**Data Mining Analysis on Student's Academic Performance through Exploration of Student's Background and Social Activities**

**Abstract**

Educational data mining techniques are widely used in academic prediction on student performance in classroom education. However most of the existing researches were studied and evaluated student coursework performance against the passing grade of the exam. In this paper, we performed analysis to identify the significant and impact of student background, student social activities and student coursework achievement in predicting student academic performance. Supervised educational data mining techniques, namely Naïve Bayesian, Multilayer Perceptron, Decision Tree J48 and Random Forest were used in predicting mathematic performance in secondary school. The prediction was performed on 2-level classification and 5-level classification on final grade. The experimental results have shown that student background and student social activities were significant in predicting student performance on 2-level classification. The model can be used for early predicting student performance to help in improving student performance on the subject.

**Existing System**

Existing literature works primarily focus on the student coursework performance, teaching quality and learning activities in predicting student performance. However, student performance might be impacted by other factors such as study habits, attendance of school, social activities, student family background and others. Understanding the impact of these factors might be able to improve student performance in a subject as early as possible.

**Proposed System**

we aim to identify and analyse the impact of student background attributes and student social activities attributes on student performance. Supervised educational data mining namely, Naïve Bayesian, Multilayer Perceptron, decision tree J48 and random forest are applied to build prediction model. The significant and impact of student background and social activities attributes can be visualized and defined from the decision tree structure generated by the models.

**Implementation**

**Modules**

**The modules are:**

**1.2-Level Classification and 5-Level Classification**

This experimental result also has shown that student coursework results are significant attributes in predicting student performance in mathematic final grade as it has highest precision accuracy 0.924 in 2-level classification and 0.791 in 5-level classification. In overall, accuracy of algorithms in 2-level classification are out performed models in 5-level classification. The models accuracy are > 0.5, this indicated that student background and student social activities are viable to be used to perform early analysis and prediction of at-risk student to determine whether it pass or fail the subject.

**2. Clustering**

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

**3.Data visualization**

Data visualization is an important preprocessing task, which used graphical representation to simplify and understand complex data. Visualization techniques have been recently used to visualize online learning aspects. Instructors can utilize the graphical representations to understand their learners better and become aware of what is occurring in the distance classes.

**Conclusion**

we have demonstrated that student background and social activities are significant to be applied for early prediction on student performance and also can be used to identify at-risk student. Hence, early prediction with these models may be helpful for the teachers and students. Students may be able to perform better in the academic performance. Meanwhile, teachers can do early preparation to perform education interventions in teaching the subject.

**Future Work**

In future work, unsupervised education data mining techniques will be applied to discover correlation and impact of the attributes in clusters. In addition, we will discover the correlation and impact using attributes analysis and feature selection to provide more accurate prediction models for predicting early at-risk students’ performance.