ABSTRACT

In education system evaluation and prediction of student performance is a challenging task. In this paper, a model is proposed to predict the performance of students in an academic organization. The algorithm employed is a machine learning technique called Naïve Bayes and KNN. Further, the importance of several different attributes, or "features" is considered, in order to determine which of these are correlated with student performance. Finally, the results of an experiment follow, showcasing the power of machine learning in such an application. In perspective of this project we are going to predict the student development and examine the greater result through machine learning algorithm. We foresee the student performance by scanning their previous academic details. To execute this prediction we have created a dataset, by using this we can predict student details.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ML	Machine Learning
NB	Naive Bayes
KNN	Kth nearest neighbour

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

There are many studies in the learning field that investigated the ways of applying machine learning techniques for various educational purposes. One of the focuses of these studies is to identify high-risk students, as well as to identify features which affect the performance of students. Students are the major strength for numerous universities. Universities and students play a significant part in producing graduates of superior calibers with its academic performance accomplishment. However, academic performance achievement changes as various sort of students may have diverse degree of performance achievement. Machine learning is the ability of a system to consequently gain from past experience and improve performance. Nowadays machine learning for education gains more attention. Machine learning is used for analyzing information based on past experience and predicting future performance.

1.2 MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure information of knowledge of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process approaches.

In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning

algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular vary. thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

1.3 OBJECTIVE

The main objective this paper introduces a ML model that classify and predict student performance by utilizing supervised ML algorithms like Naïve Bayes and K-Nearest Neighbour. Thus, the proposed approach offers a solution to predict performance efficiently and accurately by comparing several ML model.

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORK

Abdallah Namoun et.al[1] proposed accomplishing learning results is estimated principally by the exhibition of the grades (f i.e., level) and the accomplishment scores (i.e., grades). Audit and investigation of AI machines were regularly used to exhibit understudy execution. At long last, online understudy exercises, evaluation scores, and understudy feelings were the main indicators of accomplishment. Distinguish the attributes and strategies that decide the exhibition of understudies utilizing the PICO strategy. This evaluation has had many difficulties, as it is by and large wide, not zeroed in on the utilization of understudy input as a benchmark for understudy execution, experienced quality issues, and has not been distributed in the most as often as possible posed inquiries.

Durgesh Ugale et.al[2] discussed the handling step will be applied to the crude informational index to appropriately utilize the extraction calculation. The "execution" of guiding an understudy can assist with working on their presentation.

The answer for the issue is very distinct. He didn't remark on the understudies' reactions. You didn't talk about the cutoff. You didn't analyze the idea of the prescience.

Ali Salah Hashim et.al[3] In this review, we analyzed the exhibition of machine-controlled calculations. As per the consequences of the overview, the coordination's grouping was the most right in picking the last score of the understudy (68.7% of champs and 88.8% of washouts). Depict the strategies utilized in insightful exploration to anticipate understudy execution. You have not appraised the nature of the preparation. You have not thought about the models.

Alnassar, Fatema et.al[4] talks about the connection between understudy association law calculation, K-key calculation, and choice tree. This review looks at understudy execution dependent on various attributes. The design incorporates questions and replies to illustrations, little and last experimental outcomes, schoolwork, and lab work.

We talk about dynamic trees, data mining techniques, and a blend of strategies that empower understudies to anticipate understudy execution, and educators can make significant strides in creating understudy information. execution. The presentation of different tree calculations can be investigated dependent on its exactness and timing of tree conveyance. Speculations eliminated from the framework assisted the educator with distinguishing understudies with incapacities and further develop execution. Knowing the quantity of gatherings ahead of time. It is hard to know the quantity of gatherings when there is a slight change in the information.

Saifuzzaman et.al[5] discussed outcomes show that calculations and different techniques are utilized to acquire data from instruction. A significant number of these calculations are significant for characterizing and controlling data. In this review, the calculations C5.0, C4.5, and K-esteem were distinguished. At 48% of the locales, these three uncovering calculations are broadly utilized for information examination, particularly in instruction. He wrote about the general accomplishment of the understudies. The gauge study is featured with practically no huge clarification.

Leena H. Alamri et.al[6] proposed consequences of the SVM and RF calculations utilized in the two segments show that the honesty of the two parts is up to 93%, while the least RMSE is 1.13 for the RF. The request in which the examination report is set up is simple and quick to foresee. It functions admirably in numerous classes of speculations. Since SVMs are known as terrible number crunchers, yield as expectations ought not be viewed as significant.

C. Verma et.al[7] discussed the outcomes show that understudy interest and GPA in the principal semester drove the whole choice interaction, and the Bayesian organization is better than the choice tree because of its high precision. It works better as far as pay level than the quantity of factors. In the event that you attempt a little example, the information might be excessively high or excessively high.

Ferda Ünal[8] proposed this paper exhibits the utilization of information recovery innovation to distinguish eventual outcomes dependent on understudy history. In the exploration, three techniques for information mining (choice tree, non-repeating memory, and naive Bayes) were utilized in two lines of information science and Portuguese. The outcomes show that mining strategies are valuable in choosing understudy execution. Build up clear guidelines for anticipating preparing data. With the progression of innovation, e-learning as a web-based learning webpage and progressed mixed media innovation, preparing costs have been decreased and time and difficulties have been eliminated.

Perfomance prediction of students using distributed data mining

The performance of students in higher education in India is a turning point in the academics for all students for their brightest career. In today's generation the amount of data stored in educational database increasing at a great rate. These databases contain secret information for improvement of students' performance; these data can be located at different nodes in distributed system. Classification and prediction are among the major techniques in Data mining and widely used in various fields. In this paper classification techniques are used for prediction of student performance in distributed environment. Data mining methods are often implemented at many advance universities today for analyzing available data and extracting information

and knowledge to support decisionmaking. While it is important to have models at local level, their results makes it difficult to extract knowledge that can be useful at the global level. Therefore, to support decision making at this area, it is important to generalize the information contained in those models, specific classifier method can be used to generalize these rules for global model.

Predicting Student Academic Performance

Engineering schools worldwide have a relatively high attrition rate. Typically, about 35% of the first-year students in various engineering programs do not make it to the second year. Of the remaining students, quite often they drop out or fail in their second or third year of studies. The purpose of this investigation is to identify the factors that serve as good indicators of whether a student will drop out or fail the program. In order to establish early warning indicators, principal component analysis is used to analyze, in the first instance, first-year engineering student academic records. These performance predictors, if identified, can then be used effectively to formulate corrective action plans to improve the attrition rate.

Data Mining Approach For Predicting Student Performance

This work proposes a novel approach - personalized forecasting - to take into account the sequential effect in predicting student performance (PSP). Instead of using all historical data as other methods in PSP, the proposed methods only use the information of the individual students for forecasting his/her own performance. Moreover, these methods also encode the "student effect" (e.g. how good/clever a student is, in performing the tasks) and "task effect" (e.g. how difficult/easy the task is) into the models. Experimental results show that the proposed methods perform nicely and much faster than the other state-of-the-art methods in PSP.

A novel approach for upgrading Indian education by using data mining techniques

Education is the backbone of all developing countries. Upgrading of the education system, upgrades the country to the world top ranking level. One of the major problems that the education system facing is predicting the behaviour of students from large database. This paper focus on upgrading Indian education system by using one of the techniques in Data mining namely clustering. Cluster analysis solves the given data into some meaningful groups. Normally the performances of the students can be classified into different patterns as normal, average and below average. In this paper we attempt to analyze student's data in different angle beyond the above indicated patterns through newly proposed UCAM (Unique clustering with Affinity Measures) clustering algorithm.

A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration

Educational Data Mining (EDM) is an interdisciplinary ingenuous research area that handles the development of methods to explore data arising in a scholastic fields. Computational approaches used by EDM is to examine scholastic data in order to study educational questions. As a result, it provides intrinsic knowledge of teaching and learning process for effective education planning. This paper conducts a comprehensive study on the recent and relevant studies put through in this field to date. The study focuses on methods of analysing educational data to develop models for improving academic performances and improving institutional effectiveness. This paper accumulates and relegates literature, identifies consequential work and mediates it to computing educators and professional bodies. We identify research that gives well-fortified advise to amend edifying and invigorate the more impuissant segment students in the institution. The results of these studies give insight into techniques for ameliorating pedagogical process, presaging student performance, compare the precision of data mining algorithms, and demonstrate the maturity of open source implements.

Data Mining: A prediction of performer or underperformer using classification

paper is an attempt to use the data mining processes, particularly classification, to help in enhancing the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. For this purpose, the CRISP framework for data mining is used for mining student related academic data. The classification rule generation process is based on the decision tree as a classification method where the generated rules are studied and evaluated. A system that facilitates the use of the generated rules is built which allows students to predict the final grade in a course under study.

An Empirical Study of the Applications of Data Mining Techniques in Higher Education

Few years ago, the information flow in education field was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. This paper addresses the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples.

CHAPTER 3

METHODOLOGY

3.1 EXISTING SYSTEM

In this research on existing methods of prediction is still insufficient to determine the most appropriate methods for predicting student performance in institutions. Second, is the absence of inquiry of the specific courses. There are various types of