

# Identify Students at Risk Using Graph Representation and Conventional ML Methods

Balqis Albreiki<sup>1</sup>, Tetiana Habuza<sup>2</sup>, Nazar Zaki<sup>3</sup>

United Arab Emirates University

## Abstract

As in many other areas of society, education has been significantly affected by technological advancement, creating many online learning platforms like Virtual Learning Environments (VLEs) and Massive Open Online Courses (MOOCs). While these platforms offer a variety of features, none of them incorporates a module that accurately predicts students' academic performance and commitment. Consequently, it is crucial to design Machine Learning (ML) methods that are both accurate and reliable in predicting student performance and identifying students at risk as early as possible. The graph representation of students' data may provide new insights into this particular area. In this research, we present a non-complex but highly accurate technique for converting tabulated data into graphs. We employ distance measures (Euclidian and Cosine) to calculate similarities between students' data and to build a graph. From the graph, we extract topological features (TF) to enhance our data. This allows us to capture structural correlations between data and gain deeper insight than with isolated data analysis. Original features (OF) and TF can be used as it is or jointly to improve the predictive power of the ML method applied. The proposed method was tested on the educational dataset and returned superior results. A comparison has been made between OF and OF + TF fed to the ML classification model. Model segregates students into three classes: "failed", "at risk", and "good". The AUC ROC fed with OF and OF + TF reached 0.948 and 0.964 respectively. TF improved the performance by 2.019%. The proposed solution may serve as a tool for early detection of students at risk. This will benefit universities and may allow them to better predict performance, improving their effectiveness and reputations.

**Keywords:** Graph representation; classification; machine learning; student performance; self-graph topological feature