UNIVERSITY ADMIT ELIGIBILITY PREDICTOR

TEAM ID: PNT2022TMID48794

Prior Knowledge:

Prior knowledge is the amount and quality of prior knowledge positively influence both knowledge acquisition and the capacity to apply higher-order cognitive problem-solving skills.

We increase the likelihood that our students will be able to recall and use what we teach by helping them engage their prior knowledge and connect new information to their prior understanding. New learning is constructed on prior knowledge. The more we understand about what students already think, and the more we help them engage their prior understandings, the more likely they are to learn well – and the less likely they are to misinterpret the material in our courses.

Prior knowledge is defined as 'the knowledge, skills, or ability that a student brings to the learning process' Other theorists have also provided vague definitions, using numerous terms to refer to prior knowledge [as current knowledge, world knowledge, expert knowledge and pre-knowledge]. Prior literature reviews have also identified the large number of terms available as a problem with the knowledge literature.
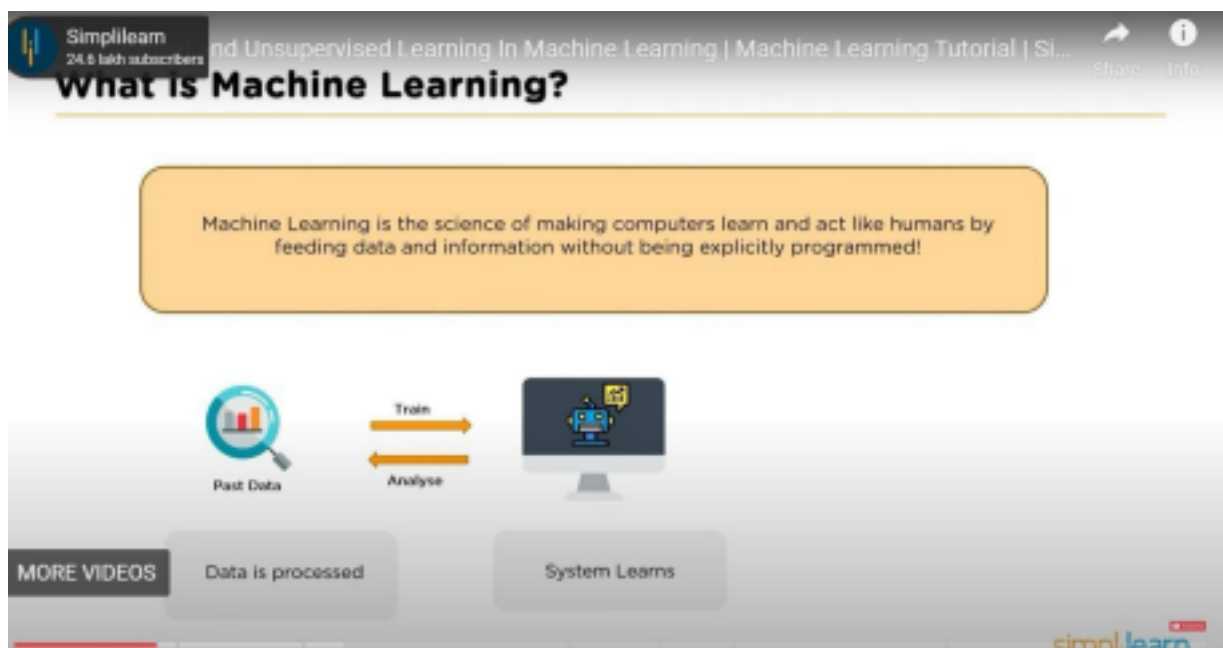
Supervised and unsupervised learning:

In Supervised Learning, a machine is trained using 'labeled' data. Datasets are said to be labeled when they contain both input and output parameters. In other words, the data has already been tagged with the correct answer.

So, the technique mimics a classroom environment where a student learns in the presence of a supervisor or teacher. On the other hand, unsupervised learning algorithms let the models discover information and learn on their own.
Supervised machine learning is immensely helpful in solving real-world computational problems. The algorithm predicts outcomes for unforeseen data by learning from labeled training data. Therefore, it takes highly-skilled data scientists to build and deploy such models. Over time, data scientists also use their technical expertise to rebuild the models to maintain the integrity of the insights given.

Common ML Problems

- **Classification**
- **Regression**
- **Clustering**

Unsupervised learning, also known uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, and image recognition.

EXPERIMENT AND RESULTS

Since the experimental outcome is very significant in any type of research, all the researchers want to achieve the highest level of accuracy according to their work. This level of accuracy can be different for using different algorithms and methodologies. Researchers must select the algorithm and approach that will provide the highest level of accuracy for the relevant study. In this investigation, the authors predict the admission of the examinees by using different types of supervised learning algorithms. This study applies some advanced algorithms, i.e., XGBoost, LightGBM, Gradient Boosting Machine to train and validate the predictive model. The authors used k-fold cross-validation (k=5) for validating the model. Therefore, error percentages are lower. Throughout this experiment, the authors use two specific different measures or metrics to evaluate the quality of classification: accuracy and F1-score based on the following Equations (1) and (4), respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

A. Classification Results before the Exam

This study used three machine learning techniques mentioned above to predict the possibility of getting admission in the engineering faculty at BSMRSTU before participating in the admission test. In this case, the obtained marks feature was not in the dataset because this investigation was before the test. The evaluation results to predict admission on test data are summarised in Table II. Note that the accuracy and F1 score are not high, the maximum value is nearly 60 in this case. It is justified as this model is trained and evaluated before the admission test and the authors do not have the admission test score. Nevertheless, the applicant can assess himself to some extent using this model before the admission test.

B. Classification Results after the Exam

This investigation then performs to predict admission opportunities in the engineering faculty at BSMRSTU after participating in the admission test. In this case, expected obtained marks in the exam is used because it is now known to the applicants. The evaluation results are given in Table III. GBM model achieves the highest score 95%. It means the proposed model using the GBM algorithm can accurately predict the admission chance of the applicants after the admission test .

C. Feature Importance before the Exam

The features' importance of the proposed model can be found by using the feature importance property of the model. Feature importance gives a score for each feature of our data, the higher the score is more important or relevant to our output variable. Fig. 2 shows the important features for predicting the admission opportunity of the applicants before the test using the XGBoost learning algorithm. Fig. 3 shows the important features for predicting the admission opportunity of the applicants before the test using the LightGBM learning algorithm. Fig. 4 shows the important features for predicting the admission opportunity of the applicants before the test using the GBM learning algorithm. The authors combined the feature importance of the aforementioned three learning algorithms. This study did the average feature importance (score) of each feature and plotted in Fig. 5. It shows the features that affect most for getting an opportunity of admitting in engineering faculty before participating in the admission test and these are as follows:

➢ College name
➢ Town/Village name
➢ District name
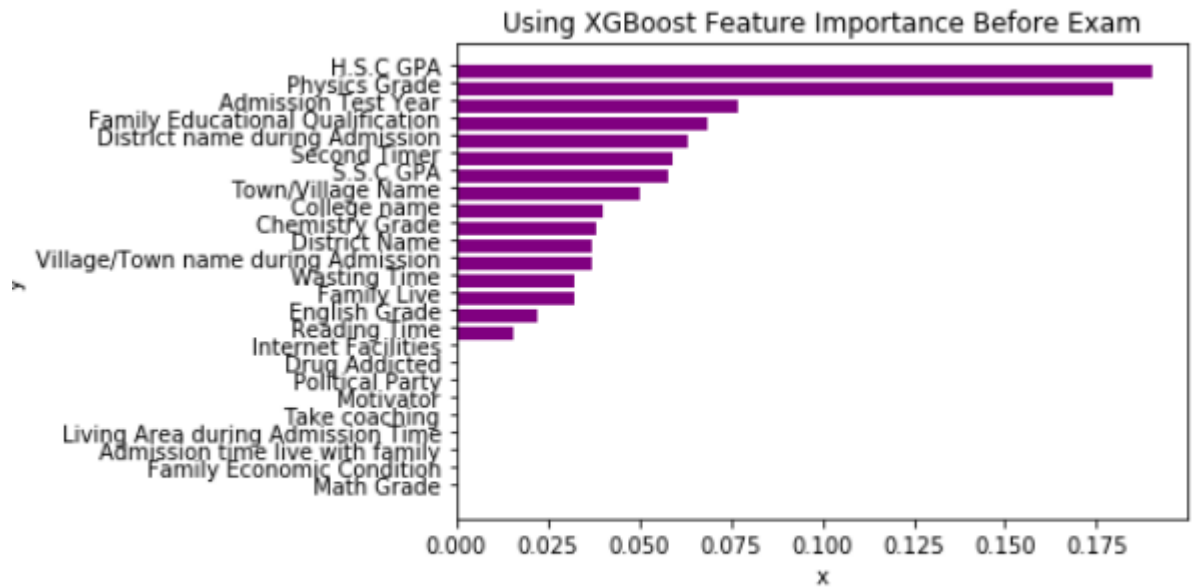➢ H.S.C GPA
➢ Reading time

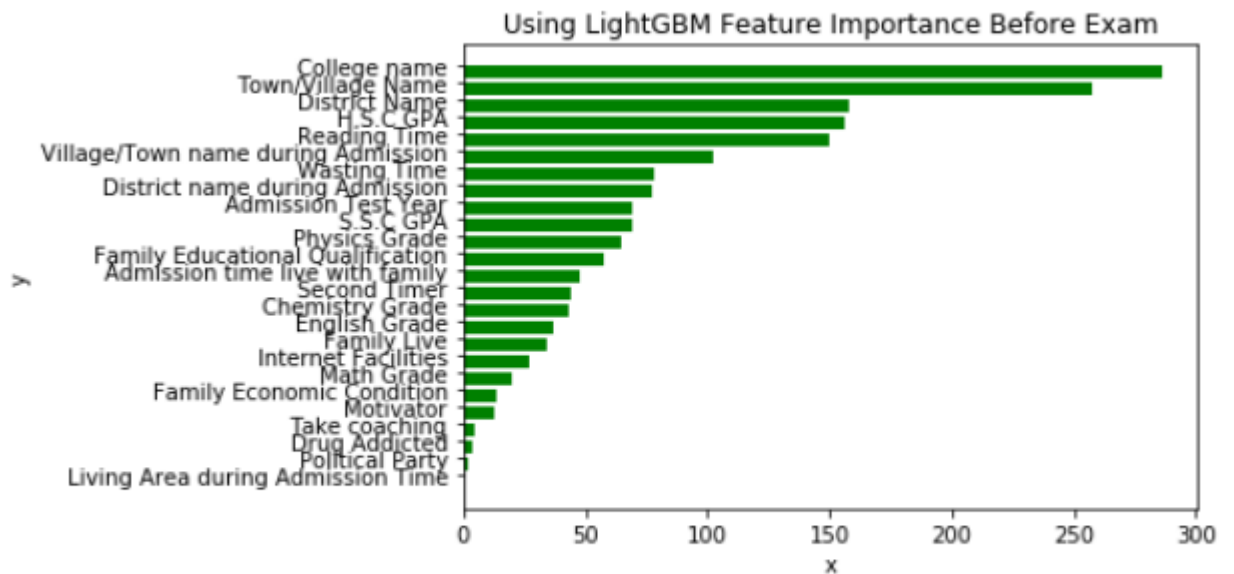Fig. 2.   Feature Importance using XGBoost (before the exam).

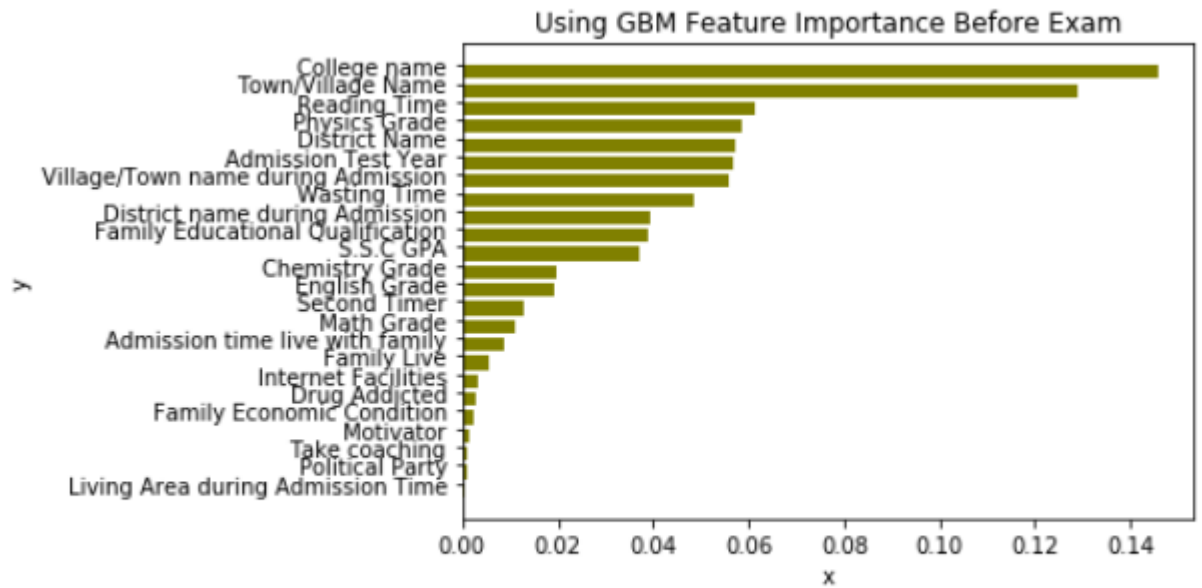Fig. 3.   Feature Importance using LightGBM (before the exam).

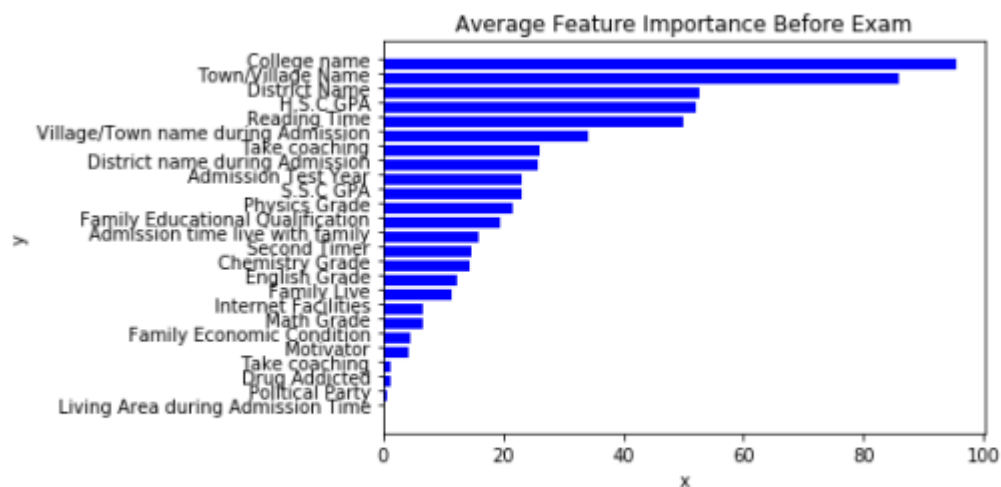Fig. 4. Feature Importance using GBM (before the exam).

Fig. 5. Average Feature Importance (before the exam).

Here, the academic performance of a candidate e.g. H.S.C GPA and Reading time is an obvious reason for better chances of admission. As we mentioned in Section III, since Bangladesh is a developing country and the development of the country is not equally distributed based on location the address and nativity of the candidate i.e. College name Town/Village name, District name are discovered as important features to predict undergraduate admission before the examination.

Feature Importance after the Exam

The important features for predicting the admission opportunity of the applicants after the test using the XGBoost learning algorithm. Fig. 7 shows the important features for predicting the admission opportunity of the applicants after the test using the LightGBM learning algorithm. Fig. 8 shows the important features for predicting the admission opportunity of the applicants after the test using the GBM learning algorithm.

The authors combined the feature importance of the aforementioned three learning algorithms. This study did the average feature importance (score) of each feature and plotted it in Fig. 9. It shows the features that affect most for getting an opportunity of admission in engineering faculty after participating in the admission test and these are as follows:

➢ Obtained marks
➢ Admission test year
➢ Town/Village name
➢ College name
➢ H.S.C GPA

Here, the academic performance of a candidate e.g. Obtained marks in the admission test and H.S.C GPA is understandable as the important features to predict undergraduate admission after the examination. Along with these, the address and nativity of the candidate i.e. Town/Village name and College name are discovered as important features because the development of Bangladesh is not equally distributed based on geographical factors. The students who live in town usually have more facilities than rural students. Admission test year is also found as an important feature because it is observed that some years' questions were comparatively easier and some years' questions were comparatively hard which may affect getting a chance in the admission