

Chapter 20

An Exploratory Study on the Use of Machine Learning to Predict Student Academic Performance

Patrick Kenekayoro

 <https://orcid.org/0000-0003-0021-6584>

Mathematics / Computer Science Department, Niger Delta University, Amassoma, Nigeria

ABSTRACT

Optimal student performance is integral for successful higher education institutions. The consensus is that big data analytics can be used to identify ways for achieving better student academic performance. This article used support vector machines to predict future student performance in computing and mathematics disciplines based on past scores in computing, mathematics and statistics subjects. Past subjects passed by students were ranked with state of art feature selection techniques in an attempt to identify any connection between good performance in a particular discipline and past subject knowledge. Up to 80% classification accuracy was achieved with support vector machines, demonstrating that this method can be developed to produce recommender or guidance systems for students, however the classification model will still benefit from more training examples. The results from this research reemphasizes the possibility and benefits of using machine learning techniques to improve teaching and learning in higher education institutions.

INTRODUCTION

Teaching is one of the main goals of Higher Education Institutions (HEIs), and part of the objectives of teaching is guiding students towards specialties that they may thrive in. If a student's strong areas can be determined from his/her previous performances, it takes us closer to automating the process of guiding students to specialties. Supervised learning techniques may also be used to identify those attributes (performance in certain subjects) that may influence success in a particular specialty. Accurately

DOI: 10.4018/978-1-6684-6291-1.ch020

identifying these attributes and adding significance to ensuring that those students who may want to advance to a particular specialty have a good grasp of its required (previously identified) subjects may increase the likelihood of students' success in their chosen specialties. This may also be reflected in the design curricula that adequately shows prerequisites or desired knowledge for respective specialties.

Traditional methods that have been used to investigate factors that affect student academic performance involve correlation and regression approaches (Figlio & Kenny, 2007; Myller, Suhonen, & Sutinen, 2002; Pritchard & Wilson, 2003), but data mining techniques may also be used for this purpose. Pahwa, Arora and Thakur (2011) succinctly defined data mining as "...efficient discovery of non-obvious information from a large collection of data..." Data mining is becoming an increasingly important tool for knowledge discovery, as there is an exponential increase in the amount of information available. Data mining deals with this information overload by finding patterns or relationships in data that would otherwise be unidentifiable, thereby simplifying decision making for policy makers (Govindarajan & Chandrasekaran, 2012).

The use of data mining techniques such as supervised learning to explore educational data in order to improve student performance is an aspect of Educational Data Mining. This research uses a case study from a real higher education institution to determine if machine learning techniques can be used to predict student academic performance.

LITERATURE REVIEW

Educational Data Mining (EDM) has been defined as "...scientific inquiry focused on developments of methods for making discoveries with data from educational settings and using those methods to better understand students and the settings which they learn in..." (Baker, 2010). Understanding students through educational data mining can give new insights to ways that can improve student academic performance. Academic success is seen as a critical factor for individual success in contemporary society (Pritchard & Wilson, 2003). If students' academic performance can be previously predicted, it gives policy makers the opportunity to introduce policies that will improve student academic success rate, thereby increasing the likelihood of successful completion of a higher education degree. Also, creating predictive models that can be used for early identification of weak students who will be at risk is beneficial for reducing failure or dropout rates in higher education institutions (Raju & Schumacker, 2016).

Techniques for predicting student performance have been researched extensively. Historically, correlation and multiple regression are the traditional methods used to investigate the extent to which socioeconomic or psychological factors can positively or negatively affect a student's academic performance. For example, Figilo and Kenny (2007) have used correlation to investigate the relationship between teacher incentives and student performance. However, machine learning techniques can use these socioeconomic, psychological or other factors to also predict how well a student will perform in a particular subject, thus it is a useful technique that can be used in the design of systems that can provide real time guidance to students.

Although early studies have used regression to investigate factors influencing student performance (Myller et al., 2002; Pritchard & Wilson, 2003), there is still a dearth of research on how effective these models are for Educational Data Mining in higher education institutions (Raju & Schumacker, 2016). Linear regression may perform poorly and sometimes more elaborate classification models perform better than linear regression for student performance prediction (Myller et al., 2002), but in other times

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/an-exploratory-study-on-the-use-of-machine-learning-to-predict-student-academic-performance/307461

Related Content

An Introduction to Audio Manipulation and Generation of a Unified High Performance Recommendation Framework for Netflix

R. Angeline, Harish Babu M., Rehith B. and Pranav B. (2023). *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications* (pp. 319-332).

www.irma-international.org/chapter/an-introduction-to-audio-manipulation-and-generation-of-a-unified-high-performance-recommendation-framework-for-netflix/316763

MHLM Majority Voting Based Hybrid Learning Model for Multi-Document Summarization

Suneetha S. and Venugopal Reddy A. (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 67-81).

www.irma-international.org/article/mhlm-majority-voting-based-hybrid-learning-model-for-multi-document-summarization/233890

Survey of Recent Applications of Artificial Intelligence for Detection and Analysis of COVID-19 and Other Infectious Diseases

Richard S. Segall and Vidhya Sankarasubbu (2022). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-30).

www.irma-international.org/article/survey-of-recent-applications-of-artificial-intelligence-for-detection-and-analysis-of-covid-19-and-other-infectious-diseases/313574

Boosting Convolutional Neural Networks Using a Bidirectional Fast Gated Recurrent Unit for Text Categorization

Assia Belherazem and Redouane Tlemsani (2022). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-20).

www.irma-international.org/article/boosting-convolutional-neural-networks-using-a-bidirectional-fast-gated-recurrent-unit-for-text-categorization/308815

A Review on Time Series Motif Discovery Techniques an Application to ECG Signal Classification: ECG Signal Classification Using Time Series Motif Discovery Techniques

Ramanujam Elangovan and Padmavathi S. (2019). *International Journal of Artificial Intelligence and Machine Learning* (pp. 39-56).

www.irma-international.org/article/a-review-on-time-series-motif-discovery-techniques-an-application-to-ecg-signal-classification/238127