STUDENTS PERFORMANCE PREDICTION USING MACHINE LEARNING

Aditya Shelar
SPPU
MMCOE, Pune, Maharashtra
adityashelar.it@mmcoe.edu.in

Mohit Patil
SPPU
MMCOE, Pune, Maharashtra
mohitpatil.it@mmcoe.edu.in

Sushmita Gatne
SPPU
MMCOE, Pune, Maharashtra
sushmitagatne.it@mmcoe.edu.in

Siddhi Bhendkahle
SPPU
MMCOE, Pune, Maharashtra
siddhibendkhale.it@mmcoe.edu.in

Prof. Jitendra Chavan
SPPU
MMCOE, Pune, Maharashtra
jitendrachavan@mmcoe.edu.in

ABSTRACT: Precisely foreseeing students' future execution dependent on their continuous scholarly records is essential for viably doing important instructive sessions to guarantee students' on-schedule and acceptable graduation. In spite of the fact that there is a rich writing on anticipating student execution when taking care of issues or contemplating for courses utilizing information driven methodologies, foreseeing student execution in finishing degrees (for example school programs) is considerably less examined and faces new difficulties: (1) Students contrast massively as far as foundations and chose courses; (2) Courses are not similarly useful for making exact expectations; (3) Students' advancement should be consolidated into the forecast. In this paper, we build up a novel ML technique for foreseeing student execution in degree programs that can address these key difficulties. The proposed technique has two noteworthy highlights. Initial, a study of students marks variations in subjects is understood, by using ML techniques. Second, information driven methodology dependent on survey for students study habits and daily routine is considered to affect the performance.

Index Terms—Student performance prediction, personalized education, impact factor, gradient.

I. INTRODUCTION

Anticipating Students Performance is a fundamental Part. The evaluation is essential to keep up student Performance dependent on course structure, Assessment marks, Final test score and furthermore Extracurricular activities. Colleges today are working in complex and highly competitive conditions. The principle challenge for colleges is to gauge their performance and to build the system for further advancement and future activity. To make school moderate, it is along these lines significant to guarantee that a lot more students graduate on time through early mediations and students whose execution will be probably not going to meet the graduation criteria of the degree program on schedule .A basic advancement towards Effective Intervention is to make a framework that can monitor students' scholastic Performance throughout the course and precisely foresee their future execution, for example, when they are probably going to graduate and their gauge GPAs, given the present advancement. In spite of the fact that

foreseeing students' performance has been widely examined with regards to taking care of issues.

With the wide utilization of PCs and web, there has as of late been an enormous increment in freely accessible information that can be broke down. Be it online deals data, site traffic, or student habits, information is produced each day. Such a lot of information presents both an issue and a chance. The issue is that it is troublesome for people to break down such expansive information. The open door is that this sort of information is perfect for PCs to process, since it is put away carefully in an all around arranged way, and PCs can process information a lot quicker than people. Current frameworks utilize just scholarly information to examine the outcome. The scholastic outcomes depend on scholarly execution as well as student behaviour that influences the execution. Henceforth for better expectation or outcome both must be considered.

The difficulties looked by existing frameworks are for the exact expectation the reliance between subjects must be distinguished, students developing advancement should be consolidated into the forecast. Subsequently to defeat every one of the disadvantages of current framework, new framework is being created and is characterized as 'Students Performance Prediction dependent on academic history and personal behaviour to improve results'.

II.   LITERATURE SURVEY

In this paper, we build up a novel ML strategy for anticipating student execution in degree programs that can address these key difficulties. The proposed strategy has two noteworthy highlights. Initial, a study of students marks variations in subjects is understood, by using ML techniques, to derive the variation factor. Second, information driven methodology dependent on survey for students study habits and daily routine is considered to affect the performance. Through broad recreations on an undergraduate student dataset gathered more than three years at UCLA. [1]

The mean square blunder of expectation is determined for an exponentially weighted moving normal (e.w.m.a.), when the arrangement anticipated is a Markov arrangement, or a Markov arrangement with superimposed mistake. The best decision of damping consistent is given; the decision isn't basic. There is an estimation of the Markov relationship po underneath which it is difficult to anticipate, with an e.w.m.a., the neighborhood varieties of the arrangement. The mean square mistake of an e.w.m.a. is contrasted and the base conceivable esteem, to be specific that for the best direct indicator (Wiener). An adjusted e.w.m.a. is developed having a mean square mistake moving toward that of the Wiener indicator. This modification will be of esteem if the Markov connection parameter is negative, and perhaps at the same time when the Markov parameter is ne [2]

Diverse strategies and procedures of information mining were thought about amid the expectation of students' prosperity, applying the information gathered from the reviews directed amid the late spring semester at the University of Tuzla, the Faculty of Economics, scholastic year 2010-2011, among first year students and the information taken amid the enlistment. The achievement was assessed with the passing evaluation at the test. The effect of students' socio-statistic factors, accomplished outcomes from secondary school and from the selection test, and frames of mind towards contemplating which can have an effect on progress, were altogether explored. In future examinations, with distinguishing and evaluating factors related with procedure of considering, and with the example increment, it is conceivable to deliver a model which would remain as an establishment for the advancement of choice emotionally supportive network in advanced education. [3]

Paper investigates the utility of bunching in lessening mistake in different expectation errands. Past work has alluded to the improvement in forecast exactness credited to bunching calculations whenever used to pre-process the information. In this work we all the more profoundly explore the immediate utility of utilizing bunching to improve expectation exactness and give clarifications to why this might be so. We take a gander at various datasets, run k-implies at various scales and for

each scale we train indicators. This produces k sets of expectations. These forecasts are then consolidated by a gullible group. We saw that this utilization of an indicator related to grouping improved the expectation exactness in many datasets. We trust this demonstrates the prescient utility of abusing structure in the information and the information pressure gave over by grouping. We additionally discovered that utilizing this technique enhances the expectation of even a Random Forests indicator which recommends this strategy is giving a novel and valuable wellspring of change in the forecast procedure. [4]

As of late Educational Data Mining (EDM) has risen as another field of research because of the advancement of a few measurable ways to deal with investigate information in instructive setting. One such use of EDM is early forecast of student results. This is fundamental in advanced education for distinguishing the "feeble" students with the goal that some type of remediation might be sorted out for them. In this paper a lot of properties are first characterized for a gathering of students studying Computer Science in some undergrad schools in Kolkata. Since the quantities of characteristics are sensibly high, highlight determination calculations are connected on the informational index to diminish the quantity of highlights. Five classes of Machine Learning Algorithm (MLA) are then connected on this informational collection and it was discovered that the best outcomes were acquired with the choice tree class of calculations. It was likewise discovered that the forecast outcomes got with this model are tantamount with other recently created models. [5]

Lately Educational Data Mining (EDM) has risen as another field of research because of the advancement of a few measurable ways to deal with investigate information in instructive setting. One such use of EDM is early forecast of student results. This is vital in advanced education for distinguishing the "frail" students with the goal that some type of remediation might be sorted out for them. Since the quantities of qualities are sensibly high, highlight determination calculations are connected on the informational collection to lessen the quantity of highlights. Five classes

of Machine Learning Algorithm (MLA) are then connected on this informational index and it was discovered that the best outcomes were acquired with the choice tree class of calculations. It was additionally discovered that the forecast outcomes acquired with this model are equivalent with other recently created models. [6]

In this period of computerization, training has additionally patched up itself and isn't restricted to old address technique. The customary journey is on to discover better approaches to make it increasingly successful and effective for students. These days, bunches of information is gathered in instructive databases, yet it remains unutilized. So as to get required advantages from such a major information, incredible assets are required. Information mining is a rising integral asset for examination and expectation. It is effectively connected in the zone of misrepresentation location, promoting, advertising, advance evaluation and forecast. Be that as it may, it is in beginning stage in the field of instruction. Significant measure of work is done toward this path, yet at the same time there are numerous immaculate territories. Also, there is no brought together methodology among these looks into. [7]

The student's execution forecast is a vital research theme since it can enable educators to keep students from dropping out before last tests of the year and recognize students that need extra help. The target of this investigation is to anticipate the troubles that students will experience in an ensuing computerized structure course session. We broke down the information logged by an innovation upgraded learning (TEL) framework called advanced gadgets instruction and structure suite (DEEDS) utilizing AI calculations. The AI calculations incorporated a counterfeit neural systems (ANNs), bolster vector machines (SVMs), strategic relapse, Naïve bayes classifiers and choice trees. The DEEDS framework enables students to settle computerized configuration practices with various dimensions of trouble while logging input information. The info factors of the present examination were normal time, all out number of exercises, normal inert time, normal

number of keystrokes and absolute related action for each activity amid individual sessions in the advanced structure course; the yield factors were the student(s) grades for every session. We at that point prepared AI calculations on the information from the past session and tried the calculations on the information from the up and coming session. We performed k-crease cross-approval and registered the recipient working trademark and root mean square blunder measurements to assess the models' exhibitions. The outcomes demonstrate that ANNs and SVMs accomplish higher exactness than do different calculations. ANNs and SVMs can without much of a stretch be incorporated into the TEL framework; in this manner, we would anticipate that educators should report improved student's execution amid the consequent session. [8]

In this paper, we tended to the evaluation challenge in the ASSISTment framework, which is an online coaching framework that fills in as an e-learning and e-appraisal condition. We - 19 - concentrated on the appraisal capacity of the framework and assessed it by mining our log information and contrasting and state sanctioned test outcomes. Some proof was displayed that the online evaluation framework completed a superior employment of anticipating student learning by having the capacity to mull over what amount mentoring help was required, how quick an student tackles an issue and what number of endeavours were expected to complete an issue. [9]

To build adequacy in conventional homeroom courses just as in Massive Open Online Courses (MOOCs), computerized frameworks supporting the educator are required. One vital issue is to consequently distinguish students that will do inadequately in a course sufficiently early to have the capacity to take therapeutic activities. This paper proposes a calculation that predicts the last grade of every student in a class. It issues an expectation for every student exclusively, when the normal exactness of the forecast is adequate. The calculation realizes online what is the ideal forecast and time to issue an expectation dependent on previous history of students'

execution in a course. We infer exhibit the execution of our calculation on a dataset acquired dependent on the execution of around 700 college students who have taken a starting computerized flag preparing in the course of recent years. Utilizing information got from a pilot course, our philosophy recommends that it is powerful to perform ahead of schedule in-class evaluations, for example, tests, which result in convenient execution forecast for every student, along these lines empowering auspicious intercessions by the teacher (at the student or class level) when vital. [10]

III.   PROPOSED SYSTEM

In this paper, an algorithm for predicting marks of current semester based on previous semester record as well as previous 2 years record for same course and behavioural analysis is proposed. As per performance of student in written exam of all previous semesters, gradient of all semesters is calculated. This is then saved in database for further evaluation. Some fixed no. of questionnaire is answered by students for behavioural analysis of student. Average of this questions leads to calculation of impact factor.

The impact factor calculated for the questionnaire is used for calculating impact from average values of obtained marks. Previous 2 years data of students studying in same class is obtained from database. An exponential moving average and linear regression is calculated and fed to marks prediction.

Gradient, impact factor, moving average and regression together predicts a result of marks for current semester.
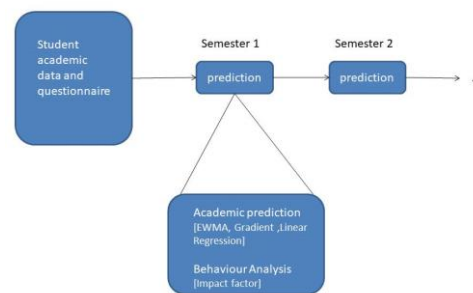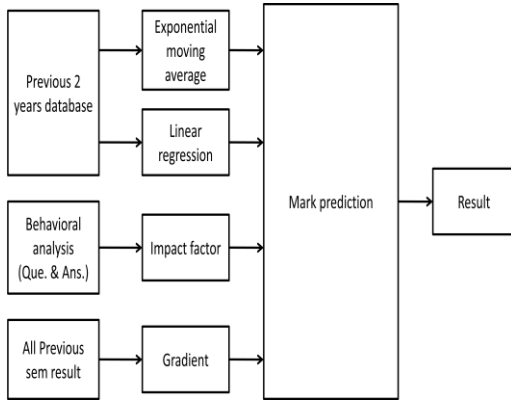

Fig. Overall system architecture

Figure proposed algorithm

Moving averages smooth the data to form a trend following indicator. They do not predict direction, but rather define the current direction. The exponential moving average is a type of moving average that gives more weight to recent values in an attempt to make it more responsive to new information.

$$EMA = (P * \alpha) + \big(previous\ EMA * (1 - \alpha)\big)$$

.... (1)

P=current price

α=smoothing factor=$\frac{2}{1+N}$

N=no. of time periods

Linear regression endeavours to demonstrate the connection between two factors by fitting a direct condition to watched information. One variable is viewed as an informative variable, and the other is viewed as a dependant variable. A linear regression line is given as

$$Y = a + bX \qquad \text{.... (2)}$$

Where X = explanatory variable

Y = dependent variable.

b= slope of the line is b

a=the intercept (the value of y when x = 0).

Gradient in python is given by numpy library to get an array with the numerical derivative for every variable. The gradient is defined as

$$\frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x} \qquad \text{.... (3)}$$

The impact factor is calculated by considering the weightage given to the answers of the question. Each answer of the question is given

some predefined value which is considered for calculation.
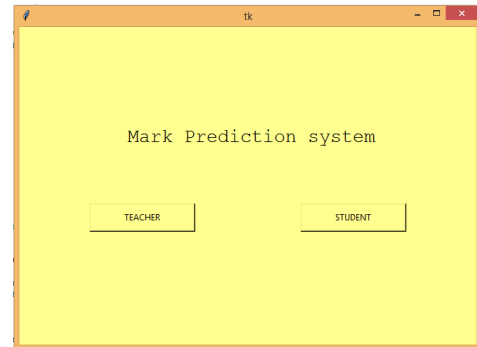
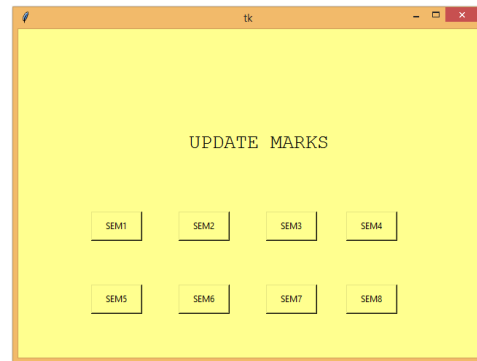## IV. RESULT

Fig GUI for marks prediction system

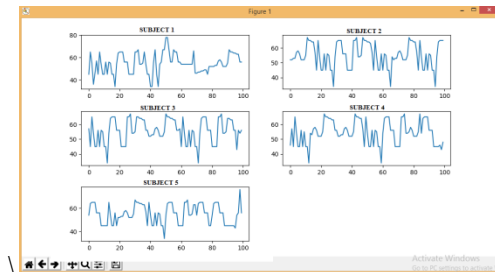Fig teacher should update marks here

Fig analysis of performance of student

The proposed system shows the result of next semester of every student. As in the figure above it give graphs to visually analyse the student performance to get better solution.

V. CONCLUSION

Present examination demonstrates that scholarly execution of the students is essentially subject to their past exhibitions. Past performance without a doubt got a huge impact over students' performance. ML has overcome much from its previous stages, and

can end up being an amazing asset in Academic In this we propose new strategy by utilizing impact factor based algorithm for anticipating Students Performance utilizing their current just as past academic records and taking a contribution of the general behaviour as well as the study habits to improve the results by considering the real time scenario. This can help to analyse the results in much effective way that actually contributes to improvement of students performance.

REFERENCE

[1]   J. Xu is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, USA." A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs" IEEE, and Mihaela van der Schaar, Fellow, IEEE 2017.

[2]   Prediction by Exponentially Weighted Moving Averages and Related D. R. Cox Journal of the Royal Statistical Society. SeriesB(methodological)   Vol. 23, No. 2(1961), pp. 414-422 , 2017.

[3]   Data Mining Approach for Predicting Student Performance , ResearchGate,EdinOsmanbegovic' &MirzaSuljic , May 2012.

[4]   The Utility of Clustering in Prediction Tasks, ShubhenduTrivedi, Zachary A. Pardos and Neil T. Haffernan, e Department of Education IES Math Centre   for Mathematics and Cognition grant. Report Date: 05 September 2011.

[5]   Student Performance Prediction using Machine Learning, HavanAgarwal ,

HarshilMavaniDept of Information Technology K.J.Somaiya College of Engineering,Mumbai, IJERT -2015.

[6]   Early Prediction of Students Performance using Machine Learning   Techniques, Anal Acharya,Department of Computer Science,St Xavier's College, Kolkata,India.   ,   DevadattaSinha, Department of Computer Science,and Engineering University of Calcutta, Kolkata, India. International Journal of Computer Applications , 2014.

[7]   Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue, PoojaThakar Assistant Professor VIPS, GGSIPU Delhi, India, Anil Mehta, Ph.D Professor University of RajasthanJaipur,India,   International Journal of Computer Applications, 2015

[8]   8. Using Machine Learning to Predict Student Performance, Murat Pojon, University of Tampere Faculty of Natural Sciences Software DevelopmentM. Sc. Thesis Supervisor: JormaLaurikkala June 2017.

[9]   Feng, M., Heffernan, N.T., &Koedinger, K.R. (2010). Addressing the assessment challenge in an Online System that tutors as it assesses. User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI journal). Volume 19, p.243-26.

[10]   Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Personalized grade prediction: A data mining approach," in Data Mining (ICDM), 2015 IEEE International Conference on. IEEE, 2015, pp. 907–912.