Journal of Autonomous Intelligence (2024) Volume 7 Issue 1
doi: 10.32629/jai.v7i1.873

ORIGINAL RESEARCH ARTICLE

Predicting student final examination result using regression model based on student activities in learning management system

Nurul Nadia Nik Pa[1], Azwa Abdul Aziz[1,2,*], Suhailan Safei[1,2], Wan Azani Mustafa[3]

[1] *Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin (UniSZA), Besut Campus, Besut 22200, Malaysia*

[2] *Data Science & Analytics (DASA), Special Interest Group (SIG), Universiti Sultan Zainal Abidin (UniSZA), Gong Badak 20300, Terengganu, Malaysia*

[3] *Advanced Computing, Centre of Excellence (CoE), Universiti Malaysia Perlis, Perlis, Arau 02600, Malaysia*

*** Corresponding author:** Azwa Abdul Aziz, azwaaziz@unisza.edu.my

ABSTRACT

Background: Learning analytics (LA) is the measurement, collection, analysis, and reporting of data about learners and their contexts to understand and optimize learning and the environments in which it occurs. Most teaching and learning (T & L) data was obtained from learning management systems (LMS), such as Moodle platform. However, this data is not utilized for teaching purposes, for example, learning how students' activities can influence the final exam marks. **Methods:** Therefore, this study aims to find a correlation and develop a prediction model between students' activities, independent variables known as LMS factors, against dependence variables, which are students' final results. Besides, four non-LMS factors (race, sponsorship, admission requirements and final coursework marks) were also included in the research to obtain the best model. The regression analysis, models are used to predict the outcomes by evaluating the accuracies of testing data. **Results:** The findings reveal that the best model utilizes Simple Linear Regression (SLR) with coursework as an independent variable, resulting in an average error difference (AED) of only 1.8. The remaining experiments produced AED results ranging from 2.74 to 7.58 using Multilinear Regression (MR). **Conclusions:** in summary, this study provides a significant finding that demonstrates the potential of utilizing LMS activities to predict final marks, enabling lecturers to enhance their students' results.

Keywords: learning analytics; learning management system; prediction; regression

ARTICLE INFO

Received: 30 July 2023
Accepted: 7 August 2023
Available online: 9 October 2023

COPYRIGHT

Copyright © 2023 by author(s).
Journal of Autonomous Intelligence is published by Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).
https://creativecommons.org/licenses/by-nc/4.0/

1. Introduction

In recent years, technologies have been advancing with various innovations and modernity. All industries optimizing the general use of information and communication technology (ICT) facilities and technologies come in many ways and applications and friendly interaction. Big data is created to address 5V (volume, variety, velocity, value, and veracity) challenges. The current data primarily exists in unstructured data types, presenting significant challenges to the researcher in turning this data into useful information. On the other hand, big data's high dimensionality lack defined with measurable techniques.

The revolution of data analytics approaches is across multiple domains such as business, healthcare, security, engineering, and others. Education is also a part of the data revolution, where information focuses on results and the behaviour towards achieving the result from students or teachers. The term educational data mining (EDM), educational intelligence (EI), and the recent one, learning analytics (LA), show the importance of data science approaches implementation towards a massive amount of learning data. However, several problems exist in implementing LA in higher educational institutions (HEI), including infrastructure.

Researchers have looked into how well LA might help students study more effectively. In contrast to traditional classroom instruction, a study by Santally et al.[1] indicated that learning analytics in a blended learning environment boosted student engagement and resulted in higher grades. Similar to this, Chen et al.[2] study discovered that students who received tailored feedback through learning analytics outperformed those who only received broad input on exams. The advanced learning analytics research of the last years converges with the industry demand to enhance the famous learning management system (LMS) with learning analytics capabilities promoting higher education critical challenge for designing personalized curricula and learning experiences. The current COVID-19 pandemic leads most educational institutions to use 100% of an online platform to replace physical face-to-face (F2F) classes.

Therefore, a tremendous amount of learning data is created daily, making LA approaches essential to understanding behaviours, especially student performance. Recent research focuses on utilizing learning data from educational systems, learning management system (LMS), or other educational resources sources from the teaching and learning (T & L) process. For example, AlJohani et al.[3] show the LA process framework in **Figure 1**.

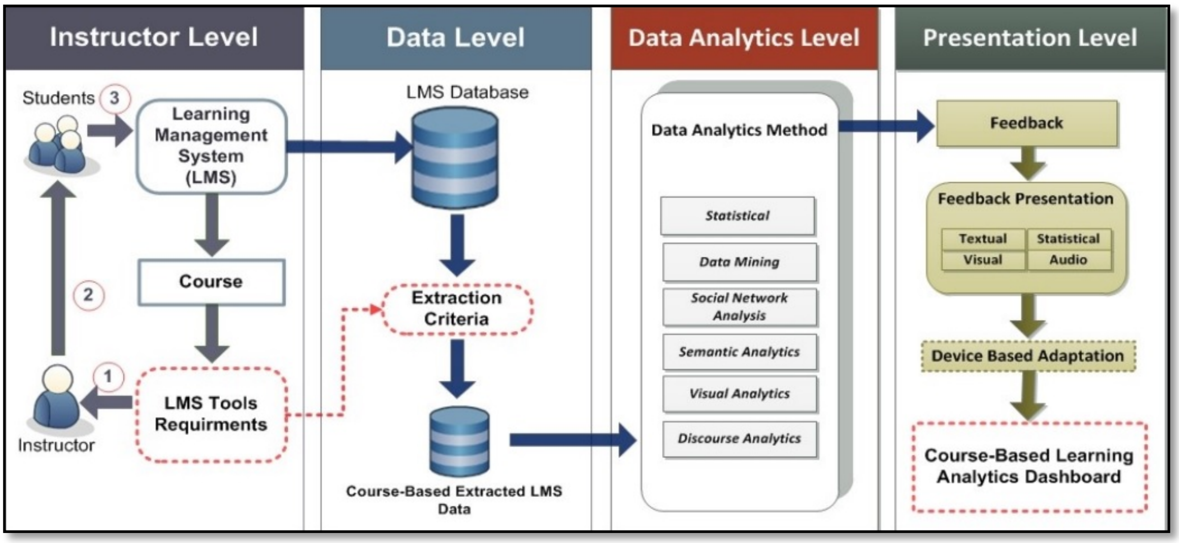

Figure 1. Learning analytics framework[3].

In **Figure 1**, the researcher proposed four primary levels of LA; instructor level, data level, data analytics level, and presentation level. Instruction level contains learning data generated from LMS, which are stored in LMS databases. Data analytics level use generated data and apply analytics algorithms to find patterns and behaviours or new knowledge of the T & L process. The final level presents the output to university management, faculty, and teachers.

Turning to universiti sultan Zainal Abidin (UniSZA) LA environment, LMS is a crucial tool in the T & L process, whether pre-pandemic COVID-19 or vice versa. Knowledge e-learning integrated platform (KeLIP) is a UniSZA LMS developed from Moodle and plays a considerable role in using technologies in the T & L process. However, there is currently no proper research utilising data from KeLIP to improve LA in UniSZA. Therefore, the project investigates a correlation between students' activities in KeLIP, such as involvement in activities (forum, quiz, link clicks), and results obtained by the end of the semester.

2. Related works

2.1. Learning analytics

LA is the measurement, collection, analysis, and reporting of data about learners and their contexts to understand and optimise learning and the environments in which it occurs. LA is intersection of learning fields, such as educational research, sciences of learning and assessment, and educational technology[4]. A study by Hasan et al.[5] develop a mobile application to enhance the video streaming server's existing infrastructure. It help the usability analysis of the mobile application with the student's learning experience for LA purposed.

The main idea is to data about learners and their environments to understand and improve learning outcomes. The four main functionalities of LA are:

- Measure key indicators of student performance
- Support student development
- Understand and improve the effectiveness of teaching practices
- Inform institutional decisions and strategies

LA refers to the process of measuring, gathering, analyzing, and presenting information about learners and their surroundings, aiming to enhance and optimize the learning experience and the surroundings in which it takes place. However, in the early stage, data are limited, and the term big data analytics (BDA) is also a new term that most of the research regarding LA on a small scale. Educational analytics can also be viewed as existing in various levels, ranging from individual classrooms, departments, universities, regions, states/provinces, and international.

However, the trend changes as academic data grows, and most universities introduce a blended-learning mode, leading to more data being analysed. The COVID-19 pandemic shows the importance of online learning, where most T & L activities use LMS and other platforms, including webex or Google classroom. It creates more academic data captures than the physical data analytics class. The advancement of data science technologies such as machine learning algorithms and BI applications, creates a different LA environment compared to the LA concept's early stages. Miller[6], states is one of the digital tactics and technologies that the EDUCAUSE learning initiative expects will soon become widely adopted, according to the horizon report: 2019 higher education edition. Therefore, skilled data analysts are required to support these analytics-driven initiatives and ensure institutional success.

Researchers adapted advanced technologies such as business intelligence (BI) to use in the academic ecosystem. For example, Aziz et al.[7] proposed educational data warehouse (EDW) architecture that integrates proprietary and open source BI tools in academic environments. The implementation utilise BI tools such as online analytical processing (OLAP) and dashboard to perform insight analysis of student performances. Thus, to improve student performances by promoting advanced reporting that usually use to predict sales performances. Mahmud et al.[8] analyse learning performance assessment using augmented reality (AR) via a mobile application for Malaysian preschool students. The AR technologies' feedback can be used to identify student strengths and weaknesses.

In addition to research on the effectiveness of learning analytics, there have also been studies on implementing learning analytics in higher education institutions. For instance, Chen, et al.[9] examined the variables that affect the adoption of learning analytics in higher education and concluded that faculty support, leadership backing, and data privacy issues were crucial in determining the adoption's success. Another study by Kim et al.[10] investigated the challenges higher education institutions face in implementing learning analytics and found that a lack of resources, technical challenges, and cultural resistance were some of the biggest obstacles in implementing learning analytics.

2.2. Learning management system (LMS)

LMS is an application that responds to managing, documenting, monitoring, reporting, and training that support the T & L process. The majority of universities use LMS as their primary medium for online teaching. A learning management system allows an instructor to create and deliver content, monitor student participation and assess student performance. A learning management system may also enable students to use interactive features such as threaded discussions, video conferencing, and discussion forums.

One of the most popular LMS in Moodle, open-source LMS written in PHP and distributed under the GNU general public license. It is a learning platform designed to provide educators, administrators, and learners with a robust, secure, and integrated system to create personalized learning environments. UniSZA has applied Moodle since 2002 and doing customisation to suit UniSZA's needs. The roles of KeLIP are crucial for the T & L process, especially during the pandemic, whereby the courses are run entirely online. More than 10,000 students and 500 teaching staff are involved in UniSZA academic environment.

2.3. Educational data mining

Educational data mining (EDM) is an emerging discipline concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to understand better students and the settings in which they learn. It is a subset of data mining that specialised in the academic domain, with most research aiming to improve students' performances. Baker and Yacef[11] summarised the following four goals of EDM:

- Predicting learners' behaviours by improving students' model.
- Discovering or improving knowledge domain structure model.
- They are studying the most effective pedagogical support for student learning that can be achieved through learning systems.
- Establishing empirical evidence to support or articulate pedagogical theories, frameworks, and educational phenomena to determine influential core components of learning to design better learning systems.

Despite the modest differences in characteristics between EDM and LA, the influential factors that change the quantity of data are the same due to relying on the secondary data generated by online learning technologies[7]. The advent of new technologies and public data repositories in pedagogical contexts increase the quantity of data and diverge data quality, such as data created by LMS. Another example is that mobile devices enable researchers to capture learners' interactions in more detail. Therefore, LA focuses on data captured in a new era whereby EDM can be a subset of LA approaches to analyse data. Most research in LA concentrates on finding a pattern to reveal students' performance. It includes investigating the relationship between activities during the T & L process and students' achievement.

2.4. Previous research

Research by Amrieh et al.[12] proposed a prediction model based on students' activities to the education tools (Kalboard) known as students' behavioural features. The classification model's four variables are discussion groups, visited resources, raised hands-on class (virtual), and viewing announcements. The model develops using ensemble methods, learning approaches that combine multiple models and obtain the result between 67–79 accuracies. The classification results are based on three categories: good, medium, and bad. Although the result is promising, the research does not conduct a correlation analysis between variables and the target class. It can support the model the approaches produce if the correlation analysis is performed. Ahmad et al.[13] propose a framework for predicting students' academic performance of first-year bachelor students in the computer science course. The data were collected from eight years period intakes. The model uses three classification approaches decision tree, naïve Bayes, and rules-based.

The highest result produces the rules-based approach with 71.3% accuracy. Further work needs to be done to improve accuracy and model abilities. Several researchers focus on LMS data to predict students' performance. Conijn et al.[14] studied the performance of students across 17 blended courses and 4898 students from a single institution. The experiment involves running correlation analysis to obtain a significant connection between numerous LMS activities such as total number of clicks, number of online sessions, time online (min), and average time per session (min). The result shows only the in-between assessment grade correlated significantly with the final exam grade in all courses, confirming that the in-between measurement of performance is in line with the performance measurement at the end of the course. Meanwhile, discussion forums and wiki usage showed significant correlations in the lowest number of courses, indicating that these variables are not very useful (and not very stable) predictors of final exam grades across courses, at least in our data.

Macarini et al.[15] try to provide an early prediction model for dropout students based on LMS interaction. Moodle's data was used to generate 13 distinct datasets based on student interactions (cognitive presence, social presence, and teaching presence) inside the virtual environment. Thirteen dataset combinations and five classification algorithms (k-nearest neighbor, multilayer perceptron, naive Bayes, AdaBoost, and random forest) were used in the experiments. Results show no statistically significant difference among models generated from the different datasets and that the counts of interactions and derived attributes are sufficient for the task.

Research by Geng et al.[16] observed the teaching activities of three classes. It looked into the connections between learning performance, learning motivation, and students' learning behaviours for readers in digital learning material. A questionnaire, tests, and learning logs were the three (3) approaches employed in this study to collect data. Nine learning behaviours students exhibited when reading instructional materials from the BookRoll system were observed in their learning logs and examined. Turning to the next page and then returning to the previous page is a behaviour pattern of a passive rehearsal approach resulting from correlation. In addition, students repeated readings for text memorization rather than to truly comprehend the material being learned, which decreased their motivation to learn and their sense of self-worth.

Shayan and Zaanen[17] identify weak/strong students throughout the course using data from a collection of courses (Moodle LMS). The outcome must be understandable and interpretable. The intention is to provide this information to lecturers so they can utilise it to enhance their course and recognise students who require more help. All 888 samples that were between the assessment grade and the final exam grade were included in the data collection. Then, the learning behaviours towards the LMS are recorded, including the overall click and online session counts. Two classification algorithms were implemented with WEKA and R, respectively: decision tree J48 and ID3 algorithms.

The output demonstrates that past GPA, a measure of student attributes, had the highest gain ratio in the first half of the course. Performance data (midterm grade) and LMS data (number of views, clicks, and sessions) were crucial during the second half of the semester, especially prior to the midterm and final test. This indicates that all of these characteristics have a significant impact on students' academic performance, and instructors need to take this into account more.

Previous researchers have proved a significant correlation between students' activities in LMS (activities) with the final result. Therefore, based on the evidence prior to the past researcher, the research investigated the relation between UniSZA's student activities with KeLIP (LMS) and the final result.

3. Method

Include the research methodology process involved five phases, starting from problem identification and review, data collection, data analysis, model development, and result evaluation. **Figure 2** shows the structure of the research methodology.

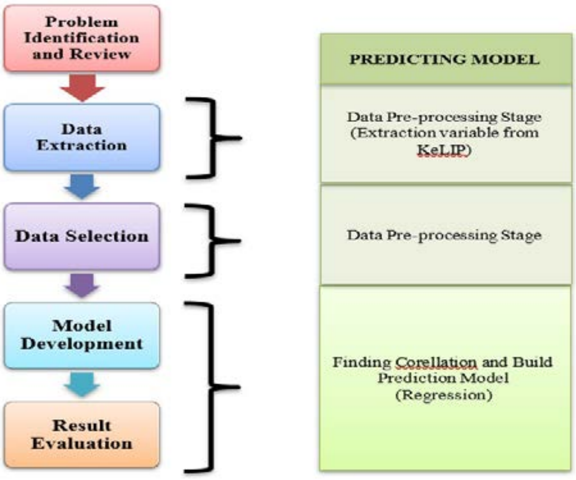

Figure 2. Research framework.

The first stage is data extraction involving the process of obtaining data regarding students' activities in LMS systems and academic systems from two subjects; compiler and data structure. These factors are divided into two main groups:

i. LMS factor–variables derived from LMS.
ii. Non-factor–variables from other systems such as academic systems.

LMS factors derived from KeLIP (Moodle) activities. Moodle activities refer to the interactive and collaborative features of the Moodle LMS. Moodle is a popular open-source platform educational institutions use to deliver online courses and facilitate e-learning. The activities in Moodle enhance the learning experience by engaging students, encouraging interaction, and promoting active participation. Some common Moodle activities include assignments, quizzes, forums, and chat systems.

Table 1 shows both factors. LMS includes activities in learning engagements and forum participation. The selection of LMS factors is based on principal component analysis (PCA) results. PCA is a statistical technique used for dimensionality reduction and data compression. The primary goal of PCA is to transform a high-dimensional dataset into a lower-dimensional space, known as the principal components.

On the other hand, non-LMS factors focus on three main features used in other studies for Malaysia cases; sponsorship, race, qualification, and coursework marks. The selection of non-LMS factors based on paper[13] and also important factors that be selected by stackholders (faculty management). Three subjects were selected to be part of the experiments with a total of 146 students are included to the project. The extraction process also involves the transformation of some data in some cases. Data transformation consists of converting data into a suitable format, aggregating data, applying calculations, and handling data inconsistencies.

Then, the pre-processing (cleansing) process to ensure data are fit to supply in algorithms as shown in **Figure 3**. Data cleansing is part of the process to ensure data quality. The following process is data transformation, whereby the format of data changes before use in the model. In the project, the data transform from text to numeric due to the regression factor only taking numeric values to the model. Therefore, categorical data such as male or female convert to 1 or 0.

Table 1. Features selection for studies.

Features (independence parameter)	Attributes	Total record
LMS-factor	i. Learning engagement (activities) -Crack the software -Quiz–introduction to compiler -Regular expression for websites -Query (using RegExp) -NFA exercises -NFA to DFA -RegExp to NFA	43
Non-LMS factor	ii. Data extracted from academic system -Sponsorship -Race -Qualification	61
Non-LMS factor	iii. Coursework mark	42

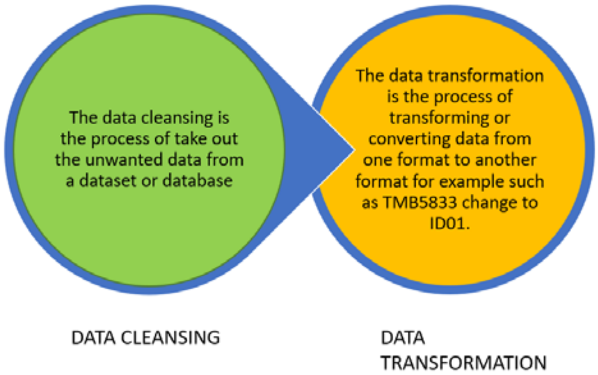

Figure 3. Data cleansing and data transformation.

Next, in the model development phase, the project proposed using regression as a method for prediction. Regression is a statistical method used in finance, investing, and other disciplines to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables). To perform the regression, the first step is to identify the correlation between variables and the target class. In this case, a correlation between students' activities in LMS with the final result. Correlation analysis is a statistical method used to evaluate the strength of the relationship between two quantitative variables. A high correlation means that two or more variables have a strong relationship, while a weak correlation means that the variables are hardly related. One of the popular correlation methods is the Pearson correlation coefficient (PCC), a statistical calculation that measures the linear correlation between two variables, X and Y.

The simple and multilinear regression model selected for this study is due to the success of these algorithms in predicting continuous values, such as market price prediction. It has proved as one of the successful ML approaches for continuous values. Yet, it is still limited studies using regression approaches for education domains. Therefore, as the nature of marks is continuous values data types and some recent success of regression models in LA, it becomes inspirational to model for KELIP data types. **Figure 4** explains how the data can run the prediction model. The data will divide into two groups in this situation. There are the training data and the testing data. The selected data will be trained using the regression model above to get the prediction model. The prediction will be evaluated using average error difference (AED). AED as a method to measure the model's accuracy.

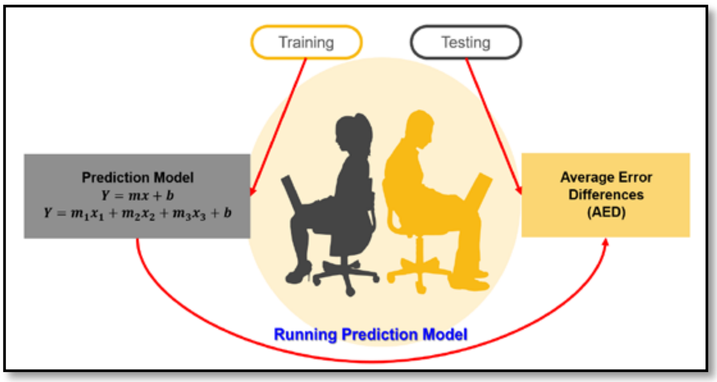

Figure 4. Running prediction model.

The formula explained that the symbol of Δ represented the difference between actual value (γ) and predicted value (γ') as depicted below:

$$|\Delta| = y - y^1$$
$$\text{avg} = \sum\Delta/N \tag{1}$$

where, *N* refers to the number of tuples for testing data.

There are four experiments process running as shown in **Figure 5**. These experiments were classified into two factors. There are LMS factors and non-LMS factors. Experiments 1 and 2, Submission task activities by student and forum participation, are part of LMS factor. Meanwhile, non-LMS factors are divided into experiment 3, student information involving qualification, race and sponsorship, and experiment 4, coursework mark.

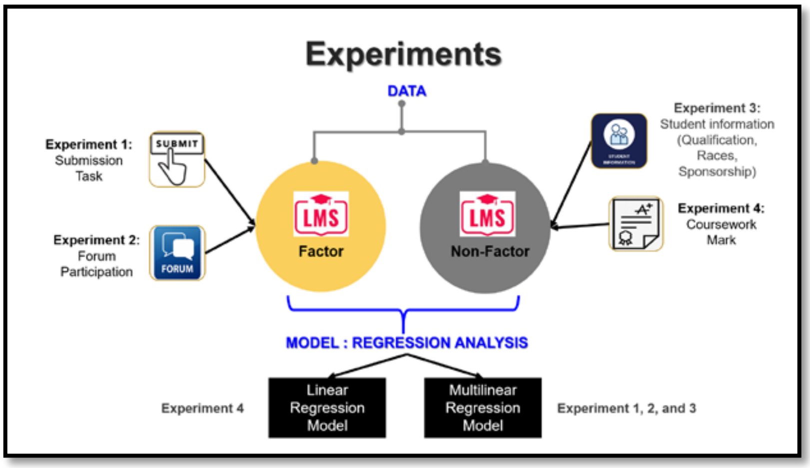

Figure 5. Experiments design.

4. Result and discussion

4.1. Experiment 1

The first experiments were carried out by looking at submission status (1 = submitted, 2 = not submit) for each task the lecturer assigns for the compiler subject. The result of implementing the MR to the training data produces a formula to perform prediction:

$$\begin{aligned} y = &-2.273x_1 + 3.011x_2 + 23.37x_3 + 0.293x_4 + (-3.421x_5) + 2.965x_6 + 0.945x_7 \\ &+ 55.302 \text{ (intercept value)} \end{aligned} \tag{2}$$

The formula is used to predict the training data's regression model, and the result is shown in **Table 2**.

Table 2. Predicting result for experiment.

Student	Actual values (Y)	Predicted values (Y')	Differences $\Delta = y - y'$	$\lvert\Delta\rvert$
1	83.1	82.34	0.70	0.70
2	80.2	82.34	−2.19	2.19
3	84.2	82.34	1.80	1.80
4	78.8	79.92	−1.12	1.12
5	90.2	79.39	10.81	10.81
6	74.6	83.05	−8.45	8.45
7	90.8	75.96	14.83	14.83
8	83.6	79.92	3.68	3.68
9	83.6	82.40	1.20	1.20
10	80.8	78.98	1.82	1.82
11	72.1	76.91	−4.81	4.81
-			Summation $\sum\Delta$	51.44
			N samples	11
			AED	4.68

AED for the first experiment is 4.8, which is a promising achievement. Six instances indicate an average between zero 0.7 to 2.5, a signal that using learning activities can predict the result, as shown in **Figure 6**.

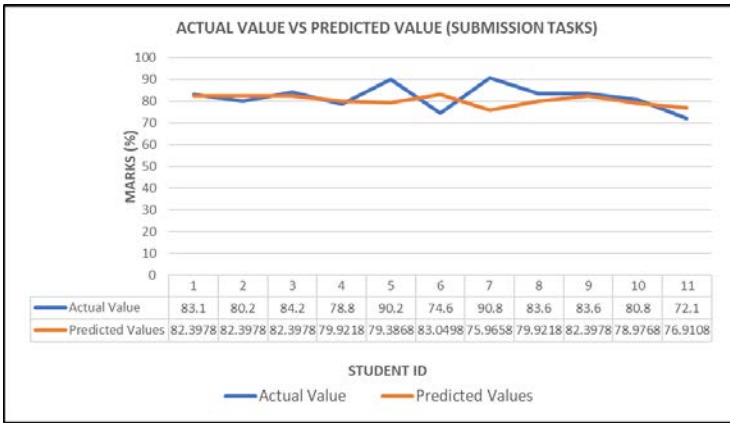

Figure 6. Actual value VS predicted value (submission task).

4.2. Experiment 2

The experiment continues by using total views of all forums during the academic semester for the compiler subjects for the 2nd experiment. There are six forums (features) extracted to develop the model. However, correlation analysis selected only two forums due to the significant value (*P*-value) affecting the final result. The formula generated are:

$$y = 0.446x_1 + 0.836x_2 + 71.583 \text{ (intercept value)} \tag{3}$$

where

$$x_1 = \text{Forum 2 and } x_2 = \text{Forum 5.}$$

Figure 7 shows predicted values compared with actual values with an AED of only 2.78. Experiment 2's results are better than experiment 1, differentiated by 1.9. However, both experiments could use LMS data to predict students' results with AED less than 5.

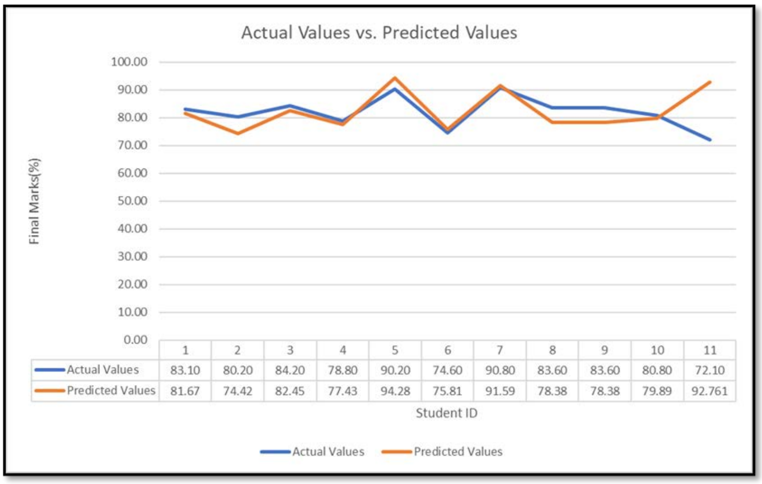

Figure 7. Actual value VS predicted value (forum participation).

4.3. Experiment 3

The experiment continues by investing in non-factors of LMS by looking at three features; qualification, sponsorship and race. Sixty-seven tuples were selected for analysis, where 66% of the data use for training purposes. MR model indicates qualification is the most significant value compared to sponsorship or race with *P*-values less than 0.05 (0.0136). Both race and sponsorship are more than 0.05, with 0.84 and 0.88, respectively. The prediction formula created from the training model is:

$$\mathrm{y} = -0.33\mathrm{x}_1 + (-0.244\mathrm{x}_2) + 3.87\mathrm{x}_3 + 64.46(\text{intercept value}) \tag{4}$$

where,

$$\mathrm{x}_1 = \text{sponsorhip}, \mathrm{x}_2 = \text{qualification}, \mathrm{x}_3 = \text{race}$$

Figure 8 depicts predicting values for experiment 3. AED for experiment 3 is 7.58, indicating the highest prediction from the three experiments.

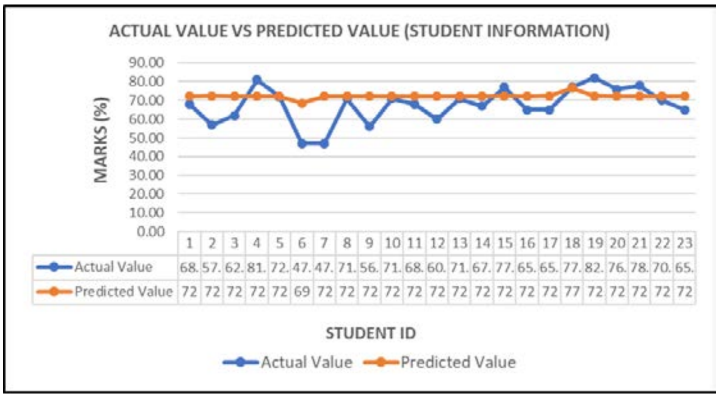

Figure 8. Actual value VS predicted value (student information).

In this case, the result consistently indicates 70% accuracy. The result may reflect the most inconsistent data factors used as training data. However, further detailed investigation must be carried out to explain the results.

4.4. Experiment 4

The last experiment uses coursework as an independent variable by applying simple linear regression (SRM). The result for correlation analysis represents a high positive correlation for the model with R square 0.945, and a *P*-value is 0.186. **Figure 9** shows a positive correlation generated from the model, which is a great prediction model.

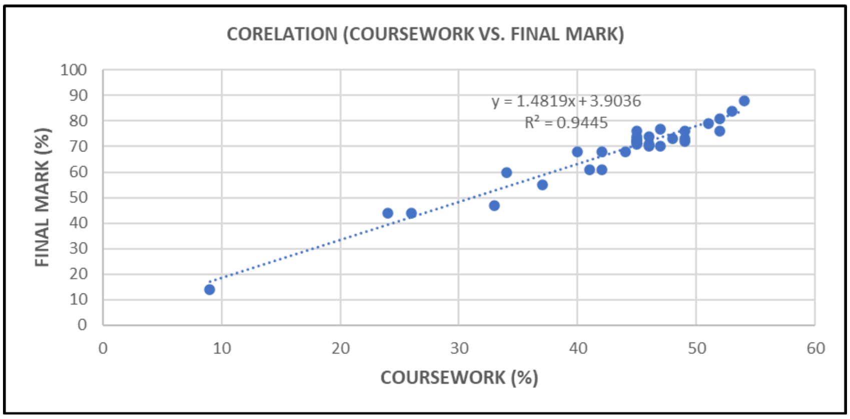

Figure 9. Correlation analysis (coursework mark VS final mark).

The formula for predicting the model is:

$$y = 1.4819x + 3.9036 \text{ (intercept value)} \tag{5}$$

where, $x =$ coursework mark.

As expected, the result obtains a good indicator with the margin average error between + −1.82. Hence, it supports the find out by Conijn et al.[14] that clearly stated the influence of coursework marks on the final result, as depicted in **Figure 10**.

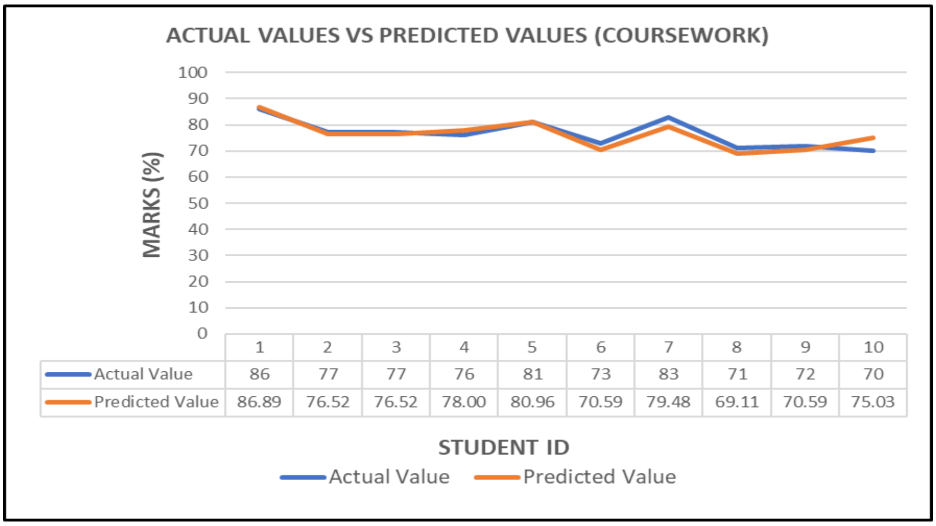

Figure 10. Actual value VS predicted value (coursework).

Four experiments (2 LMS factors + 2 non-LMS factors) have been conducted in this study. The average error differences are used to measure the accuracy of the model. **Table 2** summarises find out from the experiments.

Based on **Table 3**, the best factor indicated by the non-factor (coursework mark) with 1.82, which supported the outcomes of previous research of Conijn et al.[14]. Then it, followed by factors course forum participation (factor) and tasked submissions (factor) with 1.82 and 4.66, respectively. The last result is an experiment for non-factor variables that includes qualification, race and sponsorship with 7.58, as shown in **Figure 11**. In these cases, qualification, races and sponsorship are irrelevant factors in determining students' success. One hypothesis is that the end results are influenced by the participation of learning activities that directly relate to the subjects.

Table 3. Average error differences (AED) for all experiments.

No	Features	AED
1	Factors (Submission tasks)	4.66
2	Factors (Forums participation views)	2.78
3	Non-factor (Qualification, race, sponsorship)	7.58
4	Non-factor (Coursework mark)	1.82

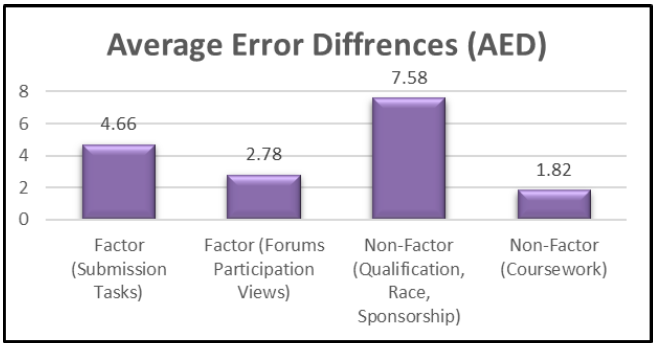

Figure 11. Comparison of average error differences (AED).

The promising result proves that using LMS activities can predict the overall outcome. Although the coursework mark (non-factor) indicates the best result, the coursework mark parameter is also a factor due to the contribution of assessments from LMS activities. It allows specific actions to be taken before students take the final exam to boost their grades.

5. Conclusions

The study proposed investigating UniSZA students' activities with LMS known as KeLIP, such as participating in KeLIP activities (quiz, forum participation, the total number of clicks) with the final result. The pre-defined variables based on literature are selected to find the correlation between variables (students' activities) and the final result. It is an essential factor that can enhance the T & L process by emphasizing each critical variable influencing students' results.

Regression models are crucial in analyzing academic data by establishing relationships between variables and making predictions. In academic research, regression analysis allows researchers to investigate the impact of independent variables, such as study time, attendance, or extracurricular activities, on academic performance, represented by the dependent variable, typically the students' grades or scores. By utilizing regression models, educators and policymakers can gain valuable insights into factors influencing students' academic achievements. Additionally, the model can help identify potential risk factors, such as low attendance, that might negatively affect student performance. Furthermore, regression models can aid in predicting future academic outcomes, enabling early intervention strategies to support struggling students and foster their progress. By analyzing historical data, educators can forecast a student's expected performance based on various factors, allowing for personalized guidance and tailored educational plans.

In our cases for factors that are taken in this case study, the result of four experiments shows the AED with the lowest indicating 1.82 and the highest being 7.58. SLR using the coursework mark as an independent parameter shows the best result with the lowest AED, followed by LMS activities (viewing forum participation). The overall result is significant with the ability of the model to predict the final result. It is the main contribution of the research projects, which support the hypothesis that we can use students' feedback in learning activities to forecast result. Hence, specific actions can be derived to improve students' marks before the final examination based on students' strengths and weaknesses. For example, giving advice focuses on detailed forum discussions for better results.

The scope of the research undertaken in this project is limited to the available datasets for two subjects. Therefore, a direction for future work will be to evaluate the consistency of the model by implementing the model in other subjects in UniSZA. Developing a robust model that can be applied to real problems is crucial. Another issue is the limitation of the study only focusing on the Moddle activities. It is because the research focuses on Moddle engagement with students' final results.

Another critical opportunity for future work is to explore other prediction models to compare with SLR or MR. One method that indicates state-of-art in any domain for prediction is deep learning (DL). Exploring DL in education is a great opportunity based on the abilities DL gives good forecasts in vast and heterogeneous data types, which is the case for the education environment.

Author contributions

Conceptualization, NNNP and AAA; methodology, NNNP; software, NNNP; validation, AAA and SS; investigation, NNNP and AAA; writing—original draft preparation, NNNP; writing—review and editing, WAM; visualization, SS; supervision, AAA; project administration, WAM; funding acquisition, WAM. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work is conducted under the smart technology and system cluster, UniSZA.

Abbreviations

AED, average error difference; BI, business intelligence; DL, deep learning; FIK, faculty informatics computing; KeLIP, knowledge and e-learning integrated platform; LA, learning analytics; SLR, linear regression model; MR, multi linear regression model; principal component analysis (PCA), UniSZA, universiti sultan Zainal Abidin.

Conflict of interest

The authors declare no conflict of interest.

References

1. Santally M, Jegathesan J, Sookhareea R. The impact of learning analytics on student engagement in blended learning environments. *Journal of Educational Technology Development and Exchange* 2018; 1(1): 1–18.
2. Chen W, Liang Y, Liang Y. The impact of personalized feedback on student learning outcomes: Evidence from a learning analytics application. *Journal of Educational Technology Development and Exchange* 2020; 3(1): 1–12.
3. Aljohani NR, Daud A, Abbasi RA, et al. An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior* 2019; 92: 679–690. doi: 10.1016/j.chb.2018.03.035
4. Tsai YS, Rates D, Moreno-Marcos PM, et al. Learning analytics in European higher education—Trends and barriers. *Computers & Education* 2020; 155: 103933. doi: 10.1016/j.compedu.2020.103933
5. Hasan R, Palaniappan S, Mahmood S, et al. eDify: Enhancing teaching and learning process by using video streaming server. *International Journal of Interactive Mobile Technologies* 2021; 15(11): 49–65. doi: 10.3991/ijim.v15i11.20245
6. Miller K. What is learning analytics & how can it be used? Available online: https://www.northeastern.edu/graduate/blog/learning-analytics/ (accessed on 15 August 2023).
7. Aziz AA, Jusoh JA, Hassan H, et al. A framework for educational data warehouse (EDW) architecture using business intelligence (BI) technologies. *Journal of Theoretical and Applied Information Technology* 2014; 69(1): 50–58.

8. Mahmud NIMM, Ismail I, Shamsuddin SNW, et al. Learning performance assessment using mobile-based augmented reality application for preschool environment. *International Journal of Recent Technology and Engineering* 2019; 8(2S3): 436–439. doi: 10.35940/ijrte.B1076.0782S319
9. Chen W, Liang Y, Liang Y. Factors influencing the successful implementation of learning analytics in higher education. *Journal of Educational Technology Development and Exchange* 2021; 4(1): 1–10.
10. Kim YJ, Lee J, Lee Y. Challenges in implementing learning analytics in higher education. *Journal of Educational Technology Development and Exchange* 2022; 5(1): 1–9.
11. Baker RS, Yacef K. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining* 2009; 1(1): 3–16. doi: 10.5281/zenodo.3554657
12. Amrieh EA, Hamtini T, Aljarah I. Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application* 2016; 9(8): 119–136. doi: 10.14257/ijdta.2016.9.8.13
13. Ahmad F, Ismail NH, Aziz AA. The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences* 2015; 9(129): 6415–6426. doi: 10.12988/ams.2015.53289
14. Conijn R, Snijders C, Kleingeld A, Matzat U. Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies* 2017; 10(1): 17–29. doi: 10.1109/TLT.2016.2616312
15. Macarini LAB, Cechinel C, Machado MFB, et al. Predicting students success in blended learning—Evaluating different interactions inside learning management systems. *Applied Sciences* 2019; 9(24): 5523. doi: 10.3390/app9245523
16. Geng X, Xu Y, Chen L, et al. Learning analytics of the relationships among learning behaviors, learning performance, and motivation. In: Proceedings of the IEEE 20th International Conference on Advanced Learning Technologies (ICALT); 6–9 July 2020; Tartu, Estonia. pp. 1–161.
17. Shayan P, Zaanen M. Predicting student performance from their behavior in learning management systems. *International Journal of Information and Education Technology* 2019; 9(5): 337–341. doi: 10.18178/ijiet.2019.9.5.1223