

ABSTRACT

Keywords: Machine Learning, Classification, Student potential, performance

Students are an important factor in any education sector because any university or institution's progress is based on student's academic performance. The digitization of the world put an immense impact on the working of the education system and the career planning of students. Generally, career counselors/ experts play an important role to evaluate and assisting in choosing appropriate career selection and planning. These conventional methods and practices are not much impactful and after a point, they proved inefficient and ineffective. This study will explore the various technologies used to provide career guidance and counselling. This study works on real-time students' data and considers different attributes to find out which factors play a major role in choosing a career by using Machine Learning techniques.

ML techniques can be implemented successfully in the education field. the data source used and their types were characterized which consists of 3 main parts: student dataset for potential analysis, student datasets based on boards and student dataset for student performance. These datasets contained student's demographic, socio-economic and academic attributes. Then, the data pre-processing was done. In this data was cleaned, sorted, filtered and transformed because this help to understanding the data and gives better predictive accuracy used by different ML techniques. Different feature selection methods were also used for the optimal feature subsets.

Different ML techniques were applied on the student datasets and generate results. These results were evaluated by measuring the five metrics (accuracy, precision, sensitivity, specificity and f-measure) of all the models. The results obtained from the different classifier models compared each other. The results showed that GB was the highest predictive accuracy for analysis of student potential. The results of student performance prediction showed that Random Forest achieves higher accuracy. Finally, the aim of this comparison to find the best classifier as selected for this study.