E-LEARNING FOR RURAL SCHOOLS

Project Reference No.: *45S_BE_2680*

College ***:*** *Sahyadri College of Engineering and Management, Mangaluru*
Branch ***:*** *Department of Computer Science and Engineering*
Guide(s) ***:*** *Dr. PRIYA R KAMATH*
Student(S) ***:*** *Mr. AKHILESH SARASIJAN*
Mr. SREEVISHNU T
Ms. DIYA SUJITH
Ms. TANISHA SINGHAL

Keywords: Data Analytics, Machine Learning, Hybrid Algorithm, Decision tree

Introduction:

It is difficult to analyze a large amount of data, enter data, and personally evaluate data. This relieves some of the faculty members' workload related to the individual student evaluations. The failure rates can be decreased by identifying the students who have performed poorly, and the at-risk pupils can gain from this. The ability to predict student achievement is essential for comprehending the pupils' current situation. Giving them direction can aid them in their academic endeavors. It can be helpful to provide career counseling for the kids by analyzing each student based on a variety of attributes. The void is filled by machine learning, which offers a method of analyzing these enormously changing data. In this article, we compare the accuracy of various machine learning (ML) techniques, and a hybrid ML algorithm is employed to find the best accurate model for further processing. A machine learning model that can predict student's final marks from the dataset collected taking into consideration various attributes.

Objectives:

- An early performance prediction would be beneficial for at-risk students
- Reducing the time and work of the faculties in evaluating the student's performance
- Determine the critical factors that affect the performance of students
- Minimize dropouts and failure ratios among students
- To provide proper career guidance

Methodology:

Data was collected from the rural school for training and testing purpose. The data consisted of the details regarding the unit test and the terminal marks that are the academic details for the 11th and 12th batches. After the collection of the dataset, data preprocessing and labeling was done. Here the data cleaning and filtering was done and only required data was taken into consideration. Various attributes were taken into consideration and data was filtered out. After the filtering of the required data was done the model was trained using various single algorithms and hybrid algorithms. The dataset were trained using various single algorithms and hybrid algorithms.

The algorithms that were trained were:

Decision Tree:
The most powerful and widely used tool for categorization and prediction is the decision tree. A decision tree is a flowchart-like tree structure in which each internal node represents an attribute test, each branch reflects the test's outcome, and each leaf node (terminal node) stores a class label. Decision Trees are designed to mirror human decision-making abilities, so they are simple to comprehend.

Logistic regression:
A supervised classification algorithm, logistic regression is a given collection of features (or inputs), X, the target variable (or output), y, can only take discrete values in a classification issue. Logistic regression, contrary to popular assumption, is a regression model. The model creates a regression model to forecast the likelihood that a given data entry belongs to the "1" category. Logistic regression models the data using the sigmoid function, just like linear regression assumes that the data follows a linear distribution.
g(z) = 1/(1 + e^-^z)

K-Nearest Neighbor algorithm:
The K-Nearest Neighbor algorithm is based on the Supervised Learning technique and is one of the most basic Machine Learning algorithms. The K-NN algorithm assumes that the new case/data and existing cases are similar and places the new case in the category that is most similar to the existing categories. The K-NN method stores all available data and classifies a new data point based on its similarity to the existing data. This means that new data can be quickly sorted into a well-defined category using the K-NN method. The K-NN algorithm can be used for both regression and classification, but it is more commonly utilized for classification tasks. The K-NN algorithm is a non-parametric algorithm, which means it makes no assumptions about the underlying data. A lower k value indicates that noise will have a greater impact on the outcome, whereas a big value indicates that it will be computationally costly. If the number of classes is two, data scientists normally choose an odd number, thus setting k=sqrt(n) is another straightforward way to find k.

Result:
For at-risk students, an early performance prediction would be helpful to reduce the faculty's time and workload when assessing student achievement. Identifying the important elements that have a bearing on pupils' performance can reduce student failure and dropout rates as much as possible. The model employed in our project compares the accuracies of different ML algorithms as well as one hybrid algorithm and it has been found that the hybrid algorithms are more accurate.

Scope for Future Work:

- The accuracies can be increased by adding more data to the dataset .
- Newer Ensemble methods can be employed to make the prediction more accurate.
- Advanced ensemble based machine learning algorithms like the extreme gradient boosting could also be used.
- Revise the model to include some more extra features in future