Women Codeathon – predicting performance of students in academics

<https://github.com/Py-Contributors/Student-s-Perform-Prediction>

-------------------------------------------------------------------------------------------------------------------------------

<https://www.kaggle.com/code/rmalshe/student-performance-prediction/notebook>

<https://garba.org/posts/2022/scoring_regression/>

<https://medium.com/nothingaholic/understanding-the-mean-squared-error-df41e2c87958>

https://realpython.com/train-test-split-python-data/

--------------------------------------------------------------------------------------------------------------------------------

To predict student performance in academics using machine learning, you can follow these general steps:

1.Collect and preprocess data: Gather academic data from various sources such as grades, test scores, attendance records, and participation in class discussions. Preprocess the data to remove missing values, outliers, and other noise that may affect the quality of the model.

2.Select relevant features: Identify relevant features or variables that may impact student performance. Feature selection is important for reducing the dimensionality of the data and improving the accuracy of the model.

3.Split data into training and test sets: Split the data into two sets: a training set and a test set. The training set is used to train the model, and the test set is used to evaluate the model's accuracy.

4.Train the model: Select a machine learning algorithm and train the model on the training set. Depending on the algorithm, you may need to tune hyperparameters to optimize the model's performance.

5.Evaluate the model: Evaluate the model's accuracy using the test set. Common evaluation metrics include mean absolute error, mean squared error, and R-squared.

6.Use the model for prediction: Once the model is trained and evaluated, you can use it to predict student performance on new academic tasks. Input the relevant features for a new student, and the model will output a predicted performance score.

7.Continuously monitor and improve the model: Monitor the model's performance over time and continuously improve it by incorporating new data or retraining the model with updated parameters.

It's important to note that the success of the model depends on the quality and relevance of the data used, as well as the appropriate selection and tuning of the machine learning algorithm. Additionally, the model should be regularly updated and improved based on new data and feedback.

----------------------------------------------------------------------------------------------------------------------------------

Machine learning frameworks that can be used for student performance prediction:

Scikit-learn: Scikit-learn is a popular Python library for machine learning. It includes a wide range of algorithms for classification, regression, and clustering, which can be used for performance prediction.

TensorFlow: TensorFlow is an open-source machine learning library developed by Google. It provides tools for building and training deep learning models, which can be used for performance prediction.

PyTorch: PyTorch is another open-source machine learning library that is widely used for building deep learning models. It offers a range of features for natural language processing, computer vision, and other applications.

Keras: Keras is a high-level neural networks API written in Python. It can run on top of TensorFlow, CNTK, or Theano, and provides a simple and easy-to-use interface for building deep learning models.

XGBoost: XGBoost is a popular open-source library for gradient boosting. It is known for its high performance and scalability, and can be used for building predictive models for large datasets.

H2O: H2O is a scalable open-source platform for machine learning and predictive analytics. It includes a range of algorithms for supervised and unsupervised learning, and can be used for building predictive models for large datasets.

Four algorithms that were used most often to predict the success of students' learning are ANN, Naïve Bayes, Logistic Regression, SVM and Decision Tree. Meanwhile, the data used in these research articles predominantly classified students' success in learning into two or three categories which are pass/fail; or fail/pass/excellent.

What is supervised learning?

<https://www.ibm.com/in-en/topics/supervised-learning#:~:text=pak%2Dfor%2Ddata-,What%20is%20supervised%20learning%3F,data%20or%20predict%20outcomes%20accurately>.

**Machine Learning Algorithm to Predict Student’s Performance: A Systematic Literature Review:**

Students' past learning achievement data (previous test scores), engagement (number of student visits to learning materials, search activities, participation in discussions), student demographics (age, expertise, date and time, location) are used to predict future test scores or assessments. Forward in the form of classification (PASS/ FAIL) and regression (range 0-100). Algorithm used: KNN, SVM, ANN, DT, Naïve Bayes, and Logistic Regression. The ANN algorithm has the highest level of accuracy when using a dataset of student engagement and learning success in the past.

The most widely used machine learning algorithms to predict student’s performance are Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, and Decision Tree algorithms.

Various machine learning algorithms used to predict student’s performance were tested for accuracy according to the data analysis stages required in machine learning algorithms and it was found that the most widely used machine learning algorithms to predict student’s performance were Artificial Neural Network (ANN) algorithms, Support Vector Machine (SVM), Logistic Regression, and Decision Tree

Machine learning algorithms implemented in predicting student’s performance include J48, JRIP, REPTree, Nnge, random tree, nave bayes, SMP, KNN, regression tree, random forest, decision tree, logistic regression, multilayer perceptron (MPL) neural network, support vector machine (SVM) and Artificial Neural Network

-------------------------------------------------------------------------------------------------------------------------------------

Paari sir (02-03-2023)

1. Outeryears removal(null values replaced with min values)
2. Normalization(types-Min max, z score) For all columns
3. correlation analysis( ranges between -1 to +1)(if 0 it doesn’t give a proper result)
4. feature selection analysis
5. modal building ( from skilearn lib algo)

Easy steps:

1. 80 %-20% model building
2. Prediction
3. Testing -New set of data
4. RMS(root mean square) value must be less for good model. Taken between Actual data vs predicted data
5. [Student Marks Prediction - Comparing Top ML Models | Kaggle](https://www.kaggle.com/code/yasserh/student-marks-prediction-comparing-top-ml-models)