Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

MODELLING THE LEARNERS' BEHAVIORS USING CLASSIFICATION ALGORITHMS IN WEKA ENVIRONMENT

S. Bhuma Devi Research Scholar, Department of Mathematics and Statistics, Banasthali Vidyapith - 304022, Rajasthan, India

Abstract
The learning environments are essential part of training and learning process for every country. Learners' study-life balance could be transformed through the effects of learning activities in learning eco-systems. Sentiment analysis could be done for adaptability of situational necessity in evaluation process, along the methods of learning in offline, online and hybrid. Machine learning classification algorithms like a rule-based, decision tree, Bayes and combination of logistic regression and decision tree classification models which are the inbuilt in WEKA-environment. The real time problems solutions or inputs could be drawn from primary datasets, it was observed that rule-based model's accuracy was more acceptable for the conditional situation for assessment process in virtual learning scenario. Knowledge from the decision tree model was most evidently indicating that the direct contact method of learning leads a positive approach in maintaining relation among the family members.

Key words: Learners' behavior, Learning environment, study-life balance, Machine learning classification algorithms.

I. Introduction

The quality of humans' life could be assessed based on their life-style which could be derived from their mental, physical fitness. The gist of these qualities could be availed from the existing learning eco-systems. Learning eco-systems should possess the experiential learning features. The development of skill among the individuals is inclined towards their learning behaviors. Higher education learners' behavioral part is one of the most significant aspects in acquiring knowledge for their betterment.

Machine learning part of artificial intelligence are inseparable catalyst in fetching information from the primary and secondary educational datasets. Simple and multiple linear regression models are very familiar in education sector. (Bhuma Devi & Jain, 2020) discussed about the simple linear regression model of the higher education learners' performance in usual and online methods of assessments for prediction, and analyzed about the factors like effectiveness of online teaching, educator's audibility, concept acquisition level and learners' recalling level during the virtual learning process. (Bhuma Devi & Jain, 2021) analyzed about the requirement of weekly preparation time for the learners depending upon number of subjects and number of online lectures during online learning using multi-linear regression model.

Learners' opinion about the online learning, expectations towards method of learning, difficulty level of assignments and compatibility of online assessments, regular learning activities like classwork, homework and relation with peer-group and family members are essential and unavoidable aspects for the humans' learning process. Insights from the ML-models will be studied, based on metrics like accuracy and kappa statistic. Comparison of classification ML-models and its metrics and special outputs of the analysis will be discussed in this study.

The main objectives of this paper are,

- To analyse the learners' preference about the online assessment using machine learning classification algorithms.
- To extract the most significant feature which affects the learners' study-life balance, especially with their relations with family using classification algorithms.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

II. Related work

The learners' learning behavior, their mental health, study-life balance and their emotions towards the virtual learning have been studied and modelled using clustering, tree-based, and many classification algorithms for getting insights for the real time learning related problems. Models' metrices like recall, precision and F1-scores had been observed and compared in the following literature.

(Ratnapala, Ragel, & Deegalla, 2014) compared two courses among which first course was non-graded second was graded course for full semester in MOODLE plate form. Authors studied about the learners' online self-directed learning behaviours using clustering algorithm and concluded that the online learning could be improvised by introducing proper assessment process.

(Golino & Gomes, 2014) discussed the essentiality of tree-based machine learning models in predictions of academic performance of medical students, from the second and third year of a private Medical School from the state of Minas Gerais, Brazil. Psychological and educational attributes like approach of learning, educational background, recalling and application ability had been considered in model creation. Sensitivity of Boosting model was very high among four models, Bagging and Random Forest were stable models in this work.

(Srividya, Mohanavalli, & Bhalaji, 2018) investigated about the mental health features of two distinct targeted groups, first was between 18 to 21, second was 22 to 26 years. Labeling the target variable had been done using clustering algorithm. Authors discussed about the scaffolding for assessing metal health and using classification algorithms like Logistic regression, Naïve Bayes, support vector machine, Bagging, Decision tree, KNN and Random Forest had been modelled the behavioral features for the predictions.

(Shen & Yuan, 2021) utilized the clustering ML algorithm to analyze learners' study-life quality, form the data collected from the digital campus platform all-in-one card system of a University from China. In this study, authors observed 2017 learners' spending, learning and living patterns with the data analysis, through which they could find a group of learners' behavioural, economical and learning issues. Considerable recommendations had been suggested for the quality improvement, for the better learning environment to the learners, educators and managements of that particular University.

(Yayla, Yayla, Ortaç, & Bilgin, 2021) proposed a classification model using natural language processing for understanding the people's emotions towards distance education. Tweets in Turkish language had been considered for this study. First, the tweet-comments will be translated into English language, secondly those statements will be analyzed as supported, not-supported and neural comments against virtual learning. Lastly, the proposed model will be classifying those comments into distinct categories. From against and neural comments' emotions had been considered for solving problems. Authors concluded that some of the problems could be solved using social media discussions with the help of NLP-modelling.

(Webb, et al., 2021) discussed about the similarity and importance of the machine learning and human learners. Machine learning systems could be converted to deep learning systems which will be useful to solve the social and technical problems. For creation of such systems the learners should be trained from the school level education. This could be possible by providing sustainable resources, creating trained educators, finding the gaps between the actual and existing systems and researches should flow in the same directions.

(Hamid, Ismail, Hamzah, & Malik, 2021) developed a student engagement model, for continuing learning activity effectively in online mode, using learning management system (LMS) in a public Malaysian University. With the help of literature review in learning analytics field, authors included the activities like collaboration, submission, view and discussion to improve the student engagement model. Data had been collected form LMS, explored, analyzed and validated the model with experts. Authors had concluded that the behavioural pattern of other Malaysian Universities students could be observed and studied using this above-mentioned model.

(Yilmaz, 2022) analyzed, the relation between the online skill management in self-regulation learning activities, acceptance of process of virtual learning and recognizing the learning. Data of 415 learners

had been collected from a particular University of country Turkey. Author studied the relations between the attributes through structural equation modelling and concluded that the proper self-regulatory aspects should be introduced for optimizing the online learning.

III. Methodology

WEKA datamining tool have been used to compare machine learning models. Classification and decision tree ML-models have been deduced by exploring the datasets in WEKA environment.

Datasets

Understanding the behavioral aspects of learners will be helpful in drawing progress steps for designing better learning environment. Insights and knowledge will be derived with the help of machine learning models.

Two data sets have been considered for this study.

- First dataset is primary data which had been collected from 219 engineering learners as their opinion. Related to their approach towards online learning, comfort level of online assessment, acceptance level in assignment and revision activities, leaners' opinion about instructors' knowledge and comfort agreement in mode of online assessments (Yes, No).
- Second dataset had been taken from Kaggle website.

Data collection

During online learning time of COVID-19 pandemic period, all the teaching-learning processes were performed in virtual mode. Receiving and mounting in those situational exercises of learning were bit difficult for the learners and educators. Maintenance of harmony in that process was a prime motive for the stakeholders. Educators were in the position to consider learners' background, metal and physical health to overcome crisis. Feedbacks and opinions were the only medium to handle the situation better, which have been taken through Google form.

After the pandemic situation the learning environments, learning activities and learners' attitude towards learning and keeping the relation with their peer and family got affected. The second dataset which was taken form Kaggle website for the study purpose for analyzing the learning-life balance.

Data pre-processing

Proper perceptions could be gained about the factors which are affecting learners' behaviors like the learning environment, learning activities and relationships with friends and family using machine learning models. Microsoft excel tool has been used to pre-process the data. The primary data which had been collected from the first- and second-year engineering learners were arranged properly. Dependent and independent variables had been set. As per the requirement of the datamining WEKA tool, for the second dataset the dependent variable had been converted as categorical data.

Selection of machine learning model

Distinct ML-models could be extracted for the processed of datasets, through WEKA-environment. ML-models could be trained, tested and validated in the same environment. The ML-models' classification accuracy, kappa statistics could be compared and studied for getting the suitable knowledge from the data.

IV. Model predictions and evaluations

Discussion about Assessment compatibility

During adaptation stage of virtual learning process, the educators were in a position to make learners comfortable for new normal learning situation. Sentiment analysis was essential to cope up the new state of learning process. Main purpose of learning is to adapt new knowledge for the better cognitive development or to use new process for the task completion or to acquire new skill for the betterment. Assessment is the unavoidable aspect in the training and learning process.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

Compatibility of online assessment among the engineering learners was studied with binary classification problem. Learners' opinion had been taken through Google-form, the following questions were the part of questionnaire

Table 1 Questions and Categories

Sr.No.	Questions	Categories		
1.	Are you comfortable with online lectures?	Easier	Neither easy nor tough	Tougher
2.	What kind of moderation do you expect in the way of teaching and learning process?	Speed of explanation of the concepts should be reduced	One to one correspondence for clarity of the concepts.	An increase in number of solved examples
3.	The instructor was very knowledgeable about the topic that was taught	Fair	Good	Excellent
4.	Difficulty level of Assignment	Easier	Normal	Tougher
5.	Are you comfortable with same evaluation pattern for the internal assessment?	Yes	No	

ML-Classification algorithms' prediction-accuracy attainment has been observed from the following figure.

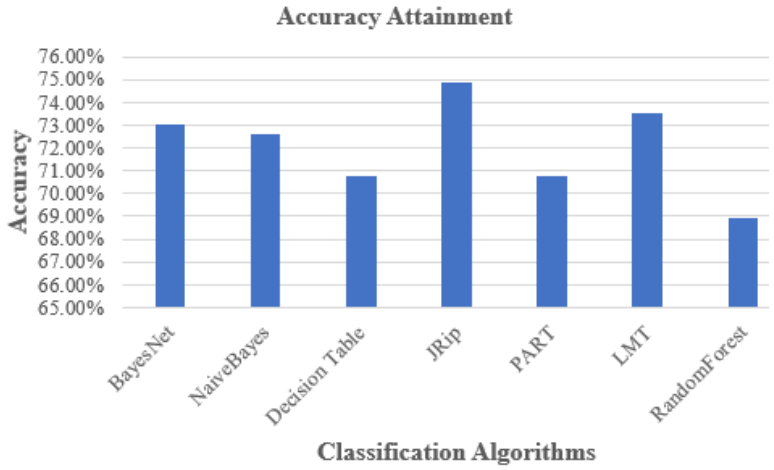

Fig.1 For Primary Dataset

Real time opinions could be studied by the machines with distinct ML-algorithms and give the favorable results to the mankind. The above table shows machine predictions of the respective ML-models. Sentiment analysis could be done using categories in WEKA environment. It is clear that decision tree ML-models' accuracies are lower in predictions. Sequential features' patterns could be learnt by the machines with proper weighted rules, measurable cropped and maximized information from the dataset. Better predictions were observed from logistic regression-decision tree model (LMT) and repeated incremental pruning to produce error reduction (RIPPER) a rule-based model (JRip). The false positive rate of acceptance class was observed in higher level compared to non-acceptance class which was for compatibility of online assessment.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

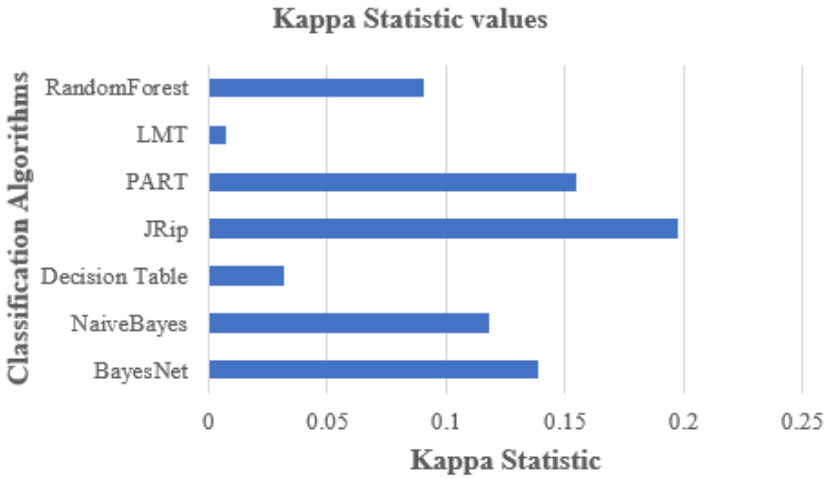

Fig.2

Discussion about Behavioral insight from the secondary dataset

The secondary dataset was containing fourteen features, the following figure describes them. The training process had been observed based on leaners' behaviour towards learning activities.

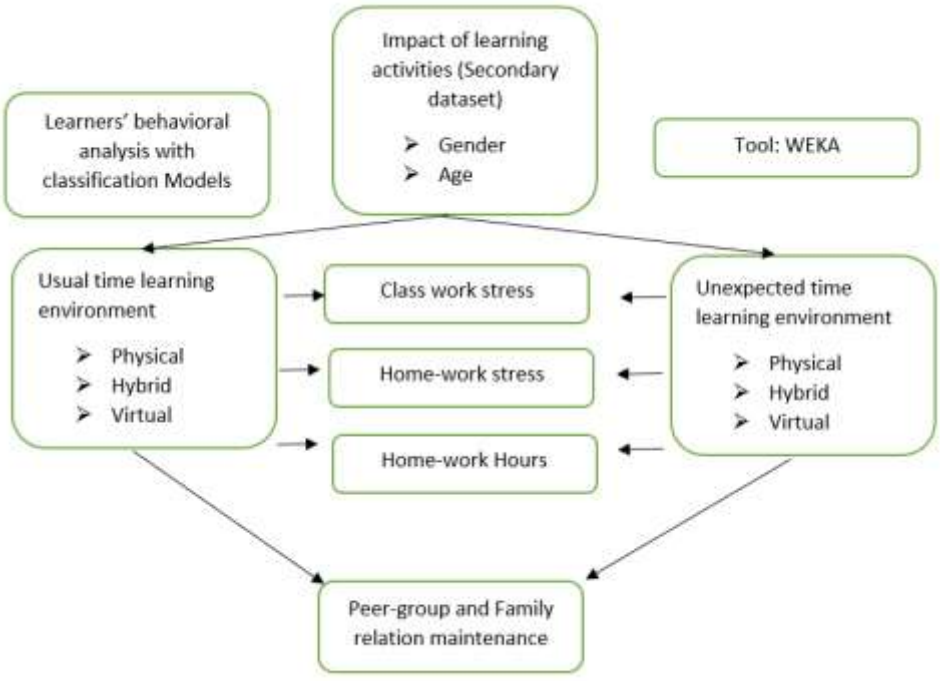

Fig.3

Many learning activities which are affecting the learners' life. Proper study-life balance among the learners could be estimated by their relations with peer group and family. The relationship index values were available in Kaggle dataset, which was taken for this study. Since the WEKA environment is good in finding insights for classification problems. The dependent variable's values had been changed from numeric to categorical values.

Machine predictions for training and testing had been compared. The split for training and testing it had been considered as 75% and 25% respectively. Rules.PART model's prediction were more accurate in training. Meta.Bagging and Trees.DecisionStump, NaïveBayes, Meta.AdaBoostM1 models' predictions for testing were 80%, 90% and 100% respectively.

The connections between features could be studied with help of decision trees and rule-based ML-algorithms. Machine predictions for the entire dataset had been observed, even if the accuracy of the

model is not up to the mark. Model's knowledge representation about study-life balance for the learners could be observed more evidently. Most contributing factor for the balance is study-environment.

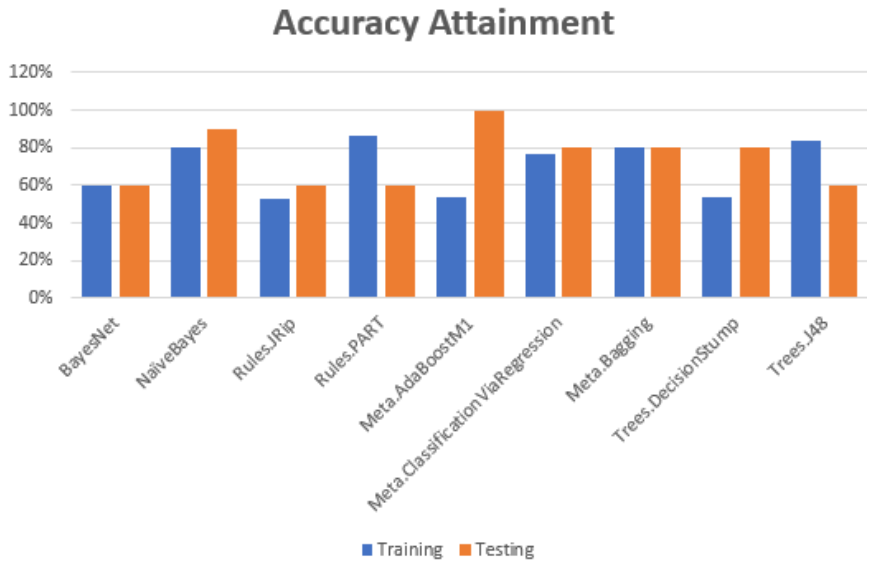

Fig.4

Conclusion

Very big virtual world has been created by the technical work-force of many countries' people. The data form many sources are collected every day. Many experts are working to extract new knowledge from distinct form of data. New perceptions could be extracted for the education sector, especially from training and learning data. Micro-level extracts from ML-models could contribute for macro-level positive change in country level. Modelling of learners' behavior with their learning activities towards their relations with stakeholders should be studied in every educational institution, to bring feasible solutions and to create acceptable or updated policies for the betterment of learning process in their respective institutions. Based on the information from ML-models, educators could be reconstructing the learning activities for the maximizing the learning among learners. Creating competence among the learners, skill acquirement part is prime factor, which could be drawn from logical and analytical subjects. Further part of research work will be carried out, for the higher education learners' skill acquirement part, which will be studied using Mathematics learning. Performance of ML-classification models and the insights from relation among features will be analyzed.

References

Bhuma Devi, S., & Jain, P. (2020). Predictive Model of Learning Analytics for On-Ground Verses Online Assessment. *International Journal of Education and Psychological Research*, 56-59.

Bhuma Devi, S., & Jain, P. (2021). Impact of Lockdown on Virtual Learning and Learning Analytics. *Educreator Research Journal, 8*(2), 16-22.

Golino, H., & Gomes, C. (2014). Four Machine Learning Methods to Predict Academic Achievement of College Students: A Comparison Study. *REVISTA ELETRÓNICA DE PSICOLOGIA, EDUCAÇÃO E SAÚDE*, 68-101.

Hamid, S., Ismail, S., Hamzah, M., & Malik, A. (2021). Developing Engagement in the Learning Management System Supported by Learning Analytics. *Computer Systems Science & Engineering*, 1-16.

Ratnapala, I., Ragel, R., & Deegalla, S. (2014). Students Behavioural Analysis in an Online Learning Environment Using Data Mining. *IEEE*, 1-7.

Shen, X., & Yuan, C. (2021). A College Student Behavior Analysis and Management Method Based on Machine Learning Technology. *Hindawi, Wireless Communications and Mobile Computing*, 1-10.

Srividya, M., Mohanavalli, S., & Bhalaji, N. (2018). Behavioral Modeling for Mental Health using Machine Learning Algorithms. *Springer Nature*, 1-13.

Webb, M., Fluck, A., Magenheim, J., Malyn-Smith, J., Waters, J., Deschênes, M., & n Zagami, J. (2021). Machine learning for human learners: opportunities, issues, tensions and threats. *Education Tech Research Dev*, 2109–2130.

Yayla, R., Yayla, H. N., Ortaç, G., & Bilgin, T. T. (2021). A Classification Approach with Machine Learning Methods for Technical Problems of Distance Education: Turkey Example. *Open Praxis*, 312–322.

Yilmaz, Y. (2022). Structural Equation Modelling Analysis of the Relationships Among University Students' Online Self-Regulation Skills, Satisfaction and Perceived Learning. *Participatory Educational Research*, 1-21.