

PERFORMANCE ANALYSIS OF MULTI FACTOR INFLUENCE IN STUDENT ACADEMIC PERFORMANCE

¹Chiranjeevi Kommula, ²Dr. B V V Shiva Prasad

¹ Student – Author , ² Guide

Dept. of Computer Science and Technology, Career Point University, kota, Rajasthan.

Email - ckommula@gmail.com

drbvshivaprasad@gmail.com

Abstract: Since more than a decade, Educational research accelerated in which various areas like learning analytics, Outcome based education, student academic performance were explored resulting innovation in education. There was a significant change in the scenario of education where tremendous progress had been noted. In realizing this remarkable progress, technologies like Data Mining, Machine learning and others aided a lot. Most importantly the techniques of Machine Learning falling under classification, prediction, learning, clustering, and dimensionality reduction paved the way to make a step up in education domain. These techniques added intelligence to the domain. In this paper, concerning student academic performance we analyzed the factors influencing the overall performance. The computations are carried out basing on the correlation coefficients of the feature vector. We made our effort and succeeded in emphasizing the need of various other factors in addition to marks and grades in achieving better performance.

Key Words: Educational analytics, Learning analytics, Student Academic performance.

1. INTRODUCTION:

Education is the primary strength of any Nation. More the literacy rate proportionate is the development. In most of the countries the learning amenities are scanty and the development in this domain during past decades is also at slow pace. But the technology changed the scenario. The progress in education and its related technologies registered is worthwhile and drastic. Learning Analytics is a predominant area nowadays which could be defined as the integration of Information Systems in education. It interprets data analytics and data mining techniques in the field of education. It is a process which comprises the measurement, collection, analysis and reporting of data related to every educational aspect so as to optimize the learning.

As per [12], Lap-Kei Lee et.al. the fields and subfields of educational technology is receiving growing attention from educational researchers and practitioners (Hui and Kwok 2019)[14]. With the evidences of students and teachers in teaching and learning process, learning analytics extracts the learning behaviors (Wong and Chong 2018)[15] and gives insights for the development and innovation. Consequently the policymakers, instructors, and learners (Hwang et al. 2014)[16] could make a revision of educational policies which empower the nation. Data mining, Machine Learning, Augmented and Virtual Reality, Deep learning and Predictive Analytics are the major technologies supplementing the educational progress in which the research community is exploring new ways in accomplishing the domain standardization. The pedagogical perspective is not limited to any certain level of education where in the research has its impact giving futuristic insights.

Data Mining is the area of Computer science which extracts the statistical information from the available raw data. On the other hand Machine learning with its vast algorithmic space, adds intelligence to the existing technology irrespective of the application into which it is being incorporated. Since there is a lot of data being generated from time to time the aforementioned technologies could be effectively used to explore the novelty and reap the benefits.

Student academic performance is influenced by many factors in which age, parents, family, personal traits, psychological aspects, habits are few amongst. Due to various reasons the nature of study among the students varies. Despite the reasons, the students' academic performance must not be lowered. Often the factors in various combinations might influence the performance. In this paper giving an introduction to what we are intended in showing that there is a certainty in our claim that multiple factors influence the student academic output, the rest of paper is organized in such a way that the second section gives the information about how various researchers had given their contributions followed by the proposed work in next section. In the fourth section the results and discussion are mentioned and the final section concludes the paper.

2. RELATED WORK:

Filippo Sciarrone et.al. in [1] explained the trend in educational paradigm shift through a multifaceted view of different disciplines. They highlighted the growth of web based education systems during the recent past and about how much they are useful to the world. They advocated on how various processes like learning, teaching and administration could be strategically improved with the help of various techniques of machine learning, data mining – especially educational data mining, teaching and learning analytics. They discussed on how the process of learning and teaching improvement is necessary and in this aspect various models were introduced and discussed. In addition, they cited regarding how the whole learning process could be optimized with the help of Analytics.

Enhancement of knowledge management and learning is the prime task in Education management. **Oyerinde O. D. et.al.** in [2] proposed a framework which administers the prediction of academic performance of students. Multiple Linear Regression was used for analysis and prediction. In their research they were able to build a model which predicts performance of a student and also derive associations in choosing the courses. Their future insights are to build a central repository of education data pertaining to students, teachers, courses, modes of learning & teaching and the experiences. Subsequently various techniques could be applied to extract the required information.

Mere Learning could not be accomplished without understanding the underlying patterns and characteristics. **Billy Tak Ming Wong et.al.** through their research work on Learning analytics [3] demonstrated to what extent the learning analytics are useful in the higher education and how to reap the benefit in doing so. For their study they collected experimental data and case studies on analytics, categorized basing on the objectives, approaches followed and outcomes. Their findings include how the learning analytics could facilitate in improving the instructional designs. Also their research outcome indicated the evaluation of pedagogies and the effectiveness involved. Close monitoring of students could be accomplished as well students at academic risk could be identified. Due to their research, areas such as Quality assurance and student support are still strengthened. It is vivid that the learning performance seemed to be better after using Learning analytics.

Predictive analytics is the concept of analyzing the data and predicting the future happenings. These analytics are very much required in any domains and education is not an exception. In [4], **V. Belsini Gladshiya et.al.** focused on the usage of predictive analytics and they play a vital role in predicting the future challenges. Through implementing prediction algorithms on the data and evidences, the risk factors could be predicted consequently with which a robust transformation can be made in higher education. The growth in education domain is phenomenal and it could be achieved only due to the usage of analytics. The overall idea of education research is to achieve success in the academics and the future life gets amplified and analytics play the necessary role in accomplishing that.

Learning outcomes are the principal outcomes of any education system which include program outcomes, program educational outcomes and program specific outcomes. Any educational institution must attain these and such attainment indicates the quality of the institution. **Abdallah Namoun et.al.** in [5] had made a thorough research on the topic for more than a decade which covers the period between 2010 and 2020. They conducted a survey using the literature to conceptualize the usage of intelligent techniques to predict student performance. Their focus is on how the learning outcomes are predicted, the model formulated to forecast student learning and the factors accountable for student outcomes. They carried out their work using 62 relevant papers from the bibliographic database in the above 3 dimensions. So as to classify the student performance few supervised machine learning models in addition to regression were used and finally they could identify that student academic emotions, term assessment grades and online learning activities are the most prevalent factors of learning outcomes.

Ghaith Al-Tameemi et.al. also carried out their work in the sphere of predictive learning analytics but they focused on the student retention factors leading to student success. The process of predictive learning analytics was explained in detail in [6] which involves various other processes like data collection, data preprocessing, data mining, and others. As a part of their study various machine learning algorithms were explored and compared which could determine the most suitable algorithms for learning analytics. Student logs in learning platforms as well student information systems are the major data sources besides Naïve Bayes, Artificial Neural Network, and Decision Tree algorithms were widely been used in predicting student performance. They illustrated how frequently the ML algorithms used in various aspects in evaluating the classification accuracy, confusion matrix and AUC-ROC curves.

In [7], **Zacharoula Papamitsiou et.al.** explained how the research community of Learning Analytics are focusing on the area for the past decade. They aimed at identifying a conceptual structure of the evolution in learning analytics and further emerging topics. An elaborative analysis was carried out in their study, considering 459 proceeding articles of

LAK conferences and 168 Journal papers, 3092 keywords which were mentioned by the authors and 4051 machine extracted key phrases. Co-word analysis was applied on the concerned literature so as to extract the essential information from the text. They constructed an intellectual map of learning analytics community with the help of strategic diagrams, network analysis and hierarchical clustering. The major focus areas of their study include MOOCs, Learning Management systems, Natural language processing. They highlighted that there was a significant paradigm shift in the research interests in the Learning Analytics community.

Learning analytics research is not confined to certain level of education. At all levels of education there must be an advancement and yield. **Erverson B. G. de Sousa et.al.** in [8] made a detailed study focusing on implementation of learning analytics at high school level. They performed systematic literature review following a four step process viz., search, select, critical assessment and extraction of relevant field.

Educational data mining is one among vital areas in the domain and it was addressed by **Chitra Jalota et.al.** in [9]. EDM engrosses ML algorithms and statistical techniques in interpreting the how learning habits of students, their academic performance and . to help the user for interpretation of student's learning habits, their academic performance and further progress. Various techniques predicting the performance measure of students were discussed in their contribution for which kalboard 360 dataset and weka tool were used for analysis. They used 5 different classifiers and performance comparisons are drawn in terms of accuracy and error measures. They could identify that Multilayer Perceptron exhibited the best performance among the five approaches.

Carolina Guzmán-Valenzuela et.al. in [10] made a detailed study on the publication patterns on the education analytics in higher education and the major challenges. The authors examined 385 papers with the intent of giving insights that distinguishes a practice based and management oriented learning community. Undoubtedly, analytics play a powerful role but the focus to be thrown upon the learning activity also. In their study it was identified that the research community mostly focused on analytics leaving the learning process aside which may not yield promising results. Although the research is being taken place worldwide mostly the countries in North are conducting a quantifiable research in education and learning analytics. Study reveals that despite the sound research, there was only a slight throw of data mining and analytical techniques on complex learning processes. As per (Ferguson 2012), (Ferguson and Clow 2017; Leitner et al. 2017; Viberg et al. 2018)[17-20] there is a deficit in research of finding out novel educational theories relating to student learning and there is a need of emphasis on educational perspectives. A reinforced framework is very much required on learning and techno perspectives augmented with audit arrangements and managerial discourse.

Social media is a double edged knife which can impact any entity in either way. Education is also seriously impacted by various social media platforms. In [11], **Yu-Sheng Su et.al.** made their efforts to give their inputs in this perspective. Sharing of material online had been trending and accordingly the flipped classroom mechanism also provisioned the students to preview the material before class and even while class is going on. Diversified interactions like sharing material and knowledge, annotating learning materials, and setting up common objectives became common with the help of social media. Few studies explored on how the student attitude, performance and engagement mode of students were impacted due to the launched social media. While in this paper the authors with the help of educational data mining, explored on the relationship between the student performance in flipped classrooms and their behavior in accessing learning material. The study is carried out by experimenting with two groups among which one group uses the media for flipped learning while the other group is a control group uses the learning management system. It was observed that the learning performance of experimental group which used the social media seems to be more than the control group.

The analysis was done using four indicators - viewing time, active viewing time, viewed amount, and actively viewed amount. The students are clustered into three groups among which the first cluster has the actively engaged group of students. The second and third clusters include engaged group and long term engaged group respectively. Subsequently various tests like Kruskal-Wallis test, Mann-Whitney U test were conducted to identify the cluster deviations. They concluded that social media platforms such as facebook might be very much useful in active engagement of flipped classrooms so as to maximize the learning achievements of students.

3. PROPOSED WORK:

In our proposed work we considered the data [13] pertaining to the student performance. The dataset contains total 649 rows where each record is defined with 33 parameters. The data consists of demographic details, parental details, student habits, grades obtained. Basically the student performance is measured in terms of grades attained in the classes. Depending upon the grade secured, students are generally classified into various categories. But the performance not

only depends on the grades obtained but also various factors like the parental occupation, habits of students and other aspects. The spread of data with respect to various highly influential parameters could be observed from the below box plots (figures 1(a) to 1(i)).

The numerical values of the parameters are as mentioned.

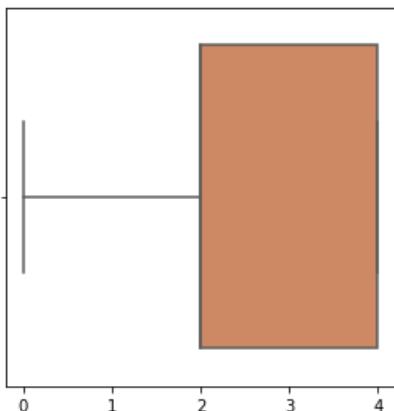


Fig 1(a): Mother education

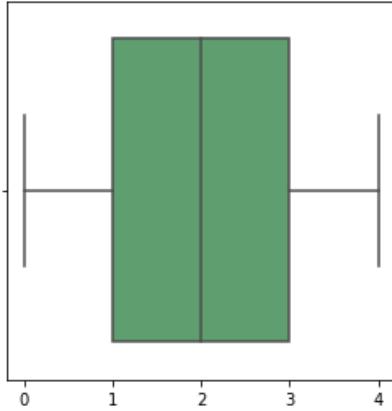


Fig 1(b): Father education

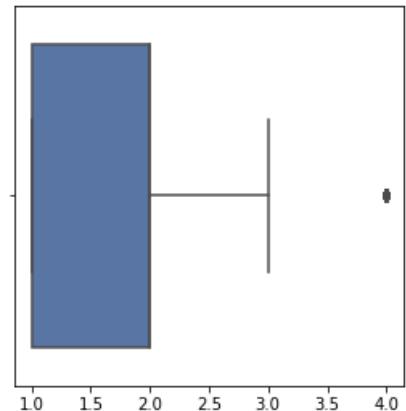


Fig 1(c): Study time

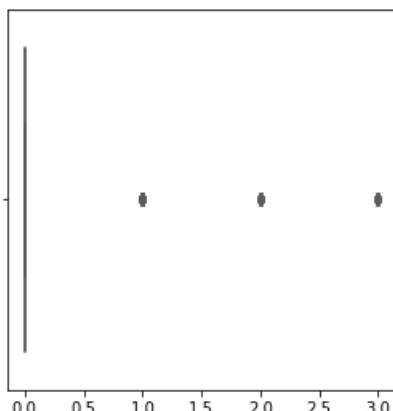


Fig 1(d): failures

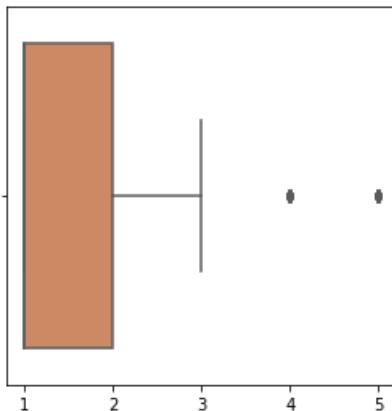


Fig 1(e): D Alc consumption

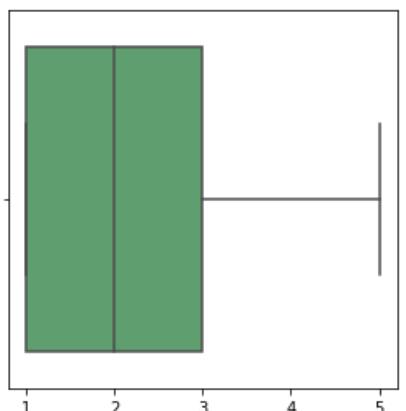


Fig 1(f): W Alc consumption

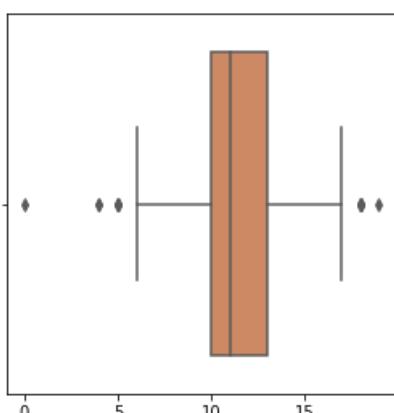


Fig 1(g): Grade 1

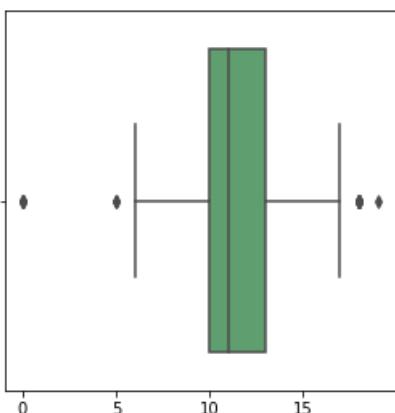


Fig 1(h): Grade 2

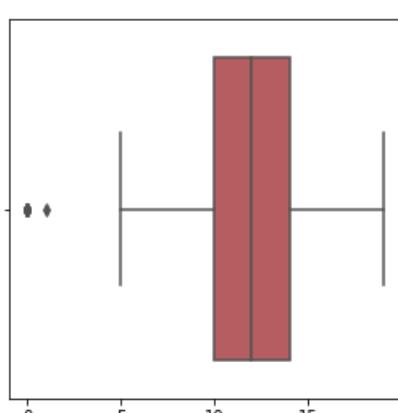


Fig 1(i): Grade 3

Parent education

: numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 – secondary, 4 – Higher education

Weekly study time: numeric : 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours

Number of past class failures : numeric: n if $1 \leq n \leq 3$, else 4

Workday alcohol consumption: numeric: from 1 - very low to 5 - very high

Weekend alcohol consumption: numeric: from 1 - very low to 5 - very high

Grade 1 to 3 : numeric: from 0 to 20

We are deliberated to identify the way the attributes impact the grade. Moreover we also are intended to make out the fact that the categorization of students is more appropriately done basing on the influential factors rather than only the grades. For this we calculated the average grade of students, thereafter the correlation coefficients of all the parameters were computed in association with the average grade.

The algorithm of our proposed mechanism is as follows.

Step 1: Start

Step 2: Consider the parameters P_1 to P_{33} and D_1 to D_{649}

Step 3: Initialize $\text{classt}_{\text{Per}}$, classt_G to 0

Step 4: for each r $G_{\text{avg}}[r] \leftarrow (G_1[r] + G_2[r] + G_3[r]) / 3$

Step 5: for each i do step 6

Step 6: $\text{Cor}[i] \leftarrow \text{CorrelationCoefficient}(P_i, G_{\text{avg}}[r])$

Step 7: $\text{Thr}_P \leftarrow 0.2$, $\text{Thr}_N \leftarrow -0.15$

Step 8: for each i do step 9 and 10

Step 9: if ($\text{Cor}[i] \geq \text{Thr}_P$) $W[j] \leftarrow \text{Cor}[i]$, $F[j] \leftarrow i$

Step 10: if ($\text{Cor}[i] \leq \text{Thr}_N$) $W[k] \leftarrow \text{Cor}[i]$, $F[k] \leftarrow i$

Step 11: Load the dataset D

Step 12: for each $D[r]$ do step 13

Step 13: for each j, k do step 14

Step 14: $\text{Per} \leftarrow \text{Per} + W[j] * P_{F[j]}[r] + W[k] * P_{F[k]}[r]$

Step 15: for each $\text{Per}[r]$ do steps 16 to 19

Step 16: if ($\text{Per}[r] / \max(\text{Per}[r]) * 100 \leq 20$) $\text{class1}_{\text{Per}} \leftarrow \text{class1}_{\text{Per}} + 1$

Step 17: if ($\text{Per}[r] / \max(\text{Per}[r]) * 100 > 20$ and $\text{Per}[r] / \max(\text{Per}[r]) * 100 \leq 40$) $\text{class2}_{\text{Per}} \leftarrow \text{class2}_{\text{Per}} + 1$

Step 18: if ($\text{Per}[r] / \max(\text{Per}[r]) * 100 > 40$ and $\text{Per}[r] / \max(\text{Per}[r]) * 100 \leq 60$) $\text{class3}_{\text{Per}} \leftarrow \text{class3}_{\text{Per}} + 1$

Step 19: if ($\text{Per}[r] / \max(\text{Per}[r]) * 100 > 60$ and $\text{Per}[r] / \max(\text{Per}[r]) * 100 \leq 80$) $\text{class4}_{\text{Per}} \leftarrow \text{class4}_{\text{Per}} + 1$

Step 20: if ($\text{Per}[r] / \max(\text{Per}[r]) * 100 > 80$) $\text{class5}_{\text{Per}} \leftarrow \text{class5}_{\text{Per}} + 1$

Step 21: for each $G_{\text{avg}}[r]$ do steps 22 to 26

Step 22: if ($G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 \leq 20$) $\text{class1}_G \leftarrow \text{class1}_G + 1$

Step 23: if ($G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 > 20$ and $G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 \leq 40$) $\text{class2}_G \leftarrow \text{class2}_G + 1$

Step 24: if ($G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 > 40$ and $G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 \leq 60$) $\text{class3}_G \leftarrow \text{class3}_G + 1$

Step 25: if ($G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 > 60$ and $G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 \leq 80$) $\text{class4}_G \leftarrow \text{class4}_G + 1$

Step 26: if ($G_{\text{avg}}[r] / \max(G_{\text{avg}}[r]) * 100 > 80$) $\text{class5}_G \leftarrow \text{class5}_G + 1$

Step 27: barplot(class1_{Per}, class1_G, class2_{Per}, class2_G, class3_{Per}, class3_G, class4_{Per}, class4_G, class5_{Per}, class5_G)

Step 28: Stop

Observing the coefficients we considered the most influential positive and negative factors by fixing positive and negative threshold values, after which a performance score is computed. The positive threshold is 0.2 while the negative threshold is -0.15. Basing on score the students are categorized and comparisons were drawn. It is observed that there is a variation in number of students categorized based on performance score and average grade. Hence it could be substantiated that the student performance not only depends on the marks/grades obtained but also other vital factors.

In the above algorithm

P_1 to P_{33} indicates the parameters,

D_1 to D_{649} indicates the dataset rows.

Thr_P and Thr_N are the Threshold values of positive and negative correlation coefficients.

$\text{Cor}[i]$ is the correlation coefficient values of each parameter with average grade.

$W[j]$ is the weights array. $\text{Per}[r]$ is the performance score computed for each row.

$G_{\text{avg}}[r]$ is the Average grade value of each row.

$\text{classt}_{\text{Per}}$, classt_G indicates the Class t with respect to performance score and Average grade

4. RESULTS AND DISCUSSION:

The implementation of our proposed work is done in Python. The correlation coefficients of all the attributes with the average grade score obtained are as shown in table 1, besides the graphical representation could be observed in figure 2.

Table 1: Correlation coefficients of the attributes

S.No	Feature	Correlation Coefficient	S.No	Feature	Correlation Coefficient	S.No	Feature	Correlation Coefficient
1	school	-0.3	12	guardian	-0.1	23	romantic	-0.092
2	sex	-0.12	13	travelttime	-0.15	24	famrel	0.071
3	age	-0.13	14	studytime	0.26	25	freetime	-0.11
4	address	0.17	15	failures	-0.41	26	goout	-0.084
5	famsize	0.046	16	schoolsups	-0.068	27	Dalc	-0.21
6	Pstatus	0.011	17	famsup	0.048	28	Walc	-0.17
7	Medu	0.27	18	paid	-0.053	29	health	-0.082
8	Fedu	0.23	19	activities	0.072	30	absences	-0.13
9	Mjob	0.17	20	nursery	0.035	31	G1	0.93
10	Fjob	0.085	21	higher	0.35	32	G2	0.97
11	reason	0.15	22	internet	0.15	33	G3	0.96

Description of the parameters:

school - student's school

sex - student's Gender

age - student's age

address - student's home address type (binary: 'U' - urban or 'R' - rural)

famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)

Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)

Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')

guardian - student's guardian (nominal: 'mother', 'father' or 'other')

travelttime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

schoolsups - extra educational support (binary: yes or no)

famsup - family educational support (binary: yes or no)

paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)

activities - extra-curricular activities (binary: yes or no)

nursery - attended nursery school (binary: yes or no)

higher - wants to take higher education (binary: yes or no)

internet - Internet access at home (binary: yes or no)

romantic - with a romantic relationship (binary: yes or no)

famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

freetime - free time after school (numeric: from 1 - very low to 5 - very high)

goout - going out with friends (numeric: from 1 - very low to 5 - very high)

health - current health status (numeric: from 1 - very bad to 5 - very good)

absences - number of school absences (numeric: from 0 to 93)

It is observed that few attributes are positively correlated while some other attributes are negatively correlated. Moreover some attributes are highly correlated and some other are least correlated. In our empirical process we considered only few attributes which are positively as well negatively correlated and those are beyond the threshold values. The thresholds considered are 20% on the positive side and 15% on the negative side. With the chosen attributes and considering their coefficients as weights the performance score of each student is computed. Basing on the performance score and average grade score the students are categorized into five classes which are as shown in table 2. The corresponding bar plot is shown in the figure 3.

Table 2: student categorization basing on score and grade

CLASS	No. of students categorized as per Score	No. of students categorized as per Average Grade score
1	8	8
2	55	50
3	240	315
4	258	241
5	88	35

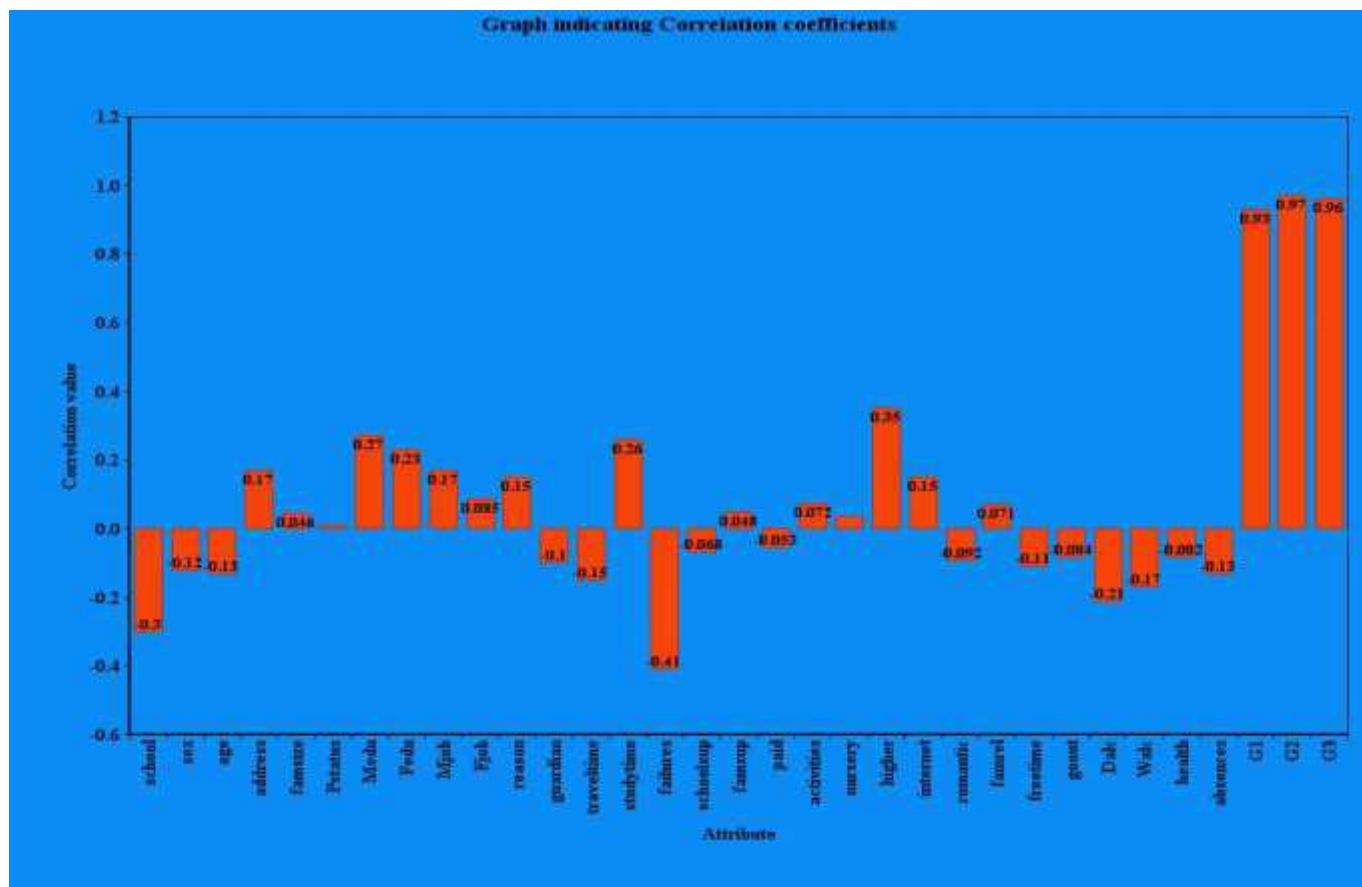


Fig 2: Correlation coefficients of the attributes

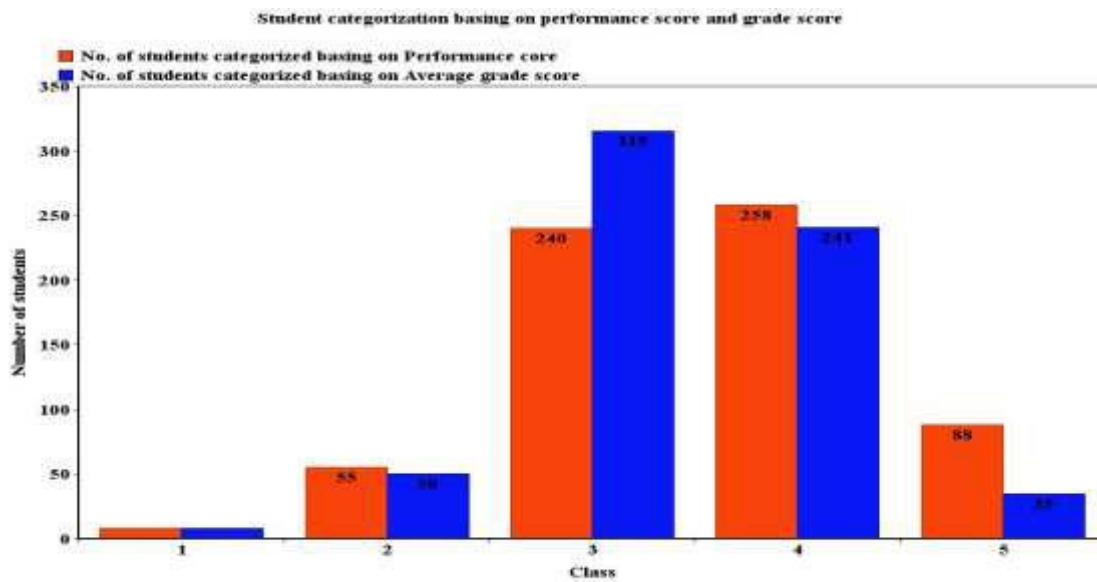


Fig 3: Categorizing the students basing on score and grade

5. CONCLUSION:

Learning Analytics is the leading area nowadays in which we can apply various machine learning algorithms to explore the new trends in education domain to shape out the new education policies. In our work we considered the student performance data and understand that the attributes are not just quoting some data values but they are very important in assessing the student performance. With the same idea, analysis was done so as to identify which factors are mostly influencing the performance measure. Firstly it was identified that with the mere grade information assessment could be done up to some extent. Later the average grade is computed to which the data is correlated and found out the factors including grades as well the parents education, higher education aspiration, study time positively influences the performance of a student. Besides factors like alcohol consumption and others shall negatively impact the performance. Assigning weights the score is computed after which the data is classified into five categories. It is a thoughtful observation with which we concluded that the classification is more appropriate when factors in addition to the grade are also included in evaluating the performance.

REFERENCES:

1. F. Sciarrone, "Machine Learning and Learning Analytics: Integrating Data with Learning," 2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET), 2018, pp. 1-5, doi: 10.1109/ITHET.2018.8424780.
2. Oyerinde, Yinka & P., Chia. (2017). Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression. International Journal of Computer Applications. 157. 37-44. 10.5120/ijca2017912671.
3. Wong, B.T.M. (2017), "Learning analytics in higher education: an analysis of case studies", *Asian Association of Open Universities Journal*, Vol. 12 No. 1, pp. 21-40. <https://doi.org/10.1108/AAOUJ-01-2017-0009>
4. V. Belsini Gladshiya , Dr. K. Sharmila, 2019, A Review Study on Predictive Analytical Tools and Techniques in Education, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 11 (November 2019)
5. Namoun, Abdallah & Alshanqiti, Abdulllah. (2020). Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review. Applied Sciences. 11. 1-28. 10.3390/app11010237.
6. G. Al-Tameemi, J. Xue, S. Ajit, T. Kanakis and I. Hadi, "Predictive Learning Analytics in Higher Education: Factors, Methods and Challenges," 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2020, pp. 1-9, doi: 10.1109/ICACCE49060.2020.9154946.

7. Papamitsiou, Zacharoula & Giannakos, Michail & Ochoa, Xavier. (2020). From childhood to maturity: Are we there yet? Mapping the intellectual progress in learning analytics during the past decade. 10.1145/3375462.3375519.
8. Sousa EBGd, Alexandre B, Ferreira Mello R, Pontual Falcão T, Vesin B and Gašević D (2021) Applications of Learning Analytics in High Schools: A Systematic Literature Review. *Front. Artif. Intell.* 4:737891. doi: 10.3389/frai.2021.737891
9. C. Jalota and R. Agrawal, "Analysis of Educational Data Mining using Classification," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 243-247, doi: 10.1109/COMITCon.2019.8862214.
10. Guzmán-Valenzuela C, Gómez-González C, Rojas-Murphy Tagle A, Lorca-Vyhmeister A. Learning analytics in higher education: a preponderance of analytics but very little learning?. *International Journal of Educational Technology in Higher Education*. 2021;18(1):23. doi:10.1186/s41239-021-00258-x
11. Su Y-S and Lai C-F (2021) Applying Educational Data Mining to Explore Viewing Behaviors and Performance With Flipped Classrooms on the Social Media Platform Facebook. *Front. Psychol.* 12:653018. doi: 10.3389/fpsyg.2021.653018
12. Lee, LK., Cheung, S.K.S. & Kwok, LF. Learning analytics: current trends and innovative practices. *J. Comput. Educ.* 7, 1–6 (2020). <https://doi.org/10.1007/s40692-020-00155-8>
13. <https://archive.ics.uci.edu/ml/datasets/Student+Performance>
14. Wong, A., & Chong, S. (2018). Modelling adult learners' online engagement behaviour: Proxy measures and its application. *Journal of Computers in Education*, 5(4), 463–479.
15. Hui, Y. K., & Kwok, L. F. (2019). A review on learning analytics. *International Journal of Innovation and Learning*, 25(2), 197–222.
16. Hwang, G. J., Hung, P. H., Chen, N. S., & Liu, G. Z. (2014). Mindtool-assisted in-field learning (MAIL): An advanced ubiquitous learning project in Taiwan. *Educational Technology & Society*, 17(2), 4–16.
17. Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304–317. <https://doi.org/10.1504/IJTEL.2012.051816>
18. Ferguson, R. & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In Proceedings of the seventh international learning analytics & knowledge conference (pp. 56–65).
19. Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—a literature review. *Learning analytics: Fundaments, applications, and trends* (pp. 1–23). Springer, Cham.
20. Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110.