Abstract

In recent years Educational Data Mining (EDM) has emerged as a new field of research due to the development of several statistical approaches to explore data in educational context. One such application of EDM is early prediction of student results. This is necessary in higher education for identifying the "weak" students so that some form of remediation may be organized for them. In this paper a set of attributes are first defined for a group of students majoring in Computer Science in some undergraduate colleges in Kolkata. Since the numbers of attributes are reasonably high, feature selection algorithms are applied on the data set to reduce the number of features. Five classes of Machine Learning Algorithm (MLA) are then applied on this data set and it was found that the best results were obtained with the decision tree class of algorithms. It was also found that the prediction results obtained with this model are comparable with other previously developed models.

References

Early Prediction of Students Performance using Machine Learning Techniques

- Castro, F., Vellido, T., Ángela Nebot, and Mugica F., Applying Data Mining Techniques to e-Learning Problems.
- Domingos, P., A Few Useful Things to Know about Machine Learning, 
- Stefanowski, J., An Experimental Study of Methods Combining Multiple Classifiers - Diversified both by Feature Selection and Bootstrap Sampling.
- Zhao, Y., and Zhang, Y., Comparison of decision tree methods for finding active objects

Index Terms

Computer Science

Information Sciences
Early Prediction of Students Performance using Machine Learning Techniques

Keywords
Educational Data Mining  College Education  Machine Learning  Result Prediction
Kappa Statistic
F-Measure
WEKA.