تحليل أداء الطلبة في التعليم الإلكتروني باعتماد الانحدار

Analysis of Students' Performance in E-learning Based on Regression

د. ثريا احمد خالص الشعلان

مدرس

العلوم التربوية والنفسية، التربية، جامعة الحمدانية، نينوى، العراق

Thurya Ahmed Khalis Shaalan Al-Shimary
lecturer
Department of Educational and Psychological Sciences, Education, University of Al-hamdaniya, Mosul, Iraq

thurayaalshmaree@gmail.com

تاريخ الاستلام ٢٠٢٢/١٠/٢٣ **تاريخ القبول** ٢٠٢٢/١١/٢٨

الكلمات المفتاحية: اداء الطلبة، التعليم الالكتروني، خورزميات تعلم الالة، الانحدار

Keywords: E-learning, Students' performance, machine learning, regression.

الملخص

اصبح التعليم الالكتروني اكثر شيوعا بسبب انتشار جائحة كورونا والتقدم التكلنوجي في مجال التواصل والانترنت. يعتبر الحفاظ على مستوى اداء الطلبة ومشاركتهم وتحفيزهم من التحديات التي تواجه التعليم الالكتروني وبالاخص الصلة مابين مدى مشاركة الطلبة والاداء الاكاديمي. ولتحقيق هذا الهدف ، اقترحت هذه الورقة استخدام خوارزميات الانحدار للتنبؤ باداء الطلبة وايجاد العلاقة مابين مشاركة الطلبة ونتيجة الاختبار النهائي في بيئة التعلم الالكتروني وتاثير كل فعالية على الاداء. ولغرض تحليل اداء الطلبة والتنبؤ بادائهم تم استخدام مجموعة البيانات Open University Learning Analytics (OULAD) والتي تحتوي على مفردات تفاعل الطلاب أثناء التعلم عبر الإنترنت من حيث تحرير النص والوقت الذي يقضيه كل نشاط وما إلى ذلك ، بالإضافة إلى درجة الاختبار التي تم تحقيقها في كل جلسة. والتي تم تصنيفها لغويًا في ثلاث فئات واسعة بناءً على معايير مختلفة: (١) نوع النشاط ، (٢) إحصاءات التوقيت ، و (٣) عدد الأنشطة الطرفية. يهدف نموذج Machine learning المقترح إلى توقع ما إذا كان أداء الطالب سيكون منخفضًا أم مرتفعًا. تم استخدام اربع مصنفات شائعة في دراستنا ، وهي: XGBoost ، LASSO ، اشجار القرار ومتجه دعم الالة وباستخدام مقاييس مختلفة. أظهرت النتائج أن LASSO حقق أفضل أداء لدقة التنبؤ.

Abstract

Due to the spread of the Corona pandemic, e-learning has grown in popularity. Maintaining studenzt performance, participation and motivation are considered challenges were facing e-learning, especially the link between student participation and academic performance. This paper suggested using regression algorithms to predict students' performance and find the relationship between students' participation, the final exam result in the e-learning environment, and each effect's attribute on performance. To analyze students' performance and predict their performance, the Open University Learning Analytics (OULAD) dataset was used, which shows the students' interaction during the online laboratory work in terms of text editing, time spent on each activity, etc., in addition to the test score achieved in each session. Which are divided into three major groups depending on a variety of factors: (1) the kind of activity, (2) time statistics, and (3) the number of peripheral tasks. The suggested ML model forecasts whether a student will do poorly or well. Four popular prediction algorithms—XGboost, LASSO, Decision Trees, and Support Vector Regression—were applied and tested in the investigation. The results showed that LASSO performed the best in terms of prediction accuracy.

1. Introduction

Due to the spread of the Covid-19 virus, e-learning has recently been accepted in most nations and for all educational levels. This has prompted study into the processes and methods of e-learning and tests of the effectiveness of this approach to education (Almossa & Alzahrani, 2022; Maatuk, Elberkawi, Aljawarneh, Rashaideh, & Alharbi, 2022). The foundation is the management of all educational and scientific events and their requirements, as well as the provision of educational content and the communication of skills and concepts to the learner via information and communication technologies and their interactive media in a way that enables the student to interact with the content, the teacher actively, and peers concurrently or asynchronously at a time, place, and speed that suits the learner's circumstances and ability (Yang, 2017).

Along with the rapid advancement of information technology over the past ten years, the instability of health conditions, and new changes in line with changes, e-learning offers rich educational resources that are useful for both fulfilling educational requirements and maintaining communication between educational staff and students. The majority of educational institutions worldwide have shifted to e-learning (Weerathunga et al., 2021). Any country's educational system must be given considerable consideration if it is to grow. To improve the e-learning system, research is conducted into many of its mechanisms and evaluating students' performance in light of this type of education (Coman, Țîru, Mesesan Schmitz, Stanciu, & Bularca, 2020).

Data analysis and modeling have spread in many areas to find hidden patterns and help in decision support systems. Statistical methods and machine learning algorithms can help educational institutions in many areas (Sarker, 2021). By applying information mining algorithms, information mining in the academic field aims to show the hidden relationships in databases to benefit from them in providing advice to those in charge of the educational process, professors or students alike (Manjarres, Moreno Sandoval, & Suárez, 2018). Finding the association rules makes it possible to find the characteristics that affect the student's performance, such as attendance, participation, and the activities he performs (Yagci, 2022). To predict the final result of one of the subjects or the general result based on similar previous cases, sort students into groups according to their levels. It is also possible to know the distortions in performance, one of the biggest problems students and professors face. Mining educational data can offer a lot in the case of e-learning.

Nowadays, a framework that allows analysis and monitoring of student performance are needed, especially in the e-learning system. To evaluate students' performance by modeling students' interaction with the website and adherence to a schedule and specific assignments. The main objective of this paper is to analyze student data and build e-learning classification models for a "student performance" data set based on student behaviors and interaction with the website.

2. Related Work

Numerous strategies were examined to predict student performance; most were created for machine learning (ML) models. By predicting an underlying relationship between input variables, the models attempt to identify patterns in the data. K-NN and Support Vector Machine (SVM) ML techniques were employed by (Al-Shehri et al., 2017) to forecast students' performance on the final test. To validate the effectiveness of ML models, a bespoke dataset created by the University of Minho in Portugal comprising 395 data samples was employed. The dataset contains information on each student's family background and personal data elements. According to the study, SVM's accuracy was marginally higher than k- NN's. According to (Atherton et al., 2017), there is a link between the utilization of course materials and student grades; those who accessed course materials more often performed better on tests and other evaluations. (Soni, Kumar, Kaur, & Hemavathi, 2018) created a dataset of roughly 2000 students with 50 characteristics and used classification algorithms to examine the student performance from their previous performances. Results show that extracurricular activities and personal habits are just as important to a student's achievement as their grades. (Ünal, 2020) utilized two publicly accessible datasets to forecast student results using feature subset selection and classification techniques. The research compared the accuracy rates of categorization systems, including decision trees, random forests, and Naive Bayes. (Dewan, Murshed, & Lin, 2019) proposed classifications for engagement detection from the existing approach for online learning learners. There are three major types: automatic, semiautomatic, and manual. These classifications depend on the method and the level of user participation in the engagement identification procedure.

Research has been conducted to predict students' academic performance using data mining techniques. However, few have highlighted the essential traits/characteristics influencing students' educational performance and found relationships between traits. This paper will use some essential machine learning and artificial intelligence algorithms to predict students' academic and overall performance.

3. Preliminaries

Machine Learning Technique. Numerous ML techniques have been used as different kinds of predictive models. The following section describes the current study's ML techniques evaluated as predictive models.

XGBoost

The XGBoost method is used to solve issues involving supervised learning. It may be used for problems involving classification, regression, and ranking, among other applications (Budholiya, Shrivastava, & Sharma, 2020). In most cases, while dealing with supervised issues, the forecasting of the target class is done by using the training data, which may include many characteristics. XGBoost is a prominent algorithm used in various applications, including educational and commercial ones (Sukarsa, Pinata, Rusjayanthi, & Wisswani, 2021). XGBoost is built with an excessive amount of ambition (Jegananathan, Lakshminarayanan, Ramachandran, & Tunze, 2022). The building process begins at the root node, and the mean square error determines where leaf splits should occur.

Decisions Tree

A Decisions Tree (DT) is organized like a tree, with internal nodes and leaves. The child nodes of an internal node might range from two to an unlimited number of nodes (Benediktus & Oetama, 2020). The internal nodes represent the dataset's characteristics, while the branches indicate the values of these features (Arumugam, G, & Chandrasekaran, 2018; Benediktus & Oetama, 2020). Each leaf in the tree has a class associated with the dataset, and DT is trained using a training set with tuples. In the last step, the DT is used to classify a dataset with unknown class labels. The processing of information for decision-making is the primary function of DTs. The decision tree is formed from the dataset by identifying which characteristics may best separate the input features at the child nodes. In this particular instance, we used the idea of information gain, which is founded on information theory (Charbuty & Mohsin Abdulazeez, 2021). When a node has the lowest possible entropy and the most significant possible information gain, we refer to that node as a split node. The criteria for DTs are straightforward to comprehend and interpret and know which classifier leads to a decision (Varade & Thankanchan, 2021). This makes DTs a crucial component of analysis when the analysis aims to establish which aspects of a student prediction model are most significant.

SVR

A generalized linear classifier called a Support Vector Machine (SVM) uses supervised learning to conduct binary classification on data, and its decision boundary is to solve the training sample (Uemoto & Naito, 2022). The fundamental principle of SVM is to locate some data at the edge of the set (referred to as the Support Vector) and use these points to locate a plane (referred to as the decision surface) to maximize the distance between the Support Vector and the Plane (Farooqui & Ahmad, 2020). Support vector regression (SVR) is a method for solving regression issues using the SVM (support vector machine). SVR is a generalization of the support vector classification approach to resolve regression issues (Karlsson, 2021; Uemoto & Naito, 2022). Support vector classification models only use a portion of the training data since training points outside the boundaries are not considered by the cost function used to build the model. Similarly, models created by support vector regression only employ a portion of the training data since any training data that is similar to the model's predictions is disregarded by the cost function that builds the model (Uemoto & Naito, 2022).

Lasso

Compression estimation is done using the Lasso (Least Absolute Shrinkage and Selection Operator) approach (Lee & Kim, 2020). A more refined model is obtained by creating a punishment function, compressing certain coefficients, and setting other coefficients to zero. As a result, the benefit of subgroup shrinking is still there, and the estimate is skewed when dealing with intricately connected data (Mwikali, Mwalili, & Wanjoya, 2019; Pardede, Sumargo, & Rahayu, 2022; Tardivel & Bogdan, 2022). The fundamental principle of Lasso is to minimize the residual sum of squares under the restriction that the total of the regression coefficients' absolute values is less than a constant to produce certain regression coefficients that are strictly equal to 0 and achieve an understandable model.(Lee & Kim, 2020; Mwikali et al., 2019)

4. Materials and Methods

4.1. Data Description

To analyze students' performance, the Open University Learning Analytics (OULAD) dataset was used (Kharis & Zili, 2022; Kuzilek, Hlosta, & Zdrahal, 2017), the most comprehensive data set regarding student information. The Open University's role in developing this dataset is to support research in learning analysis by collecting and analyzing learner data to provide personalized guidance and improve learning resources. The data set includes the information of 486 students and 14 features, including 13 features of student data, and the last feature is the student's result in the final exam. In this study, two types of prediction are adopted; the first is to find the student's class, where the students' scores were divided into three classes: superior, average, and weak, and in the second type, the student's final score is predicted using regression. Table 1 displays the newly derived desirable metrics included in this research, their description, kind, and range of values. Engagement metrics are selected based on log file data to extract the most information.

Table (1)

Description of the attributes of the dataset

Feature	***Interaction Type***	***characterization***
Student ID	-	Identifier
Num. of Logins	Interact	Student's LMS Course site access count
Num. of Content Reads	Interact	How often a student downloaded course material
Num. of Forum Reads	Interact	Student's amount of forum reads

Feature	***Interaction Type***	***characterization***
Num. of Forum Posts	Interact	Student forum posts
Num. of Quiz Reviews	Interact	Student's quiz review frequency
Assign.1 lateness indicator	achievement	An indication of whether Assignment 1 is late.
Assign.2 lateness indicator	achievement	A binary indication whether Assignment 2 is late.
Assign.3 lateness indicator	achievement	An indication of whether Assignment 3 is late.
Assign.1 duration to	achievement	Assignment 1 posting-to-submission time (in hours)
Assign.2 duration to	achievement	Assignment 2 posting-to-submission time (in hours)
Assign.3 duration to	achievement	Assignment 3 posting-to-submission time (in hours)
Average Assign. duration to	achievement	Average hours between posting and submitting assignments

4.2. Proposed Framework

The proposed framework consists of several stages. In the first stage, the data set is identified, described, and then analyzed to find the features' effect on the target value (student performance). The target value that has been adopted is the student's score in the final exam in the case of regression, and the division of students' scores into classes for the classification case has been adopted to identify the importance of features and measure the performance of prediction and algorithms in predicting students' performance in e-learning.

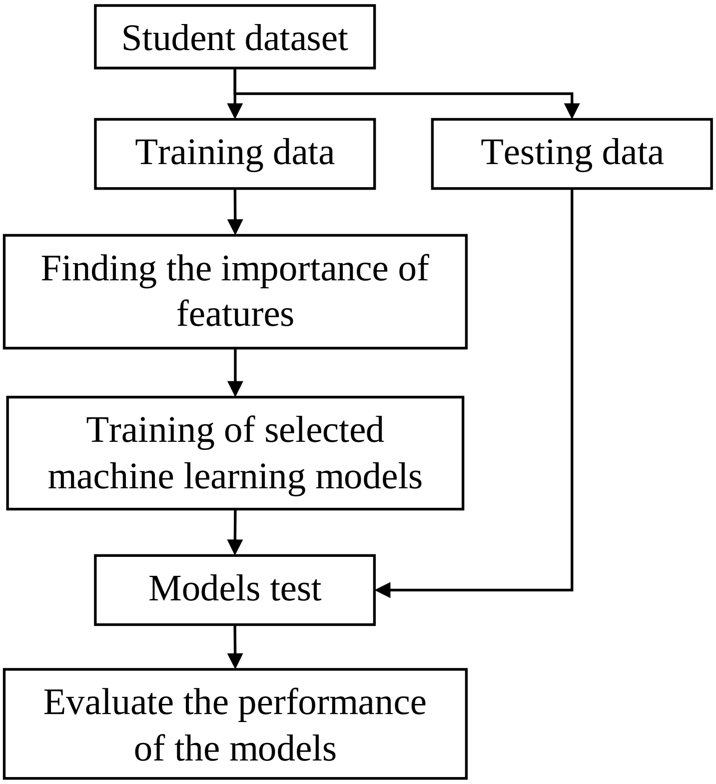

Figure (1)

Student Academic Performance Prediction Framework for Online Learning

After obtaining and preparing the data set, the students' data set is divided into a training set and a test set (80:20). In the next stage, the significance of the features concerning the final exam score is found using regression and classification. This is followed by building the selected machine learning models (Decisions Tree, XGBoost, SVR, Lasso). After training, the models are tested using (Root mean squared error, mean absolute error, mean squared error, Median absolute error, and R2 score) and compared between them (Adekeye, Adewara, Aako, & Olaomi, 2021; Ahmar, 2020; Chicco, Warrens, & Jurman, 2021; Karunasingha, 2021).

4.3. Performance Evaluation

This research will use some performance metrics to determine the minimum forecast error to measure forecast performance, and the used metrics can be expressed.

$$MAN = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

(1)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

(2)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|}$$

(3)

$$R2\ score = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2}$$

(4)

Were,

y : is an actual value

$\hat{y}$: predictive value

$\bar{y}$: is the mean value of y

5. Results and discussion

5.1. The Effect of Features on Student Performance

Determining the importance of features is important in the process of e-learning and determining the effect of each attribute on students' performance in final exams. Figure 2 shows the ranking of features in terms of importance.

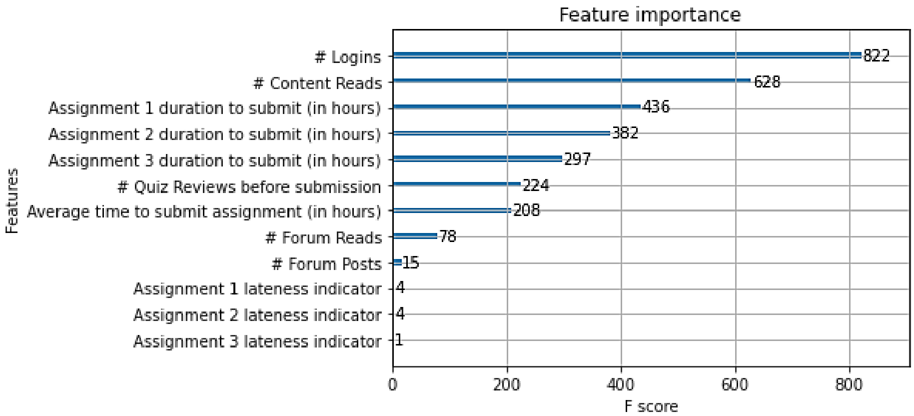

Figure (2)

Ranking of Features in order of importance Based on Regression

Figure 2 also notes that logging in to the study website is one of the essential features in determining the student's level in the final exams, followed by the number of times the content has been read and the duration of delivery of the required tasks. At the same time, when finding the importance of the features in the case of classification, the performance of students based on classification found that the importance took a different arrangement. Figure 3 shows the importance of the features by adopting the classification of categories.

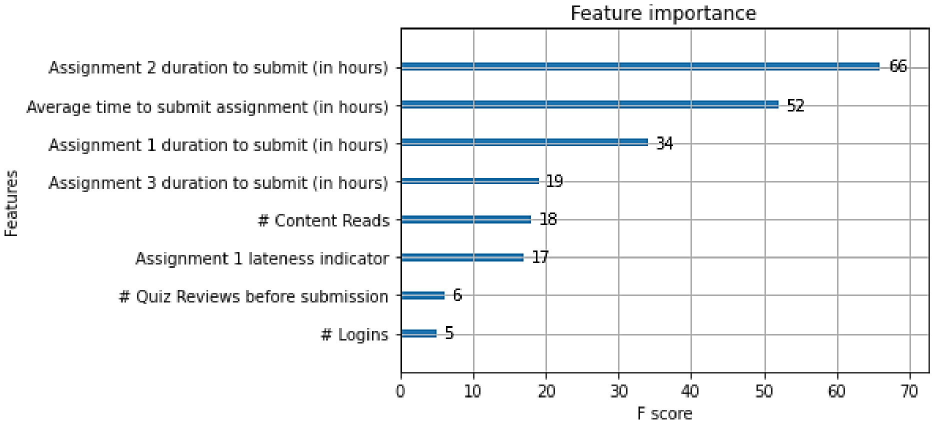

Figure (3)

Ranking of Features in order of Importance Based on the Classification

As in Figure 3, the importance of features concerning the classification of students based on the degree of tasks assigned to them, in addition to the final exam, found that the process of logging in to the site was the least important, while the period of tasks assigned to them was the highest.

5.2. Performing models to predict the performance of students based on regression.

To predict students' performance in final exams, regression algorithms were used. After building and training the tested models, use the test data. Table 2 compares the trained models based on the percentage of error.

Table (2)

Compare Regression Models to Predict Student Performance

	XGBOOST	LASSO	DT	SVR
RMSE	16.601454	**15.907939**	19.476307	16.409676
MEAN ABSOLUTE ERROR	12.15	**11.37**	15.02	11.64
MEAN SQUARED ERROR	275.61	**253.06**	379.33	269.28
R2 SCORE	0.04	**0.12**	-0.32	0.06

The Lasso model achieved the lowest error rate on different scales, where the error rate according to the RMSE scale was 15.9, 11.37 according to the MAE scale, and 253 according to the MSE scale. At the same time, the highest was 0.12 according to the scale of R2, followed by SVR, then XGBoost; Finally, the decision trees with Low performance in modeling this type of data.

6. Conclusion

To predict students' performance under e-learning using machine learning, the data set and the impact of features on students' performance were analyzed, and regression models (Decisions Tree, XGBoost, SVR, Lasso) were applied to predict students' scores in the final exam. The study found that the effect of the features varies according to the approved degree of the student, as the results showed that the processes of logging in and reading the content in the case of the approval of the student's grade in the final exam, while the completion of the tasks with the highest impact was in the approval of the exam grade and participation. To predict student performance, the Lasso algorithm achieved the best using regression measures. The importance of features highlighted behavioral features as the highest determinant of student success. To radically mitigate the risks of failure or dropout in e-learning, irregularities in students' behavioral features are indications for teachers or institutions to intervene on time to prevent massive damage to the student's learning path. Based on the predictive models used in this study and their assessment metrics, it concludes that student participation can play a vital role in predicting student success.

References

- Adekeye, K., Adewara, J., Aako, O., & Olaomi, J. (2021). Performance of median absolute deviation and some alternatives to median absolute deviation control charts for skewed and heavily tailed process. *Quality and Reliability Engineering International, 37*. doi:10.1002/qre.2926
- Ahmar, A. (2020). Forecast Error Calculation with Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE). *JINAV: Journal of Information and Visualization, 1*. doi:10.35877/454RI.jinav303
- Al-Shehri, H., Al-Qarni, A., Al-Saati, L., Batoaq, A., Badukhen, H., Alrashed, S., . . . Olatunji, S. (2017). *Student performance prediction using Support Vector Machine and K-Nearest Neighbor*.
- Almossa, S., & Alzahrani, S. (2022). Assessment practices in Saudi higher education during the COVID-19 pandemic. *Humanities and Social Sciences Communications, 9*. doi:10.1057/s41599-021-01025-z
- Arumugam, S., G, V., & Chandrasekaran, D. (2018). Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers. *Journal of Computer Science*, *14*, 654-662. doi:10.3844/jcssp.2018.654.662
- Atherton, M., Shah, M., Vazquez, J., Griffiths, Z., Jackson, B., & Burgess, C. (2017). Using learning analytics to assess student engagement and academic outcomes in open access enabling programmes. *Open Learning: The Journal of Open, Distance and e-Learning, 32*, 1-18. doi:10.1080/02680513.2017.1309646
- Benediktus, N., & Oetama, R. (2020). The Decision Tree C5.0 Classification Algorithm for Predicting Student Academic

Performance. *Ultimatics : Jurnal Teknik Informatika, 12*, 14-19. doi:10.31937/ti.v12i1.1506

❖ Budholiya, K., Shrivastava, S., & Sharma, V. (2020). An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University - Computer and Information Sciences, 34*. doi:10.1016/j.jksuci.2020.10.013

❖ Charbuty, B., & Mohsin Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends, 2*, 20-28. doi:10.38094/jastt20165

❖ Chicco, D., Warrens, M., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. doi:10.7717/peerj-cs.623

❖ Coman, C., Țîru, L., Mesesan Schmitz, L., Stanciu, C., & Bularca, M. (2020). Online Teaching and Learning in Higher Education during the Coronavirus Pandemic: Students' Perspective. *Sustainability, 12,* 10367. doi:10.3390/su122410367

❖ Dewan, M., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: a review. *Smart Learning Environments, 6*. doi:10.1186/s40561-018-0080-z

❖ Farooqui, M., & Ahmad, D. (2020). DISEASE PREDICTION SYSTEM USING SUPPORT VECTOR MACHINE AND MULTILINEAR REGRESSION. *International Journal of Innovative Research in Computer Science & Technology, 8*. doi:10.21276/ijircst.2020.8.4.15

❖ Jegananathan, S., Lakshminarayanan, A. R., Ramachandran, N., & Tunze, G. (2022). Predicting Academic Performance of Immigrant Students Using XGBoost Regressor. *International Journal of*

Information Technology and Web Engineering, 17, 1-19. doi:10.4018/IJITWE.304052

❖ Karlsson, N. (2021). Comparison of linear regression and neural networks for stock price prediction. In.

❖ Karunasingha, D. (2021). Root Mean Square Error or Mean Absolute Error? Use their ratio as well. *Information Sciences, 585*. doi:10.1016/j.ins.2021.11.036

❖ Kharis, S., & Zili, A. (2022). Learning Analytics dan Educational Data Mining pada Data Pendidikan. *JURNAL RISET PEMBELAJARAN MATEMATIKA SEKOLAH, 6,* 12-20. doi:10.21009/jrpms.061.02

❖ Kuzilek, J., Hlosta, M., & Zdrahal, Z. J. S. d. (2017). Open university learning analytics dataset. *4*(1), 1-8.

❖ Lee, H., & Kim, H.-S. (2020). Logistic Regression and Least Absolute Shrinkage and Selection Operator. *Cardiovascular Prevention and Pharmacotherapy, 2,* 142. doi:10.36011/cpp.2020.2.e15

❖ Maatuk, A., Elberkawi, E., Aljawarneh, S., Rashaideh, H., & Alharbi, H. (2022). The COVID-19 Pandemic and E-learning: Challenges and Opportunities from the Perspective of Students and Instructors. *Journal of Computing in Higher Education, 34*. doi:10.1007/s12528-021-09274-2

❖ Manjarres, A., Moreno Sandoval, L. G., & Suárez, M. (2018). Data mining techniques applied in educational environments: Literature Review. *Digital Education Review,* 235-266. doi:10.1344/der.2018.33.235-266

❖ Mwikali, J., Mwalili, S., & Wanjoya, A. (2019). Penalized Poisson Regression Model Using Elastic Net and Least Absolute Shrinkage and Selection Operator (Lasso) Penality. *International Journal of Data Science and Analysis, 5*, 99. doi:10.11648/j.ijdsa.20190505.14

- ❖ Pardede, T. T., Sumargo, B., & Rahayu, W. J. J. S. d. A. (2022). Penerapan Regresi Least Absolute Shrinkage And Selection Operator (LASSO) Untuk Mengidentifikasi Variabel Yang Berpengaruh Terhadap Kejadian Stunting di Indonesia. *6*(1), 37-48.
- ❖ Sarker, I. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*. doi:10.1007/s42979-021-00592-x
- ❖ Soni, A., Kumar, V., Kaur, R., & Hemavathi, D. (2018). Predicting student performance using data mining techniques. *International Journal of Pure and Applied Mathematics, 119*, 221-226.
- ❖ Sukarsa, M., Pinata, N., Rusjayanthi, D., & Wisswani, N. (2021). Estimation of Gourami Supplies Using Gradient Boosting Decision Tree Method of XGBoost. *TEM Journal, 10*, 144-151. doi:10.18421/TEM101-17
- ❖ Tardivel, P. J., & Bogdan, M. J. S. J. o. S. (2022). On the sign recovery by least absolute shrinkage and selection operator, thresholded least absolute shrinkage and selection operator, and thresholded basis pursuit denoising.
- ❖ Uemoto, T., & Naito, K. (2022). Support vector regression with penalized likelihood. *Computational Statistics & Data Analysis, 174*, 107522. doi:10.1016/j.csda.2022.107522
- ❖ Ünal, F. (2020). Data Mining for Student Performance Prediction in Education. In.
- ❖ Varade, R., & Thankanchan, B. (2021). Academic Performance Prediction of Undergraduate Students using Decision Tree Algorithm. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, 13*, 97-100. doi:10.18090/samriddhi.v13iS1.22

- ❖ Weerathunga, P., Samarathunga, W., Rathnayake, H., Agampodi, S., Nurunnabi, M., & Madhunimasha, M. (2021). The COVID-19 Pandemic and the Acceptance of E-Learning among University Students: The Role of Precipitating Events. *Education Sciences*, *11*, 436. doi:10.3390/educsci11080436
- ❖ Yagci, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*. doi:10.1186/s40561-022-00192-z
- ❖ Yang, D. (2017). Instructional strategies and course design for teaching statistics online: perspectives from online students. *International Journal of STEM Education, 4*. doi:10.1186/s40594-017-0096-x