STUDENT GRADE PREDICTIONUSING MACHINE LEARNING

NAKKA YAMINIDURGA [1], K.VENKATESH [2]

1.PG student, D.N.R. COLLEGE, P.G. COURSES (AUTONOMOUS), BHIMAVARAM-534202.

Emai id:nakkayamini2000@gmail.com

2.Assistant Professor in DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS, BHIMAVARAM-534202

Email id:kornalavenkatesh@gmail.com

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.

1 INTRODUCTION

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.

1.1MACHINE 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 and that is 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.

1.2 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 Neigh bour. Thus, the proposed approach offers a solution to predict performance efficiently and accurately by comparing several ML model.

2. LITERATURE SURVEY AND 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.

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.

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.

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.

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.

3 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 classification are implemented in existing framework. The datasets are uploaded in WEKA tool with any WINDOWS OS configuration. K-means clustering algorithms provides reduce number of error rate values.

EXISTING ALGORITHM

- K-MEANS CLUSTERING ALGORITHM

4 PROPOSED WORK AND ALGORITHM

In the proposed system Machine Learning algorithm is used for the classification. An automated evaluation system has been proposed to evaluate student performance and to analyze. A prediction system has been proposed by using their marks, staff opinion, attendance and ragging. The study is evaluated using machine learning classifier. Analysis of the student behaviour has been proposed by using intellectual parameters of the student which affect their study. Various mining techniques are used to determine the educational data covering some factors. In this paper, a novel approach based on KNN significant academic attributes for performance predictions. The experiment displays good performance of the proposed algorithm and was compared to similar approaches over the same dataset. By analyzing the experimental results, it is observed that the Naïve Bayes and KNN algorithm turned out to be best classifier for student performance prediction because it contains more accuracy and least error rate.

ADVANTAGES OF PROPOSED SYSTEM

We did same steps to build cumulative predicting model and there may be some change of syntax due to different technology as in paper author used JAVA.

We have used same dataset given by you to implement algorithms given in paper and in paper author has compare cumulative predictor with Naïve Bayes and KNN.

5 METHODOLOGIES

MODULES

DECISION TREE ALGORITHM

Decision Tree algorithm belongs to the family of supervised learning algorithms. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too.

The general motive of using Decision Tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data(training data).

The understanding level of Decision Trees algorithm is so easy compared with other classification algorithms. The decision tree algorithm tries to solve the problem, by using tree representation. Each internal node of the tree corresponds to an attribute, and each leaf node corresponds to a class label.

DECISION TREE ALGORITHM PSEUDOCODE

1. Place the best attribute of the dataset at the root of the tree.
2. Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.

3. Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

STRATEGIC APPROACH TO SOFTWARE TESTING

The software engineering process can be viewed as a spiral. Initially system engineering defines the role of software and leads to software requirement analysis where the information domain, functions, behavior, performance, constraints and validation criteria for software are established. Moving inward along the spiral, we come to design and finally to coding. To develop computer software we spiral in along streamlines that decrease the level of abstraction on each turn.

A strategy for software testing may also be viewed in the context of the spiral. Unit testing begins at the vertex of the spiral and concentrates on each unit of the software as implemented in source code. Testing progress by moving outward along the spiral to integration testing, where the focus is on the design and the construction of the software architecture. Talking another turn on outward on the spiral we encounter validation testing where requirements established as part of software requirements analysis are validated against the software that has been constructed. Finally we arrive at system testing.

UNIT TESTING

Unit testing focuses verification effort on the smallest unit of software design, the module.

WHITE BOX TESTING

This type of testing ensures that

- All independent paths have been exercised at least once
- All logical decisions have been exercised on their true and false sides
- All loops are executed at their boundaries and within their operational bounds
- All internal data structures have been exercised to assure their validity.

To follow the concept of white box testing we have tested each form .we have created independently to verify that Data flow is correct, All conditions are exercised to check their validity, All loops are executed on their boundaries.

BASIC PATH TESTING

Established technique of flow graph with Cyclomatic complexity was used to derive test cases for all the functions. The main steps in deriving test cases were:

Use the design of the code and draw correspondent flow graph.

Determine the Cyclomatic complexity of resultant flow graph, using formula:

V(G)=E-N+2 or

V(G)=P+1 or

V (G) =Number Of Regions

Where V (G) is Cyclomatic complexity, E is the number of edges,

N is the number of flow graph nodes, P is the number of predicate nodes.

Determine the basis of set of linearly independent paths.

CONDITIONAL TESTING

In this part of the testing each of the conditions were tested to both true and false aspects. And all the resulting paths were tested. So that each path that may be generate on particular condition is traced to uncover any possible errors.

DATA FLOW TESTING

This type of testing selects the path of the program according to the location of definition and use of variables. This kind of testing was used only when some local variable were declared. The method was used in this type of testing. These were particularly useful in nested statements

LOOP TESTING

In this type of testing all the loops are tested to all the limits possible. The following exercise was adopted for all loops:

- All the loops were tested at their limits, just above them and just below them.

- All the loops were skipped at least once.
- For nested loops test the inner most loop first and then work outwards.
- For concatenated loops the values of dependent loops were set with the help of
- connected loop.Unstructured loops were resolved into nested loops or concatenated loops and tested as above.

6 RESULTS AND DISCUSSION

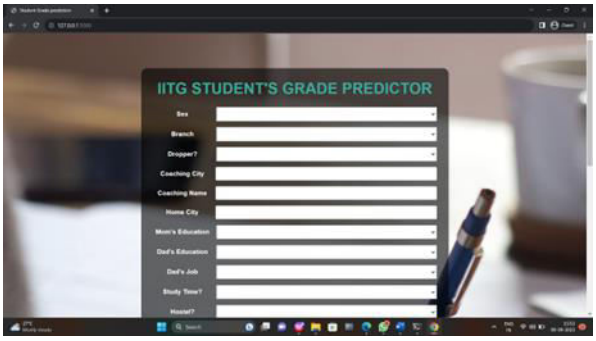

Fig 1: HOME SCREEN

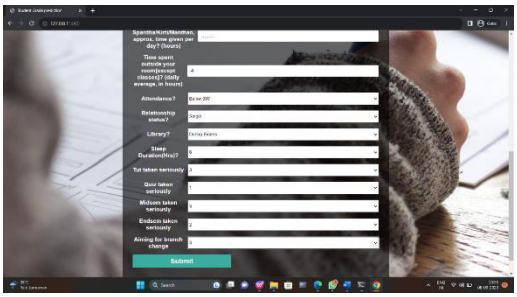

Fig 2: REGISTRATION PAGE

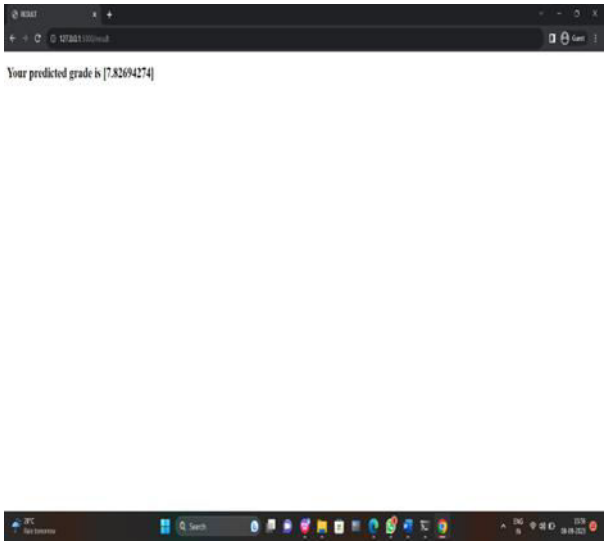

Fig 3: Predicted Page

7.CONCLUSION AND FUTURE SCOPE

The paper focuses on the student academic growth analysis using machine learning techniques. For analysis Naïve Bayes and KNN classifier are used. This process can help the instructor to decide easily about performance of the students and schedule better method for improving their academics. In future additional features are added to our dataset to acquire better accuracy.

8 REFERENCES

1. Hashim, Ali Salah, Wid Akeel Awadh, and Alaa Khalaf Hamoud. "Student performance prediction model based on supervised machine learning algorithms." In IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 032019. IOP Publishing, 2020.
2. Alnassar, Fatema, Tim Blackwell, Elaheh Homayounvala, and Matthew Yee-king. "How Well a Student Performed? A Machine Learning Approach to Classify Students‟ Performance on Virtual Learning Environment." In 2021 2nd International Conference on Intelligent Engineering and Management (ICIEM), pp. 1-6. IEEE, 2021
3. Saifuzzaman, M., Parvin, M., Jahan, I., Moon, N.N., Nur, F.N. and Shetu, S.F., 2021, June. Machine Learning Approach to Predict SGPA and CGPA. In 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) (pp. 211-216). IEEE.
4. H. Alamri, Leena, Ranim S. Almuslim, Mona S. Alotibi, Dana K. Alkadi, Irfan Ullah Khan, and Nida Aslam. "Predicting Student Academic Performance using Support Vector Machine and Random Forest." In 2020 3rd International Conference on Education Technology Management, pp. 100-107. 2020.
5. Verma, C., Illés, Z. and Stoffová, V., 2019, February. Age group predictive models for the real time prediction of the university students using machine learning: Preliminary results. In 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-7). IEEE.
6. Ünal, Ferda. "Data mining for student performance prediction in education." Data Mining-Methods, Applications and Systems (2020)