# Smart Mechanism for Assessment of Basic Schooling Systems in Developing Countries Using Data Mining Models

M. Zulqarnain Siddiqui[1], M. Zamin Ali Khan[2], Kamran Khan[3], Abdul Jameel Khan[4] and Atia Elahi[5]

## Abstract

*The research highlights an aspect of the current education system, aiming to review and address criticized educational difficulties while proposing resolutions. The study underlines the significant relationship between action and development, highlighting the diverse school policies derived from different time intervals, regimes, and political objectives to cope with educational challenges. The main findings are related to data mining strategies in the school environment using information technology on a college dataset. The study explores the demonstration of useful latent performance metrics and detects extreme, significant, and insightful classes from unstructured information gathered during the monitoring process. The solutions are presented to enhance the expertise of current schooling infrastructure by providing immediate guidelines for policymakers.*

**Keywords:** Data Mining, Education, Institutional Evaluation, Data Accuracy.

## Introduction

As living in a growing world of knowledge and technology, we know that today education is the most essential key in a growing world, but nowadays, it's very challenging for parents in developing countries to select an educational institution for primary education for their children. Some school fees are costly, some schools don't support extracurricular activities, and some have other problems. These are the only problems private institutions face, but what about public schools/institutions? They have multiple issues, such as curriculum, unsuitable facilities, etc. Through this, the parents can easily select the most suitable school for their child. Parents can easily get an idea about their particular school online, and they don't need to visit. After all, the child will also be motivated and convinced to attend school regularly. In this research, we tried to provide or create a better mechanism for parents to choose an institution. Now, the definition of a good school is not just to look at the top scorers' merit results but also to view the history of those graduates who were best at extracurricular activities (Ramageri, 2010).

According to the report Published in Dawn (a renowned newspaper), the literacy rate has not improved over time. It stated in Dawn newspaper (2018) "Despite making significant improvements in the educational budget, the incumbent PML-N government could not improve the literacy rate in the country. This year, the literacy rate remained at 58pc, the same as last year." It further explained the unfortunate scenario regarding the apparent unsuccessful situation of the country by stating,

[1]*Department of Computer Science, Iqra University Karachi, Pakistan. Email: zulqarnain@iqra.edu.pk*
[2]*Department of Computer Science, Iqra University Karachi, Pakistan. Email: muhammad.zamin@iqra.edu.pk*
[3]*Department of Computer Science, Iqra University Karachi, Pakistan. Email: kamrankhan@iqra.edu.pk*
[4]*Department of Computer Science, Iqra University Karachi, Pakistan. Email: a.jamil@iqra.edu.pk*
[5]*Department of Business Administration, Iqra University Karachi, Pakistan. Email: atia.elahi@iqra.edu.pk*

Last year, the literacy rate saw an annual drop of 2pc from 60pc to 58pc, and this year, there is no change in the rate, according to the recently released Economic Survey 2017-18. The survey says that the overall education condition is based on key performance indicators such as enrolment rates, number of institutes and teachers who have experienced minor improvement.

Furthermore, it stated,

The total number of enrolments at the national level during 2016-17 stood at 48.062 million as compared to 46.223 million during 2015-16. This shows a growth of 3.97 pc, estimated to reach 50.426 million in 2017-18. The Economic Survey stated that the total number of teachers during 2016-17 was 1.726 million compared to 1.630 million during the last year, showing an increase of 5.9pc.

A significant factor that makes it difficult for countries like Pakistan to streamline the education policy is a division of education quality at all levels. Pakistan is amongst the most densely populated lands on the list of developing countries. There is no automated system for collecting real-time school data to monitor and assess the progress. Due to the lack of such a system, all the processes are carried out manually and depend on the data the institutions share. Compiling the data and analyzing it is pretty challenging.

Following is the statistical representation of the overall population in terms of literacy.

**Figure 1: Statistical representation of the overall population**



## Literature Review

Data mining techniques are used to extract beneficial know-how from records. The information gained is precious and substantially influences decision-makers. Educational statistics mining (EDM) is a way to acquire helpful statistics that could impact an organization. The boom in using an era in academic structures has brought about massive quantities of pupil statistics. Consequently, it's very critical to use EDM to improve coaching and learning techniques (Tiwari & Pandey, 2012).

EDM allows educators to expect situations of early dropout or less interest in a course, analyze inner factors affecting their performance, and create statistical techniques to expect scholar instructional overall performance. Academic establishments often keep all available data about students on a digital medium. The statistics are saved in databases for processing. These facts may be of many kinds and volumes, from scholar demographics to their instructional achievements (Yağcı, 2022).

Academic statistics mining is printed as a place of scientific research that specializes in growing techniques for developing innovations within particular kinds of information that come from the instructional environment, after which the use of the methods and strategies to discover how college students study and the surroundings in which they stay. Instructional establishments continuously acquire scholarly facts. In the digital era, e-mastering mediums have also helped produce information in massive portions (Rizwan, 2022).

Correctly predicting college students' development and ability at the beginning of their studies is critical to figuring out susceptible students and stopping them from losing out. Instructional records Mining (EDM) is the application of facts mining strategies to this unique type of dataset that comes from academic surroundings to address crucial academic functions (Asad et al., 2022).

EDM has labored to develop techniques for assessing and using extraordinary styles of statistics from mastering environments to better understand students and their studying environments. He tells us about offline training. It's miles of try-and-talk records and skills based on face-to-face verbal exchange as well as bodily and intellectual research on how people analyze, whilst e-gaining knowledge gives virtual schooling, even as a learning management system provides connectivity, coordination, control, and monitoring tools (Akhtar et al., 2019).

There are a total of three steps to explore the knowledge or information from data mining known as exploration, pattern identification, and deployment. First, the data is cleaned and converted into another form to determine the problems; next, they make patterns in similar variables to identify and choose the best prediction. Then, the patterns are deployed for the expected output. The techniques used in data mining algorithms are classification, clustering, and prediction (Ramageri, 2010).

The vast amount of historical data stored in a database system can provide a variety of technical and data mining applications. Creating a data warehouse and a mining model for economic indicators are two of the most critical steps in data mining. The model can perform various analytical procedures such as clustering, association, and anomaly analysis.

The mining model can provide a basis for scientific analysis and knowledge discovery. It can also help predict the future results of various studies. Before you start working on a project, ensure you have at least five paragraphs and over a hundred words in the text box (Chang & Hu, 2022).

The most crucial application that displays records mining is applications on kinetic gene expression facts evaluation and protein motif discovery tasks. Motif utility is used in biology, where large amounts of protein molecule shape, and sizes are examined through co-clustering techniques. The motif recognition problem takes as enters a fixed of recognized styles or capabilities that, in some manner, outline a category of proteins. The aim is then to look unmonitored or supervised for other instances of equal patterns. Many applications depend upon ordered data evaluation and may benefit from such restrained co-clustering methods (Pensa et al., 2010).

Another research mainly focuses on business fields and growing industries through data mining and business intelligence. Embedded Systems are also derived from business intelligence. Many multinational companies use significant data architecture where they need the help of heavy databases to store and extract data. Big data are also used to apply data mining and artificial intelligence algorithms (Yang & Wu, 2006).

Facts mining era has been used for decades in lots of fields such as groups, scientists, and governments. Used to sift via volumes of statistics along with airline passenger journey statistics, the data mining era facilitates the design of efficient item transportation and distribution rules and decreases enterprise prices. Statistics mining inside the telecommunications industry can assist in

understanding the relevant business, discovering telecommunication patterns, locating fraudulent sports, and better using assets. This file presents a famous creation to statistics mining, learning exciting understanding from large quantities of information saved in statistics repositories (Asad et al., 2022).

The studies found that the authors substantiated and built a methodology for an ensemble classification of character students' overall performance and collective overall performance quantification.

We found that the J48 algorithm provided better outcomes, and, in reality, we intend to increase our studies by using one-of-a-kind parameters in an exclusive analytic environment. Comparisons of student effects for non-public and public colleges are not unusual within the economics of education literature.

The two essential analytical questions within the literature are: 1. Does performance in private schools exceed performance in public colleges, all else being the same? And a couple of. If non-public faculties outperform public colleges, is it due to better colleges or better students? Each question is tricky. Due to the fee of personal colleges, wealthier families can better join their kids. Additionally, private faculties can also have stricter academic admissions criteria.

Private higher faculties must provide an equal curriculum to the public sector to confer provincial secondary diplomas. The reality is that personal faculties usually rely upon lesson charges to cover their working prices, which may also result in self-selection problems if wealthier families are in a higher role to pay. In addition to training prices, personal faculty's admission criteria can be extra stringent, which may additionally lead to choice issues (Frenette & Chan, 2015).

The evaluation suggests that there is far a need to enhance first-rate schooling by predicting the instructional overall performance of the students and assisting folks who are within the threat organization, but while the point of interest is on forecasting, neural networks, aid vector machines decision trees is more efficient and deliver extra accurate outcomes. Statistical strategies are to create a version that can correctly expect output values based on available entered facts.

They were designing a version with artificial neural networks on college students' records related to their navigation through the LMS. The consequences showed that demographics and student clickstream sports considerably impacted student overall performance (Ibrahim et al., 2022).

The fine of education is the concern of each kingdom. Schooling performs a critical function in the improvement of a country. One desires some competencies to live to tell the tale in this international. Most effectively, the elite class can come up with the money for such a first-class education for their children. The tradition of setting up "public faculties" and "cadets", as indicated in advance, which emphasize the improvement of management qualities, these "unique establishments" cannot meet the desires of the people and are very steeply priced. Handiest cell people can find the money for their youngsters. So, efforts should be made to increase the quality of average secondary and better secondary schools by improving their physical and material centers and educational surroundings (Amir et al., 2020).

A system that will rank students' performance using data mining techniques will collect data from various colleges. Then, it applies data mining algorithms to the data to find students' progress. Educational data mining is the most advanced data processing technology:
• Predicting academic results before accepting a student.
• It is possible to specify an acceptance criterion.
• The technique used here is known as linear regression.
• Most accurately predicts the future academic year of applicants.

We can use many techniques for data mining; for example, if we want to work on a small group data set, we can use ANN and decision trees (Mengash, 2020).

We observe academic facts mining to explore the connection between students' viewing behaviors in having access to study materials and their overall performance in flipped school rooms. The converted lecture room is a concept wherein students are given materials at home, after which they practice them in school. There are numerous issues with a flipped study room. For example, no longer all college students experience getting to know earlier than magnificence. Secondly, the educational portal isn't utilized by many college students, so to tackle that, Facebook is getting used as a platform to share records with students as it's miles most regularly used platform educational statistics mining includes facts, visualization, category, clustering, associative evaluation, anomaly detection, and text mining (Su & Lai, 2021).

Statistics mining facilitates companies' leveraging their existing reporting abilities to discover and recognize hidden styles in extensive collections. These styles are then included in information mining fashions and used to as it should be predicting the conduct of people. Because of this knowledge, institutions are better prepared to allocate resources and personnel. For instance, information mining can offer a group the knowledge it desires to do so before a pupil leaves, or it could assist an organization in allocating sources more successfully by providing a correct forecast of what number of college students will attend a given route (Arcinas et al., 2021).

This paper's proposed information mining techniques use multidimensional rotation, translation, reflection, and transformation alongside random tuple shuffling and randomized enlargement. These submitted paintings propose perturbation with records mining to address utility problems, usability, privacy, scalability, and efficiency (Haoxiang & Smys, 2021).

## Methodology

In this study, we collect data from different schools and then put that data in our cloud database to store it. We have applied a data mining algorithm to the data to find the school's rating. For data collection, we have identified some factors based on which we will collect and rate data. In the future, we will be deploying this cloud-based model in schools, and after that, real-time data will be entered into the database, and then the data mining operations will be applied to the data.

Selected primary and secondary schools of private and public sectors from Sindh were considered, and a questionnaire was developed specifically for these institutions to gather information. After the questionnaire had been circulated and the responses had been received, the collected data needed to be refined to reduce the margin of error. The resultant data revealed that little attention had been given to the centralized evaluation policy for educational institutions, specifically schools of public and private sectors (Asim et al., 2018). This "attention" was mainly directed towards Higher Education Institutes (HEIs) by the Higher Education Commission of Pakistan. Not only did schools and colleges get entirely overlooked by the system, but the universities that did get all this attention ended up hurting the overall growth of the system. This is because ranking systems started becoming more and more relevant, and the performance indicators used to determine the "quality" of a university, or the ones used to evaluate the university's overall performance, varied from system to system. The world's most renowned ranking systems are arguably the QS Ranking System and the Academic Ranking of World Universities (ARWU) also commonly known as Shanghai Ranking. Each system uses a different set of performance indicators, and their yearly ranking of the top universities is entirely different for the same reason. This has created an anomaly in evaluating these institutes and can often result in students making a somewhat confusing decision. For Pakistan, however, the evaluation systems followed are far worse in performance.

The evaluation is conducted manually, and unlike the QS/THE ranking systems, which use surveys to receive most of their evaluation data from neutral sources, bodies like HEC in Pakistan gather their information through visits to the campus premises and personally evaluate every aspect of the performance indicators. This leaves room for error and dishonesty since the institute being assessed is granted time and knowledge of what is to be evaluated.

This research focused on the evaluation of schools in particular. The identified factors according to which assessments are evaluated are:

## Data Collection Methods

The method we are using to collect data is an interview. We will be interviewing different schools and collecting data from their teachers and admin; an interview is a lengthy process, but it ensures that the data collected is relevant, and the more relevant the data, the more accurate the results. Later on, after the deployment of the model, data will be inserted by the institutions themselves, and then the authenticity will increase more.

### Analysis Method

The analysis method that we will be using in this project is data mining. Data mining is used to find patterns in data, and we will apply it to our database. The data collected and stored in the database will be analyzed, including the type and quantity, and then data mining will be applied to it.

### Why Use Data Mining

In this project, we are using data mining because it is one of the best techniques for analyzing and rating data. As we are working on an educational dataset, educational data mining has proposed many data models and algorithms that we can use and implement quickly in our project compared to other methods.

Factors that contribute in building data models and algorithms are:
- Total no. of students
- Total no. of Classes/ Batches/ Sections
- Total no. Of Qualified Faculty Members
- Faculty Rotation
- (Stability)
- Results
- Grading
- Total no. of classes per week
- Weightage of practical activities
- Economics
- Delivery of lectures
- Automated attendance system
- Total no. of events per year
- Infrastructure
- Lecture courses
- Library book
- Playground

## Description of Factors

**Total no. of students**
This factor is based on the number of students that will be enrolled in a particular school

**Total no. of Classes**
This factor is based on how many numbers of classes are in a particular school including sections and occurrence of classes in a week

**Total no. Of Qualified Faculty Members**
To go through academic skills, check the qualifications of faculty members, and their education and to see if they are eligible for that position or designation.

**Faculty Rotation**
This includes the Stability of the members who are already enrolled in that institution. Might include shifts of evening or mornings which have different slots and faculty members too.

## Results

The institute has grading sessions too, to tell how many marks students have scored. Grading tells which student is achieving more and we got to know which one is the best candidate. The highest marks and grading give benefits to students such as scholarships and concessions, which might help them to work harder.

**Total no. of classes per week**
Every session has a limited time limit means a 12-week session after that they will have their mid-term exam, so class is accordingly set per week to convey lectures. There might be 2 days of the same lecture per week or 3. It depends on the target that the academic teacher has set.

**Weightage of practical activities**
It depends on Economics activities, laboratories, practical, labs, and social works. Delivery of lectures in a way that shows pictures art, graphic designing work, PowerPoint presentations, or introducing it by some innovative and creative activities, will make students learn more easily.

**Automated attendance system**
It includes attendance by machine system, automatic registration and marking of attendance, and advanced technology of AI techniques making our work easier and less time-consuming.

**Total no. of events per year**
Every institution has some criteria which they work on accordingly. Mainly it does have 2 main events and the rest are organized according to the weather and seasons.

**Infrastructure**
It adds lecture courses, and library books which give extra knowledge about our course. We can get references from books related to our subjects. It will help students to get to know more and more about the course that they are learning. Secondly, it includes Playground (1/0) which allows students to play, have karate, race, and sports. After collecting data according to these factors from various schools we will analyze it for the rating and then the algorithm will decide which school

has scored more points and then will rate it to the top similarly the school with fewer points will be at the bottom of the rating.

## Statistics Mining-Knowledge Discovery in Databases (KDD) process

KDD (know-how Discovery in Databases) is a technique that entails extracting useful, formerly unknown, and probably precious facts from huge datasets. The KDD system in statistics mining typically entails the subsequent steps:

1. *Selection:* Choose a relevant subset of records for analysis.
2. *Preprocessing:* Clean and transform the statistics to make it prepared for analysis. This can include tasks together with information normalization, handling missing values, and facts integration.
3. *Transformation:* Remodel the statistics right into a format appropriate for information mining, such as a matrix or graph.
4. *Information Mining:* Apply facts mining techniques and algorithms to records to achieve useful statistics and insights. This will include duties that include clustering, classification, association rule mining, and anomaly detection.
5. *Interpretation:* Interpret the outcomes and extract understanding from the facts. This could include duties such as visualizing consequences, evaluating the fine of determined styles, and figuring out relationships and institutions between records.
6. *Evaluation:* Compare the consequences to ensure that the extracted expertise is useful, accurate, and meaningful.
7. *Deployment:* Use the discovered understanding to clear up an enterprise's trouble and make a choice.

### C4.5 Algorithm

C4.5 builds selection trees from schooling information in the same way as ID3 the use of the idea of statistics entropy. The schooling statistics are fixed on pre-categorized samples S=s1, s2, each pattern SI consists of a p-dimensional vector (x1i, x2i, XPI, wherein XI represents values of attributes or homes of the sample as well as the class it belongs to).

At each node in the tree, C4.5 selects the facts attribute that most efficiently divides its sample set into subsets wealthy in one elegance or some other. The distribution criterion is the normalized statistics benefit (distinction in entropy). The attribute with the highest not unusual statistics is chosen for selection. Algorithm C4.5 is then repeated at the partitioned subsystem.

This algorithm has numerous base cases.

• All samples inside the listing belong to the equal class. While this occurs, it creates a leaf node for the choice tree that says to pick that class.
• Not one of the capabilities offers any data benefit. In this situation, C4.5 creates a selection node higher in the tree for the use of the expected class fee.
• Located an example of a formerly unseen magnificence. Again, C4.5 creates a choice node higher within the tree the use of the expected cost.

It divides the facts set into smaller and finer subsets and is steadily evolving. The cease result is a tree next to decision nodes and leaf nodes. Each decision node has extra divisions, and a leaf node embodies an association or decision. The highest choice node within the tree that corresponds to the excellent predictor is referred to as the unique node.

**J48 Algorithm**

The J48 algorithm is one of the maximum widely used systems studying algorithms to examine the data categorically and continuously. The C4.5 algorithm (J48) is ordinarily used in many fields for classifying facts for example deciphering the clinical facts for the prognosis of coronary heart sickness, classifying E-governance statistics, and plenty more. This algorithm is widely known for its accuracy and simplicity of knowledge for this reason it is utilized in our studies for the rating of colleges.

Attributes like the library and playground, for example, if a school has a playground, we mark it as true and if it doesn't have it so we mark it as false, similarly, we have applied the same approach to all the collected data. Here we can see that the correctly classified instances are 80 percent and the incorrectly classified instances are 20 percent, which means that the results of our prediction are 80/2.

As per concern to our task, we undertake to make available or generate an improved optimal choice for the parents to choice an upright school for their child., having research for a good school might help children to decrease his or her insecurities with their classmates as some children are good with education, some are good with the sports, rare love history, limited love arts, etc., Indorsing a brilliance education system for the forthcoming youth of Pakistan primary to high school is a problem of ultimate importance but, as we comprehend nowadays in Pakistan that the condition of schools and colleges of government sector is underprivileged as equated to private schools and colleges. Our task is based on a portal that will be accessed by operators of all age groups and all subdivisions. In the future, we will be organizing this cloud-based model in schools, and afterward, factual time data will be moved into the database, and then the data mining processes will be applied to the data.

**Figure 2: Represents the results of the J48 algorithm**
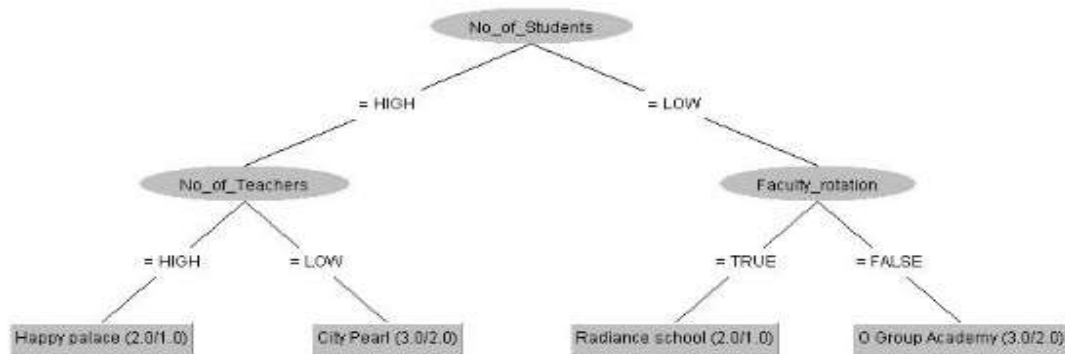
**Figure 3: Represents factors defining the project**

```
=== Run information ===

Scheme:         weka.classifiers.bayes.NaiveBayes
Relation:       Tree
Instances:      10
Attributes:     15
                School_id
                School_Name
                No_of_Students
                No_of_Teachers
                No_Classes
                Faculty_rotation
                Overall_Result
                Classes_per_Week
                Practical_Activities
                Automated_attendence
                Total_Events
                Library_books
                Playground
                Area
                Sugession
Test mode:      10-fold cross-validation
```

## Conclusion

This research defined the mechanism to improve the efficiency of the educational sector by introducing technology in regular activities and processing educational data to get actual trends so that they can be analyzed for improvement of the overall education sector of any third-world country. This study is based on Pakistan's education system, focusing on primary-level education in both the public and private sectors. It includes a selected batch of students from 10 schools. Institutions in Pakistan are administrated either by local government or private parties, hence segmenting them as public or private sector institutions. There are many differences when we compare the overall progress of institutions and student academic performance in both sectors. This study helped highlight their differences and develop the urge for improvement. Automating the whole system regarding data collection helped to accentuate and observe the overall performance of institutions and individual students. From admission to graduation, complete data helped discover the trends towards education.

A system has been designed and deployed at selected institutions to collect the records for a batch of students so that this research can be carried out. Successful implementation started giving the real-time data at a centralized server through which trends have been observed. Student academic and institution performance has been analyzed using statistical modelling and data mining techniques. Factors have been identified, and selected institutions were ranked based on the data collected. Trend analysis and predictions have been made to identify the new admissions and dropout ratios. Overall objectives of the planned study have been achieved, and through results, areas for improvement can be identified. Hence, after technology implementation, the general

assessment of institutions is made easier for the authorities and bodies related to ranking institutions with transparency.

Similar to other third-world countries, the bodies that oversee the education sector in Pakistan need to pay more attention to schools and colleges. The related literature revealed that not only in Pakistan but worldwide, little attention had been given to educational institutions that did not fall under the category.

This study offers beneficial latent performance metrics after using statistics technology on student-centric datasets. We present our vital outcomes concerning the prospective of data mining strategies in an educational setup. This venture pointed in the direction of the feature by using the existing education cataloguing. To stabilize the instructional impediments, many academic guidelines can be implemented.

## References

- Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big data-savvy teams' skills, big data-driven actions, and business performance. *British Journal of Management, 30*(2), 252-271.
- Amir, S., Sharf, N., & Khan, R. A. (2020). Pakistan's education system: An analysis of education policies and drawbacks. *Electronic Research Journal of Social Sciences and Humanities, 2.*
- Arcinas, M. M., Sajja, G. S., Asif, S., Gour, S., Okoronkwo, E., & Naved, M. (2021). Role of data mining in education for improving students' performance for social change. *Turkish Journal of Physiotherapy and Rehabilitation, 32*(3), 6519-6526.
- Asad, R., Arooj, S., & Rehman, S. U. (2022). Study of educational data mining approaches for student performance analysis. *Technical Journal, 27*(1), 68-81.
- Asim, M., Shamshad, F., Awais, M., & Ahmed, A. (2017). *Introducing Data mining for Predicting trends in School Education of Pakistan: Preliminary results and future directions*. In Proceedings of the Ninth International Conference on Information and Communication Technologies and Development (pp. 1-5).
- Chang, Q., & Hu, J. (2022). Research and Application of the Data Mining Technology in Economic Intelligence System. *Computational Intelligence and Neuroscience*, Doi: https://doi.org/10.1155/2022/6439315
- Frenette, M., & Chan, P. C. W. (2015). *Academic Outcomes of Public and Private High School Students: What Lies behind the Differences*? Analytical Studies Branch Research Paper Series. Statistics Canada. 150 Tunney's Pasture Driveway, Ottawa, ON K1A 0T6, Canada. https://files.eric.ed.gov/fulltext/ED585228.pdf.
- Haoxiang, W., & Smys, S. (2021). Big data analysis and perturbation using data mining algorithm. *Journal of Soft Computing Paradigm, 3*(1), 19-28. DOI: https://doi.org/10.36548/jscp.2021.1.003
- Ibrahim, W., Abdullaev, S., Alkattan, H., Adelaja, O. A., & Subhi, A. A. (2022). Development of a model using data mining technique to test, predict and obtain knowledge from the academics results of information technology students. *Data, 7*(5), 67.
- Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *Ieee Access, 8,* 55462-55470.
- *No improvement in literacy rate*, (2018). The Dawn. https://www.dawn.com/news/1404082

- Pensa, R. G., Boulicaut, J. F., Cordero, F., & Atzori, M. (2010). Co-clustering numerical data under user-defined constraints. Statistical Analysis and Data Mining. *The ASA Data Science Journal, 3*(1), 38-55. Doi: http://doi.wiley.com/10.1002/sam.10064
- Ramageri, B. M. (2010). Data mining techniques and applications. *Indian journal of computer science and engineering, 1*(4), 301-305.
- Rizwan, M., Kiyoshi, T., & Rie, H. (2022). *Access Challenges to Education in Pakistan.* Asian Development Bank.
- Tiwari, R., & Pandey, P. (2012). Pattern Analysis in Education Data Mining. *International Journal of e-education, 2*(97).
- Su, Y. S., & Lai, C. F. (2021). Applying educational data mining to explore viewing behaviors and performance with flipped classrooms on the social media platform Facebook. *Frontiers in Psychology, 12*, 653018. Doi: 10.3389/fpsyg.2021.653018
- Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making, 5*(4), 597-604.
- Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1), 11.