The OCEM Journal of
Management, Technology & Social Sciences

Recieved: Nov 06, 2024 | Revised: Nov 17, 2024 | Accepted: Nov 26, 2024 | Published: Jan 20, 2025

DOI: https://doi.org/10.3126/ocemjmtss.v4i1.74754 **Paper Type: Research Article**

Impact of Algorithms and Big Data on Educational Field

Sita Devi Dhakal
Assistant Professor, Department of Computer Application, Oxford College of Engineering and Management,Gaindakot, Nepal
Corresponding e-mail: sitadevimca@gmail.com

Volume 4, Issue 1

ISSN Print:2705-4845

ISSN Online:2705-4845

Abstract

Using the potential of big data in education is the best option for students and institutions to effectively anticipate student data in this competitive environment. By analyzing this data, educators can identify students at risk of academic underperformance and provide targeted interventions to improve their outcomes. Techniques such as clustering and classification are pivotal in segmenting students based on criteria like academic achievements, interests, and socio-economic backgrounds. This allows institutions to effectively tailor teaching methods and resources for diverse student needs, which also investigates the challenges associated with big data analytics and machine learning algorithms. Issues like data quality management, the complexity of integration from various sources, and resource allocation are explored. Proposed solutions include organizing advanced training programs, leveraging expert consultations, and employing robust data models to address these obstacles efficiently. The study conducts a comparative analysis of algorithms used in educational settings, particularly linear regression and logistic regression, to forecast student performance. Logistic regression emerged as more accurate, demonstrating its effectiveness in predicting grades and identifying dropout risks. The methodology used offline and online datasets from Oxford College of Engineering and Management, Gaindakot-2, Nawalparasi to analyze internal marks and predict semester outcomes, highlighting the transformative role of machine learning in education.

Algorithms and big data are revolutionizing education by enabling personalized learning, improving decision-making, and optimizing resource allocation. They help track student progress, predict outcomes, and identify areas needing attention. However, the reliance on data can also raise concerns about privacy, bias, and the dehumanization of educational experiences.

Keywords: *Algorithms, education, big data, machine learning, prediction, logistic, linear.*

How to cite this paper:
Dhakal,S.D.(2025). Impact of Algorithms and Big Data on Educational field.
The OCEM Journal of Management , Technology & Social Sciences, 4(1),125-131.

Introduction

Predictive analysis can help the continuous intake of enormous measures related to data of students; these arrangements help reveal important examples and future outcomes, giving the dice player and break news clues in a small amount of time against manual counts. Predicting student exposure can help recognize students who are in danger of disappointment, and this way, teachers can give timely assistance and find a way to mentor the students to improve their presence. (Devi & Muthu Selvam, 2022). Nowadays, educational databases have evolved rapidly because of the enormous extent of information that has taken off in them. The dedicated students inspire advanced educational settings, to perform well; the most ideal way is to use the authentic administration and handling of the student database. The ability to anticipate a student's appearance is important in instructive situations. School students' exposure depends on various factors such as individual, social, mental, and others. (Halde, R. R. (2016)) The information in the training area develops intermittently and can be used to predict the exposure of students in the coming semesters. From this, students who are at risk of disappointment stand, and legitimate leadership can be given to students for a better future. The foundation will be instructive and also benefit by improving their overall results. Cheng and Shwe,(2019). Grouping is the path to the collection arrangement as a given vision into disjointed groups.

This is done with the ultimate goal of designing a similar group in the same way, and examples of having a place with two separate groups are extraordinary. Clustering is commonly used to mean segmentation. A group can be in the hierarchy of classes of the college students' comparable organization. By grouping students can be grouped mainly according to school age, areas of interest, background, specialization, and so on. Software training, in combination, can help the student members of the group of institutes in classes that exhibit similar behavior. Researchers are partitioned into groups so that researchers within a group are all different, while several students from other groups use a hierarchical classification plan to shape the largest groups (Hlosta et al., 2022). Despite using the same algorithm in a lot of evaluated results, no student pooled outcomes to recognize similarities to optimize courses at university. As the findings directly analyze the efficacy of courses, the resulting model clusters will make it easier for the university administration faculty to design or program with results similar to those of students. (Cheng & Shwe, 2019) Different authors suggest many existing algorithms to improve and forecast the performance of education development. The review of different algorithms, which have already been proposed, is discussed in this paper.

Literature Survey

(Devi & Muthu Selvam, 2022). highlight that big data is an enormous collection of data from various sources, including social media, which includes all structured, unstructured, and semi-structured data. Deep learning is an artificial intelligence (AI) function that mimics how human skill functions when it comes to processing information and creating decision-making patterns. Text is being translated using deep learning techniques from one language to another. These techniques are actual techniques and outcomes utilizing artificial intelligence, and deep natural language processing has been utilized recently for speech recognition translation. (da Silva et al., 2019), the authors claimed that breaking up is a complex phenomenon related to the failure of hope. This cause is related to the socio-economic, political, and cultural context of education structure and educational organizations.

Dropping out of school is a process of exclusion that is determined by internal and external factors for educational organizations. Both public institutions and private educational institutions feel this phenomenon. The author proposes three different models to predict dropouts. It is based on a regression Ensemble. The proposed set of regression models to obtain comparable performance models literature better. The model set based on linear regression pocket has a lower prediction error (Cheng & Shwe, 2019). The authors addressed the knowledge of educational

data via the results of grouping students for application to the distillery design course.

Changes in technology and other aspects of the call from the working environment for continued progress in the education sector. Understanding these results allows teachers and administrators to improve the program effectively. The basis of this study is to group students' achievements and distinguish them with a higher similarity. These results can be combined to design the optimized approach-directed courses. (Halde, 2016) The author proposed different manners by which AI is utilized in instructive foundations and foresee the exhibition of students and some significant highlights to be considered in anticipating various things.

Challenges in Big Data and Machine Learning Algorithms

Big data is misinterpreted and underappreciated, which is the problem. Big data adoption via an undefined initiative is doomed to fail. The author proposes that a number of workshops and training sessions should be held in order to address this issue. There are many different big data technologies available. The issue is that many technical options are available. The solution for this is to resolve this issue; professional guidance is required. The complexity of data quality management makes it challenging: The challenge with data integration is that the information to be assessed comes from various sources, in a variety of forms, and could contain inaccurate information. The solution is to solve any issues; a suitable big data model or technology is required (Wang et al., 2021). Algorithm challenges include storing many data and requiring more memory. Development is required in order to accomplish one-shot learning. Instead of viewing static films, machine learning algorithms can better comprehend video datasets by listening and observing. To address the issues facing the world, intelligence is required. There is still a shortage of classification and localization in machine learning and computer vision. A thorough understanding of machine learning and the capacity to rapidly apply and solve problem-based algorithms is essential for overcoming these issues. (Devi & Muthu Selvam, 2022).

Algorithms Proposed by different Authors for Education

Education is the most important factor for the growth of the nation. There is a massive and enormous amount of knowledge available in the educational field, but no one uses it properly. New algorithms must be proposed to solve the various problems in the educational field or to predict the information using algorithms to use this information properly.

Figure 1: Educational Algorithms Requirement

This study aims to list the algorithms proposed to improve, predict, or analyze the student's and teacher's activity in the field of education. An analysis that compares algorithms created by various authors, such as Zheng et al.(2023) proposed DBN Model; Ouatik et al. (2022) proposed KNN SVM- Gulzar et al. (2018) proposed hybrid methodology, and Waheed et al. (2020) proposed the deep learning model.

Table 1: A comparative investigation

Title	Year	Author/Author's	Findings	Methodology	Algorithms
Evolutionary Machine Learning builds smart education big data platform: Data-driven higher education	2023	Lu Zheng, Cong Wang, Xue Chen, Yihang Song, Zihan Meng, Ru Zhang	machine learning models to build a personalized course recommendation model	DBN Model and PSO algorithms	Supervised learning and Unsupervised learning
Predicting Students Success Using Big Data and Machine Learning Algorithms	2022	Farouk Ouatik, Mohammed Erritali, Fahd Ouatik, Mostafa Jourhmane	Predication of the academic success of students	methods of artificial intelligence and educational data mining	KNN, C4.5 and SVM algorithms
Predicting academic performance of students from VLE big data using deep learning models	2020	Hajra Waheed, Saeed-U1 Hassan, Naif Radi Aljohani, Julie Hardman, Salem Ale Lyani, Raheel Nawaz	a) VLE portal and demographic data b) Quarterly clickstream data for each students in each new course	Deep learning model	Deep artificial neural network, logistic regression, support vector machine
PCRS: Personalized Course Recommender System Based on Hybrid Approach	2018	Zameer Gulzar a, A_Anny Leema b, Gerar d Deepak c	To make accurate recommendation courses expansion- based information retrieval for course recommendation along with ontology support	Hybrid methodology	N-gram query classification

Methodology

The purpose of this study is to forecast student performance using datasets from Oxford College of Engineering and management in order to forecast grades and increase the accuracy of other algorithms' output, with off-line data. Online data was used in the author's earlier research Abirami T & R. Vadivel, (2023). The analysis-related variables were used in this study to view the grade and result for the current semester, The students' initial internal mark to the two different predictive models such as linear regression and logistic regression.

Result and Discussion

Students database is used for different purposes with different algorithms, classification algorithms to predict student results, clustering algorithms to improve course and performance, and regression algorithms to find dropout rates. There are several issues in the field of education. Only algorithms are needed to solve these issues. There are several algorithms accessible from which the major task is to recognize and use the problem to find a solution. This study also uses some sample data to predict the Subject Mark using the Logistic Algorithm and Linear Regression based on the methodologies covered above. These two algorithms will forecast the students' final semester grade based on their internal marks.

Table 2: Sample Students Dataset

```
df.head()
```

	S.N	roll	name	regno	CAO	gender	Internal	Sem	hrs
0	1.0	20530046.0	Alisha Shrestha	2019-1-53-0140	27.0	F	0.0	F	48.0
1	2.0	20530047.0	Amrit Puri	2019-1-53-0141	27.0	F	32.0	C	45.0
2	3.0	20530048.0	Anisha Bhatta	2019-1-53-0142	7.0	M	36.0	B-	44.0
3	4.0	20530049.0	Anita Mahato	2019-1-53-0143	11.0	F	34.0	C+	40.0
4	5.0	20530050.0	Anjum Angdembe	2019-1-53-0144	12.0	F	33.0	C	46.0

The following Table contains the final sample data after the data has been cleaned and after using label encoder.

Table 3: Students Dataset

	sno	Regno	Name	gender	Internal	Sem
0	1	1	0	0	0	7
1	2	2	1	0	18	2
2	3	3	2	1	30	1
3	4	4	3	0	26	3
4	5	5	4	0	22	2

Table 4: Train and test data

42	43	42	1	16
24	25	24	0	10
6	7	6	0	7
23	24	23	0	34
36	37	36	1	17
21	22	21	1	12
19	20	19	1	16

Board exam score prediction with internal scores and data split using 25% for the test and 75% for the train utilizing both linear and logistic regression.

```
[34] from sklearn.linear_model import LinearRegression
     model1=LinearRegression()
     model1.fit(x_train,y_train)
     pred1=model1.predict(x_test)

     print(pred1)
```

```
[0.84110396 2.5532005  0.78462762 2.6909336  5.83509272 1.52643582
 5.03894552 5.60670141 3.21175895 2.14627475 2.24294056 2.89190067]
```

```
diff=abs(pred1-y_test)
print(diff)
```

```
[0.84110396 1.5532005  0.78462762 0.6909336  1.16490728 0.52643582
 0.96105448 0.39329859 1.21175895 0.85372525 0.24294056 0.89190067]
```

```
[36] diff.mean()
```

```
0.8429906075815716
```

```
[38] from sklearn.metrics import r2_score
     r2_score(y_test,pred1)
```

```
0.8407912121222727
```

Figure 2: Linear regression model result

This figure shows the result of linear regression model.

```
[26] from sklearn.linear_model import LogisticRegression
     model4=LogisticRegression()
     model4.fit(x_test,y_test)
     pred4=model4.predict(x_test)
     print(pred4)
```

```
[0 1 0 2 7 1 6 6 2 2 2 2]
```

```
[27] from sklearn.metrics import accuracy_score
     accuracy=accuracy_score(y_test,pred4)
     print(accuracy)
```

```
0.9166666666666666
```

Figure 3: Logistic Regression model

This figure shows the result of logistic regression model.

The outcome of this research has provided insight into the big data challenges and Algorithm use in current technology for the better result or prediction. The algorithm locates an important role in many real-world problems, a special algorithm always solves the problem quickly. Big Data introduces a

novel and powerful processing algorithm that can be used to improve teaching strategies based on the outcomes of each student group

Conclusion

The research reveals that the algorithm is used efficiently in education to predict student activities using classification, clustering, and regression algorithms. The algorithm's main advantage is that it can process large sets of data elements. Studies show that the algorithm used in education analyzes, improves, identifies, and predicts student activities such as dropout rates, performance, the risk of failure, and, of course, the completion rate for improving education. As time passes, it will be important to create new algorithms to assist students' practical understanding. This will lead to the conclusion that logistic regression, with 91%, is preferable over linear regression, with 84%, for this type of data. Future work will focus on recommendation engines that employ a variety of algorithms to improve big data recommendations.

References

Abirami T, & R. Vadivel. (2023). Student semester marks prediction using linear regression algorithms in machine learning. World Journal of Advanced Research and Reviews, 18(1), 469–475. https://doi.org/10.30574/wjarr.2023.18.1.0591

Cheng, W., & Shwe, T. (2019). Clustering Analysis of Student Learning Outcomes Based on Education Data. 2019 IEEE Frontiers in Education Conference (FIE), 1–7. https://doi.org/10.1109/FIE43999.2019.9028400da Silva, P. M., Lima, M. N. C. A., Soares, W. L., Silva, I. R. R., de A. Fagundes, R. A., & de Souza, F. F. (2019). Ensemble Regression Models Applied to Dropout in Higher Education. 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), 120–125. https://doi.org/10.1109/BRACIS.2019.00030Devi, B. S., & Muthu Selvam, M. (2022). SoloDB for Social Media's Big Data Using Deep Natural Language with AI Applications and Industry 5.0 BT - Ubiquitous Intelligent Systems (P. Karuppusamy, I. Perikos, & F. P. García Márquez (eds.); pp. 279–294). Springer Singapore.

Gulzar, Z., Leema, A., & Deepak, G. (2018). PCRS: Personalized Course Recommender System Based on Hybrid Approach. Procedia Computer Science, 125, 518–524. https://doi.org/10.1016/j.procs.2017.12.067

Halde, R. R. (2016). Application of Machine Learning algorithms for betterment in education system. 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 1110–1114. https://doi.org/10.1109/ICACDOT.2016.7877759

Ouatik, F., Erritali, M., Ouatik, F., & Jourhmane, M. (2022). Predicting Student Success Using Big Data and Machine Learning Algorithms. International Journal of Emerging Technologies in Learning (IJET), 17, 236–251. https://doi.org/10.3991/ijet.v17i12.30259

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189. https://doi.org/https://doi.org/10.1016/j.chb.2019.106189

Zheng, L., Wang, C., Chen, X., Song, Y., Meng, Z., & Zhang, R. (2023). Evolutionary machine learning builds smart education big data platform: Data-driven higher education. Applied Soft Computing, 136, 110114. https://doi.org/10.1016/j.asoc.2023.110114

Xu, B., Song, S., Wang, D., Borshukov, I., Georgieva, O., Wang, X., Wang, C., Li, X., Leung, V. C. M., Taleb, T., Nida Uzel, V., Sevgi Turgut, S., Ayşe Özel, S., Al-Nadabi, S. S., Jayakumari, C., Wati, M., Haeruddin, Indrawan, W., Utomo, M. N. Y., … Syamsuddin, I. (2020). Predict the selection of mathematics subject for 11th grade students using Data Mining technique. Proceeding - 2017 3rd International Conference on Science in Information Technology:, 9441–9455. https://doi.org/10.1016/j.ssci.2020.104608

Wang, H., et al. (2021). Predictive analytics in education: Trends and challenges. Educational Data Journal, 35(4), 45–67.