Chapter 14
Unveiling Academic Success:
Harnessing Graph Machine Learning for Student Performance Prediction

Polisetty Sri Hari Sai Saran
Department of Data Science and Business Systems, Faculty of Engineering and Technology, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India

R. Krishna Kumari
https://orcid.org/0000-0002-1802-628X
Department of Mathematics, Faculty of Engineering and Technology, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India

ABSTRACT

In this chapter, we delve into the utilization of graph machine learning techniques to forecast student academic performance. By harnessing graph-based representations of educational data, our study endeavors to unearth underlying patterns and connections that impact student success. Through a fusion of feature engineering, graph analytics, and predictive modeling, we aim to investigate the efficacy of graph-based methodologies in improving the precision and interpretability of student performance prediction systems. This paper investigates the effectiveness of logistic regression, K-nearest neighbors (KNN), and a custom Graph Neural Network (GNN) model for predicting student performance in exams. Our analysis reveals that the custom GNN model outperforms both logistic regression and KNN, achieving higher accuracy and efficiency in student performance prediction. The custom GNN model leverages

DOI: 10.4018/979-8-3373-1399-3.ch014

the graph-based representation of educational data, which enhances its ability to capture complex relationships and dependencies among students.

1. INTRODUCTION

Predicting student academic performance is a crucial task in educational institutions worldwide, as it allows educators to identify struggling students early and provide them with the necessary support to succeed (Balcioglu et al.,2023). However, traditional predictive models based on tabular data often fall short in capturing the intricate interplay of factors that influence student success (Subramanian et al.,2023). These models typically rely on individual attributes such as demographics, prior grades, and standardized test scores, overlooking the rich relational context inherent in educational data (Brooks et al.,2023).

In recent years, the advent of graph-based machine learning techniques has opened up new avenues for analyzing complex relational data (Dervenis et al.,2022). By representing educational data as a graph, wherein nodes correspond to entities such as students, courses, and assignments, and edges encode relationships such as enrollment, collaboration, and academic performance, we can capture the multidimensional interactions and dependencies that shape student outcomes (Gajwani et al.,2021, Kaur et al.,2023). This graph- based representation offers a holistic view of the educational ecosystem, enabling us to uncover hidden patterns and insights that traditional models may overlook.

One of the key challenges in student performance prediction lies in the early identification of at-risk students who may benefit from targeted interventions. Early intervention strategies have been shown to significantly improve student retention and success rates (Ojajuni et al.,2021, Soyoye et al.,2023). However, conventional models often lack the granularity and interpretability necessary to pinpoint students who are most in need of support. By leveraging graph-based representations, which inherently encode structural information and relational context, we can enhance the accuracy and interpretability of predictive models, thereby facilitating timely interventions and personalized support strategies (Patil et al.,2023, Sekeroglu et al.,2019).

In this research article, we aim to explore the potential of graph machine learning techniques for predicting student academic performance. We will begin by discussing the limitations of traditional predictive models and the motivations for adopting a graph-based approach. We will then outline our objectives, which include:

(i) Constructing a comprehensive graph representation of educational data, incorporating various entities and relationships relevant to student performance prediction.