International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol. 9, Issue.9, September 2021 E-ISSN: 2347-2693

A Critical Analysis of Techniques Used For Learning Analytics Corresponding to Buiness

Kajal Devi[1*], Harjinder Kaur[2]

[1,2]Department of Computer Science, Swami Sarvanand College of Engineering and Technology, Punjab Technical University, Jalandhar, India

[*]*Corresponding Author: dkajal077@gmail.com, Tel.: +91-9501497689*

DOI: https://doi.org/10.26438/ijcse/v9i9.3944 | Available online at: www.ijcseonline.org

Received: 12/Sept/2021, Accepted: 20/Sept/2021, Published: 30/Sept/2021

Abstract- Data Mining plays an important role in the Business world and it helps to the marketing institution to predict and make decisions related to the business' academic status. Predicting business' performance becomes more challenging due to the large volume of data in marketing databases. Currently in Malaysia, the lack of existing system to analyse and monitor the performance of the business is not being addressed. There are two main reasons of why this is happening. First, the study on existing prediction methods is still insufficient to identify the most suitable methods for predicting the performance of the business in Malaysian's institutions. Second, Due to the lack of investigations on the factors affecting student's achievements in particular courses within Malaysian context. Therefore, a systematically literature review on predicting student performance by the proposed system is a web based which makes use of the mining techniques for the extraction of useful information.

This work is dig insight into state and event-based approaches for predicting student performance. Comparative analysis is conducted to suggest regression-based algorithms of state-based framework lack accuracy and correlation-based algorithms under event driven approach outperforms classical regression algorithms. It is also concluded from pedagogical point of view, higher engagement with social media leads to higher final grades

Keywords— Performance Prediction, Learning Analytics, Regression algorithm, correlation algorithms, social media

I. INTRODUCTION

Student performance analytics becomes important for predicting performance of business in course of study. Learning analytics in that regards is growing research area that selects, analyse and report student data, find patterns from student behaviour, display information in suggestive formats with end goal to predict student performance, achieving optimization of prediction system and customization of personalized intervention. Learning analytics is framework that is used for measurement, analysis and reporting of data about learner for understanding and optimizing learning and environment in which it is considered. Distinct marketing tasks are supported by the use of learning analytics. These tasks are categorized into seven categories: Monitor and analysis, intervention and prediction, feedback, modification, customization and recommendation, reflection.

Predicting performance of student is one of most important objective of learning analytics. Learning analytics not only predict future performance but also set performance indicators to improve performance of business. Data processing is critical aspect of learning analytics as it is required for monitoring learning progress of business, instructor can be advised to change study patterns for business who need more assistance. In addition, individual strategies for poor business with personalized intervention and feedback can be set for guidance of business using application of predictive modelling. The utilization framework for predictive modelling along with used metrics is elaborated in figure 1.

The primary component associated with LA is learning that could be distinguished as supervised and unsupervised mechanisms. These mechanisms form a effective part of performance and monitoring module of LA. Effectiveness of learning analytics is judged in terms of classification accuracy. The classification accuracy of learning analytics must be considerably high otherwise wrong predictions could have adverse effect on career of business

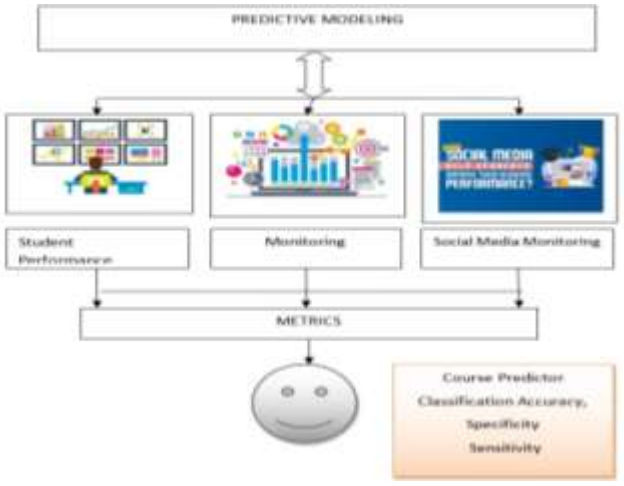

Figure 1: Framework for learning analytics

Rest of the paper is structured as under: section 2 present related work of techniques used for performance monitoring, section 3 gives the algorithm analysis corresponding to state and event driven algorithm, section 4 compare state and event driven algorithm, section 5 gives conclusion and future scope.

II. METHODOLOGY

The review process included the following steps:

1. Describing inclusion and exclusion criteria
2. Defining the work analysis of student performance monitoring
3. Explaining the existing techniques that are used for student performance monitoring
4. Defining the problems with existing techniques and gives comparative analysis

1. Describing inclusion and exclusion criteria

After an initial selection of 24 papers, the relevant studies were determined following the inclusion/exclusion criteria as follows.

Inclusion Criteria

• Include the papers discussing longitudinal analysis of evolution of student performance systems' structure, and giving empirical evidence thereof;

• Include the research studies that combine evolution of student performance monitoring software structure along with evolution

• Include studies which propose different strategies to analyse and/or predict student performance monitoring;

• Include studies on various techniques used for student performance evolution for complete the picture wherever need be. Though such studies were not part of the initial selection, they were included by following the references in the selected set of studies.

Exclusion Criteria

• Exclude research studies focusing only on the monitoring;
• Exclude the studies which don't report any empirical results;
• Exclude Review Studies.

Figure 2: Inclusion and exclusion mechanisms

2. WORK Analysis

The significant part of learning analytics is predictive modelling for learning and teaching; main motive of this is to forecast success of student in terms of achievement in academic(Thai-Nghe et al. 2010)[1]. Knowledge, score or grade, performance; can be the predicted values; classification approaches are basically used for values that are discrete/categorical, and approaches of regression for values that are numerical/continuous.

For predictive models two types of data is used that is (i) state-based data- e.g. Past performance, demographics, psychological traits; (ii) event –driven data- it is based on activity of student, it is derived from interaction of student with marketing system and resources(Osman Begovic, E., Suljic 2012)[2]. The latter can be in the form of structured(e.g., sever logs) or unstructured(e.g., forum postings)(Kaur et al. 2018)[3].It can be derived from marketing system that are centralized(e.g., LMS) or learning environment that is distributed (e.g., formal and informal platforms, spread across space, media and time)(Lopes 2017)[4]. The sources of data include MOOCs(massive open online courses), social media or wearable sensors, the integrity of these are directing to more accuracy of the models of learner(Kroft et al. 2014)[5].

A additional classification of the indicators of performance, given in (Khder 2018)[6], includes three types:
i) dispositional indicators (e.g., age, gender, previous learning experiences); ii) activity and performance indicators (e.g., no. of logins, time exhausted, number of discussion posts); iii) student artifacts (e.g., essays, blog posts, forum discussions) (Jasmer 2015)[7]. About 200 indicators were recognized in a review done in (Kakkar 2016)[8]; within them, frequently used are: demographic characteristics, prior grades, portfolios, multimodal skills, levels of partaking and commitment, mood and affective states (Wang and Graduate 2015)[9].

So far as computational techniques are concerned, large no. of methodologies have been applied for predicting performance of business(Gitinabard et al. 2014)[10], such as linear regression(Zhang et al. 2018)[11], logistic regression (Loh 2006)[12], neural network models (Feng et al. 2015)[13], support vector machines and k-nearest neighbours (Eashwar and Venkatesan 2017)[14], Bayesian networks(Bekele and Menzel 2017)[15], decision trees(Rosenfeld et al. 2018)[16] or genetic algorithms(Ix et al. 2016)[17]. While provided that a whole review of literature is ahead of scope of this paper, in what follows we express a few (recent) proposals in academic performance prediction, which are more narrowly connected to our work.

(Hoic-Bozic et al. 2015)[18] explored business' usage data in Moodle LMS as a forecaster for their grade of exam. 438 business from seven engineering courses were included in the study.

Eight attributes related to learner action on quizzes, assignments and forum messages were calculated for each and every student. For classifying business with similar grades (statistical classifier, decision tree, rule induction, fuzzy rule learning, neural networks), authors applied various data mining techniques. Performance comparisons were carried out, with various pre-processing techniques. On the whole, the accuracy obtained is not very high (around 65%), representing the complexity of the prediction task.

(Kabra and Bichkar 2016)[19] also analysed business' online activity in a LMS, in the framework of a blended medical course, aiming to show a relationship it with the learners' _nil performance. 133 business used LMS for six weeks and numerous data is collected: logins, resource views, forum posts and reads, time spent using marketing materials, formative evaluation results. Five engagement indicators were calculated depending on business' traces.

For grade prediction automatic linear modelling was used, leading to a 63.5% accuracy. Additionally, binary logistic regression was employed for predicting business at danger, with an accurateness of 80.8%.

(Pinjuh 2015)[20] focused on the use of business' contribution in a conversation forum as an displayer of learner performance. Data was taken from 114 business that are enrolled in an introductory computer science course. They used the forum that are included in Moodle LMS for discussing the contents of course, asking questions or providing help to peers and took a _nil exam at the last of the semester. The authors intended to forecast whether business passed or failed the course based on their forum usage, in conditions of quantitative, qualitative and social network indicators. A contrast between traditional classification and clustering algorithms implemented in Weka was performed, together with various approaches for instance and attribute selection

.

(Chen et al. 2019)[21] investigated business' interaction patterns with digital textbooks as predictors of _nal course grades. Data was collected from 233 business from 11 courses (such as Introduction to Accounting, American Judicial Process, Human Resource Management etc.) they take into account a digital textbook presented by Course Smart supplier. The authors performed linear regression analysis on textbook practice metrics and found out that time used up reading was a strong forecaster of _nal course grade. The Engagement Index score (computed by Course Smart based on various usage metrics) was also a good display of the course result (better than prior academic achievement). As it can be seen, most of the above studies were performed in the context of Moodle LMS; the social learning environments and business' social media traces were much less explored in the literature.

(Carter 2017)[22] used mixed effects analysis of variance (ANOVA) models to evaluate the impact of using Twitter on college student engagement and learning outcome. Engagement was measured with a dedicated instrument called National Survey of Student Engagement. Results showed a significant increase in both engagement and grades for the experimental group, in which business used Twitter for various types of academic discussions. (Popescu and Leon 2018)[23] investigated whether business' engagement with a collaborative wiki tool can forecast marketing performance. Significant correlations were found between wiki activity indicators (number of page edits, number of different articles edited, number of days on which the student edited the wiki) and the _nal grade. On the whole the business who were busy with the wiki (both high- and low-grading ones) obtained higher exam scores, with an average increase of 5 percentage points. The originality of our existing work consists in the utilize of an novel algorithm, called Large Margin Nearest Neighbour Regression (rather than classic algorithms, available in a variety of data mining engines, as mentioned in the related works). A initial study depending on one student only cohort yielded hopeful results (Amra 2017)[24]; this paper extends the pilot study to a much outsized number of business (six cohorts, over the course of six years), also providing a requirement of the LMNNR technique, as described next

3. BACKGROUND ANALYSIS OF ALGORITHMS

State and event driven algorithms are considered for evaluation through this literature. The state driven algorithms are dependent upon historical data and strictly static in nature.

The comparison of techniques using state and event driven approaches is given in table 1.

The mechanism followed in state driven approaches is generally historical data based and hence convergence rate is generally a problem but with even driven approach data is real time in nature and convergence rate is also good. This is demonstrated within table 1 also.

Table 1: Parametric comparison of state and event driven approach

Parameter	State driven approach	Event driven approach
Dataset	Historical	Real time
Convergence Rate	Low	High
Specificity	Specificity measure degree of false positive rate that is high in state driven approaches	Low specificity indicates better performance
Sensitivity	Sensitivity indicating degree of true positive is low	High sensitivity is accomplished through this approach
Accuracy	Classification accuracy is hampered by noisy data	Classification accuracy is rectified using pre-processing mechanisms
Learning rate	Learning rate is about 60-70%	Learning rate is about 80-90%
Coupling	High coupling indicating high dependency on modules	Low coupling indicting modules can be changed without affecting other modules
Cohesion	Cohesion is high	Cohesion is low causing performance degradation in changeable entity
Mobility	Modifiable data does not degrades accuracy	Modifiable data causes classification accuracy to decay considerably.

The classification of state and event driven approaches is presented in figure 2

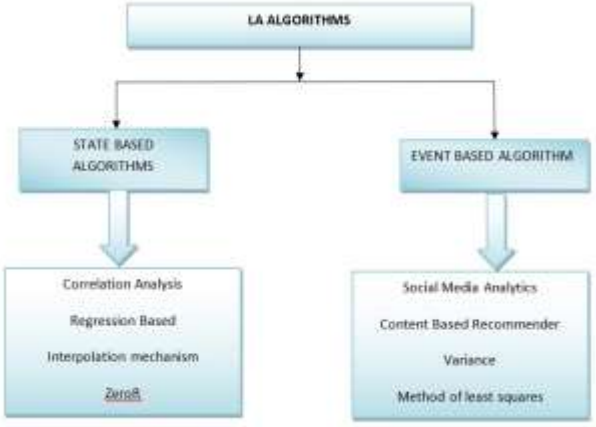

Figure 3: Algorithms for performance prediction

The various algorithms that are used for performance prediction is as given below

3.1 State Based Algorithm

This is the one which is based on the previous data analysis. In this the data that are stored utilized for further analysis.

3.2 Correlation Analysis

It is widely used technique that is used to identify relationships in data that helps in predicting the target classes. This measures how the variable is predicted using another set of values. It measures the relationships between two variables and the result shows the effect of one variable change to another variable. The comparison of two variables correlation is to be done and the changes are seen. It analysis the dependent variable and independent variables that help in predicting the values.

It decides the quality of a relationship between two sets, which can be a dependent and a free factor or even two autonomous factors. In such a case, the quality can be distinguished dependent on bearing, structure, and scattering quality. Numerically, this relationship is normally dictated by a decimal esteem, known as the correlation coefficient. The correlation coefficient is resolved under a certain predefined run (contingent upon the calculation). In light of the estimation of the coefficient in the given range, its quality and course can be resolved. The coefficient having positive sign shows that the two factors are decidedly related, though negative sign demonstrates negative relationship. Higher number of coefficients demonstrates that the two factors have solid connection and lower esteem shows generally. For instance, Lift is one relationship measure with a coefficient going around one. However, that the esteem is more prominent than one, at that point the two factors are decidedly associated; else, they are contrarily connected. On account of the esteem being 1, there is no relationship. This relationship encourages us distinguish thee independent and dependent variable.

Consider two variables A and B. Then the Pearson's correlation coefficient can be calculated using the following formula

$$C_{A,B} = \frac{Covariance(A,B)}{\sigma_A \sigma_B}$$

3.3 Regression Analysis

Regression analysis is statistical based technique that is used to establish relationship between dependent variables and the predictors set. It is data mining technique that utilizes supervised learning for prediction. In supervised learning the database is divided into validation and training data that are used in regression analysis. Regression Analysis is a measurable apparatus that uses the connection between at least two quantitative factors with the goal that one variable (subordinate variable) can be anticipated from the other(s) (autonomous factors).

In any case, regardless of how solid the factual relations are between the factors, no circumstances and logical results example is fundamentally suggested by the regression model. Regression analysis comes in numerous flavours, including direct, different direct, curvilinear, and various curvilinear regression models, just as strategic regression. Strategic Regression is utilized when the reaction variable is a twofold or subjective result. Also, it calculated regression finds a "best fitting" condition similarly as direct regression does, the standards on which it does as such are somewhat unique.

3.4 Interpolation Model

The way through which the unknown values are found from known values is known as interpolation. In this the knowledge of two points are require and also the rate of change must be constant. The tabulated functions are expensive in order to store large set of values so the arbitrary values of the arguments are calculated.

For example: the value of (X_i, Y_i) is calculated using function Y=f(x). where i -0,1,2,3……. N. the value of y is estimated using the intermediate values of x. this process is known as interpolation.

3.5 Regression based analysis Zero

This technique uses analysis to build classifier and then classification is done. In classification data are predicted by using class labels.

These datasets are further divided into test set and training set. Further analysis is done using this test set and it randomly sampled dataset. The tuples remaining that are not used to build classifier are independent of training set and dataset. The classifiers accuracy is estimated using test set. It will give the test tuples that are classified correctly. It uses cross validation to predict the higher accuracy.

3.6 Event based Analysis

These analyses are based on the real time data set values. It is also used to predict the values. The various techniques that are used is as given below:

3.7 Social media analytics

Social media analytics (SMA) consist the methodology of gathering information from social media destinations and writes and assessing that information to settle on business choices. This procedure goes past the typical observing or an essential investigation of re-tweets or "preferences" to build up a top to bottom thought of the social purchaser. This is viewed as the essential establishment for empowering a venture to:

• Execute centred commitment like balanced and one-to-numerous

• Enhance social joint effort over an assortment of business capacities, for example, client administration, promoting, support, and so on.

• Maximize the client experience

Social media is a decent medium to see continuous buyer decisions, aims and opinions. The most pervasive utilization of social media analytics is to become acquainted with the client base on a progressively enthusiastic dimension to help better target client administration and promoting.

3.8 Content Based Recommender

The recommender system is based on the data that is provided by the users using implicit or explicit methodology. On the basis of this data the profile of user is generated that are further used for making suggestions. According to these suggestions users takes action and the recommendation becomes more or more accurate. It utilises the terms term frequency and inverse document filtering that are used to find importance of documents within system.

3.9 Variance

It is used to measure spread between the numbers in dataset. In this first of all mean value is calculated by taking the difference of each number in set and then mean of squaring of values are taken after that sum of all values in the set is done.

3.10 Variance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - mean(x))^2}{n}$$

where:

xi=the it data point

$\bar{x}$=the mean of all data points

n=the number of data points

3.10 Method of least square

It is a statistical technique that is used to find line of best fits for the set of values. This provides the visual representation of the relationships between various data points. It describes the relationship between dependent and independent variables using each point in variable.

The least squares method gives the general justification to the situation of the line of best fit among the information focuses being examined. The most widely recognized use of the least square's method, which is alluded to as direct or standard, expects to make a straight line that limits the aggregate of the squares of the mistakes created by the aftereffects of the related conditions, for example, the squared residuals coming about because of contrasts in the watched esteem and the esteem foreseen dependent on the model.

III. PROBLEM DEFINITION

The existing SVM technique is uses prediction rate that are used for predicting the performance of student. But it does not handle misleading data properly and the classification accuracy is less. The state-based approaches uses previous data set for prediction but the current values ignored so the prediction rate is not good. The event driven approaches utilize real time values and gives relationships among these values so that predict rate accuracy improved but it does not considered historical data therefore the performance of student cannot be accurate.

The comparative analysis of these techniques is given as below:

Technique	Advantages	Disadvantage
State driven approaches	It is used to handle any problem with the extracted values from dataset and also predict relationships among these values	Missing values causes the problem and classification accuracy is a problem
Event driven approaches	Useful marketing data are extracted for predicting the student evaluation at early stage using real time data .	There is no mechanism present that group both real time and historical data to accomplish greater classification accuracy

IV. CONCLUSION AND FUTURE SCOPE

In this paper we analysis various techniques used to predict student performance. It gives detailed review of state driven and event driven algorithms that are used for prediction. Also, comparative study is done that elaborate limitations of existing techniques. It explained that regression-based algorithms of state-based framework lack accuracy and correlation-based algorithms under event driven approach outperforms classical regression algorithms. It is also concluded from pedagogical point of view, higher engagement with social media leads to higher final grades. For resolving this problem, we proposed least square method that has more accuracy of predicting performance of business.

REFERENCES

[1] Amra IAA. Business performance prediction using KNN and Naïve Bayesian. Res Gate. **2017** doi: 10.1109/ICITECH.2017.8079967

[2] Bekele R, Menzel W. A BAYESIAN APPROCH TO PREDICT PERFORMANCE OF STUDENT(BAPPS): A CASE WITH ETHIOPIAN BUSINESS. IJAER, **2017.**

[3] Carter BJ (2017) An ANOVA Analysis of Education Inequities Using Participation and Representation in Education Systems

This is to certify that the doctoral dissertation by. IEEE Access
[4] Chen J, Feng J, Sun X, Wu N, Yang Z, Chen S (2019) MOOC Dropout Prediction Using a Hybrid Algorithm Based on Decision Tree and Extreme Learning Machine. IEEE Access 2019:
[5] Eashwar KB, Venkatesan R (2017) STUDENT PERFORMANCE PREDICTION USING SVM. IJMET 8:649–662
[6] Feng W, Tang J, Liu TX (2015) Understanding Dropouts in MOOCs. IEEE Access
[7] Gitinabard N, Khoshnevisan F, Lynch CF, Wang EY (2014) Your Actions or Your Associates ? Predicting Certification and Dropout in MOOCs with Behavioral and Social Features. IEEE Access
[8] Hoic-Bozic N, Holenko Dlab M, Mornar V (2015) Recommender System and Web 2.0 Tools to Enhance a Blended Learning Model. IEEE Trans Educ 1–1 . doi: 10.1109/TE.2015.2427116
[9] Ix DQG, Dojrulwkp J, Dpvd D, Kdnnhwkrwwdp QL (2016) student academic performance prediction model using decision tree and fuzzy genetic algorithm. Elsevier 25:326–332 . doi: 10.1016/j.protcy.2016.08.114
[10] Jasmer AM (2015) DISPOSITIONAL EMPLOYABILITY AND THE RELATIONSHIP TO CAREER SUCCESS : A META-ANALYSIS. CSUSB Sch Work
[11] Kabra RR, Bichkar RS (2016) BUSINESS ' PERFORMANCE PREDICTION USING GENETIC ABSTRACT : IJCEA VI:19–29
[12] Kakkar H (2016) The Dispositional Antecedents of Promotive and Prohibitive Voice. J Appl Psychol 101:1342–1351
[13] Kaur A, Umesh N, Singh B (2018) Machine Learning Approach to Predict Student Academic Performance. IJRASET 6:734–742
[14] Khder M (2018) A Classification and Prediction Model for Student ' s Performance in University A Classification and Prediction Model for Student ' s Performance in University Level. Res Gate. doi: 10.3844/jcssp.2017.228.233
[15] Kloft M, Stiehler F, Zheng Z, Pinkwart N (2014) Predicting MOOC Dropout over Weeks Using Machine Learning Methods. IEEE Access 60–65
[16] Loh W-Y (2006) Logistic Regression Tree Analysis. Handb Eng Stat 537–549
[17] Lopes AP (2017) LEARNING MANAGEMENT SYSTEMS IN HIGHER EDUCATION. Res Gate
[18] Osmanbegovic, E., Suljic M (2012) Data mining approach for predicting student performance. J Econ Bus X:3–12
[19] Pinjuh A (2015) Learning Management system logs. Res Gate. doi: 10.1109/SOFTCOM.2015.7314114
[20] Popescu E, Leon F (2018) Predicting Academic Performance Based on Learner Traces in a Social Learning Environment. IEEE Access 6:72774–72785 . doi: 10.1109/ACCESS.2018.2882297
[21] Rosenfeld A, Sina S, Sarne D, Avidov O, Kraus S (2018) A Study of WhatsApp Usage Patterns and Prediction Models without Message Content. IEEE Access 1–24
[22] Thai-Nghe N, Drumond L, Krohn-Grimberghe A, Schmidt-Thieme L (2010) Recommender system for predicting student performance. Procedia Comput Sci 1:2811–2819 . doi: 10.1016/j.procs.2010.08.006
[23] Wang W, Graduate I (2015) Deep Model for Dropout Prediction in MOOCs. IEEE Access
[24] Zhang S, Yao L, Sun A, Tay YI (2018) Deep Learning based Recommender System : A Survey and New Perspectives. CSIR 1:1–35

AUTHORS PROFILE

Miss kajal Devi pursed Bachelor of Science from Swami Sarvanand college of Engineering and Technology,Dinanagr since 2018 and currently pursuing Master of Science from Swami Sarvanand Group of Institutes, DinaNagar Punjab since 2018. This is my first paper that I am publishing in international journals computer science and Engg..

Mrs Harjinder Kaur pursed Bachelor of Science from PTU campus, Kapurthala and Master of Science in CSE from BCET, Gurdaspur. She is currently working as Assistant Professor in Dept. of CSE, Swami Sarvanand Institute of Engineering and Technology, Dinanagar Punjab. She has 12 years of teaching Experience in the Field of Computer Science and Department

