EXAMINING STUDENT PERFORMANCE BY USING MACHINE LEARNING

Parvez Hunkuntee[1], Sangita Mali[2], Shivanjali Bhosagi[3], Kishor Mane[4], Prof. U.S. Dhodmise[5]

[1,2,3,4,5]Brahmdevdada Mane Institute of Technology, Belati, Solapur, India.

ABSTRACT

To understand the student's rate of progress, it is crucial to forecast their performance. "Prevention is better than cure," goes the saying. The success of students can be significantly increased with early identification of at-risk students and preventive actions. The recommended task is utilized to assess a student's performance right now and forecast their outcomes in the future. Every year, many kids fall behind because of inadequate supervision and assistance from Staff. Based on the results, teachers can concentrate on the students who are more likely to receive lower grades in the final semester and can also help the student by identifying needs for the final exams. This project's major goal is to show how likely it is to train and model the dataset and how feasible it is to create a predictive model for student performance with a dependable accuracy rate by using linear regression algorithm **Keywords -** Marks Prediction, Linear Regression Algorithm, User interface Model.

1. INTRODUCTION

Machine learning is a part of man-made reasoning (AI) and software engineering which centers around the utilization of information and Algorithm to mimic the way that human learn, bit by bit working on its precision. ML algorithms fabricate a model in light of test information, known as preparing information, to make forecasts without being expressly modified to do as such. ML algorithms are utilized in a wide assortment of utilizations, like medication, discourse acknowledgment and PC vision. ML approaches are customarily isolated into three general classifications, contingent upon the idea of the sign or input accessible to the learning framework. Administered learning, solo learning, and Reinforcement learning are the three classifications of machine learning. ML is a part of artificial intelligence (AI) and software engineering which centers around the utilization of information and Algorithm to mimic the way that human learn, bit by bit working on its precision. ML algorithms fabricate a model in light of test information, known as preparing information, to make forecasts without being expressly modified to do as such. AI calculations are utilized in a wide assortment of utilizations, like medication, discourse acknowledgment and PC vision. ML algorithms are customarily isolated into three general classifications, contingent upon the idea of the sign or input accessible to the learning framework. Administered learning, solo learning, and Reinforcement learning are the three classifications of AI. Principal objective for this task is to help instructors to examinations understudy execution effectively and if necessary, they can help her/him to work on their understudy's exhibition by making a few moves like expanding their understanding hours, giving a few tasks. This research depends on the classification procedure, order by and large alludes to the planning of information things into predefined gathering and classes. The preparation information is investigated by classification algorithm and during arrangement stage the test information are utilized to gauge the exactness of the grouping rules. In this paper, different Machine learning algorithms on the notable consequences of a course being shown in single men in software engineering frameworks program to figure out the expectation precision

2. LITERRATURE REVIEW

2.1) Implementation of Student SGPA Prediction System: Case Study

[1] In today's world, there is competition in education institution every student plays a major role in the growth of the institution. An algorithm such as Logistic Model Tree, Random tree, and REP tree is used, the data set collected from the university may contain errors and noises which make the model less effective so data cleaning is done and the data set will reduce to 236 instances from 260 records. The REP tree algorithm has given more accuracy with 61.70%

2.2) Machine Learning Algorithm for Student's Performance Prediction: Case Study

[2]. The performance can be improved by predicting their marks by using the previous year's marks and can groom the students to improve themselves. By using machine learning techniques, we can improve the performance of every student the dataset of 1170 data was collected from three subjects. Algorithm such as K Nearest Neighbors, SVC, Decision Tree Classifier, and Linear Discriminant Analysis. The decision tree classifier model has given the highest accuracy of 94.44%.

2.3) Prediction of Student's Performance by Modelling Small Dataset Size: Case Study

[3] An educational institution's major objective is to give its students a high-quality education. Early performance forecasting for students can help them earn better grades and get into prestigious schools. The machine learning classification algorithm such as Naïve Bayes, Support vector machines, K-nearest neighbor, and Linear discriminant analysis. The Linear discriminant analysis has given accuracy of 79%.

2.4) Prediction of Student Academic Performance Using Neural Network, Linear Regression, and Support Vector Regression: Case Study

[4] Institutions have a significant impact on academic and pupil success. In the final year, pupils' academic standing has a big impact on their future jobs. The algorithm used is Neural Network (NN), Support Vector Regression (SVR), and Linear Regression (LR). The dataset of 134 data was collected, the linear regression has shown more accuracy compared to other algorithms.

3. WORK CARRIED OUT

The main purpose of the project is to help student in their academics and help teacher to predict performance of student based on there study time. Let us consider an example if a student is studying 1 or 2 hour a day by using this project based on the dataset the machine learning will predict the performance and it will recommend for further improvement. We have made changes in the UI interface to make it more attractive and User Friendly.

3.1 APPLICATION DEVELOPMENT

The implementation aims to develop a user-friendly predictive model for assessing student performance, benefiting both students and teachers. The model includes key features such as a homepage, an 'About Us' page, and a reference page, each designed with appealing CSS aesthetics. The framework employed for integrating the user interface (UI) with machine learning functionality is Flask, chosen for its effectiveness in facilitating seamless connections between components

Students can view to the web application where they can view their marks obtained from first to seven semesters and can predict their grade by clicking on the button 'predict'. Admin is also able to view the uploaded data in the web application and has the access to edit or delete the student details.

3.2 COMPARISON WITH OTHER ALGORITHMS

This step comprises of model training, pattern identification, testing, assessment results. As referenced prior dataset was separated into testing and training sets. In the training set, the model is built from the classification techniques. Testing set is utilized to evaluate the model. After those results will be assessed. To really look at which of the algorithm will best suit in prediction, the following algorithms have been tested:

1. Naïve-Bayes algorithm
2. Linear regression
3. Random Forest Support Vector Machine (SVM)
4. K Nearest Neighbor (KNN)

The drawn-out key pointers, which was extracted from the visualization analysis, were augmented in the five chosen classification algorithms. However, it's worth focusing on that since the picked classification algorithms have the capacity to train two distinct attributes' types, that is nominal and numeric. After training the model with these algorithms, the accurate results were assessed to see which variable type can work effectively with every classification algorithm in training the datasets of interest. As a result, Linear regression model reported the highest accuracy rate in predicting students' grades.

4. IMPLEMENTATION DETAILS

4.1. Data Collection: We collect data based on the time spent studying and the resulting study outcomes.

The data used in this research is collected from K S School of Engineering and Management. Real historic data of the students who studied undergraduate course in computer science and engineering is collected across 2010, 2011 ,2012 ,2013 ,2014 ,2015 ,2016 ,2017 batches. The dataset constitutes a total of 662 instances. The information of each student include percentage procured in all the 7 semesters (Sem_1=60, Sem_5=80) as attributes and grade, this particular student is likely to secure in the final semester, as target.

4.2) Data Cleaning

Irrelevant perceptions are information we don't need and can discard right away. We also remove any mistakes that happen during calculations, data transfer, or similar situations

4.3) Data Preprocessing

This phase in this study deals with the missing instances and other problems associated with the dataset. Initially the missing values are filled by calculating the mean with respect to the particular row. If a particular row has missing values of more than two columns, then this row is deleted All the rows containing less than six non-NA values are discarded.

	study_hours	student_marks
0	6.83	78.50
1	6.56	76.74
2	NaN	78.68
3	5.67	71.82
4	8.67	84.19

Let us consider the above dataset which has (2 column) (5 row) the 2 row study hours is missing it can cause execution error. To overcome such type of error the mean value is calculated

```
df.mean()

study_hours      6.995949
student_marks   77.933750
dtype: float64
```

The mean value is computed from a dataset of student records and is substituted with a null value to mitigate potential execution errors, such as encountering undefined or unexpected data during computational processes.

	study_hours	student_marks
0	6.830000	78.50
1	6.560000	76.74
2	6.995949	78.68
3	5.670000	71.82
4	8.670000	84.19

4.Data Scaling

There are numerous methods for doing this data scaling, this study uses Min-Max scaler method. In this method, the lowest of feature is made equivalent to zero and the highest of feature is equivalent to one. Min-Max Scaler compresses the data inside the given limit, typically of 0 to 1. It changes data by scaling features to a given limit. It scales the values to a particular value limit without changing the state of the primary distribution of the data.
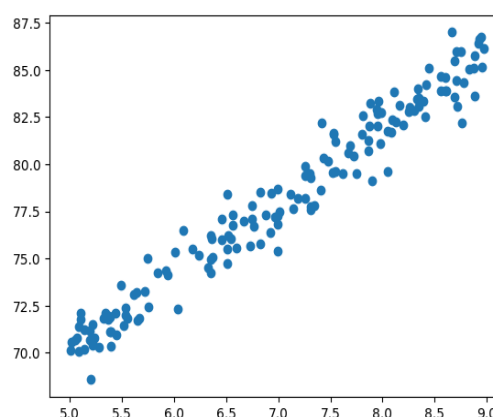
This is Scatter-plot for the dataset of student study time and obtain percentage.
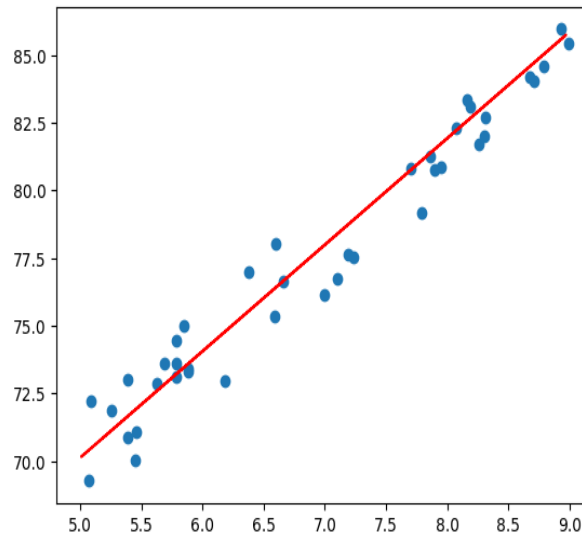
This scatter plot is generated by applying a linear regression algorithm, where the resulting data points represent the accuracy rates obtained from the regression analysis.

5. RESULTS

Prediction Result in UI Interface:

6. CONCLUSIONS

In this paper, we discuss the challenges faced by both students and teachers. To address these issues, we reference and implement a model aimed at enhancing student performance. The objective of this implementation is to demonstrate skills in machine learning and web development. This approach involves problem-solving methods to predict student performance based on their study time and assist them in achieving better results in exams.

7. REFERENCES

[1] Kaur, P. and Singh, W., 2016, August. Implementation of student SGPA Prediction System (SSPS) using optimal selection of classification algorithm. In 2016 International Conference on Inventive Computation Technologies (ICICT) (Vol. 2, pp. 1-8). IEEE.

[2] Hasan, H.R., Rabby, A.S.A., Islam, M.T. and Hossain, S.A., 2019, July. Machine learning algorithm for student's performance prediction. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[3] Zohair, A. and Mahmoud, L., 2019. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education, 16(1), pp.1-18.

[4] Obsie, E.Y. and Adem, S.A., 2018. Prediction of student academic performance using neural network, linear regression and support vector regression: a case study. International Journal of Computer Applications, 180(40), pp.39-47.

[5] Xu, J., Moon, K.H. and Van Der Schaar, M., 2017. A machine learning approach for tracking and predicting student performance in degree programs. IEEE Journal of Selected Topics in Signal Processing, 11(5), pp.742-753.

[6] Bujang, S.D.A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H. and Ghani, N.A.M., 2021. Multiclass prediction model for student grade prediction using machine learning. IEEE Access, 9, pp.95608-95621.

[7] Gull, H., Saqib, M., Iqbal, S.Z. and Saeed, S., 2020, November. Improving learning experience of students by early prediction of student performance using machine learning. In 2020 IEEE International Conference for Innovation in Technology (INOCON) (pp. 1-4). IEEE.

[8] Shah, M.B., Kaistha, M. and Gupta, Y., 2019, November. Student performance assessment and prediction system using machine learning. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 386-390). IEEE.

[9] Turabieh, H., 2019, October. Hybrid machine learning classifiers to predict student performance. In 2019 2nd international conference on new trends in computing sciences (ICTCS) (pp. 1-6). IEEE.