International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, March-April-2021, ISSN (Online): 2395-566X

Student Performance Prediction using Machine Learning Techniques

M. Tech. Scholar Nandini Sahu, Prof. Ashutosh Khemariya, Prof. Aditi Tiwari
Department of Computer Science Engineering,
Malwa Institute of Science & Technology,
Indore, India.
sahunandini2224@gmail.com, ashutosh.khemariya07@gmail.com, aditi0863@gmail.com

***Abstract*- The capacity to screen the advancement for students' academic execution is an essential issue of the educational Group from claiming higher Taking in. An arrangement to dissecting students' comes about dependent upon group dissection and utilization standard measurable calculations with organize their scores information as stated by the level about their execution may be portrayed. In this paper, we also actualized the k-mean grouping algorithm for examining students' consequence information. Those model might have been consolidated for those deterministic model should dissect those students' effects of a private foundation clinched alongside % Iberia which is a great benchmark with screen the progression of academic execution about people for higher institutional to the reason for making an successful choice by those academic organizers.**

***Keywords*- K – Mean, Clustering, Academic Performance, Algorithm.**

I. INTRODUCTION

Graded perspective Normal (GPA) may be a regularly utilized pointer about academic execution. A significant number Europe, hypothetical orders had more distinction than difficult work, and speculative chemistry was situated at least GPA that if a chance to be looked after so as will keep in the degree program. Over a few Universities, those least GPA prerequisite set for the understudies may be 1. 5. Nonetheless, for whatever graduate program, A GPA from claiming 3. 0 Also over may be recognized a pointer about handy academic execution. Therefore, GPA even now stays those the vast majority regular element utilized toward those academic organizers with assessing progression to academic earth [1].

Several variables Might go about as obstructions with people achieving Also administering An helter smelter GPA that reflects their, Generally speaking, academic execution Throughout their residency, clinched alongside college. These variables Might get a chance to be focused on toward the teaching staff to Creating methodologies on the move forward learner taking in also move forward their academic execution. Eventually, Tom's perusing method for checking that progression from claiming their implementation.

Therefore, Execution assessment is a standout amongst the bases on screen the progression of scholar execution over higher institutional for Taking in. Build once this incredulous issue, grouping for understudies under diverse classifications as stated by their execution need turn into An confounded errand. With universal grouping about understudies In light of their Normal scores, it is was troublesome to get a thorough perspective of the state of the students' execution Also all the while find imperative subtle elements starting with their the long haul to the long run execution.

With that assistance about information mining methods, for example, grouping algorithm, it will be workable to find those way qualities starting with the students' execution also conceivably utilization the individual's aspects for future prediction. There bring been A percentage guaranteeing comes about from applying k-means grouping calculation with the Euclidean separation measure, the place the separation will be registered Eventually Tom's perusing finding the square of the separation the middle of each scores, summing the squares and finding the square root of the entirety of cash [6].

This paper displays k-means grouping algorithm Concerning illustration a straightforward What's more proficient device around will screen those progression of students' execution clinched alongside higher foundation. Bunch examination Might a chance to be isolated under hierarchic grouping What's more non-hierarchical grouping systems.

Samples from claiming hierarchic strategies would single linkage, complete linkage, Normal linkage, median, What's more Ward. Non-hierarchical systems incorporate k-means, versatile k-means, k-medics, also fluffy grouping. Will determine which calculation is handy is An work of the sort of information accessible and the specific reason for examination. In that's only the tip of the iceberg destination way, those solidness from claiming groups

© 2021 IJSRET

camwood a chance to be investigated in reenactment investigations [4].

The issue from claiming selecting the "best" algorithm/ parameter setting will be a troublesome particular case. A great grouping calculation ideally ought to transform Assemblies for different non-overlapping boundaries, in spite of the fact that a flawless detachment camwood not normally make attained done act.

Figure for value measures (indices) for example, those profile width [4] alternately those homogeneity list [5] could a chance to be used to assess the caliber about detachment got utilizing An grouping calculation. The idea for solidness of a grouping calculation might have been viewed as over [3]. Those thought behind this acceptance methodology will be that an calculation ought further bolstering make rewarded to consistency. In this paper, we executed conventional intends grouping algorithm [6] and Euclidean separation measure of comparability might have been decided to be utilized within those Investigation of the students' scores.

II. METHODOLOGY

A. Development of k-mean clustering algorithm Given a dataset of n data points x1, x2, …, an such that each data point is in Rd , the problem of finding the minimum variance clustering of the dataset into k clusters is that of finding k points $\{m_j\}$ (j=1 , 2..... K) in R^d such that

$$\frac{1}{N}\sum_{i=1}^{n}[min_j d^2(x_i, m_j)] \quad (1)$$

is minimized, where d(x_i,m_j)denotes the Euclidean distance betweenx_i and m_j. The points {m_j } (j=1 , 2..... K) are known as cluster cancroids'. The problem in Eq.(1) is to find k cluster cancroids', such that the average squared Euclidean distance (mean squared error, MSE) between a data point and its nearest cluster cancroids is minimized.

Those k-means calculation gives a simple strategy should execute estimated answer for eq. (1). The purposes behind the Notoriety from claiming k-means are simplicity Also Straightforwardness for implementation, scalability, pace from claiming merging What's more versatility to meager information.

Those k-means algorithm camwood be considered perfect Likewise An gradient plummet procedure, which starts during beginning bunch cancroids', and iteratively updates these cancroids' should diminishing the destination capacity Previously, eq. (1).

The k-means dependably meet with An neighborhood base. Those specific nearby least found relies on the beginning group cancroids'. The issue for finding the worldwide least will be NP-complete. Those k-means calculation updates group cancroids' till neighborhood least is found. Fig. 1 indicates the summed up pseudo codes of k-means algorithm; and customary k-means calculation is introduced clinched alongside fig. 2 separately.

Preceding the k-means calculation converges, separation What's more cancroids calculations need aid carried out same time loops need aid executed An amount about times, say l, the place the certain basic l may be known as the amount from claiming k-means iterations. The exact esteem for l differs relying upon the starting beginning bunch cancroids' much on the same dataset.

With the goal those computational period intricacy of the calculation is O(nkl), the place n will be those aggregate number about Questions in the dataset, k will be those required amount of groups we recognized and l will be the amount from claiming iterations, k≤n, l≤n [6].

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values
Step 2: Initialize the first K clusters
- Take first k instances or
- Take Random sampling of k elements
Step 3: Calculate the arithmetic means of each cluster formed in the dataset.
Step 4: K-means assigns each record in the dataset to only one of the initial clusters
- Each record is assigned to the nearest cluster using a measure of distance (e.g Euclidean distance).
Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all
the clusters in the dataset.

Fig 1. Generalized Pseudo Code of Traditional K-Means.

```
1 MSE = large number;
2 Select initial cluster cancroids' {mj}j
K = 1;
3 Do
4 OldMSE = MSE;
5 MSE1 = 0;
6 For j = 1 to k
7 mj = 0; NJ = 0;
8 end for
9 For i = 1 to n
10 For j = 1 to k
11 Compute squared Euclidean
Distanced2 (xi, mj);
12 end for
13 Find the closest cancroids mj to xi;
14 mj = mj + xi; nj = nj+1;
15 MSE1=MSE1+ d2(xi, mj);
16 end for
17 For j = 1 to k
18 NJ = max (NJ, 1); mj = mj/nj;
19 end for
20 MSE=MSE1;
while (MSE<OldMSE)
```

Fig 2. Traditional K-Means Algorithm.

© 2021 IJSRET

International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 2, March-April-2021, ISSN (Online): 2395-566X

III. OUTCOMES

We connected the model on the information situated (academic effect about one semester) of a school to Nigeria. Those result created will be demonstrated to tables 2, 3, and 4, individually. In table 2, for k = 2; to group 1, those group measure will be 25 and the general execution is 62. 22. Also, the group sizes and the in general exhibitions to group numbers 2 Also 3 need aid 15, 29 and 45. 73 and 53. 03, consciously. Comparative analyses also hold for tables 3 what's more 4. The graphs would createon figures 3, 4 what are more 5, respectively, the place the generally speaking execution is plotted against the bunch size.

Table 5 indicates the measurement of the information set (Student's scores) in the structure n by m matrices, the place n is the rows (# from claiming students) Also m will be those section (# about courses) advertised by each learner.

The in general execution is assessed toward applying deterministic model on eq. 2 [7] the place the aggregation evaluation to every of the bunch size is assessed toward summing the Normal of the distinct scores clinched alongside every bunch.

$$\frac{1}{N}\left(\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)\right) \qquad (2)$$

Where;
N = the total number of students in a cluster and
n = the dimension of the data.

Table 1. Performance Index.

70 and above	Excellent
60-69	Very Good
50-59	Good
45-49	Very Fair
40-45	Fair
Below 45	Poor

Over figure 3, the in general execution to bunch extent 25 may be 62. 22% same time the general execution for bunch extent 15 will be 45. 73% also bunch measure 29 need those in general execution for 53. 03%.

This investigation demonstrated that, 25 out about 79 people needed a "Very Good" execution (62. 22%), same time 15 out about 79 scholars needed execution in the locale about extremely "Fair" execution (45.73%) and the remaining 29 people required an "Good" execution (53. 03%) Concerning illustration delineated in the execution list clinched alongside table 1.

Figure 4 indicates the patterns for execution examination as follows; general execution to group size 24 is 50. 08% same time the generally execution for group span 16 may be 65. 00%.

Group extent 30 needs those in general execution of 58. 89%, same time group extent 09 is 43. 65%. Those patterns in this dissection shown that, 24 people fall in those area from claiming "Good" execution list. On table 1 over (50. 08%), same time 16 scholars need execution in the locale for "Very Good" execution (65. 00%). 30 scholars need An"Good" execution (58. 89%) Also 9 people required execution for "Fair" effect (43. 65%).

For figure 5, those in general execution for bunch span 19 will be 49. 85%, same time those in general execution to bunch size 17 may be 60. 97%. Bunch span 9 needsthat general execution of 43. 65%, same time those bunch extent 14 need generally execution of 64. 93% what's more group measure 20 need generally execution about 55. 79%.

This execution examination shown that, 19 people crossed through with "Good" execution locale (49. 85%), same time 17 understudies needed "Very Good" execution comes about (60. 97%). 9 learners fall in the locale from claiming "Fair" execution list (43. 65%), 14 understudies were in the district for "Very Good" execution (64. 93%) and the remaining 20 understudies required "Good" execution (55. 79%).

Table 2. K = 2.

Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	K_means_cluster_labels	points belonging to cluster
61.34375	59.46875	63.125	55.1875	47.15625	0	32
91.83333	86.83333	91.33333	92	92.5	1	6

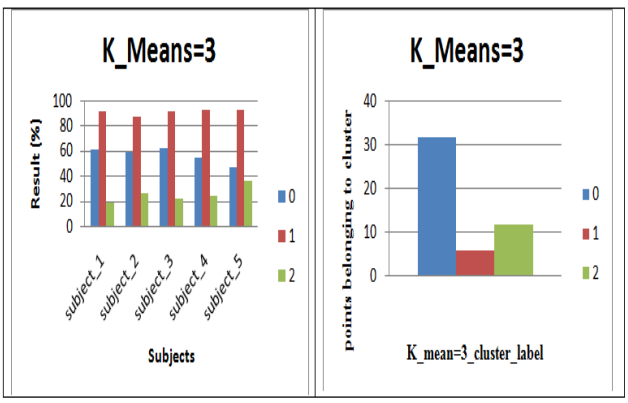

Fig 3. Overall Performance Versus Cluster Size. (# of Students) k = 3

© 2021 IJSRET

Table 3. K = 4

Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	K_means_cluster_labels	points belonging to cluster
72.5	62.5	75.5	36	24	0	24
91.833333	86.833333	91.333333	92	92.5	1	6
57.625	58.458333	59	61.583333	54.875	2	12
18.5	26.75	22.25	24.666667	36.083333	3	8

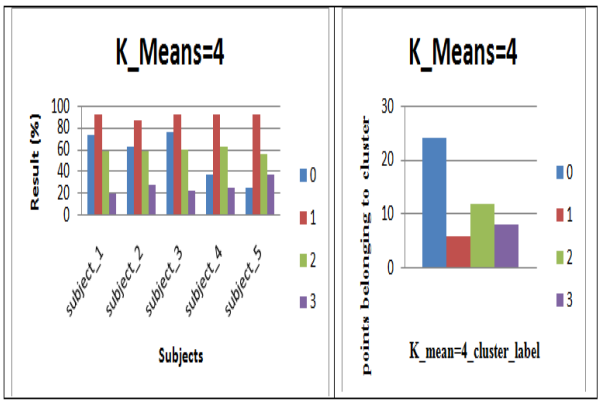

Fig 4. Overall Performance Versus Cluster Size (# of Students) k = 4

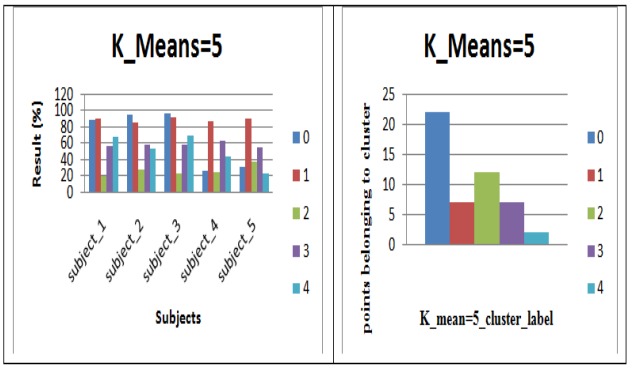

Fig 5. Overall Performance Versus Cluster Size (# of Students) k = 5

Table 4. K = 5.

Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	K_means_cluster_labels	points belonging to cluster
88	94	96	25	31	0	22
90.285714	84.285714	90.857143	86.571429	89.571429	1	7
18.5	26.75	22.25	24.666667	36.083333	2	12
56	57.818182	57.227273	61.636364	55.181818	3	7
67.714286	53.428571	68.714286	43.714286	23	4	2

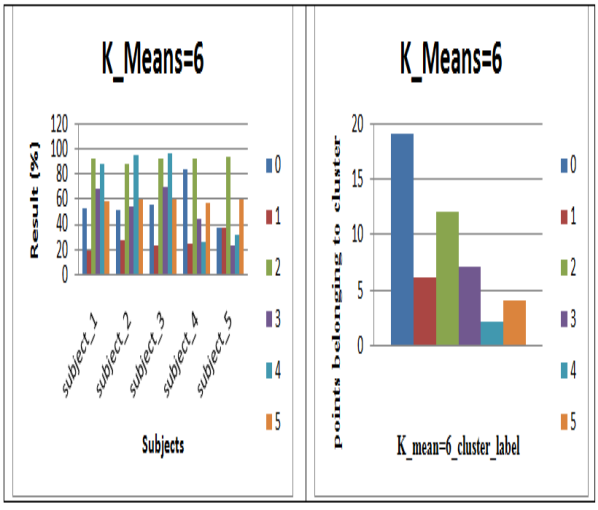

Fig 6. Overall Performance Versus Cluster Size (# of Students) k = 6

Table 5. K = 6

Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	K_means_cluster_labels	points belonging to cluster
52	51	54.5	83.5	36.5	0	19
18.5	26.75	22.25	24.666667	36.083333	1	6
91.833333	86.833333	91.333333	92	92.5	2	12
67.714286	53.428571	68.714286	43.714286	23	3	7
88	94	96	25	31	4	2
58.157895	59.842105	59.421053	56.631579	60	5	4

3.1 Graph between Id and sgpa:

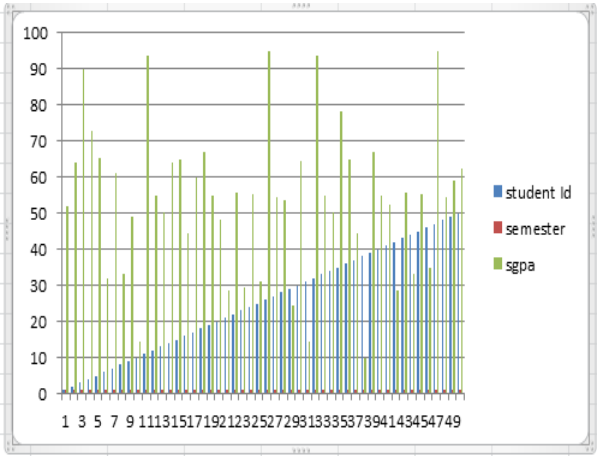

Fig 7. Bar Graph between Id, Semester, Sgpa.

The figure 7 shows the bar graph between id, sem, sgpa. These shows that for different it different sgpa are there in semester 1, so semester is same and green line shows the sgpa foe every student. According to these it can be found out the number of students in between the different sgpa.

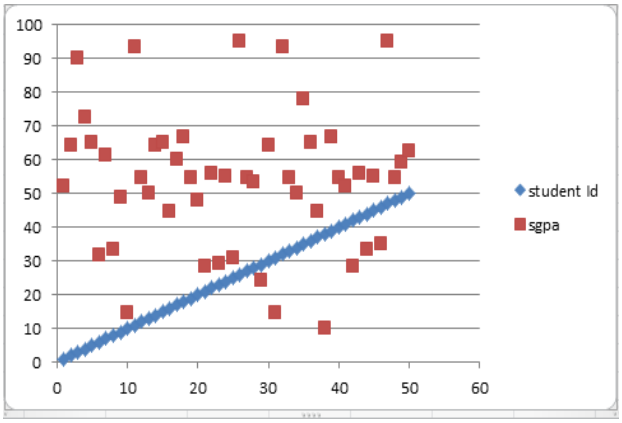

Fig 8. Scatter graph between Students Id and sgpa.

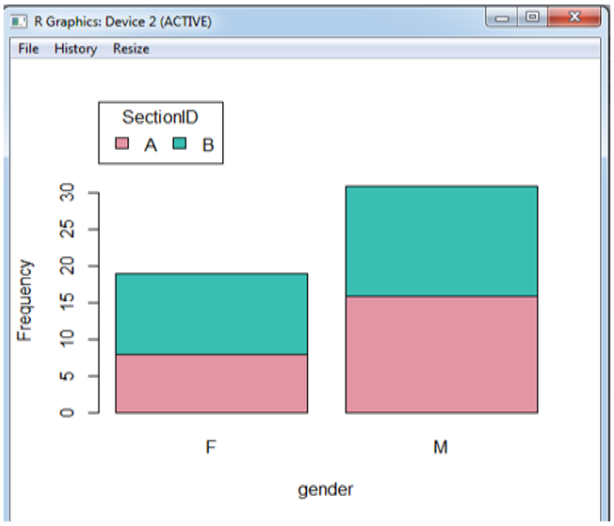

Fig 9. Bar graph for gender and section Id.

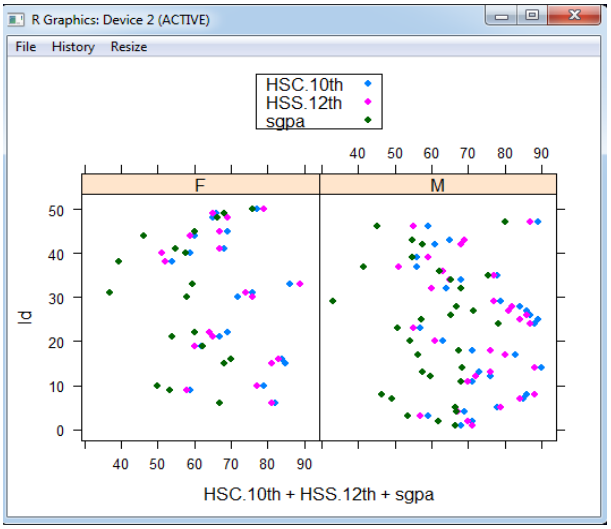

Fig 10. XY Conditioning Plot.

© 2021 IJSRET

3.2 Result Analysis:

Table 6. Compression Study Existing Work versus Proposed Work.

Parameters	Existing Work	Proposed Work
	Decision Trees	Modified K-Mean
Accuracy	Above 58%	Above 71 %
Source Code Execution		10.32 Seconds

IV. DIALOG FURTHER MORE CONCLUSION

In this paper, we given a basic Furthermore qualitative technique on analyze those predictive force of grouping algorithm and the Euclidean separation concerning illustration and measure of comparability separation. We showed our techno babble utilizing methods grouping calculation [6] Also consolidated for those deterministic model in [7] on a information situated from claiming private school outcomes with nine courses advertised for that semester to each scholar for downright amount about 79 students, What's more produces those numerical under standing of the outcomes to the Execution assessment. This model progressed around a few of the impediments of the existing methods, for example, model created Eventually Tom's perusing [7] Furthermore [8].

These models connected fluffy model to foresee students' academic execution on two dataset best (English dialect what's more Mathematics) for auxiliary Schools comes about. Additionally those exploration fill in by [9] best gives information mining structure to Students' academic execution. The examination by [10] utilized harsh set hypothesis as a arrangement methodology on examine scholar information the place the Rosetta toolkit might have been used to assess the learner information will portray different dependencies between those qualities and the person status the place the uncovered examples would clarified for plain English.

Therefore, this grouping calculation serves as a great benchmark on screen that progression of students' execution on higher organization. It also enhances the choice making toward academic organizers should screen the candidates'execution semester by semester toward enhancing on the future academic brings about the subsequence academic session.

REFERENCES

[1] S. Suitsanger, M. Bole, Furthermore K. Sail, "Factors that influence academic execution "around drug store students," American diary of pharmaceutical Education, 2006.

[2] Susitna Data Furthermore Senath Data, "Comparisons and acceptance about measurable grouping systems to microarray gene interpretation data," Bioinformatics, vol. 19, pp. 459–466, 2003.

[3] Rousseau P. J, "A graphical help of the translation Furthermore acceptance for bunch analysis," diary about computational Apple Math, vole 20, pp. 53– 65, 1987.

[4] Shamir R. What's more Sharon R.,"Algorithmic methodologies should grouping gene interpretation data," Previously, current Topics clinched alongside computational atomic science mitt Press; pp. 53-65, 2002.

[5] Mocha H. J.,"Adaptive bunch analysis, order What's more multivariate Graphics,"Weirstrass foundation to connected examination Furthermore Stochastic, 1992.

[6] Fathom An. M. ,Salem a. M. , Turkey f. A. What's more Ramadan m. A.,"A proficient improved k-means grouping algorithm," diary about Zhejiang college science An., pp. 1626–1633, 2006?

[7] J. O. Omolehin, J. O. Overlade, O. Ojeniyi and K. Raff, "Application from claiming fluffy rationale clinched alongside choice making around students' academic performance," announcement from claiming immaculate Furthermore connected Sciences, vol. 24E(2), pp. 281-187, over 2,800 doctor look assignments led from April 1, 2009 to March 31, 2010.

[8] J. O. Omolehin, An. O. Enikuomehin, R. G. Jimson And K. Raff, "Profile of conjugate gradient system algorithm on the execution examination for An fluffy system,"African diary about math and PC science Research," vol. 2(3), pp. 030-037, 2009.

[9] N. V. An And Kumar Furthermore G. V. Umea, "Improving Academic execution from claiming people Eventually Tom's perusing Applying information mining Technique,"European diary from claiming exploratory Research, vol. 34(4), 2009.

[10] Varapron P. et al. ,"Using harsh set hypothesis for programmed information Analysis," 29th congress ahead science and innovation organization of Thailand, 2003.

© 2021 IJSRET