Student Grade Analysis and Prediction Using Machine learning Algorithms

[1]Syed Abubakar S, [2]Raghavendra R

[1]Student, [2]Assistant Professor
School of CS & IT
Jain (Deemed-to-be) University

***Abstract*: The possibility to predict the basic performance of the students is very necessary to enhance their training skills. It has become a beneficial statistics that may be used for quite a number purposes. This paper encloses the advantage of documents mining techniques to predict the final grades of students especially based on their necessary total data. In the practical studies, three data mining techniques had been working on two academic datasets associated to arithmetic lesson and Portuguese language lesson. The outcomes validated the success of data mining getting to be aware of the techniques all through the prediction of student's performances.**

Performing the prediction of scholar's performance grew to become an important wish in many of the schools and institutes. The needed command to assist the risk of the students and provide guarantee to their maintenance, as long as the first class tutoring assets and knowledge , and enhancing the institute's rating and popularity. But, which will be hard to execute for beginning to medium institute, in particular those who are limited in graduation and publish graduating programs, and they may have less number of students' data for processing. So, the most purpose of the estimate is to point out the chance of coaching and modelling the dataset of small dimension and achievable of developing a template of prediction with reliable perfection outcome rate. The lookup search out of the key indications within the less number of dataset, it can be utilized in developing the forecast model, the usage of schematic and grouping algorithms. The first class symptoms had been needed into a numerous algorithms to think about the foremost correct model. With the chosen algorithms, the outcomes showed the capacity of grouping algorithm in deciding key symptoms in less number of datasets.

***Keywords*: prediction, student's performance, machine learning, artificial neural networks and deep learning.**

I. INTRODUCTION

Student's usual overall presentation is an necessary area in high level universities. This one is due to the reality that one of the standards for a greater rank organization is specifically primarily on the basis of fantastic achievements of academic records. Students overall performance can be bought with the aid of measuring the teaching evaluation. Yet, many researchers bring up about commencement judged before the student's victory. Recently, on line tutoring have a huge up going mark, and student's computerized information has gone up with the large information size. This tends to obtain the insurance policies and predictions about the college students with the aid of way of processing tutorial information with records mining techniques. Generally, most tutoring establishments makes use of the final grades to think about the student's achievement. These final grades are primarily based on the sequence procedure, evaluati006Fn marks, ultimate examination ranking and additionally extracurricular matters to do. The evaluation is necessary to preserve student's presentation and the fee of tutoring. Having reading students' presentation, an approach may be good think about for the duration of the time of learning in the organization. Now a days, many strategies suggest to reflect on consideration on university student's performance. Data mining is the widely used approach to examine student's overall performance.

II. Literature Survey

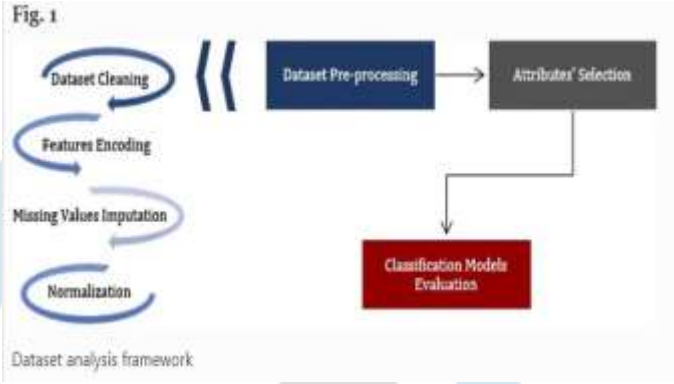
Much lookup has been administered within the location of educational data processing the place a projecting framework is made to forecast the general performance of school students to become conscious of hazard students. The hassle are often regarded the challenging trouble due to the fact the overall performance relies upon on many traits associated to school scholars. The student's GPA is that the widely significant attribute want to perform the prediction of the performance. The student's GPA can be symbolize the actual cost for the longer term instructional, profession probabilities and development. Additionally, tutorial potentials are often corrected by means of the scholar's GPA. The social economic facts that amount of the household additionally viewed as a necessary attribute. The lookup produces two new attributes focusing that on the utilization of graphic elements associated to the online and social economic community utilization and the impact on the overall student's performance. On various hands, many computer gaining knowledge of and records mining strategies have used to perform the prediction on students' overall performance.

The thinking attracts the procedure of ANN is employed to perform the prediction for the educational outcome of scholars. Results of the fashions have been measured with the use of the Pearson's correlation of coefficient .The result from the getting to know about the confirmed topics have sturdy have an have an effect on in the CGPA.

This commonly make use of sorting techniques of two types for an everyday to perform an prediction to the student's achievements above an enormous data. Learning Analytics affords in most cases on the statistics sequence and data pre-processing section the area the required information for setting up.

III. REASEARCH METHODOLOGY

To achieve the project's level, computable reproduction search for techniques has undertaken the counselled in the plan phases demonstrated in Fig. 1. In this step the dataset will be geared up for surpassed by using capacity of visualization and clustering techniques, to get the effect of correspond measure. Then, the measures signs and symptoms will be utilized in distinct classifying algorithms and the fit model is used for analyzing the outcome of the Student's performance in thesis duties and every publications grades. In middle, and until now than the classifying models' evaluation step, the datasets will go by a pre-processing stage to make it geared up for the contrast phase.



STUDENT DATA

In this process, the dataset have been utilized to predict the overall student performances. The dataset from the secondary school of two Portuguese schools. In the dataset process, two grading structures are used. The dataset consists of the attributes to analyze the scholar records that can predict the repute of the individual student how nicely the student can operate in the examine process. The place zero is the lowest grade and 20 is the perfect score. During the faculty year, college students are evaluated in three durations and the remaining assessment corresponds to the last grade.

DATA PRE-PROCESSING

On the process of pre-processing stage, some aspects have been discarded because less number of partial value. For example, few answered the earning in the home, even as nearly a hundred percent of the students remain with their parents and own a laptop computer in house. The ultimate attributes are demonstrated the place those last four line designate the variables get hold from the records.

Attribute	Description (Domain)
sex	sex of student's (female or male)
age	age of student's (from 15 to 22)
school	school of student's (Gabriel Pereira or Mousinho da Silveira)
address	address type of home (urban or rural)
Pstatus	cohabitation status of parent's (together living or apart)
Medu	mother's education (numeric: from 0 to 4a)
Mjob	mother's job (nominalb)

Fedu	father's education (from 0 to 4a)
Fjob	father's job (nominalb)
guardian	guardian of student's (mother, father or other)
famsize	student's family size (_ 3 or > 3)
famrel	quality of family relationships (from 1 – very bad to 5 – excellent)
reason	Main reason to choose this school (close to home, school reputation, course preference or other)
Travel time	Time from home to school (1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour).
study time	weekly study time (1 – < 2 hours, 2 – 2 to 5 hours, 3 – 5 to 10 hours or 4 – > 10 hours)
failures	number of pervious class fail (n if 1 _ n < 3, else 4)
schoolsup	extra support for school education (yes or no)
famsup	family support for education (yes or no)
activities	extra-curricular activities (yes or no)
paid class	extra paid classes (binary: yes or no)
internet	Having access to Internet at home (yes or no)

nursery	attended nursery school (yes or no)
higher	wants to pursue higher education (yes or no)
romantic	Maintaining any romantic relationship (yes or no)
free time	after school (from 1 – very low to 5 – very high)
gout	going out with Dudes (from 1 – very low to 5 – very high)
Walc	weekend having alcohol (from 1 – very low to 5 – very high)
Dalc	workday having alcohol (from 1 – very low to 5 – very high)
health	current health status (from 1 – very bad to 5 – very good)
absences	Total number of absents in school (from 0 to 93)
G1	first term grade (from 0 to 20)
G2	second term grade (from 0 to 20)
G3	final grade (numeric: from 0 to 20)

a 0 – none, 1 – primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education. b teacher, health care related, civil services (e.g. administrative or police), at home or other.

IV. RESULT

The demographic traits of student and the grade in some duties can construct a properly coaching for a supervised studying algorithm. Adding different traits namely the collective score factor mainly of the scholar's, the scores received at different guides to the scores of numerous assessment, construct correct types.

Mean Squared (RMSE) is a famous metric. An excessive PCC (i.e. close to 100%) suggests a good classifier, whilst a regressor ought to current a low. This metrics can be computed the use of the equations:

Where byi denotes the envisioned price for the i-th example.
G3 will be modeled the use of three supervised approaches:

1. Binary classification – ignore if G3_10, else fail;
2. 5-Level classification – primarily placed on Erasmus1 score transformation order;

3. Regression – the G3 fee output between 0 and 20.

There are three inputs for these approaches, (e.g. there might also be the school assigned grades or may be not) and the four data modelling algorithms will be examined. Moreover, an explanatory assessment will be carried out over the good models, in order to know the most applicable features.

Conclusion: As we see both Model Mean Absolute Error & Model Mean Squared Error that the linear regression is performing the best in both cases.

Conclusion

The overall performance of students benefits to predict the lectures and the learners getting to comprehend the enhancing ability for learning and instructing manner if needed. The paper has been forwarded in the formerly lookup on predicting students overall performance with a vary of analytical methods. The paper factors the utility of knowledge discovery of data (KDD) to perform a prediction the final grades of college students especially on the usage of the data. Three regularly used classification methods (RF, NB and DT) had been used in the process. In conclusion, evaluation on predicting student's performance has entreated to in a similar model search for further researches in our environment. In the future, special function determination methods can be used.

ACKNOWLEDGEMENT

I desire to exhibit my gratitude and dedication to Dr M.N Nachappa and Asst. prof. Raghavendra R for all the advice, Encouragement and consistent help that has given throughout the assignment work. Once again I would heart fully like to thank Asst. prof. Raghavendra R for the treasured recommendation and guidance that has made the assignment achievable.

REFERENCES

[1]Navamani, J.; Kannammal, A. Predicting overall performance of faculties with the aid of making use of facts mining strategies on public examination outcomes.

[2] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, A reference figure for mastering analytics,"

[3] Z. Ibrahim, D. Rusli, Predict college scholar's educational presentation: evaluating ANN, DT and LR in: twenty first Annual SAS Malaysia Forum, fifth September, 2007.

[4] Pauziah Mohd Arsad, Norlida Buniyamin and JamalullailAb Manan. "ANN Scholar's Presentation Prediction type (NNSPPM)"

[5] Midhun Mohan and Dr. Kumari Roshni V S "Massive data point of view to Classifying and indicator of Scholar's outcome by the use of MapReduce"

[6] Madhav S. Vyas and Reshma Gulwani. "Forecasting of Student's Presentation by the use of CART point of view in Data engineering"

[7] Alana M. de Morais, Joseana M. F. R. Araújo and Evandro B. Costa. "Observing Scholar Presentation by Using Data grouping and Predictive description"