Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

Relationship Between Learner Engagement and Performance based on The User Behavioral Factors in E-Learning Environments

Ranasinghe N. K
Department of Software Engineering
Sri Lanka Institute of Information Technology,
Sri Lanka

Dr. Lakmal Rupasinghe
Department of Cyber Security
Sri Lanka Institute of Information Technology,
Sri Lanka

***Abstract*:- Adaptive E-learning systems make a personalized e-learning environment according to the identified learning styles. The traditional way of detecting those learning styles is mainly based on asking learners to fill in a questionnaire or survey related to the learning process. This approach has some limitations including the excess time taken to fill the forms and the lack of self-awareness of learners regarding their own interests in learning. Furthermore, those traditional approaches lead learners to get bored when asking them to fill out some questions. Even if they are willingly fill out the questionnaire, they may give some irrelevant answers without knowing the importance of filling out those forms. Therefore, the results might be inaccurate by processing those irrelevant answers.**

Besides that, traditional approaches assume that learning styles cannot be changed over the time. But learning styles are not fixed. They are dynamic because, learning styles are derived from the learners' preferences and those preferences can be evolved due to some psychological and physical reasons. This paper explains an experimental study which has been taken place with the use a real-world dataset which is obtained from a e-learning platform in order to identify the relationship between the engagement of learners and the performance based on learners' behavioral factors in e-learning environments. Three machine learning algorithms – Decision tree, Random Forest and Support Vector Machine have been used to obtain the best predictive model.

Keywords – Adaptive E-Learning, Web usage mining, learning style, Academic Performance Prediction, Learning Analytics

I. INTRODUCTION

Nearly, the advancement of technologies has helped to change teacher centered education to student centered education. Students have different types of ways to engage in learning. Either they may be aware of their learning patterns or they may not be aware of their learning patterns. Therefore, every student should have an ability to identify their learning styles to make a good learning plan. Besides that, knowing the learning styles help students to observe, process, distinguish and comprehend learning materials. Many technological mechanisms have been introduced to detect and categorize the learning styles of students. The aim of many adaptive e-learning platforms is to identify the learning patterns of each student and adapt the deliverables accordingly.

Traditional way of getting aware of the learning behaviors of students basically depends on questionnaires and surveys. Although, filling in questionnaires have many drawbacks. Firstly, it may be an exhaustive task for students which also wastes time. Secondly, many learners do not know about their learning styles and they might not know the importance of those surveys and questionnaires, that may encourage them to give incorrect answers. Therefore, the results received through these kinds of traditional ways can be incorrect and might not exactly represent the learning patterns of learners.

To overcome the above-mentioned limitations, many automated approaches have been proposed which expects to determine learning behaviors of each student according to their activities while they are collaborating with the e-learning platforms. These automatic approaches have many benefits over the traditional approaches. Therefore, by using automated approaches, excess time usage, which was for filling in questionnaires will be reduced, since the user behavioral patterns can be detected from the user collaboration with the system. On one side, as the learning styles, which are identified by traditional approaches are static. In other words, those learning styles do not change over the time. But there is a chance that students' behavior in learning may change over the time. Therefore, students' behaviors should also be detected regularly. Previous automatic approaches have used LS (Learning Styles) models to detect learning styles of students. A learning style model assists to categorize the learners based on the specific behavior patterns in learning. It consists of some special variables related to the medium of the deliverables and the type of user interaction with the learning materials. Eventually, LS models give an idea about the way of each learner prefers to learn with. There are many LS models in literature [1] [2] [3].

Based on the literature [4] [5], the FSLSM can be recognized as the most common and the most applicable for implementing an adaptive e-learning platform. According to the FSLSM, each student can be uniquely identified based on their learning patterns. It also helps teachers, evaluators, and examiners to provide personalized learning materials and to use a personalized evaluation plan.

This work does not depend on learning styles model as there are some identified limitations in those models according to literature. As most of the previous research were carried out only according to one aspect of measurement such as the user behavior (Learning styles) with the system, this approach emphasizes the value of user behavior analysis for predicting the learners' performance. A machine learning

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

model is being used to predict the results of students based on their collaboration with the e-learning platform. The results will be predicted considering some important features which have affected the most to the performance of students.

Learning Behaviors

The term "Learning behaviors" is the awareness of that each learner learns differently and each of them have their own preferred way of absorbing, processing, comprehending, and preserving their knowledge. For example, some learners can understand a lesson by doing it practically by themselves while some learners just gain the knowledge by listening to the lesson and the instructions. This logic leads to a point that each learner is created equally and differently. Because everyone has a natural preferred way of learning, more research have gone into discovering the different learning styles. The individual learning patterns depend on many factors. According to [6], learning styles may vary according to cognitive, psychological, and environmental factors as well as learners' prior experience in learning. The understanding of different learning behaviors of learners is important for educators to understand the different learners, so that they can focus upon implementing a learner-friendly learning environment.

II. LITERATURE REVIEW

Various approaches using machine learning techniques have been proposed to automatically identify learners' learning patterns. Most of them are data-driven approaches which use data mining algorithms on existing learners' behaviors to construct a model. Then that model has been used to detect the learning pattern of a new learner. Based on [7], the BN classifier which is known as Bayesian Network can be recognized as one of the most common classifiers to presume the learning patterns. Authors of [8] have used Bayesian Networks to illustrate the relationship between the learner model and the features of a learning profile. To evaluate the link between above mentioned models, they have used the collaboration of Bayesian Network and the Overlay Model.

Decision tree is another most adopted classification algorithm which is used for automated detection of learning styles. Aijaz in [9] proposed an automatic approach for detecting learning styles of e-learners by using a decision tree as the classification algorithm. Through his solution, he has addressed the "One size fits all" issue of e-learning systems. Kolb's learning style theory has been used to understand learning styles from web logs of learners using data mining techniques. In [10], Pantho and Tiantong have addressed a solution to classify learning styles according to VARK [10] model of learning styles by using Decision Tree C4.5 algorithm. A questionnaire had been used to collect data from 1205 people. The gathered data were then analyzed using the Decision Tree C4.5 algorithm.

Addition to that, NW which is known as Neural Networks, is also one of the often used algorithms in automatic learning styles detection. [11] Proposed an approach to detect and track students' learning patterns in order to provide recommendations of relevant learning resources. That model was based on Neural Networks and Felder Silverman Learning Style model. Hmedna also introduced an approach in his work [11], to detect the learning styles of students in learning management system. That work was incorporated with fuzzy cognitive maps FCMs (Fuzzy c-means) which is a colloboration of Fuzzy logic and neural networks.

Apart from the above-mentioned algorithms, KNN (K-nearest neighbors) is often used to automatically identify the learning. [12] proposed a solution to detect learning styles by improving K-nearest neighbor (KNN) classification and combining it with genetic algorithms (GA).

Similarly, Mohomed Alloghani [13], carried out a research to develop predictive models for detection of learning styles based on the decision tree, neural network and naïve Bayes algorithms. But the dataset was previously collected not from an e-learning system.

Based on the literature, most of the works which were carried out to detect learning styles, relied on a specific learning style model. Consequently, most of the systems used FSLSM considering that there are 8 learning styles. In this component, the relationship between user engagement and performance is analyzed according to various user behavioral factors. This approach considers two aspects of learner attributes such as: learner attention towards the learning materials and the learner real intention in the learning platform.

The learning style of each learner is identified using data mining techniques and machine learning algorithms on the e-learning platform's log file. Consequently, by analyzing the log file each learners' performance level can be obtained. Higher the engagement level of learners, Greater the grades they obtain for the final examination is a theory. This model is based on that theory.

III. METHODOLGY

Data and Sources

The Open University is the one of the largest universities in the United Kingdom. VLE delivers the learning materials for each module (subject), and each user's activities per day are recorded in the VLE logs. Those learning materials in that VLE are delivered through HTML, PDFs and videos. Activity types can be identified as dataplus, forumng, glossary, oucollaborate, oucontent, resource, subpage, homepage, and URL and demographic data of users.

Since the teacher-student interactions in the VLE are limited when the number of users is getting increased, it leads to a difficulty for the academic instructors to support to all users real-time. The research focused on this approach involves reducing the dropout rate of students and optimize their performance in final examinations by predicting the results beforehand.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

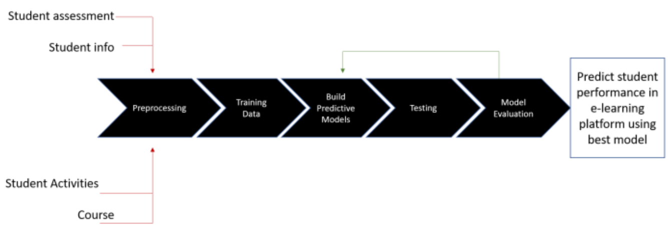
id_student	url	homepage	subpage	resource	oucontent	oucollabor	glossary	foruming	dataplus	he_level	gender	final_results
11391	1	35	10	9	0	0	0	27	0	HE Qualific	M	Pass
11393	1	3	4	0	0	0	0	0	0	A Level or	M	Fail
11400	2	14	9	2	2	2	0	8	0	A Level or	F	Fail
11501	2	24	8	0	0	0	2	40	0	A Level or	F	Withdrawn
11602	1	6	2	0	0	0	0	2	0	A Level or	M	Withdrawn
11714	2	16	5	0	0	0	0	3	0	A Level or	F	Withdrawn
11802	6	44	11	1	0	0	1	11	0	A Level or	M	Withdrawn
11900	6	27	26	0	0	0	0	4	0	Lower Tha	M	Withdrawn
11980	4	24	9	0	0	0	0	41	0	A Level or	F	Withdrawn
12125	1	4	4	2	0	0	1	4	0	HE Qualific	M	Withdrawn
12300	2	14	5	0	0	0	0	3	0	A Level or	F	Withdrawn
12350	10	70	26	3	1	1	0	10	0	A Level or	M	Withdrawn
12400	0	15	3	0	0	0	0	2	0	Lower Tha	F	Withdrawn
12450	9	26	25	0	0	0	2	4	0	Lower Tha	M	Withdrawn
12456	21	740	50	6	3	3	0	574	0	Post Gradu	M	Distinction
12653	2	23	30	2	0	0	0	15	0	Lower Tha	M	Distinction
12890	47	278	104	0	0	0	0	352	0	A Level or	M	Distinction
12890	6	112	17	0	0	0	0	134	0	HE Qualific	M	Distinction
13120	2	2	4	2	0	0	0	6	0	HE Qualific	M	Withdrawn
13200	3	24	14	1	0	0	0	5	0	Lower Tha	F	Withdrawn
13250	10	73	26	3	1	1	0	8	0	A Level or	M	Withdrawn
13300	1	2	2	0	0	0	1	1	0	A Level or	M	Withdrawn
13400	2	15	3	0	0	0	1	2	0	Lower Tha	F	Withdrawn
13478	10	88	22	3	0	0	1	230	0	A Level or	M	Distinction
14004	12	72	39	5	0	0	0	118	0	A Level or	M	Distinction
14237	2	10	5	1	0	0	1	0	0	HE Qualific	F	Distinction
14321	45	245	103	0	0	0	0	352	0	A Level or	M	Distinction

```
[4] dataset['final_results'].value_counts()
```

```
Pass           258
Fail           245
Withdrawn      149
Distinction    144
Name: final_results, dtype: int64
```

Predictive Models

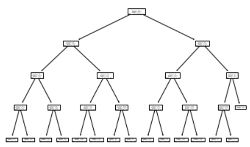
Three machine learning models which were used to obtain the most suitable predictive model for predicting learner performance are Random Forest Classifier, Decision Tree classifier and Support Vector Machine classifier. Those algorithms were chosen because they are applicable for both categorical and domain attributes. Data preprocessing is an important step in developing a model. The selected dataset which has been used to develop the model has 797 learner records. It also contains 12 variables including the target variable: Result. Three of them are categorical while rest of the variables are numeric. The target variable consists 4 values: Pass, Fail, Distinction and Withdrawn. Data preprocessing consists the steps of cleansing the dataset to obtain more accuracy in a predictive model. It also includes feature extraction and taking care of the categorical data. From 12 variables 5 variables were chosen based on their impact to the target variable. The chosen variables are homepage, url, subpage foruming and he_level. The variables he_level and final_results are categorical. Therefore, those variables have been encoded.



Decision Tree Classifier

Decision tree algorithm is based on entropy to classify the variables which is known as a mathematical technique. Decision tree has a tree-like design. Rectangles and ovals represent the internal nodes and leaves, respectively. Those nodes represent the features of dataset while each leaf has a class according to the dataset. This classifier has been developed by deciding which variables best split the input variables. As mentioned earlier, for this case, theory of information gain was used which means the node that has minimum entropy (Highest Information Gain) is used as split node.

```
[Text(181.8310344827586, 195.696, 'X[20] <= 38.5\nentropy = 1.947\nsamples = 597\nvalue = [107, 182, 196, 112]'),
 Text(92.35862068965517, 152.208, 'X[15] <= 8.5\nentropy = 1.827\nsamples = 399\nvalue = [43, 163, 128, 65]'),
 Text(46.17931034482758‍4, 108.72, 'X[14] <= 0.5\nentropy = 0.879\nsamples = 59\nvalue = [0, 48, 5, 6]'),
 Text(23.089655172413792, 65.232, 'X[20] <= 2.5\nentropy = 1.123\nsamples = 24\nvalue = [0, 17, 5, 2]'),
 Text(11.544827586206896, 21.744, 'entropy = 1.406\nsamples = 8\nvalue = [0, 3, 4, 1]'),
 Text(34.63448275862069, 21.744, 'entropy = 0.669\nsamples = 16\nvalue = [0, 14, 1, 1]'),
 Text(69.26896551724138, 65.232, 'X[16] <= 4.5\nentropy = 0.513\nsamples = 35\nvalue = [0, 31, 0, 4]'),
 Text(57.72413793103448, 21.744, 'entropy = 0.742\nsamples = 19\nvalue = [0, 15, 0, 4]'),
 Text(80.81379310344828, 21.744, 'entropy = 0.0\nsamples = 16\nvalue = [0, 16, 0, 0]'),
 Text(138.5379310344827‍7, 108.72, 'X[18] <= 2.5\nentropy = 1.875\nsamples = 340\nvalue = [43, 115, 123, 59]'),
 Text(115.44827586206895, 65.232, 'X[19] <= 0.5\nentropy = 1.89\nsamples = 321\nvalue = [43, 97, 122, 59]'),
 Text(103.90344827586206, 21.744, 'entropy = 1.816\nsamples = 277\nvalue = [29, 86, 117, 45]'),
 Text(126.99310344827586, 21.744, 'entropy = 1.908\nsamples = 44\nvalue = [14, 11, 5, 14]'),
 Text(161.6275862068965‍5, 65.232, 'X[15] <= 107.5\nentropy = 0.297\nsamples = 19\nvalue = [0, 18, 1, 0]'),
 Text(150.08275862068965, 21.744, 'entropy = 0.0\nsamples = 18\nvalue = [0, 18, 0, 0]'),
 Text(173.17241379310343, 21.744, 'entropy = 0.0\nsamples = 1\nvalue = [0, 0, 1, 0]'),
 Text(271.303448275862‍1, 152.208, 'X[17] <= 6.5\nentropy = 1.873\nsamples = 198\nvalue = [64, 19, 68, 47]'),
 Text(230.8965517241379, 108.72, 'X[14] <= 2.5\nentropy = 1.859\nsamples = 179\nvalue = [64, 19, 62, 34]'),
 Text(207.80689655172412, 65.232, 'X[18] <= 2.5\nentropy = 1.273\nsamples = 30\nvalue = [2, 0, 16, 12]'),
 Text(196.26206896551724, 21.744, 'entropy = 0.961\nsamples = 26\nvalue = [0, 0, 16, 10]'),
 Text(219.3517241379310‍3, 21.744, 'entropy = 1.0\nsamples = 4\nvalue = [2, 0, 0, 2]'),
 Text(253.98620689655172, 65.232, 'X[20] <= 133.5\nentropy = 1.836\nsamples = 149\nvalue = [62, 19, 46, 22]'),
 Text(242.44137931034481, 21.744, 'entropy = 1.885\nsamples = 94\nvalue = [30, 19, 34, 11]'),
 Text(265.5310344827586, 21.744, 'entropy = 1.398\nsamples = 55\nvalue = [32, 0, 12, 11]'),
 Text(311.7103448275862, 108.72, 'X[12] <= 0.5\nentropy = 0.9\nsamples = 19\nvalue = [0, 0, 6, 13]'),
 Text(300.1655172413793, 65.232, 'X[15] <= 99.5\nentropy = 0.994\nsamples = 11\nvalue = [0, 0, 6, 5]'),
 Text(288.6206896551724, 21.744, 'entropy = 0.0\nsamples = 4\nvalue = [0, 0, 4, 0]'),
 Text(311.7103448275862, 21.744, 'entropy = 0.863\nsamples = 7\nvalue = [0, 0, 2, 5]'),
 Text(323.2551724137931, 65.232, 'entropy = 0.0\nsamples = 8\nvalue = [0, 0, 0, 8]')]
```



```
DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='entropy',
                       max_depth=3, max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort='deprecated',
                       random_state=0, splitter='best')
```

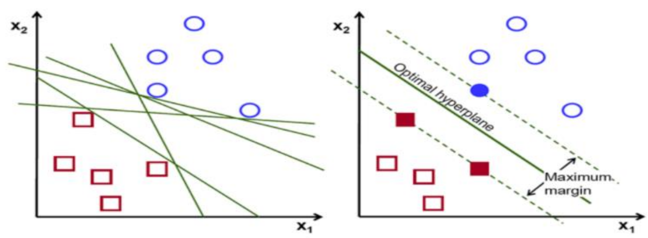
Random Forest Classifier

Random forest classifier is an ensemble machine learning algorithm. It consists of individual decision trees which work together as an ensemble. Each of those individual decision trees gives a prediction and among those predictions most voted value becomes the output of the model.

```
classifier = RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
             max_depth=None, max_features='auto', max_leaf_nodes=None,
             min_impurity_decrease=0.0, min_impurity_split=None,
             min_samples_leaf=1, min_samples_split=2,
             min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=1,
             oob_score=False, random_state=None, verbose=0,
             warm_start=False)
```

Support Vector Machine

The concept behind the support vector machine is to find out the optimal hyperplane in an N-dimensional space which classifies the data points. In other words, hyperplanes represent as decision boundaries.



```
from sklearn.svm import SVC
classifier = SVC(C=1.0, break_ties=False, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma='scale', kernel='linear',
    max_iter=-1, probability=True, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

IV. RESULTS AND DISCUSSION

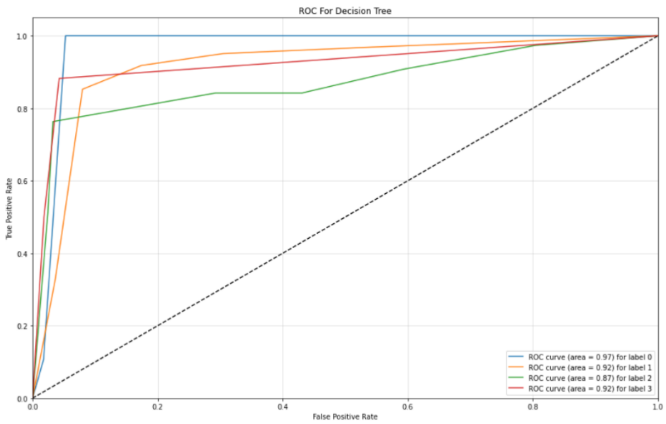
Evaluation Measures

A. Decision Tree Accuracy by Class

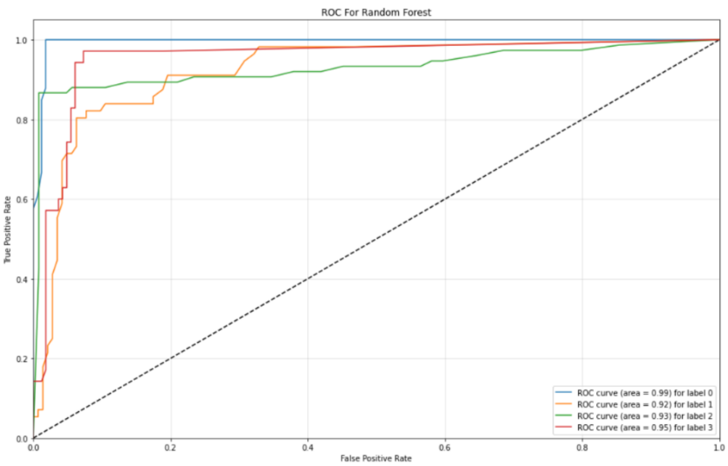
```
Accuracy: 0.8492462311557789
Precision 0.859974480985281
```

```
[22] from sklearn.metrics import classification_report
     print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.76	1.00	0.86	28
1	0.83	0.87	0.85	61
2	0.94	0.76	0.84	76
3	0.83	0.88	0.86	34
accuracy			0.85	199
macro avg	0.84	0.88	0.85	199
weighted avg	0.86	0.85	0.85	199



B. Random Forest Accuracy by Class




```
Accuracy: 0.8542713567839196
Precision 0.8566152189519023
Recall 0.8542713567839196
```

```
[21] from sklearn.metrics import classification_report
     print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.91	0.97	0.94	33
1	0.82	0.73	0.77	56
2	0.90	0.87	0.88	75
3	0.76	0.91	0.83	35
accuracy			0.85	199
macro avg	0.85	0.87	0.86	199
weighted avg	0.86	0.85	0.85	199


```
Accuracy: 0.8542713567839196
Precision 0.8566152189519023
Recall 0.8542713567839196
```

```
[21] from sklearn.metrics import classification_report
     print(classification_report(y_test,y_pred))
```

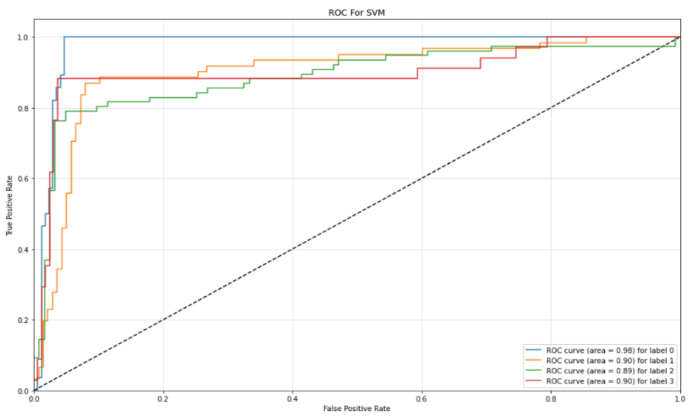
	precision	recall	f1-score	support
0	0.91	0.97	0.94	33
1	0.82	0.73	0.77	56
2	0.90	0.87	0.88	75
3	0.76	0.91	0.83	35
accuracy			0.85	199
macro avg	0.85	0.87	0.86	199
weighted avg	0.86	0.85	0.85	199

C. SVM Accuracy by Class

```
Accuracy: 0.8442211055276382
Precision 0.8442211055276382
Recall 0.8442211055276382
```

```
[ ] from sklearn.metrics import classification_report
    print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.76	1.00	0.86	28
1	0.83	0.85	0.84	61
2	0.94	0.76	0.84	76
3	0.81	0.88	0.85	34
accuracy			0.84	199
macro avg	0.83	0.87	0.85	199
weighted avg	0.86	0.84	0.84	199



International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 12, December-2020

V. CONCLUSION

Considering the descriptive statistics, Random forest Classifier was chosen to develop prediction model.

Classifier	Accuracy	Precision	Recall
SVM	0.8492	0-0.76 1-0.83 2-0.94 3-0.83	0-1.00 1-0.87 2-0.76 3-0.88
Decision Tree	0.84	0-0.76 1-0.83 2-0.94 3-0.81	0-1.00 1-0.85 2-0.76 3-0.88
Random Forest	0.8542	0-0.91 1-0.82 2-0.90 3-0.76	0-0.97 1-0.73 2-0.87 3-0.91

REFERENCE

[1] L. K. S. R. M Felder, "Learning and Teaching styles," vol. 78, pp. 674-681, 1988.
[2] H. K Laschinger, M. W Boss, Learning Styles of nursing students and career choices, 1984.
[3] Biggs, "Study Process Questionnaire Manual Student Approaches to Learning and Studying," Australian Education Research, 1987, p. 53
[4] F. L. J. Kuljis, "A Comparison of Learning Style Theories on the Suitability for elearning," in *Web technologies, Applications and services*, 2005.
[5] K. S. Graf, "Advanced adaptivity in learning management systems by considering learning styles," in *IEEE/WIC/ACM International Conference on Web Intelligent Agent Technology - Workshops*, 2009.
[6] J. Keefe, "Learning style: An overview. NASSP's Student learning styles: Diagnosing and proscribing programs," pp. 1-17, 1979.
[7] F. J, M. A and A. A, "Automatic detection of learning styles," in *State of the art*, Artificial Intelligence, 2015, pp. 157-186.
[8] M. T. A, R. N. C and M. K, "The determination of learning styles in a learner model using the combination of bayesian network and the overlay model," 2019.
[9] Aijaz Ahmed Kalhoro, "Detection of E-Learners' Learning Styles: An Automatic Approach using Decision Tree," *International Journal of Computer Science and System Analysis,* vol. 14, no. 8, pp. 420-425, 2016.
[10] O. Pantho and M. Tiantong, "Using decision tree C4.5 algorithm to predict VARK learning styles," *International Journal of the Computer, the Internet and Management,* vol. 24, no. 2, pp. 58-63, 2016.
[11] B. HMEDNA, A. E. Mezouary, O. Baz and D. Mammass, "Identifying and tracking learning styles in MOOCs: A neural networks approach," *International Journal of Innovation and Applied Studies,* vol. 19, no. 2, pp. 267-275, 2017.
[12] C.-H. Chiu, C.-P. Chua, W.-Y. Kao and Y.-C. Chang, "A learning style classification mechanism for e-learning," *Computers & Education,* vol. 53, no. 2, pp. 273-285, 2009.
[13] M. Alloghani, D. Al-Jumeily, A. Hussain, J. Aljaaf and E. Petrov, "Application of Machine Learning on Student Data for the Appraisal of Academic Performance," in *11th International Conference on Developments in eSystems Engineering* , 2018.
[14] J. E. Villaverde, D. Godoy and A. Amandi, "Learning styles' recognition in e-learning environments with feed-forward neural networks," *J. Comp. Assisted Learning,* vol. 22, pp. 197-206, 2006.
[15] O. E. AISSAOUI, Y. E. A. EL MADANI, L. OUGHDIR and Y. E. ALLIOUI, "Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles," *Procedia Computer Science,* vol. 148, pp. 87-86, 2019.