A comparative analysis of four classification algorithms for university students performance detection

Dipta Das 1, *Asif Khan Shakir 1, Sah Golam Rabbani 1, Mostafijur Rahman 1, Syamimi Mardiah Shaharum 2, Sabira Khatun 2, Norasyikin Binti Fadilah 2, Khandker M Qaiduzzaman 1, Md. Shariful Islam 3, Md. Shohel Arman 1

1 Department of Software Engineering, Daffodil International University, Bangladesh.
2 Faculty of Electrical and Electronics Engineering, University Malaysia Pahang, Malaysia.
3 Institute of Information Technology, University of Dhaka, Bangladesh.
Email: asif.swe@diu.edu.bd

Abstract:
The student's performance plays an important role in producing the best quality graduate who will responsible for the country's economic growth and social de-velopment. Labor market also concern with student's performance because the fresh graduate students are considered as an employee depends on their academic performance. So, identification the reason behind student's performance variation provides a valuable information for planning education and policies. Many researchers try to find out the reason with different types of data mining approaches in different countries. But none of them worked with Bangladeshi students. This paper proposed a model for identifying the key factors of variation Bangladeshi students' academic performance and predicts their results. This paper proposes a model which able to identify the students who need special attention. Different types of feature selection methods were used such as Co-relation, Chi-Square and Euclidean distance to select valuable features. And showing the comparison of feature selections result through decision tree, Naive Bayes, K-nearest neighbor and Artificial Neural Network classifiers algorithm. The performance analysis is done by using student SGPA and review on given facilities from a university. From the performance analysis result it is found that, decreasing number of classes in dataset, the Artificial Neural Network(ANN) (93.70%) performs better than Decision Tree(DT)(92.18%), K-Nearest Neighbors (KNN)(77.74%) and Naïve Bayes (NB)(68.33%). However, increasing number of classes in dataset the DT perform better then ANN, KNN, NB.

Acknowledgments