

Activity Based Analysis and Prediction Strategy for the Class Room Performance Improvement

M. Parvathi

ECE Dept., BVRIT HYDERABAD College of Engineering for Women, Hyderabad, Telangana, India
parvathi.m@bvrithyderabad.edu.in

Abstract: Class room assessment is one of the key issues where in individual performance gives a great impact in the overall performance of the class. In general, various activities will help in identifying the student's ability and weak strengths in understanding the subject as well. Though various online learning platforms are present, an overwhelmed usage is occurring due to the pandemic in this period. Nevertheless, very few online resources are giving their utmost use in efficient way. At the same time class room activities are also need to shape up accordingly that supports to the virtual platforms. Few activities like chat, polling, and turn around were performed through virtual platform over a class group, to address the complex engineering problems that were identified in Embedded Systems Design, an undergraduate subject. Further a common strategy was found to apply while dealing with that subject. In this process, collected the outcomes of the complete learning activities conducted during the course and using few data-mining techniques the classroom performance has been analysed. For initial analysis, k means clustering was used in which the parameters chosen are like individual marks, total class strength. The outcome of this implementation resulted in identification of clusters in the name of weak, average and best groups along with an accuracy of 46.51%. The analyses were further carried out using other measuring parameters like spontaneity, delays in answering, best answer, etc., have been chosen during the activities and Naïve Bayes Classifier was chosen for predicting the performance improvement and was observed as 70%. Finally, based on the experimental results, classification and regression supplemented by logistic was applied for further evaluation of learning performances and a remarkable accuracy progress of 96.96% is investigated in this study.

Keywords: Class room assessment, complex engineering problems, k means clustering, Naïve Bayes Classifier, logistic regression, learning performance.

Dr. M. Parvathi

Dept. of Electronics and Communication Engineering
BVRIT HYDERABAD College of Engineering for
Women, Hyderabad, Telangana, India
parvathi.m@bvrithyderabad.edu.in

1. INTRODUCTION

Activity based learning is one of the approaches in active learning methods that implicates engaging the students

actively within the course instructions. Activities are majorly group discussions, problem solving, content preparation, case studies, role plays, and many other methods. Active learning is the origin of constructive learning approach. The benefits of active learning are: students can be involved much in the classroom activity than passive listening; also they can be engaged in various activities like reading, discussions, writing etc. The best way to improve the students' knowledge, understanding, and application of information can be achieved by conducting more number of classroom activities. The average student can be paired with best student in an activity like think-pair-share. Sometimes purposeful questioning and pause for discussions makes the students more alert in the class. Demonstrations, clicker quizzes kind of activities leads to promote the level of lower-order cognition in traditional learners.

Advancements in society, media, technology, and communication have made the class room performance more important than ever for technical professional educators to understand their student community and develop instructional methods accordingly including style of lecture delivery that will produce in effect learning outcomes. In general "good teaching" will come from the instructor by applying the active learning in their classes. Further, the majority of the instructors apply the active learning approaches at least once per week and believe they are "somewhat successful" or "very successful" with their implementations. In traditional teaching learning methods, the quantitative measurement and analysis of phenomena such as class room engagement or instructional practices are inherently limited. Therefore it is necessary to consider any data analytic methods that provide appropriate insight of assessment mechanism (Allyson Barlow and Shane Brown 2020). This paper highlights the importance of active learning in the teaching-learning practices.

Learning analytics is a systematic computational analysis used as an important measuring tool for observing the impact and outcomes in learning environments, further that enables the providers to develop new ways of achieving excellence in teaching and learning (Beth Dietz-Uhler & Janet E. Hurn, 2013). This also helps in providing students with new information to make the best choice about their education. Learning analytics has become appropriate choice in quantifying various parameters like student retention ratio, student success rate, and measure of the burden of accountability. Various learning management system (LMS) tools are helpful in addressing these large-

scale issues in order to improve the student's success. The use of LMS system over a period of time, can effectively characterizes the learning performance of the students (Ya-Han Hu Et.al. 2014) and eventually useful in building the early warning system that helps in predicting the at-risk students while progressing the course.

Now a day, the process of quality assessment in learning analytics and engineering education research has become the major issue due to application of advanced technologies. Researchers are interested in measuring students' attitudes towards technology for a myriad (Piet Ankiewicz 2019). Technology advancements brought lot of active learning practices that can be implemented via online mode for facilitating both student and faculty community. Few of the active learning platforms are Jamboard, an interactive touchscreen display/monitor use just like a white board, Padlet, an online notice board that can be used by students and teachers to post notes on a common page, Mind mapping that helps in assessing student innovative thinking, remembrance, idea creation and knowledge level. Few more interactive online tools for activity based learning are: chart presentation, slido quiz, online quiz, mentimeter, quizlet, quizzizz, etc. Past literature indicates that (Susanna Hartikainen et.al. 2019) the parameters that are used to measure the impact of active learning on students' learning outcomes were mostly based on students' self-report data and focused on course specific development in subject-related knowledge. Further, thorough descriptions and theoretical justifications, as well as the deliberation of learning outcomes with appropriate research methods, were used in emphasizing the requirement of transparency of empirical interventions through the application of active learning.

Applying the complex engineering design activities are most helpful for engineering students especially for improving the problem solving skills, along with communication and teamwork skills. Collaboration is an often used instructional strategy for engaging students in solving open-ended problems (Brenda R. Brand 2020). Sometimes, certain engineering problems need child-like thinking through flexible processes and imaginative tools were needed in the form of play (Elkin Taborda et.al 2012). In addition, if the experience of learning is made "fun", the outcomes will be in progressive way. The students are more engaged in a learning process that encourages play and having fun. Certain engineering concepts need creative and innovative ways to express and explain. For that, detailed design tools are required. The available design tools are much comfortable for the students to get proficient in use, but are lacking in conceptual design and ideation. Hence it is required to reframe the curriculum that needs to develop a more effective framework for teaching visual thinking, and design thinking practices.

It is also important to measure the outcome of the student performance despite of many implementations while improving their design and analytical thinking. Learning outcomes really helps in shape the formats of curricula, courses, teaching, learning and assessment. The

learning outcomes must be well defined before the course starts. These are to be implemented in a process of team-based co-creation, but not as an administrative obligation. It is essential to make learning outcomes relevant by linking them to national and international qualification frameworks. However, matching and benchmarking learning outcomes to qualification frameworks should not stand in the way of innovating education and exploring new field of knowledge. In this process, the instructors will follow many strategies in fulfilling the course by meeting the course learning outcomes (CLO) and mapping them with the program outcomes (PO). Courses designed and the culture of educating students is to be framed in such a way that it is measurable and have the outcomes from it as outcome-based education (OBE).

Using various active learning strategies namely Think-Pair-Share, Peer-Instruction, Debate, Jigsaw, Problem based Learning, Game based Learning, Project based Learning, Team-Pair-Solo, etc, can be used in engage the students in learning process. Role play is an activity where students perform the real time case study, to express the concept. Core subjects, in which generally students get less score at first year level than the rest of the years. Hence, the class room activities must be conducted based on their level of study and relevant to the subject as well (Salumari Madhu, G. Prabhakara Reddy (2018).

Literature suggested few active learning techniques such as game based learning, flipped classroom, team based learning, case based learning (Nicolette Harris et.al 2019), problem based learning (Kevin Morgan 2011), and simulation based learning (Hyunsook Shin et.al 2014). Recent reviews on activity based learning shows the importance of machine learning algorithms in the analysis of classroom performances (Stephen J.H. Yang 2018).

In this paper, the learning outcomes in the subject of Embedded Systems Design pertain to engineering final year has been assessed. The assessment was done based on the various activities that were conducted among 120 student members from two sections. Section 2 discuss about background work of assessment, section 3 discuss on methodology of assessment and section 4 on results and discussions, finally section 5 gives conclusions.

2. Background work

The BVRIT HYDERABAD College of Engineering for Women is UG College, have an intake of 120 students every year in the stream of Electronics and Communication Engineering. Each year the students will undergo various subjects based on the curriculum that is prescribed by JNTU Hyderabad affiliation. Each year the student's semester performance in each subject will be measured in terms of attainment and the target will be fixed for the upcoming batches based on the previous attainment level. Instructors will be assigned with the subjects typically one or two in a semester hence the allotted sections will one or two. On an average each instructor will get total number of students as 120 from any two sections. The classwork load on each instructor will be 20 per week which includes 16 instructions and 4 tutorials. The teaching orientation within

the campus follows student centric that follows summer internships, institutional trainings, industrial visits, projects and field trips, various competitions like hack a thons, student chapters like Computer Society of India (CSI), Institute of Electrical and Electronics Engineering (IEEE) etc.

Prior to the start of the semester, Instructor along with the students undergo 4 hours orientation class, where they will be exposed to the ongoing institutional practices, and opportunities along with teaching-learning responsibilities. Instructors are provided with various workshops and Faculty Development Programs (FDPs) to upgrade their teaching skills. These workshops incorporate active learning teaching methods that meet the goals for a particular session so that the facilitators can not only model such approaches as the think-pair share, case studies, role playing, and jigsaw; but they can also deliberately articulate the purpose for using a particular technique during the session. In addition, instructors or faculty will get trained effectively by participating in an interactive sessions, that are conducting by Vishnu Educational Development and Innovation Centre (VEDIC), a unique privately funded residential campus belongs to Sri Vishnu Educational Society (VES). In this campus, performance of all the faculty members will be focussed on the issues of classroom climate. Participants are exposed to various learning methodologies along with packet full of resources describing on-campus teaching resources including academic integrity, teaching-learning challenges, social responsibilities and issues pertain to women students. They meet briefly with their teaching peer mentors, which is a group of experienced teaching professionals (TP) who are trained by the teaching centre to observe classes, gather student feedback and consult with their peers. The TPs learn more about this program and the teaching-related services that are available to them. Practice teaching (also known as microteaching) is a practice that is recommended by Prieto, Yamokoski, & Meyers (2007) for graduate student development because the practice helps in elevation of instructor's self-efficacy. During the main orientation, specifically, the program organizer who is a faculty developer with an engineering background presents two research studies in order to highlight the performance benefits of the ongoing approaches and the range of active learning techniques that can be used to achieve required goals. Few examples for ongoing practices under active learning are, 2-minute pauses during a 45-min class, where students will be instructed to review the notes and develop questions during these breaks, role play, minute paper, think-pair-share, brainstorming, case studies, inquiry learning, and teaching square etc.

During the active learning presentation, the participants will be suggested to watch a short video clips related to the subject, then the participants brainstorms the ways to make the class session more interactive. At this point, the presenter briefly describes the six active learning teaching methods they will use during their advanced practice teaching session at the end of the semester. Finally, the instructors will receive a handout specifying engineering

specific examples of how the teaching methods are enacted in engineering courses and are provided with video links showing instructions on conducting microteaching lessons with active learning. On every weekend Saturday, during the semester, each instructor will participate in an advanced practice teaching session i.e. knowledge sharing session where they prepare a 10-minute lesson or any advanced topic incorporating active learning teaching methods. The instructor presents a lesson to a small group of their peers and receives feedback.

3. Methodology of Assessment

During the semester, students are exposed to various assessment tests along with regular instructions. Based on the instructional direction, the student must undergo internal exam for twice, along with assignment submission. If the students are given directions accordingly for attaining good performance in internal exams, is alone not sufficient to achieve better academic performance and understanding the subject as well. Hence, the instructor will conduct subject related activities and performance analysis will be carried out time to time.

In the subject of Embedded Systems Design, apart from the regular assignment practice, few complex topics were selected and conducted three activities. The activities are like open book exam, chit-chat and polling. The first activity i.e. open book exam, in which students were given thought provoking questions and were asked to answer relevantly using suggested or prescribed book. The main purpose of this activity is to make the students habituate to read standard books. In the second activity, i.e. padlet, in which students were allowed to take a topic of their choice from the given, and speak about the content in the given time. The third activity was quizlet, a quiz like activity, in which students were asked for various questions on a particular selected topic. Based on their explanation and relevance to the topic, scores were awarded. These three activities were analysed using k means clustering ML algorithm. In this analysis, based on their performances in the activities as well as in the subject too, the students were made into clusters like weak, average and best groups.

The cluster analysis alone is not sufficient in the evaluation of classroom performance. Further the analysis was made by conducting few more events on the same class group. The second round activities conducted were chit-chat, topic-tray, seminar presentations, turn around, objective test, polling, mentimeter, and surprise test in addition to class attendance regularity. Some of these activities were conducted by taking the students as groups. For example, seminar presentations, the topics were allocated to group of three students, in which a best, average and below average students were identified to make them as a sub group. Chit chat is an event where students will be paired up, and allowed to take a chit with content in it. Within the pair of students, one will be allowed to talk on the topic in a pro way whereas the other should contradict it. This kind of event is a bit of fun and play. For this analysis Naïve bays classifier was used and

observed the improvement in the class performance. Final analysis was made by taking the parameters like attendance percentage, assignment average, and few more activities like chit-chat, open book exam, and score in the semester mock exam. In this analysis, the student’s expected performance was included in the name of ‘Grade Status’. This third round class activities were analysed using logistic regression where in all these parameters were taken as inputs to train the model. As an observation, the overall performance has been improved.

A. Analysis using k-means clustering ML algorithm

The analysis was an outcome of activities that were conducted on the entire two sections of 129 students. The database includes their roll numbers, activities conducted along with the scores achieved is included in Table I. Padlet environment for quiz activity conducted is shown in Fig.1.

Table I. Dataset including activities like open book, padlet and quizlet activities

st.roll	Activity_1_Score	Activity_2_Score	Activity_3_Score	Sem_total_Score
17WH1A0401	14	17	22	53
17WH1A0402	20	17	25	62
17WH1A0403	23	18	25	66
17WH1A0404	21	23	23	67
17WH1A0405	15	23	21	59
17WH1A0406	25	25	20	70
17WH1A0407	21	22	20	63
17WH1A0408	21	18	0	39
17WH1A0409	22	16	24	62
17WH1A0410	24	18	23	65
17WH1A0411	26	24	5	55
17WH1A0412	10	23	7	40
17WH1A0413	0	22	14	36
17WH1A0414	21	20	13	54
17WH1A0415	26	21	21	68
17WH1A0416	17	23	9	49
17WH1A0417	19	25	14	58
17WH1A0418	20	21	20	61
17WH1A0419	24	23	23	70
17WH1A0420	25	22	21	68
17WH1A0421	23	25	15	63
17WH1A0422	22	25	25	72
17WH1A0423	21	23	21	65
17WH1A0424	15	21	21	57
17WH1A0425	16	20	22	58
17WH1A0426	16	20	24	60
17WH1A0427	19	0	26	45
17WH1A0428	23	24	10	57
17WH1A0429	22	23	0	45
17WH1A0430	21	5	21	47

After conducting first round of activities that are mentioned in section 3, the scores obtained in open book, padlet, and quizlet were shown in Fig.2 along with total score.

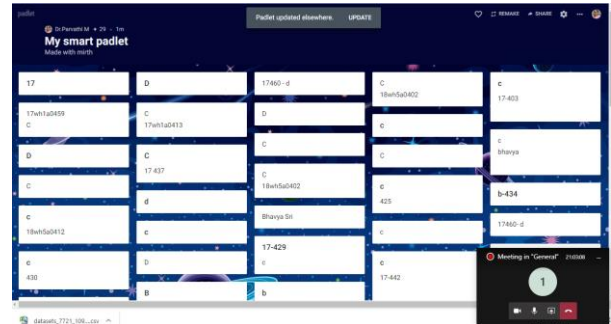


Fig.2. Padlet environment for quiz activity

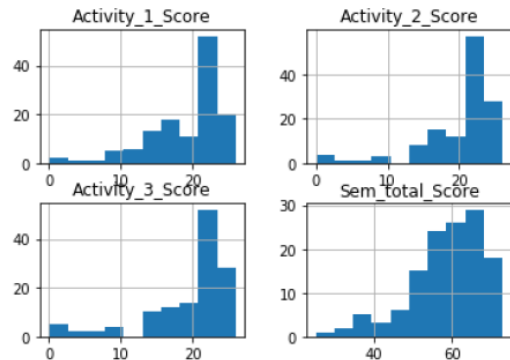


Fig.1. Activities with scores observed using k-means

In order to get optimal number of clusters and k, one should observe the graph between assumed number of clusters and Within-Cluster-Sum-of-Squares (WCSS). Where WCSS is obtained as a measure of squared average distance calculated from every point within a cluster to the cluster centroid. This was possible by Elbow method and is shown in Fig.3. K-Means stuck in an optimal local solution. Therefore, the determination of the starting point value of the clustering centre will greatly determine the results obtained by the K-Means Clustering algorithm (Kamson Sirait et.al. 2017).

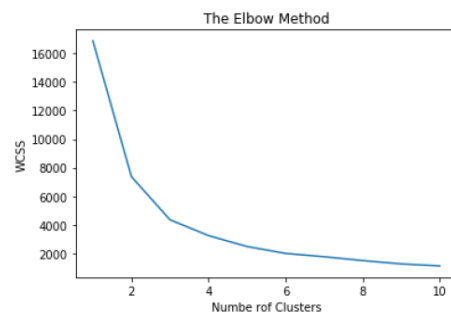


Fig.3. Optimizing k value

From the curve, the optimal value of k is chosen as ‘9’ with minimal WCSS value. To calculate WCSS, Euclidean distance to be found initially, by drawing a line between a given point and the centroid to which it is assigned. Then iterate this process for all the points in the cluster, and then sum the values for the cluster and divide by the number of points. Correlation between two objects can be determined by using the following Euclidean distance formula:

$$d_{euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_j)^2} \dots \dots \dots (1)$$

where
 d(x,y) = The distance of data to x to the centre of the cluster y.
 x_i =data i in n data.
 Y_j =data j in n data.

Finally, the average across all clusters is calculated. This will give the average WCSS. The chosen k value was applied with the dataset values that are given in Table I. The resulting clusters and the student groups within the clusters along with the marks obtained is shown in Fig. 4.

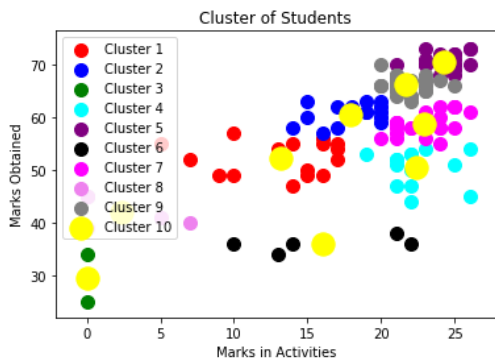


Fig.4. Clusters and student groups based on marks obtained

In general using k means clustering the accuracy with one iteration results in less value and is observed as 46.51%. Yet, using permutations and the repeated iterations for nine times, the accuracy raised and was observed as 91.47%. Both the observed values are shown in Fig. 5.

```
In [372]: from sklearn.metrics import accuracy_score
acc=accuracy_score(y_kmeans, kmeans.labels_)
print("Accuracy score is", acc)

Accuracy score is 0.046511627986976744

In [373]: import scipy
def find_permutation(n_clusters, real_labels, labels):
    permutation=[]
    for i in range(n_clusters):
        idx = labels == i
        new_label=scipy.stats.mode(real_labels[idx])[0][0] # Choose the most com
        permutation.append(new_label)
    return permutation

In [374]: permutation = find_permutation(9, kmeans.labels_, y_kmeans)
print(permutation)

[5, 7, 3, 0, 6, 2, 4, 8, 1]

In [375]: new_labels = [ permutation[label] for label in y_kmeans] # permute the labels
print("Accuracy score is", accuracy_score(kmeans.labels_, new_labels))

Accuracy score is 0.9147286821705426
```

Fig. 5 Accuracy observation using k means clustering

If the permutations are not considered, the traditional k-means algorithm converges and the centroids were calculated accordingly that the data points will be assigned to their nearest centroids. Using permutations, as the algorithm converges, the number of data points moving away from their cluster decreases with respect to each time the iteration takes place, and finally this leads to better

accuracy performance. The students were grouped up as best, average, and weak classes easily using k-means algorithm. This grouping process is helpful for further monitoring on their performances very closely.

B. Analysis using Naïve Bays Classifier

One of the classifier algorithms in ML is Naïve bays. By considering the possible existing conditions into account such that if a student get less marks in one of the activities and good marks in other activity like assignments etc., Through Naïve bays one can predict whether the student will get good score in the semester end exams. Table II shows possible activity combinations and marks obtained by the students participated in various activities. Class regularity parameter given as '1' for attendance percentage above 60 and '0' for less than that. Various activities were conducted and given various score points. For example, topic tray was conducted for 200 marks, seminar presentations were conducted for 100 marks. Another activity 'Turn around' that comprises series of questions, will be asked to the students one after the other and the students will grab their turn to score more, and this one was conducted for 50 marks. Open book test was conducted for 25, polling in which student will pick their interested question out of set of given questions, and was conducted for 50 total marks. Mentimeter, a quiz was conducted using online tool for 100 marks. A surprise test using objective questions was conducted for 50 total marks.

Table II. Dataset including activities like chit-chat, topic-tray, etc. including class regularity

chit-chat	topic-tray	pres	turnarnd	obtest	polling	mentim	test	classreg
6	148	72	35	20	33.6	67	50	1
1	85	66	29	25	26.6	51	31	0
8	183	64	0	22	23.3	67	32	1
1	89	66	23	25	28.1	56	21	0
0	137	40	35	21	43.1	88	33	1
5	116	74	0	20	25.6	87	30	0
3	78	50	32	15	31	48	26	1
10	115	0	45	25	35.3	41	29	0
2	197	70	45	18	30.5	58	53	1
8	125	96	0	15	0	52	54	1
4	110	92	0	18	37.6	91	30	0
10	168	74	0	23	38	57	34	1
10	139	80	0	25	27.1	41	57	0
1	189	60	23	15	30.1	98	59	1
5	166	72	19	17	25.8	87	51	1
7	100	0	0	18	30	84	32	1
0	118	84	47	14	45.8	51	31	1
7	107	74	0	22	29.6	54	31	1
1	103	30	38	23	43.3	18	33	0
1	115	70	30	25	34.6	59	32	1
3	126	88	41	15	39.3	70	27	0
8	99	84	0	17	35.4	88	50	0
7	196	90	0	18	39.8	51	41	1
9	119	80	35	14	29	65	29	1

The probability of achieving good score by a particular student in various activities will be observed through this Naïve base classification as shown in Fig.6, through which the overall performance of the student can be predicted. Small circle indicates the probability less than half, and big circle indicates the probability greater than half.



Fig.6 Probability of getting good score in the activities by a particular student

According to Naïve bays the posterior probability that is whether the good score is achieved or not can be observed as follows:

$$P(c|x) = \frac{p(x|c)p(c)}{p(x)} \dots\dots\dots(2)$$

- Where p(c|x) is posterior probability,
- P(x|c) is likelihood
- P(c) is class prior probability
- P(x) is predictor prior probability

Where x is student variable and c is good score attempt of activity

Table III. Frequency of participation in the activity by Student_1

Frequency of attempt in the activity by Students	Act	
	Yes	No
Student_1	2	5
	5	1

Table IV. Frequency of participation in the activity by Student_2

Frequency of attempt in the activity by Students	Act	
	Yes	No
Student_2	6	5
	2	2

Table V. Frequency of participation in the activity by Students

Frequency of attempt in the activity by Students	Act		
	Yes	No	
Students	Student_1	5	1
	Student_2	3	0
	Student_3	4	5

Table VI. Likelihood of participation in the activity by Students

Frequency of attempt in the activity by Students		Act		
		Yes	No	
Students	Student_1	5/12	1/6	6/18
	Student_2	3/12	0/6	3/18
	Student_3	4/12	5/6	9/18
		12/18	6/18	

$P(x|c) = p(\text{Student}_1|\text{Yes}) = 5/12 = 0.416$
 $P(c) = P(\text{Yes}) = 12/18 = 0.66$
 $P(x) = P(\text{Student}) = 6/18 = 0.33$

Likelihood of participation in the activities is:

$$\begin{aligned}
 P(c|x) &= P(\text{Yes}|\text{Student}) \\
 &= \mathbf{P(\text{Student}|\text{Yes}) * P(\text{Yes}) / P(\text{Student})} \dots\dots\dots(3) \\
 &= (0.416 \times 0.66) / 0.33 \\
 &= 0.832
 \end{aligned}$$

Table III to V represents the frequency of student participation in the activities along with the possibilities of getting high score. Values in ‘Yes’ and ‘No’ indicates the level of attempts made in getting high scores out of 10 scales. From Table VI, the first column indicates that 5+3+4 is the total attempts made in activities by student’s i.e.12 for getting good score, out of 18 observations. Similarly 1+0+5 i.e. 6 is attempts under poor score observed out of 18 observations. From first column itself, 5/12 is understood that 5 attempts were made for good score out of 12 attempts by student_1. That is the probability of attempts for good score made by a particular student is 0.416. Similarly 12/18 is the total probability of attempts for good score were made out of all attempts. Similarly, P(Student) is the probability of total attempts a particular student out of all attempts. According to Bays theorem, from the likelihood table, the prediction can be made to know whether the attempt is made to get good score in the activities, and is observed as 0.832. The same procedure will be repeated for total number of students present in the dataset.

The major steps to be followed while implementing the above process using machine learning are: splitting dataset in to k folds, calculating accuracy percentage, calculating mean and standard deviation of list of numbers, calculating Gaussian probability distribution function, calculating the probabilities of predicting each class for a given row, applying Naïve Bayes algorithm with train and test predictions, finally testing the model with the given dataset. After running the program, it was resulted the mean classification accuracy scores as well as the mean accuracy score. The mean classification accuracy scores were observed as 79.1%, 62.5%, 70.8%, 66.6%, and 70.8%. The mean accuracy observed is 70% as shown in Fig. 7.

```

145 # Test Naive Bayes on classperformance Dataset
146 seed(1)
147 filename = 'Desktop\ATAL_WORKSHOPS\KMEANS_PROJECT\classactperfdataset.csv'
148 dataset = load_csv(filename)
149 for i in range(len(dataset[0])-1):
150     str_column_to_float(dataset, i)
151 # convert class column to integers
152 str_column_to_int(dataset, len(dataset[0])-1)
153 # evaluate algorithm
154 n_folds = 5
155 scores = evaluate_algorithm(dataset, naive_bayes, n_folds)
156 print('Scores: %s' % scores)
157 print('Mean Accuracy: %.3f%%' % (sum(scores)/float(len(scores))))

Scores: [79.16666666666666, 62.5, 70.83333333333334, 66.66666666666666, 70.83333333333334]
Mean Accuracy: 70.000%
    
```

Fig.7 Mean classification accuracy and mean accuracy

From this analysis, the observations were: Improvement in the student’s activities involvement, that helped in emphasizing the overall class performance as each student is attaining almost good score.

C. Analysis using logistic regression ML algorithm

Based on the activity scores and assignment scores further analysis was carried out using logistic regression. Corresponding dataset is shown in Table 5. This includes attendance percentage, average assignment score, activities scores and expected grades. The grades were calculated using ‘A’, ‘B’ or ‘C’, represents best, average and respectively. While using this dataset in logistic regression, for simplicity the grade status was considered as ‘1’ for grade ‘A’ and ‘0’ for lower grades in binary value .

Table VII. Dataset including activities, assignment average, and estimated grade

Roll_No	Attendan ce%	Assignm ent_Avg	Act_1_chi tchat	Act_2_OB E	Act_3_Se m	Total_Sc Wact	Score_No _Act	Grade_st atus	Grade_TF
17WH1A0401	85	5	10	25	15	95	65	A	1
17WH1A0402	99	4	10	25	15	88	70	A	1
17WH1A0403	72	5	10	25	15	75	45	c	0
17WH1A0404	82	5	9	23	15	88	75	A	1
17WH1A0405	87	5	8	25	15	78	70	A	1
17WH1A0406	88	5	10	25	15	99	88	A	1
17WH1A0407	89	5	10	25	15	98	78	A	1
17WH1A0408	76	5	10	22	15	99	50	B	0
17WH1A0409	77	5	10	25	15	99	55	B	0
17WH1A0410	78	5	10	25	15	89	73	A	1
17WH1A0411	90	5	10	25	15	98	75	A	1
17WH1A0412	99	5	10	25	15	97	88	A	1
17WH1A0413	100	5	10	25	15	98	68	A	1
17WH1A0414	100	5	10	25	15	97	55	B	0
17WH1A0415	87	5	10	25	15	96	45	c	0
17WH1A0416	67	5	10	21	15	95	54	B	0
17WH1A0417	79	5	10	25	15	92	55	B	0
17WH1A0418	80	5	10	25	15	91	67	A	1
17WH1A0419	89	5	10	25	15	91	67	A	1
17WH1A0420	93	5	10	25	15	92	67	A	1

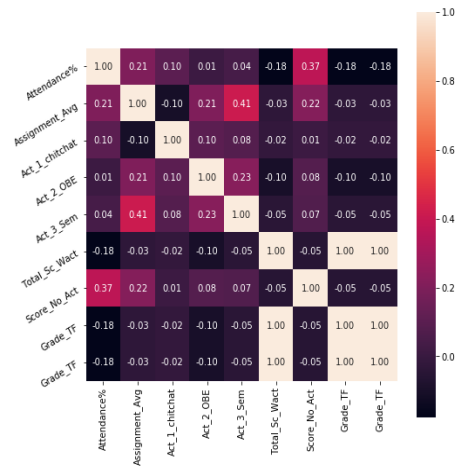


Fig.8. Correlation among x- parameters in the dataset

The correlation among the parameters is shown in Fig. 8 & 9. Each value in the box represents the correlation coefficient and is a measure of linear association between the variables. If the correlation coefficient is near ± 1, then a perfect correlation results i.e. if one variable increases, the other variable tends to also increase (if positive) or decrease (if negative). Similarly, if the coefficient value lies between ± 0.50 and ± 1, leads to a strong correlation.

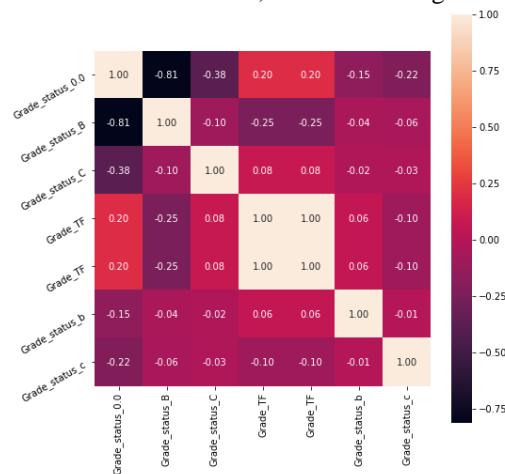


Fig.9. Correlation among y- parameters in the dataset

After the model trained and tested with the dataset values, the expected grade outcomes is shown in Fig. 10 and the performance accuracy of the model is observed is 96.96% as shown in Fig.10. Accuracy of the model can be calculated using the following relation:

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \dots \dots (4)$$

```

In [47]: 1 output_values=np.dot(x[:50],parameters_out["weight"])+parameters_out["bias"]
         2 predictions=sigmoid(output_values)>=1/2
         3 predictions

Out[47]: array([ True,  True, False,  True,  True,  True,  True, False, False,
                True,  True,  True,  True, False, False, False, False,  True,
                True,  True,  True,  True,  True,  True,  True,  True,  True,
                True, False,  True,  True,  True,  True,  True,  True,  True,
                True,  True, False, False, False,  True,  True, False, False,
                False,  True,  True,  True,  True])
    
```

Fig.10. Expected grade outcomes

```

In [237]: 1 y_pred =svc.predict(x_test)
In [238]: 1 accuracy_score(y_test,y_pred)
Out[238]: 0.9696969696969697

```

Fig.10. Performance accuracy in Logistic regression model

4. Results and discussion

In this work, a methodology is proposed that establishes a prediction strategy on improvement of student's academic performance. The model was built from three datasets that are collected from the students marks, attendance and activity scores. This model took 16 weeks i.e a semester duration that includes regular instructions, video lectures, assignments, homework supplemented with various learning activities. The assessment was done based on their performance in the activities. The student response rate was varied from activity to activity, but was in progressive way as observed from the analysis shown in Fig.12.

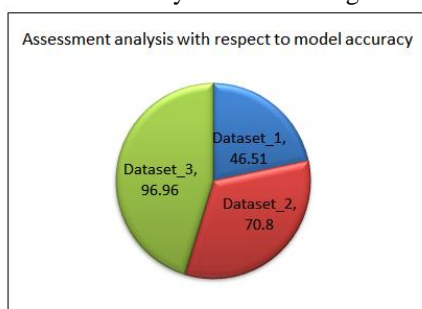


Fig.12. Assessment analysis using model accuracy

5. Conclusions

The aim of the work on activity based analysis is to enhance the learning capability through which to improve the performance of weak or average students along with their bright peers. It is observed that performance improvement in the learning environment was upgraded with various learning tools along with traditional methods of teaching. The existing traditional teaching and prediction methods are lacking in giving support to the latest competitive resourceful environment in achieving high goals. Using traditional assignment scores alone the end semester grades cannot be assessed for the improvement of performance.

Initially the prediction strategy carried out using a few number of regular and routine type of activities. Though the student response is good at those activities, but the contribution of them in the improvement in performance in weak or average students is less as per observation. The prediction analysis initially was carried out using unsupervised k-means clustering method because the initial aim was simply identification of student groups based on their performances. The true performance observed is 46.51%. Later, few more activities were added in the direction of performance improvement, and observed more student involvement in this case. This time a model is built for assessing the student performance based on their activity score. Naïve Bayes model was used, trained and

tested for this analysis and observed the improvement in performance as 70.8%. This model is more suitable for observing the improvement in student's individual performance based on their active involvement in the activities.

Further the analysis was carried out using logistic regression model, taking previous activity scores, assignment scores along with student classroom regularity. By comparing with previous model, this model is more accurate in highlighting the performance improvement and is observed as 96.96%. This study can be extended further to analyse by considering all passed out batches w.r.t to present batch, that helps in taking additional measuring steps in achieving further performance improvement.

References

1. Allyson Barlow and Shane Brown (2020). Correlations between modes of student cognitive engagement and instructional practices in undergraduate STEM courses. <https://doi.org/10.1186/s40594-020-00214-7>.
2. Beth Dietz-Uhler & Janet E. Hurn (2013). Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective, *Journal of Interactive Online Learning*, 12(1), 17-26.
3. Ya-Han Hu, Chia-Lun Lo, Sheng-Pao Shih, (2014). Developing early warning systems to predict students' online learning performance. doi: <http://dx.doi.org/10.1016/j.chb.2014.04.002>.
4. Piet Ankiewicz, (2019). Perceptions and attitudes of pupils towards technology: In search of a rigorous theoretical framework. doi: [org/10.1007/s10798-017-9434-z](https://doi.org/10.1007/s10798-017-9434-z).
5. Susanna Hartikainen, Heta Rintala, Laura Pylväs and Petri Nokelainen (2019). The Concept of Active Learning and the Measurement of Learning Outcomes: A Review of Research in Engineering Higher Education. doi: 10.3390/educsci9040276.
6. Brenda R. Brand (2020). Integrating Science and Engineering Practices: Outcomes from a Collaborative Professional development. <https://doi.org/10.1186/s40594-020-00210-x>.
7. Elkin Taborda, Lorraine Kisselburgh, Tahira Reid Karthik Ramani (2012). Enhancing Visual Thinking In A Toy Design Course Using Freehand Sketching, *Proceedings of the ASME 2012 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2012)*, August 12-15, 2012, Chicago, IL, USA.
8. Salumari Madhu, G. Prabhakara Reddy (2018). Innovations in Teaching Methodologies to Improve the Results of Core Subjects at the first Year Level: A Study under Autonomous Scheme. DOI:10.16920/jeet/2018/v0i0/120915.
9. Nicolette Harris, Cailee E. (2019). Developing Cognitive Skills Through Active Learning: A Systematic Review of Health Care Professions. <https://doi.org/10.4085/1402135>.
10. Kevin Morgan (2011). Problem-based and experiential learning: Engaging students in an undergraduate physical education module. DOI: 10.3794/johlste.101.365.
11. Hyunsook Shin, Sohyune Sok, Kyung Sun Hyun, Mi Ja Kim (2014). Competency and an active learning program in undergraduate nursing education. DOI: 10.1111/jan.12564.
12. Stephen J.H. Yang, Owen H.T.Lu, Anna Y.Q. Huang, Jeff C.H.Huang, Hiroaki Ogata, Albert J.Q.Lin (2018). Competency and an active learning program in undergraduate nursing education, DOI: 10.2197/ipsjip.26.170.
13. Kamson Sirait, Tulus, Erna Budhiarti Nababan (2017). K-Means Algorithm Performance Analysis with determining the Value of starting Centroid with Random And KD-Tree Method, doi :10.1088/1742-6596/930/1/012016.