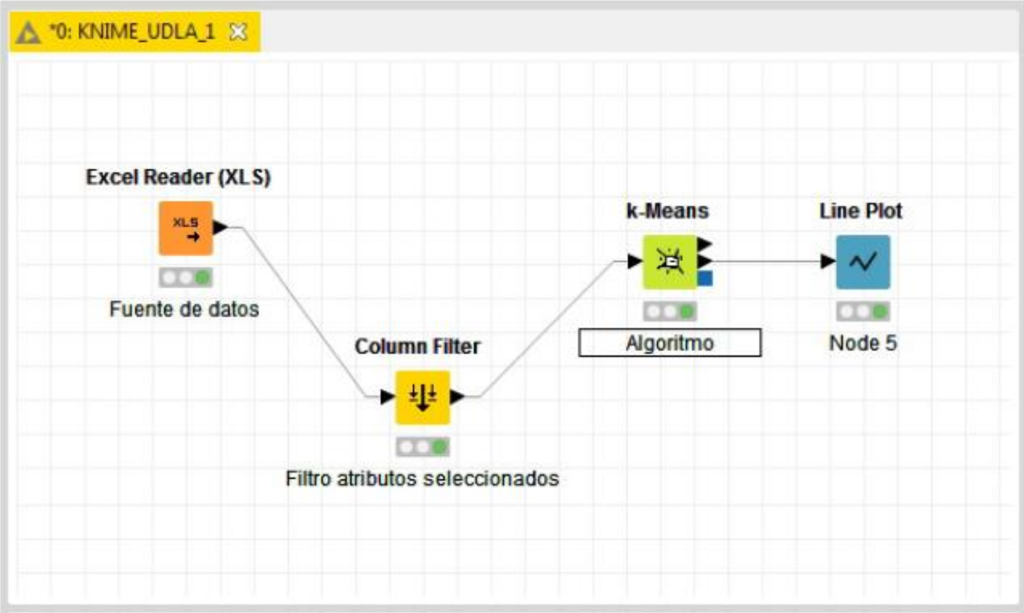

Figure 2 shows the steps involved in running Knime's K-means algorithm.

```
kMeans
======

Number of iterations: 4
Within cluster sum of squared errors: 14414.0
Missing values globally replaced with mean/mode

Cluster centroids:
                                  Cluster#
Attribute            Full Data          0          1          2          3
                        (3743)     (1760)      (855)     (1010)      (118)
==========================================================================
Cod_carrera                651        651        652        511        652
Nota_progreso_1            6.5        6.5        6.2          7          1
Nota_progreso_2              7        7.8        7.1          7          1
Nota_examen_final            1        6.9        6.5          1          1
Nota_total                 6.1        7.2        6.6        6.1          1




Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       1760 ( 47%)
1        855 ( 23%)
2       1010 ( 27%)
3        118 (  3%)
```

Weka's K-means algorithm is shown in Figure. 3.

Conclusion:

Despite its youth, educational data mining holds great promise for all those involved in the educational process. Data mining techniques can unearth patterns and insights that were previously unknowable. In educational data mining, it is possible to classify and predict students and teachers' performance levels. It can assist teachers in monitoring student progress and making instructional improvements, as well as students in making more informed decisions about which courses to take and how their educational experience should be managed.

Using educational data mining, universities can attract and retain students in order to be profitable. In order to find, detect, and comprehend which educational approaches are effective, it is critical to analyse students' data.The benefits and uses of data mining techniques in a variety of educational settings were explored in this paper. There are many educational data mining applications out there, but this study aims to show how powerful they are and encourage others to take advantage of them.

References:

1. Algarni, “Data mining in education,” International Journal of Advanced Computer Science and Applications, vol. 7, no. 6, pp. 456– 461, 2016.
2. S. Ray and M. Saeed, “Applications of educational data mining and learning analytics tools in handling big data in higher education,” in Applications of Big Data Analytics, pp. 135–160, Springer, 2018.
3. V. Manjarres, L. G. M. Sandoval, and M. S. Su´arez, “Data mining techniques applied in educational environments: Literature review,” Digital Education Review, pp. 235–266, 2018.
4. Romero and S. Ventura, “Educational data mining: a review of the state of the art,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, 2010.
5. Bharati M. Ramageri, “Data Mining Techniques and Application”, Indian Journal of Computer Science and Engineering, Vol. 1 No. 4; pp. 301-305.
6. S. Slater, et al., “Tools for educational data mining a review.,” Journal of Educational and Behavioral Statistics, pp. 1-20, 2016.
7. F. Castro, et al., “Applying data mining techniques to e-learning problems.,” In Evolution of teaching and learning paradigms in intelligent environment. Ed. Springer Link, Berlin Heidelberg, pp. 183-221, 2007.
8. Romero, C., Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40(6), 601-618. Crossref.
9. Romero, C., Ventura, S. (2007). Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications. Vol. 33, pp. 135-146. Crossref.
10. Al-Razgan, M., Al-Khalifa, A. S., Al-Khalifa, H. S. (2014). Educational data mining: A systematic review of the published literature 2006-2013. In Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013) (711-719). Singapore: Springer. Crossref.
11. Ali, M. M. (2013). Role of data mining in education sector. International Journal of Computer Science and Mobile Computing, 2(4), 374-383.
12. Baradwaj, B. K., Pal, S. (2012). Mining educational data to analyze students' performance. Retrived 25.5.2019. from https://arxiv.org/pdf/1201.3417.pdf.
13. Luan, J. (2002). Data mining and its applications in higher education. New directions for institutional research, 2002(113), 17-36. Crossref
14. Kovacic, Z. (2012). Predicting student success by mining enrolment data. Research in Higher Education, 15.
15. Kabakchieva, D. (2012). Student performance prediction by using data mining classification algorithms. International journal of computer science and management research, 1(4), 686-690.

16. Maqsood, A. M. (2013). Customer Relationship Management in B-Schools: An Overview. International Journal of Computer Sciences and Management Research, 2(4), 2108-2119.
17. C. Márquez, C. Romero.andS.Ventura, "Predicción del fracas escolar mediantetécnicas de minería de datos," IEEE RevistaIberoamericana de Tecnologias del Aprendizaje, vol. 7, nº 3, pp. 109-117, 2012.
18. S. Slater, et al., "Tools for educational data mining a review.," Journal of Educational and Behavioral Statistics, pp. 1-20, 2016.
19. A. Lausc, A. Schmidt and L. Tischendorf, "Data mining and linked open data–New perspectives for data analysis in environmental research," Ecological Modelling, vol. 295, pp. 5-17, 2015.
20. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The WEKA data mining software: an update," ACM SIGKDD explorations newsletter, vol11 (1), pp.10-18, 2009.