

# The Impact of Machine Learning On Educational Institutions: An Empirical Study

**Malik Jawarneh**

Faculty of Computing Sciences  
Gulf College, Sultanate of Oman

**Marwan Alshare**

Faculty of Computing Sciences  
Gulf College, Sultanate of Oman

**Qusay Bsoul**

Faculty of Science and Information Technology  
Irbid National University, Irbid-Jordan

**Haydar S Kalash**

Faculty of Computing Sciences  
Gulf College, Sultanate of Oman

*Abstract* - To be successful in today's competitive world, an institute must be able to predict student performance, classify people according to their abilities, and work to enhance that performance in future assessments. Prior to starting a course, students should be told to concentrate their attention on one specific subject area. An institute may be able to lower its failure rates with the support of research like this. Based on their past performance in similar courses, this research predicts students' success in a given course. A strategy to finding hidden patterns in large amounts of existing data known as machine learning is involved. These patterns could be beneficial for research and forecasting purposes in the future. In the realm of education, education data mining is a collection of data mining applications that are put to use. Data from students and teachers is analyzed by these applications. The results of the study could be utilized to make classifications or forecasting of performance. Researchers are investigating machine learning techniques such as C4.5, Naive Bayes, and SVM. UCI machinery student performance data is used in the experiment. The accuracy and error rate of algorithms are used in the evaluation of the effectiveness of the algorithm

*Keywords* - Learning Analytics, Deep Learning, Educational Data Mining, Machine Learning, Student Performance, Classification, Prediction

## 1. INTRODUCTION

Globalization has offered a massive platform for education and learning for professors and students alike. Increased competition among educational institutions and the emergence of for-profit schools have compelled public schools to reevaluate their mission and goals, which will benefit them in the long run. This has resulted in a praiseworthy line of action. The authorities have drafted a plan of action to ensure that children receive a high-quality education and that teachers play a significant role in their academic careers. There's been a huge need for the function of teacher and the interaction between teacher and student for as long as anybody can remember. [1]

Many colleges and universities are searching for hardworking students who will help them achieve their goals and portray their institutions in the best possible light. As a result of these changes in education, the next generation will be able to grow up on a higher level of education. It will also put the school system on the international stage. Detecting and preventing underachievement in students can have a long-term benefit. Students and mentors can work together to enhance academic achievement in the future because of early diagnosis. In order to help kids succeed, mentors must take the essential and critical actions. Mentors and students, therefore, endeavor to increase the value of education in a variety of educational environments. [2]

In order to provide a well-rounded education for pupils, it is essential to pay attention to their educational history. As a result, students' futures will be influenced as well as the standing of the stakeholders. The teaching method, however, is not the only factor that needs to be addressed when it comes to a student's performance. Students who don't do well in a course may be required to repeat it or given an assignment to focus on their weak areas if early academic performance prediction is used as an indicator for their future results. For the instructor, it is critical that he or she devise a strategy based on the findings of pupils. The teacher might devise corrective measures, give more work, or hold final sessions to help the student overcome their deficiencies. Students that need more attention may be inspired to study as a result of this, as well. They might study for competitive tests depending on their improved performance abilities for the kids who scored higher. [3]

Students at colleges and universities are evaluated by taking mid-term exams and taking additional tests to enhance their grades. Students who need additional attention and care to improve their academic performance should be given specific assignments and quizzes and other learning activities on a regular basis, notwithstanding all of the above-mentioned measures. I believe that knowledge is a tool to help people improve their lives, and so does Socrates. Educational institutions are tasked with creating, acquiring, and disseminating knowledge in a competitive setting. An educational institute's goal is to reduce the boundaries of teaching and learning by using data mining techniques to a data collection.

Prediction, Classification, Regression, Machine Learning, Neural Network, Association Rule Mining, and Text Mining are some of the approaches used to improve educational data. Students' academic achievement can be seen in a new light thanks to data mining tools. Educational institutions are gathering a vast amount of student data for the goal of investigating the teacher-student interaction and student performance. Researcher read and reread the data acquired in order to organize and control the educational process. As a result, much of the accessible data is utilized to create reports and queries that are easy to interpret but seldom reach those in need. Many times, school administrators are unable to process significant amounts of data because of the complexity and volume of the information. In order to promote early detection and prevention, learning algorithms can effectively translate available data into information. Prerequisites may be identified early on, resulting in a clear and meaningful data set. A teacher's performance is judged by the quality of his or her pupils, who are the ones who establish the basis for a successful future. Student-at-risk evaluation and quantification are of the highest significance for the student's better academic performance. Incorporating performance measurements gives the researcher insight into the academic achievement of the students as well as their capacity to meet predetermined goals. [4]

## 2. RELATED WORK

New types of information systems constructed with the aid of soft computing are known as computationally intelligent systems. Intelligent machines that can solve difficult real-world problems that can't be analyzed analytically are another objective. It uses tolerance for approximation, uncertainty, imprecision, and partial reality in order to achieve a likeness to human-like decision making [5].

You've found an intelligent system if you're seeking for anything that resembles nature's wisdom (IS). Awareness, adaptability, logic and the ability to handle complex information are examples of these [6]. This section provides an overview of soft computing methods used in intrusion detection. Soft computing approaches like fuzzy logic and neural networks, rather than hard computational methods like genetic algorithms, are used to identify and react to foreign invaders in an artificial immune system. Interference detection using Coupled Map Lattices is still an ongoing project.

Holland developed a flexible search technique using his genetic algorithm. Simulated evolution mimics the natural process on a computer. To identify better and better answers among a population of alternatives, one may utilize a stochastic global search approach called the GA.

New solutions are generated each generation by selecting people who have a high degree of expertise in the problem area and then having children from those individuals.. It's possible to create individuals who are more suited to a certain issue area than the ones they originated from. The fitness function [7] provides an indication of how people behaved in the problem area.

It was invented in 1995 [8] and was inspired by the social behavior of flocks of birds and schools of fish. PSO's origins may be traced back to subjects like as computer science, artificial life, psychology, and physics. Hyperspace is traversed by a "population" of particles that move at predetermined speeds. For the particle velocities, the best location of a specific particle and the best position in its immediate surroundings are employed in stochastic calibration.

The particle's best and the neighborhood's best are determined by a user-defined fitness function [9]. Each atom's movement results in a perfect or near-ideal outcome. A swarm of insects, rather than a flock of birds or a school of fish, characterizes the seemingly random movement of particles in the problem room.

An approximation-based logic known as fuzzy logic (FL) is used in problem solving and reasoning. Fuzzy logic has a number of procedures that are comparable to classical logic (two-valued or "crisp" logic). Evolutionary computing uses biological genetics and natural selection to optimize and classify machine learning algorithms. Many different topics are included in the term "evolutionary computation," such as genetic algorithms and evolutionary programming. For the most part, genetic algorithms are the most often used algorithms in applications.

The secret naive Bayes (HNB) classifier is an enhanced version of the naive Bayesian classifier. In the HNB paradigm, a new layer is created to specify each attribute's hidden parent. Naive Bayes is used to generate HNB's structural properties. In order to unite the forces of all other traits, a hidden parent is formed for each attribute. Averaging weighted one-dependence estimators [11] may be used to identify hidden parents.

## 3. METHODOLOGY AND RESULTS

Methodology consists of three machine learning algorithm. These algorithms are C4.5, Naive Bayes, and SVM Ross Quinlan created this technique. Using this method, the provided dataset is divided into training and testing data sets. In the beginning, a classification model is created to classify the data. From the root node to the leaf nodes it uses a top-down approach to classify. Data is divided into subsets based on the entropy measurements, and an efficient attribute is selected for each subset. Attribute with maximum information gain is utilized to make decisions. It is efficient because it removes unnecessary properties from the branches. It's also good at handling characteristics with numeric values and missing data. However, when dealing with numeric qualities, the decision tree gets more difficult. Naïve bayes classifier is also prominent method to predict student performance[12] [13]

Non-linear support vector machines (SVMs) are used to construct an effective recommendation system [14]. Non-linear support vector machine approaches are the most often used methodology for dealing with unlabeled data. Additionally, they're employed in a variety of industrial settings. Hyperplanes for each set of data and training samples are generated by this algorithm. In this approach, the new instances of the input data model are categorized further. The hyperplane separates one part of the hyperplane from the other. Divided into two groups, the classes sit on each side of the partitions. To put it simply, SVM divides classes. The target classes may now be neatly divided into subgroups using a hyperplane. When selecting a hyperplane, it is important that it accurately identifies the target class. The appropriate hyperplane depends on the circumstances. If there are any target classes, three hyperplanes are needed to correctly identify them. To describe this condition, "distance margin" is the phrase of reference. Changing the margin distance measurements depending on the situation yields better results. It is possible to employ SVM classification algorithms like the maximum margin classifier and the soft margin classifier that make use of measurements of margin distance. It makes advantage of this to categorize things more quickly. Most efficient classifiers in a range of disciplines are support vector machines (SVM). It can also be used to diagnose and treat cardiovascular disease. With a modest margin of separation between the two objects, this method works effectively in high-dimensional situations. Aside from samples, it works well with a wide range of other dimensions as well another drawback of this approach is that it requires a lot of time and memory to run. As a result, the performance metrics may suffer while dealing with noisy and confusing data. SVM, despite its flaws, offers a number of valuable applications because of its appropriate and cost-effective categorization.

#### 4. RESULTS AND DISCUSSION

UCI machinery student performance data set [15] is used as input to classifiers. This data set consists of 33 attributes and 649 instances. The result of different classifiers are shown below in figure 1

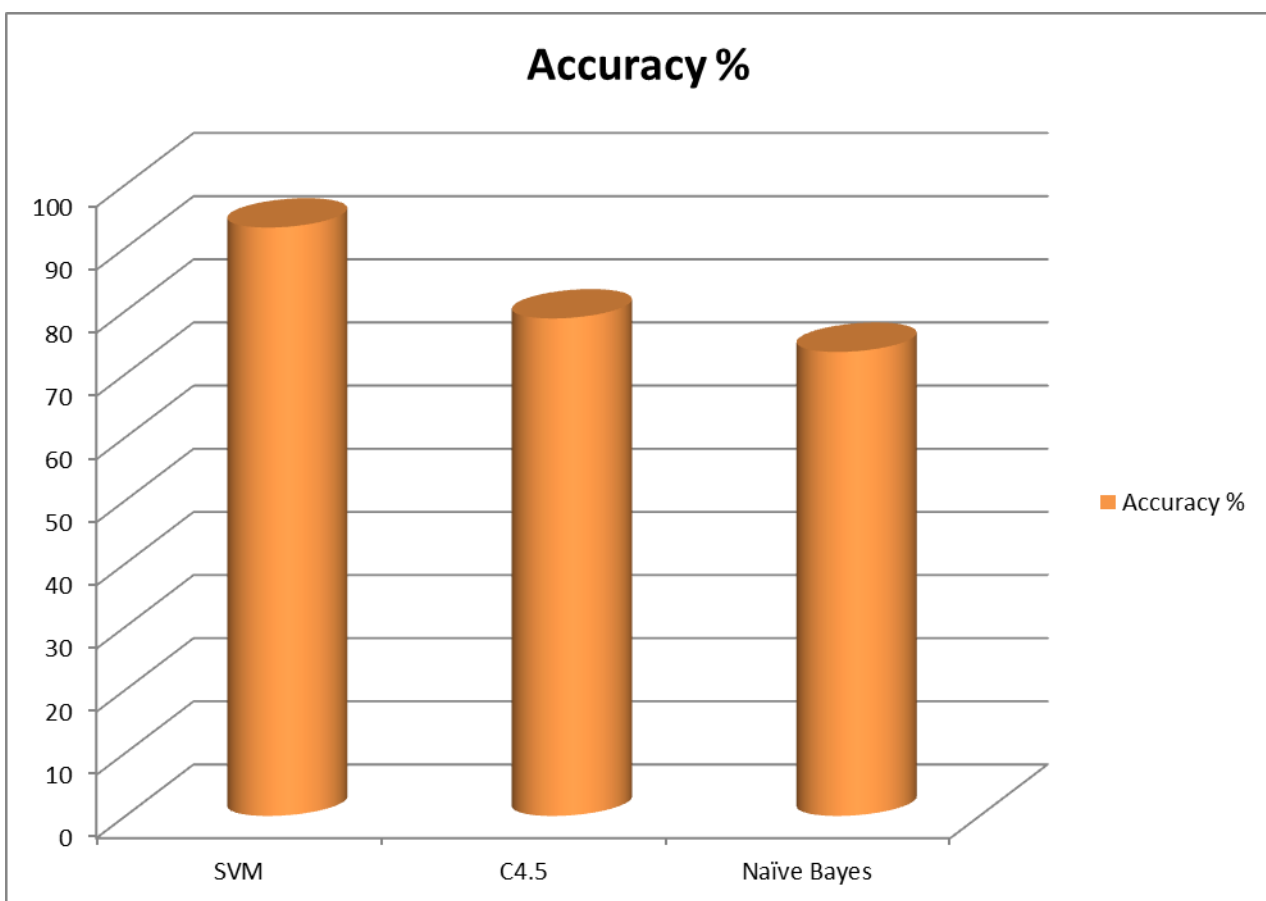


Figure 1

Student Performance Data Classification Results

#### CONCLUSION

In today's competitive world, an institute must be able to forecast student performance, classify people based on their talents, and seek to improve that performance in future exams in order to be successful. Prior to beginning a course, students should be instructed to focus their attention on one single topic area. With the use of studies like this, an institute may be able to reduce its failure rates. This study predicts students' success in a specific course based on their previous achievement in similar courses. Machine learning, an approach for discovering hidden patterns in massive volumes of existing data, is involved. These patterns may be useful in the future for study and predictions. Machine learning approaches such as C4.5, Naive Bayes, and SVM are being investigated by researchers. The experiment makes advantage of UCI equipment student performance data. Accuracy of SVM classifier is better for classifying student data.

## REFERENCES

- [1] R. Alshabandar, A. Hussain, R. Keight and W. Khan, "Students Performance Prediction in Online Courses Using Machine Learning Algorithms," 2020 International Joint Conference on Neural Networks (IJCNN), 2020, pp. 1-7, doi: 10.1109/IJCNN48605.2020.9207196.
- [2] S. Jayaprakash, S. Krishnan and V. Jaiganesh, "Predicting Students Academic Performance using an Improved Random Forest Classifier," 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), 2020, pp. 238-243, doi: 10.1109/ESCI48226.2020.9167547.
- [3] B. Sravani and M. M. Bala, "Prediction of Student Performance Using Linear Regression," 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154067.
- [4] M. Chitti, P. Chitti and M. Jayabalan, "Need for Interpretable Student Performance Prediction," 2020 13th International Conference on Developments in eSystems Engineering (DeSE), 2020, pp. 269-272, doi: 10.1109/DeSE51703.2020.9450735.
- [5] A. Patel and A. Jain, "A study of various Black Hole Attack techniques and IDS in MANET," International Journal of Advanced Computer Technology, vol. 4, no. 3, pp. 58–62, 2015.
- [6] J. Visumathi and K. L. Shunmuganathan: A computational intelligence for evaluation of intrusion detection system, Indian Journal of Science and Technology, Vol. 4 No. 1, Jan 2011.
- [7] V. Jain and M. Agrawal, "Applying Genetic Algorithm in Intrusion Detection System of IoT Applications," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 284-287, doi: 10.1109/ICOEI48184.2020.9143019.
- [8] Kunhare, N., Tiwari, R. & Dhar, J. Particle swarm optimization and feature selection for intrusion detection system. *Sādhanā* 45, 109 (2020). <https://doi.org/10.1007/s12046-020-1308-5>
- [9] Win, T.Z.; Kham, N.S.M. Information Gain Measured Feature Selection to Reduce High Dimensional Data. In Proceedings of the 17th International Conference on Computer Applications (ICCA 2019), Novotel hotel, Yangon, Myanmar, 27 February–1 March 2019; pp. 68–73
- [10] A. Chaudhary, V. N. Tiwari and A. Kumar: Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Ad Hoc Networks, International Journal of Information Technology, Vol. 6, No. 1, June 2014.
- [11] A. Chaudhary, V. N. Tiwari and A. Kumar: Analysis of Fuzzy Logic Based Intrusion Detection Systems in Mobile Ad Hoc Networks, International Journal of Information Technology, Vol. 6, No. 1, June 2014.
- [12] L. I. P. Aji and A. Sunyoto, "An Implementation of C4.5 Classification Algorithm to Analyze student's Performance," 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020, pp. 126-130, doi: 10.1109/ICOIACT50329.2020.9332088.
- [13] F. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 301-306, doi: 10.1109/CSCI46756.2018.00065.
- [14] M. A. Cano Lengua and E. A. Papa Quiroz, "A Systematic Literature Review on Support Vector Machines Applied to Classification," 2020 IEEE Engineering International Research Conference (EIRCON), 2020, pp. 1-4, doi: 10.1109/EIRCON51178.2020.9254028.
- [15] J. Moon, J. Park, E. Hwang & S. Jun, (2018). Forecasting power consumption for higher educational institutions based on machine learning. *The Journal of Supercomputing*, 74(8), 3778-3800.
- [16] Mustafa, M., & Al-Badi, A. (2022). Role of Internet of Things (IoT) Increasing Quality Implementation in Oman Hospitals During COVID-19. *ECS Transactions*, 107(1), 2229.
- [17] Alshar'e, M., Albadi, A., Mustafa, M., Tahir, N., & Al Amri, M. (2022). Hybrid User Evaluation Methodology for Remote Evaluation: Case study of Educational games for children during Covid-19 Pandemic. *Journal of Positive School Psychology*, 6(3), 3049-3063.
- [18] Alshar'e, M., Albadi, A., Mustafa, M., Tahir, N., & Al Amri, M. (2022). A Framework of the Training Module for Untrained Observers in Usability Evaluation Motivated by COVID-19: Enhancing the Validity of Usability Evaluation for Children's Educational Games. *Advances in Human-Computer Interaction*, 2022.
- [19] Arumugam, K., Swathi, Y., Sanchez, D. T., Mustafa, M., Phoemchalard, C., Phasinam, K., & Okoronkwo, E. (2022). Towards applicability of machine learning techniques in agriculture and energy sector. *Materials Today: Proceedings*, 51, 2260-2263.
- [20] Sajja, G. S., Mustafa, M., Phasinam, K., Kaliyaperumal, K., Ventayen, R. J. M., & Kassanuk, T. (2021, August). Towards application of machine learning in classification and prediction of heart disease. In 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC) (pp. 1664-1669). IEEE.
- [21] Kassanuk, T., Mustafa, M., & Panse, P. (2021). An Internet of Things and Cloud Based Smart Irrigation System. *Annals of the Romanian Society for Cell Biology*, 20010-20016.
- [22] Bsoul, Q., Abdul Salam, R., Atwan, J., & Jawarneh, M. (2021). Arabic Text Clustering Methods and Suggested Solutions for Theme-Based Quran Clustering: Analysis of Literature. *Journal of Information Science Theory and Practice*, 9(4), 15-34.

- [23] Mustafa, M., Alshar'e, M., Shariah, A., Al-Alawi, M., & Mohammad, A. (2021). Managing and analyzing factors influencing Saudi college students' entrepreneurial intention during the Covid-19 pandemic. *Turkish Journal of Physiotherapy and Rehabilitation*, 7486-7496.
- [24] Bholra, J., Jeet, R., Jawarneh, M. M. M., & Pattekari, S. A. (2021). Machine Learning Techniques for Analysing and Identifying Autism Spectrum Disorder. In *Artificial Intelligence for Accurate Analysis and Detection of Autism Spectrum Disorder* (pp. 69-81). IGI Global.
- [25] Shabaz, M., Singla, P., Jawarneh, M. M. M., & Qureshi, H. M. (2021). A Novel Automated Approach for Deep Learning on Stereotypical Autistic Motor Movements. In *Artificial Intelligence for Accurate Analysis and Detection of Autism Spectrum Disorder* (pp. 54-68). IGI Global.
- [26] Pallathadka, H., Mustafa, M., Sanchez, D. T., Sajja, G. S., Gour, S., & Naved, M. (2021). Impact of machine learning on management, healthcare and agriculture. *Materials Today: Proceedings*.
- [27] Sajja, G. S., Mustafa, M., Ponnusamy, R., & Abdufattokhov, S. (2021). Machine Learning Algorithms in Intrusion Detection and Classification. *Annals of the Romanian Society for Cell Biology*, 25(6), 12211-12219.
- [28] Mustafa, M., Virmani, D., Kaliyaperumal, K., Phasinam, K., & Santosh, T. (2021). Towards Investigation of Various Security And Privacy Issues In Internet Of Things. *Design Engineering*, 1747-1758.
- [29] Kollu, P. K. (2021). Blockchain Techniques for Secure Storage of Data in Cloud Environment. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(11), 1515-1522.
- [30] Mustafa, M., Alzubi, F. K., & Bashayreh, A. (2021). Factors Affecting Job Performance of Teaching and Non-Teaching Staff in Higher Education Levels in Oman. *Ilkogretim Online*, 20(5).
- [31] Alkhatib, K., Al-Aiad, A., Mustafa, M., & Alzubi, S. (2021). Impact factors affecting entrepreneurial intention of Jordanian private universities students: a mediation analysis of perception toward entrepreneurship. In *Sustainable and Energy Efficient Computing Paradigms for Society* (pp. 53-65). Springer, Cham.
- [32] Brahmi, B., & Mustafa, M. (2019). Impact of Knowledge Management Process on Managerial Performance in the High Tech Sector. *International Journal of Business and Management*, 14(2).